Using Machine Learning to Assist Decision Making in the Assessment of Mental Health Consumers Presenting to Emergency Departments

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Oliver Higgins

Bachelor of Technology (B.Tech.), RMIT University

Bachelor of Nursing (B.N.), University of Technology Sydney

Graduate Certificate (Clinical Redesign), University of Tasmania,

Graduate Certificate Mental Health (Nursing), University of Technology Sydney

ORCID 0000-0001-6545-145X

School of Health and Biomedical Sciences

College of Science, Technology, Engineering and Maths

RMIT University

Australia

November 2024

Using Machine Learning to Assist Decision Making in the Assessment of Mental Health Consumers Presenting to Emergency Departments

Principal Supervisor:

Prof Rhonda Wilson,

School of Health and Biomedical Sciences, College of Science, Technology, Engineering and Maths, RMIT University Australia.

Email: rhonda.wilson@rmit.edu.au

Co-Supervisor:

Prof Stephan Chalup,

School of Information and Physical Sciences, University of Newcastle, Australia.

Email: stephan.chalup@newcastle.edu.au

Declaration

I certify that except where due acknowledgement has been made, this research is that of the author alone; the content of this research submission is the result of work which has been carried out since the official commencement date of the approved research program; any editorial work, paid or unpaid, carried out by a third party is acknowledged; and, ethics procedures and guidelines have been followed.

In addition, I certify that this submission contains no material previously submitted for award of any qualification at any other university or institution, unless approved for a joint-award with another institution, and acknowledge that no part of this work will, in the future, be used in a submission in my name, for any other qualification in any university or other tertiary institution without the prior approval of the University, and where applicable, any partner institution responsible for the joint-award of this degree.

I acknowledge that copyright of any published works contained within this thesis resides with the copyright holder(s) of those works.

I give permission for the digital version of my research submission to be made available on the web, via the University's digital research repository, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Oliver Higgins 25th November 2024

Dedication

This PhD thesis is dedicated to my beloved family, Connor, Emily, Olivia, and Juliet Higgins.

Your unwavering support and understanding have been the foundation of my journey over these past four years. To Juliet, you are an amazing and beautiful wife. This achievement is as much yours as mine. Your love and encouragement have made my life extraordinary, and I am deeply grateful for every moment we've shared. I could not have reached this milestone without you.

This work is dedicated to all of you.

Acknowledgments

Acknowledgment of First Nations People

In the spirit of reconciliation this project acknowledges the Traditional Custodians of Country throughout Australia. I recognise their continuing connection to the land and waters, and thank them for protecting our coastline and its ecosystems since time immemorial, acknowledging that they never ceded sovereignty. I pay our respect to their Elders past and present and extend that respect to all Aboriginal and Torres Strait Islander peoples.

Acknowledgment of Lived Experience

I acknowledge the individual and collective expertise of those with a living or lived experience of mental health, alcohol, and other drug issues. I recognise their vital contribution to this research, valuing the courage of those who share this unique perspective for the purpose of learning and growing together to achieve better outcomes for all.

Acknowledgment of Funding

Partial financial support was received from the NSW Ministry of Health as part of the Towards Zero Suicides initiative.

Acknowledgment of Support

I wish to acknowledge the support received from Central Coast Local Health District Mental Health service as part of the Towards Zero Suicides initiative.

I wish to express my sincere appreciation to my supervisors, Prof Rhonda Wilson and Professor Stephan Chalup, for their invaluable mentorship and support throughout my PhD. Their profound knowledge and critical insights have been instrumental in guiding my research to completion. Their encouragement and faith in my abilities have helped me navigate the

complexities of this project, allowing me to explore new ideas and develop a deeper understanding of my field. Their mentorship has been both a privilege and a source of inspiration.

I wish to acknowledge that professional editor Kate Leeson provided copy editing and proofreading services.

Acknowledgment of Personal Support

David Duerden

You are first and foremost on this list because this PhD would never have happened without you. Not only did you make it possible, but you believed in my idea and had more trust in my capabilities than I did at the time. I owe so much of my career to your mentorship and your blue-sky thinking. I look back over the past 10 years that we have worked together and am amazed by what we have achieved. But I would not have taken these leaps or taken the opportunities had you not pushed. I have always followed your lead (sometimes begrudgingly), but I am always amazed at how often things come off and how you are able to make things work. I would not be the person I am today without your guidance and support. Thank you.

Dr Rachel Sheather-Reid

A long time ago, I said to you that one day I wanted to do a PhD, and you were the first person who made me feel like this was something that I could do. Without you, I wouldn't have made it this far in my PhD journey. You have supplied me with the energy, enthusiasm, and reassurance that this was something I could do. Thank you for all the time you spent listening to my ideas, whatever book I was currently reading, and all the time you spent reading my work. I am grateful and in debt to the care and time that you have provided me over the past few years. This PhD would not have been possible without you.

Dr Brooke Short

Thank you so much for your support in the genesis of this project. I left your office that first day with the belief that my idea for this PhD was in fact a good idea. If not for your help, I do not believe that we would have been able to get this project off the ground and into such an

advanced state. I enjoyed our time working together, and thank you again for your ongoing support of this research.

Dr Kevin Kellehear

Kevin, you will always have such a special place in my life. You made me enjoy learning more than I could have thought. Your constant support, encouragement and, at times, tolerance drew me to academia. I enjoyed your teaching and was not ready to leave my studies. I look back at what I learnt through my undergrad and postgrad, and they are lessons that I still use in my life and practice. I feel so fortunate to have been your student. Thank you.

I have been fortunate to have some amazing nurses in my life.

Patricia Christie (Grandma)

To the most amazing nurse, I am not one who normally makes wishes, but I wish you were still here for me to talk to about the work we have been able to achieve. You are the role model who set the standard for me as a nurse throughout my career. I strive to emulate your compassion and strength in the face of adversity. I look to you as the trailblazer and someone who would break the rules if it meant doing what was right. I hope I have pushed these same boundaries throughout my work. I miss you, and I hope my work will make you proud.

Ellen Hardcastle

Without you, this is not a path that I would have chosen. You helped me take those first steps, and I am eternally grateful for your help and guidance throughout my journey so far. When we met, I was determined to study and realise my potential, but it felt like such an impossible mountain to climb. Nursing has offered me more opportunities than I could have dreamed of. I am grateful not only to have had your guidance and help but also to have had the opportunity to walk alongside you throughout your career.

Janet Chesworth

From the moment I stepped onto the ward in M1, you threw me in the deep end. But you did it in such a way that you seemed to believe more in my capabilities than I did. Every time I have seen you since, you have reiterated your belief in my abilities and instilled a sense of self-belief for which I am grateful. Because of your guidance and support, you made me feel that I was more than capable of achieving this accomplishment.

Kylie Williams

I was reassigned to the mental health ward at Wyong at the end of my Trainee Enrolled Nursing rotation and, to be honest, it wasn't where I had wanted to be. On my very first day, you walked in with a huge stack of printed A4 pages and started talking non-stop—it felt like hours! But what stood out to me was your dedication, passion, and incredible depth of knowledge. You are one of the most caring people and amazing nurses I have ever met. I knew I was meant to be a mental health nurse from that moment on. Your unwavering commitment to my growth created an environment where I could truly flourish. I found my place in nursing because of your guidance and support. Looking back on the time we worked together and everything we've done since, I realise you've never stopped being my educator. Thank you for your patience and for helping me become the nurse I was meant to be.

Table of Contents

Declaration ..... **ii**
Acknowledgments ..... **iv**
List of Figures ..... **xv**
List of Tables ..... **xv**
Glossary and Acronyms ..... **xvi**
Abstract **1**
Preface **4**
Chapter 1: Introduction 4
Chapter 2: Background 4
Chapter 3: Critical Review of the Literature 5
Chapter 4: Theoretical Foundations of Trust 5
Chapter 5: Update to Literature Review 5
Chapter 6: Methods 6
Chapter 7: Results 6
Chapter 8: Discussion 7
Chapter 9: Conclusion 7
Complete List of all Publications from this Research 8
Complete List of all Conference Presentations from this Research 9
Chapter 1 Introduction **11**
1.1 Introduction 11
1.2 The Aims of the Research 11
1.3 Significance of the Research 13
1.3.1 Missed and Unfinished Care *15*
1.3.2 Artificial Intelligence and Missed Care *18*
1.4 Description of Research Methods 21
1.4.1 Research Question *21*
1.4.2 Research Design *21*
1.4.3 Empowering Nursing Workflows: Advancements in Health Decision Processes *24*
1.5 Researcher Perspective 25
1.5.1 Personal Reflection on the Research *30*
1.6 Alignment with National and International Strategies 34
1.6.1 Sustainable Development Goals *36*
1.6.2 Closing the Gap *37*

1.6.3 Towards Zero (NSW Health) *39*
1.6.4 Ethical Principles for Artificial Intelligence in Health *40*
1.7 Mental Health Services in the Context of COVID-19 46
1.8 Overview of Thesis 48
1.8.1 List of Publications in Chapters *48*
1.8.2 List of Further Publications Generated During this Research *49*
1.9 Chapter Summary 50
Chapter 2 Background 51
2.1 Chapter Introduction 51
2.1.1 Artificial Intelligence and Machine Learning Overview *52*
2.1.2 Example of Machine Learning *53*
2.2 The Impact of Artificial Intelligence in Healthcare and Nursing 55
2.2.1 Changes to Nursing Practice *56*
2.2.2 Challenges and Ethical Considerations *57*
2.3 Commercial Determinants and Therapeutic Chatbots: A Mental Health Nursing Perspective 58
2.3.1 Commercial Determinants and Ethical Considerations in the Integration of AI Technologies in Mental Health Care *64*
2.4 Technological Influence on Psychosis Treatment and Perception 65
2.4.1 Case Study: Distress and Delusions in the Context of Perceived Artificial Intelligence Influence *65*
2.4.2 Mental State Evaluation and Ethical Analysis *66*
2.4.3 Interpretations of Innovation: The Intersection of Technological Advancement and Psychosis *70*
2.5 Chapter Summary 87
Chapter 3 Critical Review of the Literature 90
3.1 Integrative Literature Review Protocol Paper 91
3.1.1 Introduction *92*
3.1.2 Background *93*
3.1.3 The Review *99*
3.1.4 Discussion *105*
3.1.5 Limitations *105*
3.1.6 Conclusion *106*
3.2 Artificial Intelligence (AI) and Machine Learning (ML) Based Decision Support Systems in Mental Health: An Integrative Review 106
3.3 Article Summary 122
3.3.1 The Role of Clinicians and Patients as End Users *122*
3.3.2 Key Issues *123*

3.3.3 Missed Care *123*
3.3.4 Summary of the Section *125*
3.4 Chapter Summary 125
Chapter 4 Theoretical Foundations of Trust 129
4.1 Artificial Intelligence in Nursing: Trustworthy or Reliable? 130
4.1.1 Article Summary *142*
4.2 Chapter Summary 143
Chapter 5 Update to Literature Review 145
5.1 Update to Literature Review 146
5.1.1 Introduction *146*
5.1.2 Methods *147*
5.1.3 Results *150*
5.1.4 Discussion: Advancing Artificial Intelligence-Based CDSS in Mental Health Care *160*
5.1.5 Conclusion: Integrating Artificial Intelligence with Workforce Solutions for Sustainable Care *164*
5.2 The Rise of Large Language Models (LLMs) and Their Impact on Health Care 165
5.2.1 Introduction *165*
5.2.2 Recent Development of LLMs *166*
5.2.3 Impact of LLMs on Health Care *167*
5.2.4 Ethical and Practical Challenges in Health Care *168*
5.2.5 Conclusion *170*
5.3 Chapter Summary 171
Chapter 6 Methods 173
6.1 Introduction and General Overview of the Research 174
6.1.1 Theoretical Frameworks Considered for this Study *176*
6.1.2 Design Thinking Methodology Explained *178*
6.1.3 Ethical Considerations *180*
6.1.4 Retrospective Cohort Study *181*
6.1.5 Supervised Machine Learning Method *181*
6.1.6 Research Setting and Context *184*
6.2 Research Design, Methods and Ethics Overview 186
6.2.1 Research Objectives *187*
6.3 Research Design and Methods 192
6.3.1 Study Design *192*
6.3.2 Statistical Methods *194*
6.4 Research Ethics 198
6.4.1 Data Management *198*

6.4.2 Informed Consent Process *199*
6.4.3 Independent HREC approval *202*
6.5 Study E: Investigating Mental Health Nurses' Perceptions About the Future of Mental Health Nursing and their Attitudes Towards Artificial Intelligence 202
6.5.1 Aims *202*
6.6 Methodological Rigour 204
6.6.1 Rigour and Trustworthiness *204*
6.6.2 Credibility, Dependability, and Transferability *207*
6.7 Chapter Summary 210
Chapter 7 Results 211
7.1 Introduction 211
7.1.1 Sociodemographic Factors and Presentation Features (Study A) *211*
7.1.2 Disproportionate Mental Health Presentations Among First Nations People (Study B) *211*
7.1.3 Using Machine Learning in Mental Health Care (Study C) *212*
7.1.4 Machine Learning in Mental Health Care (Study D) *212*
7.1.5 Implications and Future Research *213*
7.2 Sociodemographic Factors and Presentation Features of Individuals Seeking MH Care in EDs: a Retrospective Cohort Study 214
7.2.1 Article Introduction *214*
7.2.2 Article Summary *228*
7.3 Disproportionate Mental Health Presentations to Emergency Departments in a Coastal Regional Community in Australia of First Nation People 229
7.3.1 Article Introduction *230*
7.3.2 Article Summary *240*
7.4 Machine Learning Model Reveals Determinators for Admission to Acute Mental Health Wards from Emergency Department Presentations 241
7.4.1 Article Introduction *242*
7.4.2 Article Summary *260*
7.5 Using Machine Learning to Assist Decision Making in the Assessment of Mental Health Consumers Presenting to Emergency Departments 261
7.5.1 Article Introduction *262*
7.5.2 Article Summary *281*
7.6 Chapter Summary 282
7.6.1 Sociodemographic Factors and Presentation Features *282*
7.6.2 Disproportionate Mental Health Presentations Among First Nations People *282*
7.6.3 Machine Learning and Determinants of Admission to Acute Mental Health Wards *283*
7.6.4 Machine Learning Models to Assist Decision Making in the Assessment of Mental Health Consumers *283*

7.6.5 Implications and Future Directions *284*
Chapter 8 Discussion **285**
8.1 Introduction 285
8.2 AI in Complex Healthcare Environments 286
8.3 Overview of Sociodemographic Factors and Presentation Features 288
8.4 Implications for ED Practices and Policies 290
8.4.1 Triage and Intervention *290*
8.5 Disproportionate Mental Health Presentations Among First Nations People 293
8.5.1 Recommendations for Culturally Responsive Care Models *295*
8.6 Application of Machine Learning in Mental Health Care 296
8.6.1 Analysis of the Performance of ML Models in Predicting Admissions to Inpatient Mental Health Facilities *296*
8.6.2 Importance of Transparency, Interpretability, and Clinician Trust *298*
8.6.3 Ethical Considerations and Potential Biases in ML Applications *299*
8.6.4 Addressing Missed Care Through Machine Learning and Future Research *300*
8.7 Recommendations for Future Research, Clinical Practice and Policy 302
8.7.1 Health Service Systems *302*
8.7.2 Improving First Nations Cultural Safety in ED *303*
8.7.3 Machine Learning Recommendations *304*
8.8 Reflection on Research Limitations 307
8.8.1 Data Constraints and Generalisability *307*
8.8.2 Potential Biases in Data Collection and Analysis *308*
8.8.3 Clinical Validation and Data Quality *308*
8.8.4 Data Type and Complexity *308*
8.8.5 Legal Status and Mental Health Act Implications *309*
8.8.6 Technological and Data Availability Changes *309*
Chapter 9 Conclusion **310**
9.1 Introduction 310
9.2 Addressing the Research Question: Developing and Evaluating ML Models for Mental Health Admissions 311
9.3 Achieving the Aims: Integrating AI and ML into Mental Health Care 313
9.3.1 Complement and Augment Clinicians' Evaluation and Decision Making *313*
9.3.2 Enhance the Client's Experience of Mental Health Services While Providing High-Quality and Safe Care *313*
9.3.3 Understand What is Known About Artificial Intelligence and Machine Learning Adoption in a Mental Health Context *315*
9.4 Addressing Missed Care Through AI-Driven Solutions in Mental Health Care 315

9.5 Summary of Thesis.....317
References.....321
Appendices.....368
Appendix 1 Artificial intelligence: An eye cast towards the mental health nursing horizon: The research on clinical practice and policy.....368
Appendix 2 Editorial: The continued importance of mental health nurses engaging with social media and related emerging technologies.....375
Appendix 3 1st Workshop on Artificial Intelligence in Nursing: Advances, Methods and Path Forward.....378
Appendix 4 Top Cited Article 2022-2023: International Journal of Mental Health Nursing 382
Appendix 5 Awards.....383
Appendix 6 Ethics and Site Specific Approval.....384

List of Figures

Figure 1.1. Project Flow 23

Figure 1.2. The author's grandmother, Patricia Christie, upon graduation from her nursing training on 4 December 1955 from Sydney Sanitarium and Hospital (now known as the Sydney Adventist Hospital). She was first registered on 9 January 1956 29

Figure 1.3. Patricia Christie (centre) as a part of the Gosford District Nursing Service. Gosford Star, Wednesday, 14 May 1969, page 38 30

Figure 2.1. Visual explanation of k-nearest neighbour, an example for clarification purposes (source: Navlani, 2019) 55

Figure 5.1. PRISMA flowchart 149

Figure 6.1. Design thinking flowchart 180

Figure 6.2. Map of Central Coast Local Health District 186

Figure 6.3. Protocol machine learning data flow 197

List of Tables

Table 3.1. Scopus 102

Table 3.2. Web of Science 102

Table 3.3. Google Scholar 102

Table 3.4. Medline 102

Table 3.5. IEEE Xplore 103

Table 3.6. CINAHL 103

Table 3.7. Data 104

Table 5.1. Updated literature review article summary 157

Table 6.1. ED visit table 189

Table 6.2. CMH002 client profiles 190

Glossary and Acronyms

The table below consolidates both the glossary of terms and the list of acronyms into a single reference point, improving readability and minimising redundancy by streamlining key definitions and abbreviations in one accessible format.

Abbreviation	Term	Definition
ACC	Accuracy	Accuracy in machine learning refers to the proportion of correct predictions made by a model, out of the total predictions. It is calculated as the number of true positive and true negative predictions divided by the total number of predictions. Accuracy is commonly used as a performance metric for classification models but may not always be reliable when dealing with imbalanced datasets, where the distribution of classes is uneven. (Online Confusion Matrix, n.d.).
ACN	Australian College of Nurses	A professional organisation representing nurses across Australia, focused on advocacy, professional development, and the promotion of nursing excellence. .
AI	Artificial intelligence	The simulation of human intelligence by machines, enabling them to perform tasks like learning, problem solving, and decision making. AI includes techniques such as machine learning, natural language processing, and computer vision. It is used in various industries, including healthcare, finance, and autonomous systems, to improve efficiency and automate complex tasks (Jiang et al., 2017).
Aifred	Aifred	A mental health AI platform that assists healthcare professionals in the diagnosis and treatment of mental health disorders, such as depression, by using predictive analytics and machine learning (Benrimoh et al., 2020).
AMS	Absolute mean score	A metric used to evaluate the average absolute differences between predicted values and actual values in a dataset, commonly used in regression analysis. (Online Confusion Matrix, n.d.).
AR	Augmented reality	A technology that superimposes computer-generated content, such as images, sounds, or other sensory inputs, onto the real world in real time. AR enhances the user's perception of their environment by adding layers of digital information to what they see through devices like smartphones, tablets, or AR glasses. Unlike virtual reality (VR), which creates an entirely simulated environment, AR integrates the digital and physical worlds, allowing for interactive experiences that combine both. It is widely used in applications ranging from gaming and entertainment to health care, education, and industrial training (Bell et al., 2024; Liu et al., 2023).

AUC-ROC	Area under the ROC curve	An ROC curve shows the trade-off between true positive rate and false positive rate across different decision thresholds (Paper, 2020).
BERT	Bidirectional Encoder Representations from Transformers	A pre-trained natural language processing model developed by Google. It is designed to understand the context of words in a sentence by processing text bidirectionally, meaning it considers both the left and right context of each word. BERT is particularly effective in tasks like question answering, text classification, and language inference because it captures deeper linguistic relationships. Its bidirectional nature enables it to achieve state-of-the-art performance on a wide range of natural language understanding tasks (Devlin et al., 2018).
-	Black box AI	Models created directly from data in such a way that humans cannot understand how the variables are being combined to make predictions (Knight, 2017).
-	CatBoost	A gradient boosting algorithm designed to work well with categorical data, used for classification and regression problems (Paper, 2020).
CAMHS	Child and adolescent mental health service	
CCLHD	Central Coast Local Health District	A regional health organisation providing healthcare services to the Central Coast area in New South Wales, Australia.
CDSS	Clinical decision support system	A health information system that supports decision-making processes in clinical settings by providing healthcare professionals with patient-specific assessments or recommendations to improve care.
ChatGPT	Chat Generative Pre-Trained Transformer	An AI language model developed by OpenAI that uses deep learning to generate human-like text responses based on input. It is designed for natural language understanding and generation, enabling it to engage in conversations, answer questions, and assist with various language-related tasks (Wu et al., 2023).
CINAHL	Cumulated Index in Nursing and Allied Health Literature	A database that provides access to scholarly articles, research studies, and other literature focused on nursing and allied health professions. It covers a wide range of topics related to patient care, healthcare administration, medical ethics, and clinical practices. CINAHL is widely used by healthcare professionals, researchers, and students to access peer-reviewed journals, evidence-based care sheets, and clinical trials relevant to nursing, occupational therapy, physiotherapy, and other allied health fields. .

CL	Consultation liaison	A subspecialty in psychiatry and mental health nursing that involves providing mental health care within general medical settings, supporting the integration of physical and mental health care.
-	Cultural humility	The practice of self-reflection and lifelong learning in health care to better understand and address the cultural contexts of patients, avoiding assumptions and biases to create a respectful and inclusive healthcare environment (Ramsden, 2002).
-	Cultural safety	A framework that ensures that health care is delivered in a way that respects and acknowledges the cultural identities and needs of diverse patient populations, with a focus on minimising power imbalances in healthcare relationships (Congress of Aboriginal and Torres Strait Islander Nurses and Midwives, 2022).
-	Decision tree	A machine learning algorithm used for classification and regression tasks. It models decisions as a tree-like structure, where each internal node represents a feature or attribute, each branch represents a decision rule, and each leaf node represents an outcome or class label. The tree is built by recursively splitting the dataset based on the most significant feature at each step to maximise predictive accuracy. Decision trees are intuitive and easy to interpret but can be prone to overfitting if not properly pruned (Paper, 2020).
DL	Deep learning	A subset of machine learning that uses neural networks with many layers (hence “deep”) to model and understand complex patterns in data. These neural networks mimic the structure of the human brain, with layers of interconnected nodes (neurons) that process inputs and generate outputs. Deep learning excels in tasks like image recognition, speech processing, and natural language understanding, and is used in applications such as autonomous vehicles, health care, and AI-powered virtual assistants (Goodfellow et al., 2016).
DSS	Decision support system	An information system that supports business or organisational decision making by analysing large amounts of data to present insights, simulations or recommendations.
DT	Design thinking	Design thinking is a problem-solving approach that focuses on understanding the needs of users, challenging assumptions, and redefining problems to create innovative solutions. It is a human-centred process that involves five key stages: empathise, define, ideate, prototype, and test. This method is widely used in fields like design, business, and health care to develop user-focused products, services, and systems (Brown, 2009; Liedtka, 2015).

EBM	Explainable boosting machine	An interpretable machine learning model that combines the accuracy of ensemble methods, like gradient boosting, with the transparency of generalised additive models, allowing for clear, human-understandable insights into feature contributions and model predictions (Nori et al., 2019).
ED	Emergency department	(Harris et al., 2024).
EEG	Electroencephalography	(Harris et al., 2024).
EHR	Electronic health records	(Harris et al., 2024).
eMR	Electronic medical record	(Harris et al., 2024).
F1	F1 score	A metric used in machine learning to evaluate the accuracy of a classification model. It is the harmonic mean of precision and recall, providing a balance between the two. The F1 score is particularly useful when dealing with imbalanced datasets, as it considers both false positives and false negatives. It ranges from 0 to 1, where 1 indicates perfect precision and recall (Online Confusion Matrix, n.d.).
FDR	False discovery rate	A metric used in classification tasks to measure the proportion of false positives (incorrectly predicted positive outcomes) among all positive predictions made by a model. In other words, it calculates the likelihood that a positive prediction is actually incorrect (Online Confusion Matrix, n.d.).
	Features	Features refer to individual measurable properties or characteristics used as input variables in data science models. In machine learning, features are critical for model training and prediction (Paper, 2020).
FNR	False negative rate	Also known as the miss rate, this is a metric in classification tasks that measures the proportion of actual positive cases that are incorrectly predicted as negative by the model. It indicates how often the model fails to identify true positive cases (Online Confusion Matrix, n.d.).
FPR	False positive rate	Also known as the fall-out, this is a metric used in classification tasks to measure the proportion of actual negative cases that are incorrectly classified as positive by the model. It indicates how often the model falsely identifies negative cases as positive (Online Confusion Matrix, n.d.).

FTE	Full time equivalent	
GAAIS	General Attitudes towards Artificial Intelligence Scale	A tool designed to measure individuals' perceptions, opinions, and attitudes towards AI technologies. It typically assesses aspects such as trust in AI, perceived benefits, concerns about risks, and the overall acceptance of AI in various domains like health care, education, and everyday life. The scale helps researchers and practitioners understand public sentiment and identify factors that may influence the adoption or rejection of AI technologies (Schepman & Rodway, 2020).
GDMHN	Graduate Diploma of Mental Health Nursing	A postgraduate qualification designed for registered nurses who wish to specialise in mental health. This program equips nurses with advanced knowledge and skills in mental health care, focusing on assessment, treatment, and support of individuals experiencing mental health issues. The diploma prepares nurses to work in various settings, such as hospitals, community mental health services, and private practice, and enhances their ability to provide holistic, evidence-based care.
GPT	Generative pre-trained transformers	A class of deep learning models designed for natural language processing tasks. They are based on the transformer architecture and are pre-trained on vast amounts of text data to generate human-like text. GPT models use unsupervised learning during pre-training, followed by fine-tuning on specific tasks. They excel in tasks like text generation, translation, summarisation, and answering questions. The "generative" aspect refers to their ability to generate coherent and contextually relevant text based on a given prompt (Kaplan et al., 2020).
HDSS	Healthcare decision support system	A technology platform designed to assist healthcare providers in making clinical decisions. It uses data analysis and algorithms to provide recommendations, reminders, or insights based on patient information. HDSS can help improve diagnosis, treatment plans, and overall patient care by offering evidence-based guidance, reducing errors, and enhancing the efficiency of clinical workflows.
HIE	Health Information Exchange	
-	Iatrogenic	Iatrogenic refers to effects, conditions, or complications caused by medical treatment or intervention (Harris et al., 2024).
ICN	International Council of Nurses	

IEEE	Institute of Electrical and Electronics Engineers	
-	Informatics	The study of the structure, behaviour, and interactions of natural and engineered computational systems.
-	InterpretML	Python package for machine learning (Nori et al., 2019).
kNN	k-nearest neighbours	A simple, non-parametric machine learning algorithm used for classification and regression tasks. It works by identifying the k nearest data points (or neighbours) to a given input and predicting the output based on the majority class (in classification) or the average value (in regression) of those neighbours. kNN is intuitive and easy to implement but can be computationally expensive with large datasets, as it requires calculating distances to all points during prediction (Zhang, 2016).
LEAG	Lived Experience Advisory Group	A group or panel composed of individuals who have firsthand experience with specific health or social issues, providing insights and advice based on their personal experiences to inform policy, research or care strategies.
LLM	Large language models	Deep learning models trained on massive amounts of text data to understand, generate, and process natural language. These models, such as GPT and BERT, excel at tasks like text completion, translation, summarisation, and question answering. Due to their size and complexity, LLMs can capture nuanced meanings, context, and relationships within language, making them powerful tools for various natural language processing applications (Paper, 2020).
MAS	Mean absolute score	A statistical metric used to measure the average magnitude of errors between predicted and actual values, without considering their direction. It is calculated as the mean of the absolute differences between predictions and the actual values in a dataset. The MAS is commonly used in regression analysis to assess the accuracy of models. A lower MAS indicates better predictive performance, as it reflects smaller errors on average (Online Confusion Matrix, n.d.).
MBC	Measurement -based care	An approach to healthcare where clinicians use systematic data collection to monitor patient outcomes and adjust treatment plans accordingly to improve care effectiveness (Moggia et al., 2022).
MCC	Matthews correlation coefficient	A performance metric used for evaluating binary classification models. It measures the quality of predictions by considering all four categories of the confusion matrix: true positives, true negatives, false positives, and false negatives. The MCC value ranges from -1 to 1, where 1 represents a perfect prediction, 0 indicates no better than random guessing, and -1 signifies a completely incorrect prediction. The MCC is particularly useful

		for imbalanced datasets, as it provides a more balanced assessment of classification performance than accuracy (Online Confusion Matrix, n.d.).
MeSH	Medical subject headings	
ML	Machine learning	The use of algorithms that identify patterns in data without explicit instructions (Varone et al., 2017).
MMHN	Master of Mental Health Nursing	A postgraduate degree focusing on advanced nursing skills and knowledge in mental health practice, equipping nurses with specialised expertise in mental health care.
MH	Mental health	A state of well-being in which an individual realises their own potential, can cope with normal stresses, work productively, and contribute to their community.
MHN	Mental health nurse	A registered nurse who specialises in the care and treatment of individuals with mental health disorders, providing support, education, and clinical care in various healthcare settings.
MRI	Magnetic resonance imaging	A technique for producing images of bodily organs by measuring the response of the atomic nuclei of body tissues to high-frequency radio waves when placed in a strong magnetic field (Harris et al., 2024).
MRN	Medical record number	
NLP	Natural language processing	A field of artificial intelligence focused on enabling computers to understand, interpret, and generate human language. It combines computational linguistics and machine learning to process and analyse large amounts of natural language data. NLP techniques are used in various applications such as language translation, sentiment analysis, speech recognition, text summarisation, and chatbots. The goal of NLP is to bridge the gap between human communication and computer understanding (Paper, 2020).
NPV	Negative predictive value	(Online Confusion Matrix, n.d.).
OR	Log odds ratio	A statistical measure used to quantify the relationship between two binary outcomes. It represents the logarithm of the odds ratio, where the odds ratio compares the odds of an event occurring in one group to the odds of it occurring in another group. In machine learning and statistics, the log odds ratio is often used in logistic regression to model binary classification tasks. A positive log odds ratio indicates a positive association between the variables,

		while a negative value suggests a negative association (Tabachnick et al., 2018).
-	Patient-centred care	An approach to health care that focuses on the individual needs, preferences, and values of the patient, ensuring they are at the centre of all healthcare decisions.
PCC	Pearson correlation coefficient	The Pearson correlation coefficient measures linear correlation between two sets of data (Online Confusion Matrix, n.d.).
-	Person-centred care	Similar to patient-centred care but emphasises a holistic view of the patient as a person, including their emotional, social, and psychological well-being alongside their physical health.
PPD	Postpartum depression	(Harris et al., 2024).
PRISMA	Preferred Reporting Items for Systematic reviews and Meta-Analyses	(Page et al., 2021).
-	Python	Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. It has high-level built-in data structures, combined with dynamic typing and dynamic binding (Paper, 2020).
-	Reinforcement learning	A type of machine learning where an agent learns to make decisions by receiving rewards or penalties for actions taken, with the goal of maximising cumulative reward in a dynamic environment (Paper, 2020).
RF	Random forest	An ensemble machine learning algorithm used for classification and regression tasks. It operates by constructing multiple decision trees during training and aggregating their predictions to improve accuracy and prevent overfitting. Each tree in the forest is built from a random subset of the data and features, introducing variation that helps reduce variance and improve generalisation. The final prediction is typically made by majority vote in classification or averaging in regression, making random forests robust and effective for complex datasets (Paper, 2020).
RMIT	RMIT University	Royal Melbourne Institute of Technology
ROC	Receiver operating characteristic	A diagnostic test of ML models. An ROC curve is a plot of the sensitivity versus the specificity of a model (Paper, 2020).

ROM	Routine outcome monitoring	The process of systematically tracking patient outcomes over time to evaluate the effectiveness of treatment and inform clinical decision making (Moggia et al., 2022).
	Safe Haven	The Safe Haven provides an alternative to attending the emergency department when experiencing distress in the form of an emotional or suicidal crisis (NSW Health, 2021b).
-	Safety planning	A personalised intervention strategy developed for individuals at risk of self-harm or suicide, outlining steps and resources to reduce the risk of harm and increase safety (Nuji, et al., 2021).
-	Scikit-learn	Python package for machine learning (Paper, 2020).
SDG	Sustainable Development Goals	A set of 17 global objectives established by the United Nations in 2015 as part of the 2030 Agenda for Sustainable Development. The goals aim to address a broad range of global challenges, including poverty, inequality, climate change, environmental degradation, peace, and justice. The goals are interconnected and designed to create a more sustainable, equitable and prosperous world. Some key SDGs include ending poverty, ensuring quality education, promoting gender equality, and taking urgent action to combat climate change (United Nations, 2018).
-	Seaborn	Seaborn is a Python data visualisation library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics (Paper, 2020).
SH	Self-harm	Refers to the deliberate act of inflicting physical harm on oneself, often as a way to cope with emotional distress, trauma, or mental health issues. It may involve behaviours such as cutting, burning, or hitting oneself. Self-harm is not typically intended as a suicide attempt, but it can increase the risk of suicide. It is often a symptom of underlying mental health conditions, such as depression, anxiety, or borderline personality disorder, and requires appropriate psychological support and intervention (Perera et al., 2018).
SHA	Secure hash algorithm	A cryptographic hash function designed to securely encrypt data. SHA is used to maintain data integrity by converting data into a fixed-size hash value, ensuring that any changes to the original data are detectable.

SHAP	SHapley Additive exPlanations	A method used in machine learning to explain the output of predictive models. SHAP values are based on game theory, specifically Shapley values, to fairly distribute the contribution of each feature to a model's prediction. SHAP provides insights into how each input feature influences the model's predictions by assigning a value that indicates the positive or negative impact of each feature on the final outcome. This method is especially useful for interpreting complex models like neural networks and ensemble methods, improving transparency and trust in AI systems (Lundberg & Lee, 2017).
SMI	Severe mental illness	
SPC	Specificity	(Online Confusion Matrix, n.d.).
SQL	Structured Query Language	A programming language for storing and processing information in a relational database. A relational database stores information in tabular form, with rows and columns representing different data attributes and the various relationships between the data values.
-	Stolen Generations	Refers to First Nations Australian children who were forcibly removed from their families by government agencies and church missions between the late 1800s and 1970s, a significant event in Australian history with lasting impacts on Australian First Nations communities (Australian Institute of Health and Welfare, 2021a; Haebich, 2011).
SVC	Support vector classifier	An implementation of a support vector machine (SVM) algorithm that is designed to handle large-scale classification tasks efficiently. SVC works by finding the optimal hyperplane that separates data points of different classes with the maximum margin. SVC is particularly effective in binary classification problems and is well-known for its robustness in non-linear data when combined with kernel functions (Paper, 2020).
SVM	Support Vector Machine	A machine learning algorithm used for classification and regression tasks. It works by finding the optimal hyperplane that best separates data into different classes with the largest possible margin. SVM can handle both linear and non-linear data, using kernel functions to transform the data into a higher-dimensional space when necessary (Paper, 2020).
SWOT	Strengths, weaknesses,	A strategic planning framework used to evaluate these four aspects of a business, project, or initiative. Strengths are internal factors that provide an advantage. Weaknesses are internal limitations or areas needing improvement. Opportunities are

	opportunities, and threats	external factors that could be leveraged for success. Threats are external challenges that could hinder success. SWOT analysis helps organisations and individuals assess their current situation and develop strategies to maximise strengths and opportunities while addressing weaknesses and threats.
T5X	Text-To-Text Transfer Transformer	A natural language processing (NLP) model developed by Google that treats every NLP task as a text-to-text problem. This means that inputs and outputs for tasks such as translation, summarisation, and question answering are all represented as text. T5 is based on the Transformer architecture and is pre-trained on large amounts of data, making it highly versatile and effective for a wide range of NLP tasks. Its approach simplifies the framework for various NLP applications, allowing it to generalise well across different tasks using the same model architecture (Raffel et al., 2020; Roberts et al., 2023).
TabNet	TabNet	TabNet is a deep tabular data learning architecture that uses sequential attention to choose which features to reason from at each decision step (Yang et al., 2024).
TEN	Trainee Enrolled Nurse	
	Towards Zero	An initiative in New South Wales, Australia aimed at reducing suicide-related deaths to zero. The program promotes suicide prevention activities and initiatives such as the Safe Haven program (NSW Health, 2020).
TPR		True positive rate (Online Confusion Matrix, n.d.).
TTN	Trier Treatment Navigator	A clinical decision support tool designed to assist healthcare professionals in selecting and optimising treatments for mental health conditions, particularly depression. Developed by researchers at the University of Trier in Germany, TTN uses a data-driven approach to recommend personalised treatment plans based on patient characteristics, clinical history, and research evidence. The tool helps guide treatment choices, including psychotherapy, pharmacotherapy, or a combination of both, and is aimed at improving patient outcomes through tailored interventions (Lutz et al., 2021).
UoN	University of Newcastle	
-	Unsupervised learning	A type of machine learning that uses unlabelled data to find patterns and relationships within a dataset without explicit instructions, often used for clustering or association problems (Paper, 2020).

VC	Virtual care	Refers to the delivery of healthcare services remotely through digital technologies, such as video consultations, phone calls, mobile apps, and other online platforms. It allows patients to access medical advice, diagnosis, treatment, and monitoring without needing to visit healthcare facilities in person. Virtual care is used across various healthcare domains, including routine check-ups, mental health support, chronic disease management, and follow-up care. It enhances accessibility, convenience, and continuity of care, particularly for patients in remote areas or those with mobility challenges (Bell et al., 2024).
VH	Virtual health	A broad term that encompasses the use of digital technologies and online platforms to deliver healthcare services and manage health remotely. It includes various forms of telemedicine, remote monitoring, virtual consultations, and digital health tools, allowing patients to connect with healthcare providers from anywhere. Virtual health aims to improve access to care, enhance patient engagement, and support ongoing health management outside traditional clinical settings, offering flexibility and convenience for both patients and healthcare providers. It is particularly valuable for chronic disease management, mental health services, and preventive care (Bell et al., 2024).
VR	Virtual reality	A technology that creates a simulated, immersive environment, allowing users to interact with a three-dimensional virtual world. VR typically involves wearing a headset with screens and sensors that track head movements, providing a fully immersive experience by blocking out the physical world. In health care, VR is used for applications such as medical training, therapy for mental health conditions (e.g. exposure therapy), pain management, and rehabilitation. Outside of health care, VR is widely used in gaming, education, and simulations for various industries (Bell et al., 2024).
	Way Back	An aftercare service designed to support individuals who have attempted suicide or experienced a suicidal crisis. The program provides tailored support and care to help individuals recover and prevent further crises (Everymind, 2022).
WHO	World Health Organization	
Woebot	Woebot	A mental health chatbot (Darcy, 2023).
XAI	Explainable AI	Refers to AI systems that provide transparent, interpretable explanations of their decision-making processes. XAI aims to make AI more understandable and trustworthy for users, especially in high-stakes fields like health care (Molnar, 2020).
XGB	XGBoost	A machine learning algorithm based on the gradient boosting framework. It is widely used for both classification and regression tasks due to its speed, efficiency, and high predictive accuracy. XGBoost builds an ensemble of decision trees by iteratively

		adding new trees to correct the errors made by previous ones, optimising performance through gradient descent. It also includes features like regularisation, early stopping, and parallel processing, making it effective for large datasets and complex models (Paper, 2020).

Rationale for Terminology in Mental Health Discourse

Language in mental health care is complex and evolving, with different terms used to refer to individuals receiving care depending on context, professional discipline, and philosophical perspective. In this thesis, the terms patient, consumer, client, and individuals with lived experience appear in different contexts. The use of these terms is not intended to exclude or diminish any particular perspective but rather to reflect the diversity of language used across mental health settings (Costa et al, 2019).

The term patient is used predominantly throughout this thesis, as it remains the most commonly used term in clinical settings. This choice reflects the professional discourse in which the study is embedded and aligns with conventional terminology in mental health nursing and healthcare settings. However, it is recognised that alternative terms are increasingly preferred in recovery-oriented frameworks and advocacy contexts (Costa et al, 20109).

- Consumer is often used in policy and advocacy settings to emphasise agency and choice in mental health care.
- Client is frequently employed in community mental health and private practice settings.
- Individuals with lived experience is a term that centres the expertise of those who have personal experience of mental health challenges, particularly in co-design and peer support contexts.

While patient is the primary term used in this thesis for consistency, alternative terms are included where contextually relevant, such as in policy discussions, direct quotations, or references to specific frameworks. The inclusion of multiple terms acknowledges the diversity of language used in mental health care and research.

To support readers from a wide range of backgrounds, all key terms used in the thesis are defined in the Acronyms and Glossary section at the beginning of the document. This ensures clarity and accessibility while demonstrating an awareness of the broader discourse on contemporary mental health language.

Patient Outcomes

In this thesis, patient outcomes refer to the measurable results of healthcare interventions and service interactions within the context of emergency department (ED) presentations for mental health concerns. Specifically, this term encompasses factors such as the time taken from ED arrival to discharge or transfer, the appropriateness of discharge destinations (e.g., transfer to a mental health unit or referral to community services), and overall service efficiency. While patient outcomes can also refer to clinical improvements in health status following treatment, in this study, the focus is primarily on system-level and service-related outcomes rather than individual clinical recovery (Austin et al, 2024).

Mental Health Distress, Mental Health Challenges, and Mental Health Concerns or Problems

In this thesis, the terms mental health distress, mental health challenges, and mental health concerns or problems are used interchangeably to describe the experiences of individuals presenting to the ED with mental health-related needs. These terms encompass a range of psychological, emotional, and psychiatric experiences that prompt ED attendance, including

acute distress, situational crises, and conditions requiring assessment or intervention. The interchangeable use of the defined terms reflects the varied nature of mental health presentations in ED settings while maintaining alignment with contemporary discourse.

Abstract

Artificial intelligence (AI) and machine learning (ML) have the potential to unlock a deeper understanding of the human experience in health settings. This research investigated the integration of AI and ML into mental health care, specifically focusing on their application within emergency departments (EDs). EDs serve as critical access points for mental health services, yet they face significant challenges, including overcrowding, long wait times, and a scarcity of specialised resources. The study explored how AI and ML could be harnessed to address these challenges, improve clinical decision making, and enhance patient outcomes. The thesis begins by framing the problem within the broader context of mental health care, highlighting the significance of the study. It underscores the importance of the Sustainable Development Goals, ethical AI use, and the inclusion of lived experience in shaping mental healthcare strategies. This foundational approach emphasises the need for an integrated model of care that balances technological innovation with ethical considerations and human-centred care. The research explored the role of technology in mental health, particularly its impact on individuals experiencing conditions such as suicidality. The dual nature of technology was examined, recognising its capacity to drive innovation and empowerment while also acknowledging the potential for technology to induce anxiety, fear, or paranoia among vulnerable populations.

This exploration underscores the need for a careful and ethical approach to the integration of AI in mental health care, considering the broader implications for patients and clinicians alike. A comprehensive literature review examines the concept of trust in AI, particularly in the context of health care. The review challenges the assumption that AI systems are inherently trustworthy and explores the anthropomorphisation of AI, questioning whether these systems should be deemed deserving of trust. The study also highlights the rapid advancements in AI

and ML, particularly through the development of large language models. These models have significantly expanded the capabilities of natural language processing and offer considerable potential for healthcare applications.

However, the research identified ethical challenges associated with these technologies, including bias, privacy concerns, and the need for transparency in AI deployment. The thesis advocates the responsible development of AI systems that enhance rather than replace human clinical judgement. The methodological framework employed in this research was designed to ensure a thorough understanding of the needs and concerns of various stakeholders, including clinicians, patients, and interdisciplinary experts. The study utilised design thinking, coupled with a retrospective cohort analysis of sociodemographic factors and presentation features of individuals seeking mental health care in EDs. Ethical considerations were prioritised throughout the research, with a particular focus on cultural sensitivity and the inclusion of diverse perspectives.

The findings of the research provide a detailed analysis of the complexities involved in mental health care delivery in EDs. This research underscores the essential role of EDs in providing frontline mental health services while also identifying significant challenges, such as the need for more specialised resources and the ethical integration of AI into clinical practice. The research demonstrates the potential of ML to improve service provision, particularly through predictive modelling and clinical decision support systems. However, it also highlights limitations related to data quality, trust, and ethical considerations, emphasising the importance of a cautious and responsible approach to AI integration.

In conclusion, the thesis discusses the broader implications of the research findings for clinical practice and policy. It offers recommendations for future research, including the evaluation of safety planning interventions, the exploration of collaborative care models, and the utilisation

of ML for enhanced predictive modelling in mental health settings. The research advocates a comprehensive and integrated approach to mental health care that leverages the potential of AI while maintaining a strong ethical framework. Overall, the research contributes to the growing body of knowledge on the role of AI and ML in mental health care. It highlights the need for trust, transparency, and ethical considerations in the development of AI-based tools, ensuring that these technologies complement human clinical judgement rather than replace it. The findings suggest that, with careful and ethical deployment, AI has the potential to significantly improve mental healthcare delivery, leading to more effective, equitable, and patient-centred services.

Preface

This thesis is comprised of multiple chapters that contain manuscripts that have been published in peer-reviewed journals. Due to this approach, some content may overlap, as the articles have been published across various outlets, with consistent background information presented throughout these publications. A general introduction to the thesis and each chapter is provided, establishing the context, justification, and rationale for the research project.

Chapter 1: Introduction

This chapter introduces the key issues in delivering mental health (MH) care in emergency departments (EDs), focusing on the potential role of artificial intelligence (AI) and machine learning (ML) in addressing these challenges. It outlines the study's aims and significance, emphasising the need for innovative solutions to enhance the efficiency and effectiveness of MH services. The chapter also highlights ethical considerations, the inclusion of lived experiences, and the importance of improving clinical decision making and patient outcomes through AI and ML.

Chapter 2: Background

This chapter provides foundational context for the research, exploring the transformative role of AI and ML in MH care. A portion of this discussion is based on the published journal article by Higgins, Chalup, et al. (2023b), which focuses on both the opportunities and challenges these technologies present, particularly for individuals experiencing suicidality. Additionally, the chapter incorporates content from the article by Higgins & Wilson (2023), which examines the influence of commercial determinants in shaping AI development and their implications for patient-centred care. Together, these publications underscore the need for ethical, safe, and transparent AI deployment to improve the client's experience within MH services.

Chapter 3: Critical Review of the Literature

This chapter critically reviews the existing literature on AI- and ML-based clinical decision support systems (CDSS) in MH care. The chapter includes content from the published article by Higgins, Short et al. (2023), which focuses on the roles of trust, transparency, and clinician engagement in the successful implementation of these technologies. This article also identifies specific challenges, including algorithmic bias, accountability, and the risks of over-reliance on technology, and offers recommendations for developing AI/ML systems that align with clinical workflows and maintain human oversight.

Chapter 4: Theoretical Foundations of Trust

This chapter investigates the theoretical dimensions of trust in the context of AI adoption in MH care, incorporating insights from the published article by Higgins, Chalup et al. (2024a). This article argues that AI should be viewed as a clinical tool, requiring rigorous evaluation of its reliability and validity. The chapter draws from this publication to emphasise transparency and interpretability as essential factors in fostering clinician trust, ensuring that AI systems support rather than replace human decision-making. These insights are critical to achieving the study's goals of enhancing clinician decision-making and improving patient care.

Chapter 5: Update to Literature Review

Revisiting the original literature review, this chapter incorporates recent advancements in AI, particularly the development of large language models (LLMs). It discusses how these innovations have influenced health care, with a focus on improving clinical decision making and reducing missed care. The chapter also addresses the ethical implications of LLMs, such as transparency and trust, reaffirming the need to balance AI innovations with workforce improvements to maintain high-quality care.

Chapter 6: Methods

This chapter outlines the methodological approach used to develop, test, and evaluate ML models for predicting MH admissions from ED presentations. The use of design thinking (DT) and retrospective cohort methods integrated sociodemographic data and ML techniques to provide a comprehensive framework for understanding ED presentations and admissions. Ethical considerations, including cultural sensitivity, were central to the methodology, ensuring adherence to best practices in healthcare research.

Chapter 7: Results

The results chapter presents the findings of this research on mental MH care delivery in EDs, focusing on sociodemographic factors, presentation characteristics, and the application of ML models to enhance service provision. Key factors influencing admissions, such as suicidal ideation and triage level, are highlighted, and the chapter examines how ML can improve clinical decision-making, reduce overcrowding, and enhance patient outcomes.

The first publication included in this chapter, Higgins, Sheather-Reid et al. (2024c), identifies crucial sociodemographic factors and presentation characteristics impacting mental health admissions in EDs, with findings on age, triage category, and presenting problem, particularly highlighting suicidal ideation as a major driver of admissions. The second publication, Higgins, Sheather-Reid et al. (2024a), addresses the over-representation of Aboriginal and Torres Strait Islander populations in MH-related ED presentations, underscoring the need for culturally responsive care models and targeted interventions to address this disparity. The third article, Higgins, Chalup et al. (2024b), applies ML techniques to ED presentation data, uncovering significant patterns related to age, triage level, and prior healthcare utilisation that influence admissions to mental health inpatient facilities. These findings emphasise the potential of ML models to support clinical decision-making and improve resource allocation in EDs.

The final publication in this chapter, Higgins et al. (2024d), reports on the testing and evaluation of several ML models, including explainable boosting machines (EBMs) and support vector machines (SVMs), to predict admissions accurately. This study highlights the importance of transparency and clinician trust in the successful integration of these models into clinical workflows.

Chapter 8: Discussion

This chapter synthesises the research findings, demonstrating how ML models can be used in the prediction of admissions to MH inpatient wards from ED presentation data. It discusses the importance of transparency and clinician trust in AI integration and outlines future research directions, emphasising the ongoing need to explore AI's role in improving MH services while maintaining ethical standards and clinician engagement.

Chapter 9: Conclusion

The concluding chapter summarises the thesis' contributions to MH care, particularly in integrating AI and ML into ED workflows. It highlights the importance of trust, transparency, and clinicians' involvement in AI tool development. The chapter calls for continued research to ensure AI enhances, rather than replaces, human clinical judgement and advocates the responsible and ethical use of AI technologies in MH care.

This organisation ensures a logical progression from identifying the research problem to discussing the findings and their implications, providing a thorough exploration of AI's role in MH care.

Complete List of all Publications from this Research

Higgins, O., Chalup, S. K., Short, B. L., & Wilson, R. L. (2023b). Interpretations of innovation: The role of technology in explanation seeking related to psychosis. *Perspectives in Psychiatric Care, 2023*, Article 464934. https://doi.org/10.1155/2023/4464934

Higgins, O., Short, B. L., Chalup, S. K., & Wilson, R. L. (2023). Artificial intelligence (AI) and machine learning (ML) based decision support systems in mental health: An integrative review. *International Journal of Mental Health Nursing*, *32*(4), 966-978. https://doi.org/10.1111/inm.13114

Higgins, O., Chalup, S. K., & Wilson, R. L. (2024a). Artificial intelligence in nursing: Trustworthy or reliable? *Journal of Research in Nursing*, *29*(2), 143-153. https://doi.org/10.1177/17449871231215696

Higgins, O., & Wilson, R. L. (2023). Commercial determinants and therapeutic chatbots: A mental health nursing perspective. *International Journal of Mental Health Nursing*, *32*(6), 1509-1511.

Higgins, O., Sheather-Reid, R. B., Chalup, S. K., & Wilson, R. L. (2024c). Sociodemographic factors and presentation features of individuals seeking mental health care in emergency departments: A retrospective cohort study. *International Journal of Mental Health Nursing*. Advance online publication. https://doi.org/10.1111/inm.13414

Higgins, O., Sheather-Reid, R. B., Chalup, S. K., & Wilson, R. L. (2024a). Disproportionate mental health presentations to emergency departments in a coastal regional community in Australia of First Nation people. *International Journal of Mental Health Nursing*. Advance online publication. https://doi.org/10.1111/inm.13362

Higgins, O., Chalup, S. K., & Wilson, R. L. (2024b). Machine learning model reveals determinators for admission to acute mental health wards from emergency department presentations. *International Journal of Mental Health Nursing*. Advance online publication. https://doi.org/10.1111/inm.13402

Higgins, O., Wilson, R. L., & Chalup, S. K. (2024d). Using machine learning to assist decision making in the assessment of mental health consumers presenting to emergency departments. *Digital Health*. https://doi.org/10.1177/20552076241287364

Higgins, O., & Wilson, R. L., (2025). Integrating Artificial Intelligence (AI) with Workforce Solutions for Sustainable Care: A follow up to Artificial intelligence and machine learning (ML) based decision support systems in mental health. *International Journal of Mental Health Nursing*. Accepted (14/02/2025)

Higgins, O., Short, B. L., Chalup, S. K., & Wilson, R. L. (2022). Artificial intelligence and machine learning based decision support systems in mental health. *Prospero,* CRD42022352360. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42022352360

Complete List of all Conference Presentations from this Research

Higgins, O., Chalup, S. K., Short, B. L., & Wilson, R. L. (2021a). *Disruptions of The Familiar: Psychosis Echoes Technological Advancement* [Paper Presentation] TheMHS 2021 Hope into Action, Melbourne.

Higgins, O., Chalup, S. K., Short, B. L., & Wilson, R. L. (2021). *Interpretations of Innovation: The intersection of Technological Advancement and Psychosis* [Paper Presentation] University of Newcastle School of Nursing and Midwifery Research Symposium, Newcastle.

Higgins, O., Chalup, S. K., Short, B. L., & Wilson, R. L. (2022). *Interpretations of Innovation: The intersection of technological advancement and psychosis* [Paper Presentation] ACMHN's 46th International Mental Health Nursing Conference, https://onlinelibrary.wiley.com/doi/abs/10.1111/inm.13050

Higgins, O., Chalup, S. K., Short, B. L., & Wilson, R. L. (2022). *Interpretations of Innovation: The intersection of technological advancement and psychosis* [Paper Presentation] Research & Innovation Symposium: Central Coast Local Health District, Gosford.

Higgins, O., Chalup, S. K., Short, B. L., & Wilson, R. L. (2022). *Interpretations of Innovation: The intersection of technological advancement and psychosis* [Paper Presentation] International Association for Youth Mental Health, Copenhagen, Denmark.

Higgins, O., Chalup, S. K., Short, B. L., & Wilson, R. L. (2023). Artificial intelligence in nursing: trustworthy or reliable? [Paper Presentation] *ACMHN's 47th International Mental Health Nursing Conference, Mental Health Nursing*, https://doi.org/https://doi.org/10.1111/inm.13209

Higgins, O., Chalup, S. K., & Wilson, R. L. (2022, 17/06/2022). *Artificial Intelligence in Nursing: trustworthy or reliable?* [Paper Presentation] Forensic Mental Health Research Symposium, Middelfart, Denmark.

Higgins, O., Chalup, S. K., & Wilson, R. L. (2022, 17/06/2022). *Artificial Intelligence in Nursing: trustworthy or reliable?* [Paper Presentation] 1st Workshop on Artificial Intelligence in Nursing: Advances, Methods and Path Forward, Halifax, Canada.

Higgins, O., Chalup, S. K., & Wilson, R. L. (2024). Machine learning reveals determinators for admission to Acute Mental Health Wards from Emergency Department presentations. [Paper Presentation] *ACMHN's 48th International Mental Health Nursing Conference*, https://onlinelibrary.wiley.com/doi/abs/10.1111/inm.13209

Higgins, O., Sheather-Reid, R. B., Chalup, S. K., & Wilson, R. L. (2024). Disproportionate mental health presentations to Emergency Departments in coastal regional Australia of First Nations people. [Poster Presentation] *ACMHN's 48th International Mental Health Nursing Conference*, https://onlinelibrary.wiley.com/doi/abs/10.1111/inm.13209

Higgins, O., & Wilson, R. L. (2024). Digital innovations targeting missed mental health care. [Paper Presentation] *ACMHN's 48th International Mental Health Nursing Conference,* https://onlinelibrary.wiley.com/doi/abs/10.1111/inm.13209

Higgins, O., Sheather-Reid, R. B., Chalup, S. K., & Wilson, R. L. (2025, 13/02/25). Disproportionate mental health presentations to Emergency Departments in coastal regional Australia of First Nations people. [Paper Presentation] *Central Coast Local Health District Research and Innovation Symposium 2025.*

Higgins, O., Chalup, S. K., & Wilson, R. L. (2022, 13/02/25). Artificial Intelligence in Nursing: trustworthy or reliable? [Paper Presentation] *Central Coast Local Health District Research and Innovation Symposium 2025*.

Higgins, O., Chalup, S. K., & Wilson, R. L. (2024, 13/02/25). Machine learning reveals determinators for admission to Acute Mental Health Wards from Emergency Department presentations. [Paper Presentation] *Central Coast Local Health District Research and Innovation Symposium 2025.*

Chapter 1 Introduction

1.1 Introduction

The gap between supply and demand in health care continues to expand at an increasing rate with 2020–21 bringing 8.8 million presentations to emergency departments (ED) in Australian public hospitals (Australian Institute of Health and Welfare, 2021b). Patients are staying longer in EDs, with fewer visits completed within four hours (Australian Institute of Health and Welfare, 2021b). The ED is a noisy and unsettling environment, and for those who present with mental health (MH) issues, it can be very distressing (Clarke et al., 2007). MH presentations sometimes do not align with the typical profile of ED consumers and are thus perceived by ED staff to disrupt normal healthcare flow (Knaak et al., 2017). They often present with vague or difficult to diagnose symptomatology and require assessment by specialist MH clinicians (Fleet et al., 2003). In addition, the stigma associated with seeking help in the ED is well-documented (Knaak et al., 2017; Sukhera et al., 2017) and can negatively impact patient care and outcomes. This stigma can originate from multiple sources, including the patients themselves, who may feel ashamed or embarrassed to seek help (Corrigan et al., 2014), as well as from healthcare providers, who might inadvertently perpetuate stigma through negative attitudes or misconceptions about MH conditions (Thornicroft et al., 2016). Such stigma can lead to inadequate care, delayed treatment, and poorer outcomes for patients who are already vulnerable (Henderson et al., 2013).

1.2 The Aims of the Research

The aims of this research were to prepare, test, and evaluate the use of machine learning (ML) models to predict whether clients presenting to an ED with MH concerns or problems, particularly suicidality, will require admission to an acute MH inpatient ward. These models provide insights and data that can:

- Complement and augment clinicians' evaluation and decision making
- Enhance the client's experience of mental health services whilst providing high quality and safe care
- Build understanding of artificial intelligence and machine learning adoption in a mental health context.

This research provided an opportunity to test an approach to utilising real-time health informatics data to reduce the time MH clients spend in EDs. The research explored the application of ML to predict hospital admissions to a MH facility and to support the clinician's decision making. Expediting admissions for clients identified by the model may reduce ED wait times for MH clients and ease the clinical load on ED staff, potentially reducing the reliance on the MH Admitted Emergency Treatment Protocol (NSW Health, 2021a), increasing the number of visits completed within four hours (Australian Institute of Health and Welfare, 2021b), and commencing preparation for admission earlier. For those not to be admitted, alternative pathways can be initiated faster, for example, Way Back program (an aftercare service for people who have attempted suicide or experienced a suicidal crisis) (Everymind, 2022), or the MH Acute Care Team (Central Coast Local Health District, 2022).

This research sought to complement and augment clinical evaluation by investigating and using AI/ML-derived decision support systems. It has the potential to reduce the length of time that clients wait in the ED by providing treatment recommendations. This reduction in volume has the potential to improve bed management and workflow for clinical staff. It will enhance the client's experience of MH services by expanding treatment options and minimising the time they spend in the ED. Additionally, the research offers understanding and insight into what is currently known about the adoption and use of AI/ML in the MH setting and the role it can

play in providing support to complex clinical decision making and assisting with service delivery planning.

1.3 Significance of the Research

This study holds significant importance in the evolving field of MH care, particularly in the context of EDs, where the demand for timely and effective MH services is critically high (Australian Institute of Health and Welfare, 2021b). While EDs primarily function as sites for assessment and triage rather than ongoing mental health treatment, they increasingly serve as the first point of contact for individuals experiencing acute MH crises. Consequently, the ability to provide timely and appropriate initial care, including risk assessment, crisis intervention, and referral pathways, is crucial. The challenges of overcrowding, long wait times, and the shortage of specialised resources have become more pronounced (Australian Institute of Health and Welfare, 2021b), underscoring the need for innovative solutions that enhance the efficiency and effectiveness of MH crisis management in ED settings. These challenges underscore the need for innovative solutions that can enhance the efficiency and effectiveness of MH care delivery. One of the key contributions of this study is its exploration of the potential role of AI and ML in addressing these challenges. By investigating how AI and ML can be integrated into ED workflows, this research provides valuable insights into how these technologies can support clinical decision making, improve patient outcomes, and optimise resource allocation. The study's focus on AI/ML-driven predictive modelling and decision support systems represents a critical advancement in the development of tools that can assist clinicians in making more informed, data-driven decisions under the pressure of emergency care settings (Topol, 2019a).

Furthermore, this study is significant in its emphasis on the ethical implications of AI integration in MH care. The research highlights the importance of trust, transparency, and cultural sensitivity in the deployment of AI technologies (Floridi & Cowls, 2022). By

addressing these ethical considerations, the study contributes to the ongoing discourse on responsible AI development, ensuring that technological innovations in health care are aligned with the principles of patient-centred care and equity. The focus on ethical AI use is particularly relevant given the growing concerns about bias, privacy, and the potential for AI systems to perpetuate existing disparities in health care (Celi et al., 2022). Additionally, this research bridges a critical gap in the literature by examining the perceptions of MH professionals towards AI and its role in their practice. Understanding the attitudes and concerns of these key stakeholders is essential for the successful implementation of AI in clinical settings. The findings of this research can inform the development of training programs and guidelines that prepare clinicians to effectively and ethically use AI tools in their practice. The significance of this study extends beyond the immediate context of EDs to the broader healthcare system. As AI and ML technologies continue to advance, their integration into various aspects of health care is inevitable. This research provides a foundational understanding of how these technologies can be harnessed in MH care, offering a model that can be adapted and applied in other healthcare settings. The research's insights into the challenges and opportunities of AI integration have the potential to influence policy decisions, guiding the development of regulatory frameworks that ensure the safe and effective use of AI in health care (Floridi & Cowls, 2022). This research is significant in its contribution to understanding both the practical and ethical dimensions of AI integration in MH care. It provides a pathway for improving the delivery of MH services in EDs, while also contributing to the broader discussion on the role of AI in health care. By addressing the critical challenges facing MH care today and exploring the potential of emerging technologies, this research paves the way for future innovations that can enhance patient care and promote more equitable health outcomes.

1.3.1 Missed and Unfinished Care

The following presents the main factors contributing to increased ED presentations and their impact on healthcare delivery, particularly in the context of missed care. In healthcare systems globally, ED presentations have steadily increased, placing immense pressure on hospitals and healthcare providers. In Australia, for instance, emergency presentations rose consistently for five years prior to 2019–20, with longer patient stays in the ED and fewer visits completed within the four-hour benchmark (Australian Institute of Health and Welfare, 2021b). This trend mirrors developments worldwide, where rising patient demand is met with overstretched resources and heightened staff workloads (Mudallal et al., 2017). The surge in patient presentations exacerbates existing workforce challenges, particularly for nurses, who bear the brunt of this increasing pressure.

1.3.1.1 Clinician Burnout

Globally, the escalating demand for healthcare services often leads to clinician burnout, characterised by emotional exhaustion, depersonalisation, and a diminished sense of personal accomplishment (Mudallal et al., 2017). Burnout is particularly pervasive in the nursing profession, as nurses typically engage in prolonged direct patient care under highly stressful conditions. International literature supports the view that nurse burnout is strongly correlated with reduced productivity, inefficiency, and adverse patient outcomes. International studies have highlighted the adverse impacts of burnout, including increased absenteeism, higher turnover rates, and decreased job satisfaction among nurses (Khamisa et al., 2015; L. Wang et al., 2018). The broader implications for healthcare systems include compromised care quality and the erosion of patient trust, as nurses struggle to manage growing workloads with fewer resources.

1.3.1.2 Missed or Unfinished Care from Burnout

One of the most significant consequences of this burnout and heightened demand is the phenomenon of unfinished or missed care, which has emerged as a critical issue in acute care hospitals around the world (Jones et al., 2015). Missed care refers to any aspect of required patient care that is delayed, omitted, or incomplete (Jones et al., 2015). This issue is not limited to any one geographic region; it is a global concern, with studies from countries such as the United Kingdom, the United States, and Sweden demonstrating the prevalence of missed care and its detrimental effects on patient outcomes (Aiken et al., 2017; Ball et al., 2014; Kalisch et al., 2013). In the UK, for example, missed nursing care has been linked to increased patient mortality, longer hospital stays, and higher readmission rates (Griffiths et al., 2019). Similar trends are observed in the United States, where nurses report being unable to complete critical care tasks, leading to an increased risk of complications and patient dissatisfaction (Kalisch et al., 2012). In Sweden, research has identified that nurses' prioritisation strategies often lead to non-urgent care needs, particularly those related to emotional and psychological support, being neglected (Ball et al., 2018). These studies highlight a troubling global pattern: as the demands on nursing staff increase, essential elements of patient care are being missed, leading to adverse outcomes for both patients and healthcare systems at large.

1.3.1.3 Reasons for Missed or Unfinished Care

Missed care often arises from a combination of factors, including insufficient staffing, high patient-to-nurse ratios, and the growing complexity of patient cases (Ball et al., 2018). Nurses, particularly in acute care settings, are frequently required to prioritise their tasks, focusing on clinical and life-saving interventions at the expense of patients' emotional, educational, and psychological needs. This issue is compounded by administrative burdens and the need to meet performance targets, which can further detract from nurses' ability to provide holistic care,

creating an environment where care is missed or unfinished (Roche et al., 2016). Prioritisation of tasks is necessary, but it often results in a form of rationed care, where certain tasks are delayed or omitted altogether, depending on their perceived urgency (Jones et al., 2015).

1.3.1.4 Toll of Missed Care for Nursing Staff

International studies also demonstrate the emotional toll of missed care on nurses, who may experience moral distress as a result of being unable to fulfil their professional obligations to patients (Ball et al., 2014) and can contribute to an environment where care remains missed or unfinished, despite nurses' best efforts(Roche et al., 2016). The moral distress that arises from unfinished care can exacerbate feelings of burnout, creating a cycle in which nurses become increasingly disengaged and emotionally exhausted, leading to further inefficiencies and missed care (Mudallal et al., 2017). Moreover, missed care has been linked to higher rates of nurse attrition, as many nurses leave the profession due to the overwhelming demands and their inability to provide the quality of care they deem necessary (Aiken et al., 2017).

1.3.1.5 Interventions and Solutions to Address Missed or Unfinished Care

The international scope of missed care and nurse burnout calls for systemic interventions to address these pervasive issues. Solutions must encompass a broad range of strategies, including improving staffing levels, providing better support for nurses, and rethinking the allocation of resources to ensure that both clinical and non-clinical needs are met. For instance, research from the Netherlands suggests that the implementation of advanced technologies, such as electronic health records, automated scheduling systems (Zúñiga et al., 2015), and AI automation of documentation, scheduling, and medication management (Seibert et al., 2021) can alleviate some of nurses' administrative burdens, allowing them to focus more on direct patient care.

Ultimately, addressing the global issue of missed care requires a comprehensive approach that acknowledges the interrelated nature of nurse burnout, healthcare demand, and patient care outcomes. By addressing the root causes of missed care, such as staffing shortages, inadequate resources, and systemic inefficiencies, healthcare systems worldwide can begin to mitigate the negative impacts on both nurses and patients.

1.3.2 Artificial Intelligence and Missed Care

Integrating AI into healthcare systems can potentially address many issues surrounding missed or unfinished care. As discussed, missed care often arises from nurses being overburdened with high patient loads, complex cases, and administrative duties, leading to an inability to address all patient needs adequately (Jones et al., 2015). AI technologies can help mitigate these challenges by enhancing workflow efficiency, optimising resource allocation, and supporting clinical decision making, reducing missed care incidence (Seibert et al., 2021).

1.3.2.1 Artificial Intelligence for Optimising Nursing Workflows

One of the primary applications of AI in addressing missed care is through optimising nursing workflows and task prioritisation. ML algorithms, for instance, can analyse patient data in real time, providing recommendations to nurses on which tasks should be prioritised based on clinical urgency and resource availability (Topaz & Pruinelli, 2017). Such systems can help nurses manage their workload more effectively, reducing the cognitive burden associated with decision making and task delegation. AI-driven tools like predictive analytics can alert staff to potential patient deterioration or identify patterns that may indicate the need for immediate intervention, ensuring that critical aspects of care are not overlooked (Seibert et al., 2021). Moreover, AI can be used to automate routine tasks such as documentation, scheduling, and medication management, freeing up nurses' time for direct patient care (Seibert et al., 2021). For example, natural language processing systems can assist in automating patient records,

allowing nurses to spend less time on administrative tasks and more time attending to patients' physical, emotional, and psychological needs (Seibert et al., 2021). This redistribution of tasks could help address the problem of missed care, particularly in areas like patient education and provision of emotional support, which are often neglected when workloads are high (Jones et al., 2015).

1.3.2.2 Artificial Intelligence Assisted Decision Support Systems

Clinical decision support systems (CDSS) powered by AI are another promising tool to reduce missed care by supporting nurses and other healthcare professionals in making evidence-based decisions quickly and efficiently (Bates et al., 2014). These systems can assist nurses by providing real-time, data-driven insights into patient conditions and suggesting the most appropriate interventions based on historical data and clinical guidelines. By providing timely alerts and recommendations, AI systems help ensure that critical care decisions are made promptly, reducing the likelihood of care delays or omissions.

In acute care settings, where nurses are often required to juggle multiple patients with varying degrees of complexity, CDSS can assist in identifying the highest-risk patients, ensuring that the most urgent care needs are addressed first. For example, AI models trained on large datasets of patient outcomes can predict which patients are at higher risk of adverse events, such as falls, pressure ulcers, or infections, allowing nurses to take pre-emptive action (Seibert et al., 2021). These predictive capabilities help nurses prioritise interventions that are most likely to prevent complications, thereby reducing the chances of missed care.

1.3.2.3 Artificial Intelligence for Patient Engagement and Monitoring

Missed care often extends to the emotional, psychological, and educational needs of patients, as nurses prioritise urgent clinical tasks over these non-urgent, yet important, aspects of care

(Jones et al., 2015). AI technologies such as chatbots, virtual reality, and virtual health assistants can complement nurses' work by providing patients with real-time information and support regarding their treatment, care plans, and general health concerns (Bell et al., 2024; Haque & Rubya, 2023; Lin, 2022). These tools can be particularly useful in addressing patient education, which is frequently cited as an area of care that is neglected due to time constraints (Baddal et al., 2024). AI-driven virtual assistants can offer patients answers to their questions, reminders about medication adherence, and guidance on self-care practices, thereby helping to close the gap left by missed educational care.

Moreover, AI-based remote monitoring tools can enable continuous observation of patients, alerting healthcare providers to any deviations from the norm in a patient's condition, even when nurses are not physically present (Topaz & Pruinelli, 2017). Such systems can reduce the likelihood of adverse events going unnoticed, ensuring that critical care is provided when needed. These technologies not only assist in reducing missed care but also help alleviate some of the stress on nursing staff, enabling them to manage their workload more effectively.

1.3.2.4 Ethical and Practical Considerations

While AI presents significant potential for addressing missed care, there are practical considerations to keep in mind. AI systems must be carefully designed and implemented to ensure that they complement, rather than replace, the human aspects of nursing care, such as empathy and patient-centred communication (Floridi et al., 2018). Additionally, the integration of AI into healthcare systems requires substantial investment in training and infrastructure, as well as rigorous ongoing testing to ensure that these technologies are safe, reliable, and effective. AI has the potential to significantly reduce missed or unfinished care by improving workflow efficiency, optimising staffing and resource allocation, supporting clinical decision making, and providing tools for patient engagement and monitoring. However, the successful

integration of AI into healthcare systems requires thoughtful implementation and addressing any practical challenges that may arise.

1.4 Description of Research Methods

1.4.1 Research Question

The research question for this investigation was: *How can machine learning models be developed, tested, and evaluated to predict whether clients presenting to an emergency department with a mental health concern, particularly suicidality, will require admission to an acute mental health inpatient ward?*

The question can be split into the following three aims.

1. Complement and augment clinicians' evaluation and decision making
2. Enhance the client's experience of mental health services whilst providing high quality and safe care
3. Build understanding of artificial intelligence and machine learning adoption in a mental health context.

This complex question necessitated a robust methodological approach to ensure comprehensive and reliable outcomes.

1.4.2 Research Design

This study employed an overarching design thinking (DT) framework, which was instrumental in addressing the complex, human-centred aspects of predicting MH outcomes in an ED setting. DT was particularly relevant here as it emphasises empathy, ideation, and iterative testing, ensuring that the solutions developed are closely aligned with the needs of both patients and

healthcare providers (Brown, 2008, 2009). The framework consists of several stages—empathise, define, ideate, prototype, and test—each contributing to the development of ML models that are not only accurate but also practical for real-world application.

In addition to DT, the study utilised a retrospective cohort method to gather and analyse historical data from patients who previously presented to the ED with MH issues as their primary presenting problem. This approach allowed for the use of existing data, providing a comprehensive and rich source of information that enhanced the predictive power of the models and deepens the understanding of community needs. The retrospective cohort method was particularly valuable in uncovering trends and patterns in MH presentations, which are critical for informing the development of effective predictive models (Liamputtong, 2021, p. 212).

Central to this research was the application of supervised ML methods. By training models on the data obtained from the retrospective cohort study, the research aimed to identify and learn the patterns associated with various MH presentations to ED. Supervised learning algorithms were used to explore the data, determining the most effective approach for the specific prediction tasks (Hastie et al., 2013). The performance of these models was rigorously tested and evaluated to ensure their reliability and accuracy in predicting ED admissions for MH concerns.

Together, these methodologies form a cohesive and comprehensive research design. The combination of DT, retrospective cohort analysis, and supervised ML provided a robust framework for addressing the challenge of predicting ED admissions for patients with MH issues. The research aimed to enhance patient outcomes and optimise resource allocation in healthcare settings, contributing to more effective and responsive MH care delivery.

List of Projects

- Project 1: The Role of Commercial Determinants in AI Integration in Health Care
- Project 2: Literature Review and Examination of Trust in AI
- Project 3: Exploration of Demographic and Presentation Data in EDs (Studies A, B & C)
- Project 4: Application of Machine Learning to Predict Mental Health Admissions (Study D)

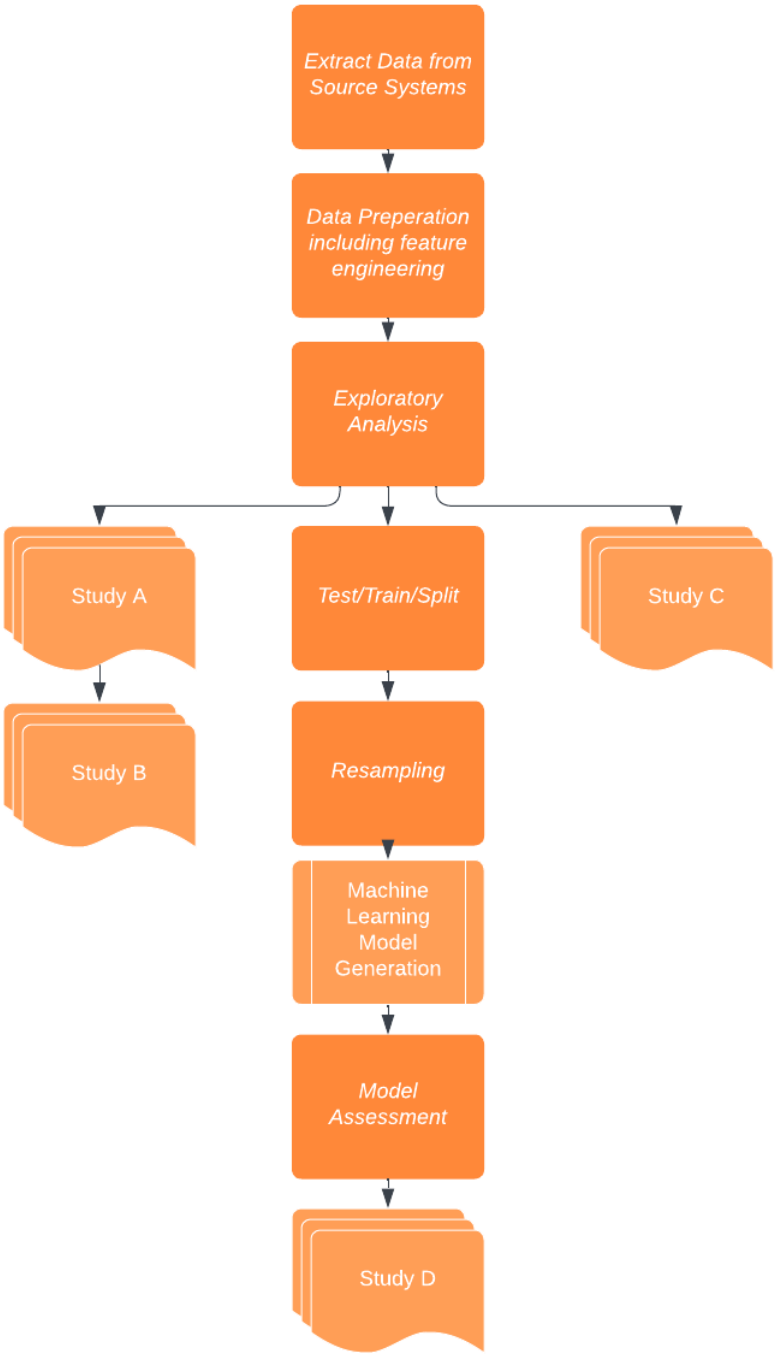


Figure 1.1. *Project Flow*

1.4.3 Empowering Nursing Workflows: Advancements in Health Decision Processes

For many industries, it is only a matter of time before automation through technologies such as AI have a significant role in how the environment is shaped (Purkis, 2018; Schoettle & Sivak, 2016). Replacing hands-on care is not as straightforward as creating a robot to replace a nurse, nor should this be the end goal. Many systems exist that do aid and assist, but they often have a precise or limited scope (Purkis, 2018).

Automation is defined as "applications of robotics, artificial intelligence, machine learning, machine vision, and similar emerging and mature digital technologies that will allow human work to be substituted by computer capital" (Willis et al., 2019). The introduction of the electronic medical record has revolutionised many parts of the health industry, allowing the clinician to process, find, and retrieve information faster than a traditional paper record. The natural next step is to allow ML and other AI systems to examine the data and establish patterns and insights that would not be discernible by humans. However the uptake of automation has been slow, and the digitisation of medical data often produces closed systems, which can make access to data extraction and real-time use exceptionally difficult (Marques & Ferreira, 2020). Additionally, some healthcare professionals have concerns regarding the impact of automation, fearing that these approaches will replace and negate the value of human contact with individuals (Witkowski et al., 2024). However, automation through AI technologies may provide the opposite, freeing up healthcare professionals and giving them more time with their patients (Marr, 2019). This research tested an approach to utilising real-time health informatics data to reduce the time MH clients spend in ED.

1.5 Researcher Perspective

I am a nurse and, although I never set out to become one, I am eternally grateful for what it has provided for me. Being a nurse has given me moments in my life that are simultaneously cherished and heartbreaking. For example, I have held a person's hand when they had no one else, been witness to intimate moments, and I have been the person who cared for someone in their final hours when there was no other. As such, in keeping within my scope of practice, I am an advocate for those who cannot advocate for themselves (Nursing and Midwifery Board of Australia, 2016).

My parents came from a low socio-economic group and were also members of a devout Seventh-Day Adventist community. As an only child, I was brought up to accept the word of the church without question, yet I observed that the world around me did not follow the rules that the church set. When I asked about the inconsistency between our beliefs and the way that I perceived the world to work, I would be told that it was simply God's plan, an answer that I never truly accepted as it lacked a logical, repeatable, and rational explanation.

I was born the grandson, and first boy, to a master carpenter. The only thing more impressive than the quality of his workmanship was his driving desire and expectation for perfectionism of himself and those around him. I always felt that I was the "heir to the empire" he had envisioned; I remember reading a book about electrical circuits at the age of five that I found in his workshop and being fascinated by how the components fitted together, meanwhile exhibiting little interest in working with wood. For me, woodwork presented little challenge, but electricity and computers were a frontier that I was compelled to understand, including both *why* and *how* they worked. This desire to understand complex machines landed me in trouble numerous times as a child for pulling apart toys or even taking to them with a hammer. My aim was not to be destructive at all, but rather, I was driven to dismantle everything down to singular

components and, not always successfully, reconstruct so that I could understand how they worked.

At 14 years of age, my parents separated, and my experience of the world was fundamentally altered. I felt a profound sense of uncertainty and struggled to find my place. This led me, at 17, to pack a single bag and travel from rural Hunter Valley to Sydney for a job interview. I was successful, and when asked when I could start, I replied, “Tomorrow”. The following years were a whirlwind, marked by both chaos and discovery, as I navigated a world that I often found bewildering and confusing. Despite the challenges, I persevered, taking on any job I could find—from sales to cleaning, construction and even garbage collecting. However, I was driven by a growing belief that I had much more to contribute to the world, even if I was not yet sure where to begin, having no qualifications and limited opportunities. I had always worked but I never had the chance to complete an education. I desired to pursue engineering at university, but my high school incompletion and inability to afford time away from work meant that I might never be able to attend any formal training. Sharing my predicament with my mother-in-law, she suggested that I pursue nursing, herself having experienced a career in the field. I replied that if I were to undertake university, it would be to become an engineer, not a nurse. She told me about the Trainee Enrolled Nurse (TEN) program, a competitive entry and paid twelve-month program. Applying meant taking a leap of faith and a reduction in what little pay I was currently earning, but I was able to pass the required tests, and I embarked on my nursing career, graduating as an Enrolled Nurse in 2005.

My grandmother was a nurse (Figures 1.1 and 1.2), the opposite of my grandfather. She was the most level-headed and non-judgemental person I have ever met. To me, she embodied what the discipline of nursing is and, while I did not recognise it until much later, she remains the blueprint for everything I imagine a nurse should be. I believe that being a nurse enabled her to view the world in a way that others could not, a perspective that I have learned through my

own nursing practice. This perspective comes only through the long hours and selfless dedication fundamental to nursing.

I vividly recall coming home from my first day of TEN training, absolutely exhausted but committed to the cause. In the year that followed I worked the hardest I had ever worked in my life, but the more I put into my development as a nurse, the more I benefited. Nursing had found me, and I was not letting go of it. Every corner I turned, I discovered a new challenge and new knowledge that had once been my "unknown unknowns". I was able to achieve my dream of an education when I attended university the following year to commence my Bachelor of Nursing.

However, while I thoroughly enjoyed nursing and the place in the world it afforded me, the desire to properly understand technology still needed to be met. Over the next few years, I undertook a Bachelor of Technology via distance education with the Royal Melbourne Institute of Technology (RMIT). After graduating with distinction, I felt that I had finally done something important for myself, and this knowledge has allowed me to work on developing technologies such as the internet of things, microcontrollers, and robotics.

I have been employed by Central Coast Local Health District and NSW Health for over 20 years. During this time, I have held various clinical and non-clinical roles, most recently as Manager of Population, Data & Outcome Measures and Deputy Director of Mental Health. I have witnessed major changes in the way that mental health care is delivered, such as the introduction of the electronic medical record (eMR), virtual health, and the volumes of data that are now generated.

The conceptualisation for this PhD project started when I attended a conference about comparing different health services data. The big numbers became so overwhelming that seeing the patient in the data was impossible. In the room next door, Nvidia, a leading computer

hardware developer, was hosting a conference on AI. I attended their keynote session and was blown away by what they were doing. I went back to my conference, where the speaker was reporting on hospital utilisation, attempting to compare large health services against one another. I asked myself, “What do we currently have the data for that, if we knew the outcome before the event, could change the way we deliver care in mental health? How much better care could we deliver if we could predict the outcome?” Health services generate large quantities of data, but it is not always clear what we need to do with it to improve the services.

I bring a unique perspective. I am a mental health nurse first, but I am also a scientist, a researcher, and a technologist. This eclectic collection of knowledge and skills allows me to peel back the layers of the data to locate the patient beneath. Instead of big data that reduces humans to numbers, I am a digital advocate for people. As clinicians, we must lead the change that is coming through technological innovation. If we do not advocate for ourselves, the digital tools constructed may not possess the crucial elements for nursing to continue to deliver high-quality, safe care and the advocacy that our patients need when they are unwell.

Figure 1.2. *The author's grandmother, Patricia Christie, upon graduation from her nursing training on 4 December 1955 from Sydney Sanitarium and Hospital (now known as the Sydney Adventist Hospital). She was first registered on 9 January 1956*

Growth of district nursing service — 34,000 patients visited annually

GOSFORD HOSPITAL GROWING TO MEET DEMANDS OF AREA

10th NURSE NEEDED SOON

Gosford Hospital's Nursing Service now has nine nurses visiting about 34,000 patients and present indications are that a 10th district nursing sister will soon be required.

The service has grown rapidly since it started in 1955 with just one nurse.

The hospital's chief executive officer, Mr. M. Boyce, has said this in a paper presented to a conference of the Australian Institute of Hospital Administration.

Mr. Boyce's subject was "the district nursing service in the planned care of the aged".

Outlining the introduction of the service at Gosford, Mr. Boyce told the conference that it was formed on the hospital in 1955 because of the acute shortage of beds and the high percentage of aged people within the hospital area.

He said the present population of the area is about 70,000 plus half this number again for about four months each year in the tourist season.

Increased ratio

Before recent extensions to the hospital were completed, it was serviced with .9 beds per 1000 of population — the addition of 65 beds has increased the ratio to 1.5 per thousand of population.

"Taking into consideration the State daily average of occupied beds of 4.2 per 1000 of population, it is estimated that today we are more than 250 beds short of our requirements.

"With the tremendous growth in population in 10 years it is anticipated that an additional 400 beds will be required on the Central Coast," Mr. Boyce said.

He said the benefits of the district nursing service to the patient and family were:—

● Treatment of the patient in his own home and environment.

● The morale of an aged patient, cared for at home, is maintained at a higher level than his counterpart in hospital.

● The treatment of possible long term patients at home obviated the necessity of constant hospital visits by the relatives.

● Following tuition by the District Nurse, responsible relatives are able to help care for the patient.

● The use of equipment lent by the service saves the patient a considerable sum of money.

● If the patient is a pensioner, free treatment is allowed under the District Nursing Service.

● With the constant supervision of the District Nursing Sister, other facilities are more readily made available, such as the provision of Meals on Wheels and the Emergency House-keeping Service.

● The District Nursing Sister is able, in many cases, to relieve the tension of the patient.

More space

Mr. Boyce added that with the provision of the District Nursing Service, the hospital was able to relieve the bed position of chronic cases and allow more space for the turnover of acute patients.

"From a recent costing of the District Nursing Service of the hospital, each home visit has cost less than $1 (95 cents). It has been estimated that because this hospital is able to conduct a major district nursing service, a saving of an estimated 100 patient beds is effected," Mr. Boyce said.

He said the nurses this year were expected to carry out about 34,000 visits, treating more than 1900 patients. Statistics disclosed that a nurse visits each patient 25 times.

"The saving figure was unbelievable, when it was considered that the overall district nursing service cost approximately $34,000 a year to maintain, he said.

Criticism

Mr. Boyce criticised the lack of contribution funds rebates for fees paid to the District Nursing Service.

"This, I feel, is completely wrong when a nurse working in a private capacity is able to make a charge and the patient is able to claim a rebate from the fund," he said.

Pointing out the problems of developing a District Nursing Service, Mr. Boyce said the service should be controlled by the hospital within its area, and the approval of the Federal and State authorities was first required.

Under present arrangements, the Commonwealth pays various rates to subsidy to district nursing services, depending on their starting dates.

"My hospital is one of the fortunate ones. We were in operation before 1956 and for all nurses appointed following this date we receive a maximum of $2200 Commonwealth subsidy per nurse.

"In our area we have voluntary district nursing associations in each district, providing the cost of motor cars' running expenses and other incidentals amounting to about $4800 annually," he said.

Doctor's request

Visit by the district nurse is made at the request of the patient's doctor.

At Gosford Hospital it is not necessary that he be a member of the honorary medical staff, Mr. Boyce said.

He said there were about 50 doctors on the Central Coast with only half this number visiting the hospital.

The District Nurse rings in daily to receive calls lodged with the hospital by the doctors and in some areas the district nursing sisters have direct contact with the local medical officer.

"Our service is a success because the medical staff support the District Nursing Service and provide the patients to be treated," he said.

Income

Mr. Boyce said income from the District Nursing Service, from the patients, was negligible as they were mostly pensioners.

Mr. Boyce concluded by summing up the benefits of the service to Gosford District Hospital. "The provision of the District Nursing Service enabled us to cope with the admission of acutely ill patients when the beds provided represented less than one bed per thousand of population."

These of the district nursing sisters attached to Gosford District Hospital stand with one of the cars provided for their use. They are (l. to r.) Sisters D. Hewitt, who covers the Wyong-Budgewoi area, and P. Christie and J. Roberts, who both visit The Entrance district. Other sisters are Sister Ellis (Gosford), Lockley (Terrigal area), Wemberall, Warnecke and Hennington (both Way Way). Relieving sister is Sister Tiller. —Masters Studio photo.

Due to expiration of

Figure 1.3. *Patricia Christie (centre) as a part of the Gosford District Nursing Service. Gosford Star, Wednesday, 14 May 1969, page 38*

1.5.1 Personal Reflection on the Research

I write this reflection as I finalise my PhD and reflect on the ways that the research has challenged and changed my personal and professional frame of reference. I approached this endeavour from a very scientific and code-driven perspective; however, I was not prepared for the knowledge that would be presented to me and the Western view of the world that I lived in. Throughout my life, I have always considered myself respectful and understanding of the needs of Aboriginal and Torres Strait Islander peoples. However, my journey of self-reflection and exposure to various experiences and literature through this PhD experience has profoundly challenged and reshaped my perspective. In this reflection I aim to explore the evolution of my

understanding and commitment to cultural safety, particularly concerning Aboriginal and Torres Strait Islander communities in Australia.

I vividly remember a conversation with my principal supervisor about my childhood Aboriginal friends in which she stated that, statistically, many of them were likely in gaol or even deceased. This was a harsh realisation, considering how disconnected I felt from these realities despite my belief in my own respectfulness. I was encouraged to read works such as *Growing up Aboriginal in Australia* (Heiss, 2018), *Dark emu* (Pascoe, 2018) and *Born-again Blakfella* (Charles, 2019) and these authors exposed me to perspectives I had never imagined, nor ever considered reading previously. These books challenged my previously dismissive attitude towards the Stolen Generations. I used to think, "Well, I did not cause this issue, so why can't these people just let it go and move on with their lives?" I now recognise the deep flaws and insensitivity in that viewpoint. My exposure to this literature, along with academic discussions that contributed to the results chapter of this thesis, profoundly challenged both my thinking and my person. I realised that my understanding of the issues faced by Aboriginal and Torres Strait Islander peoples, such as cultural safety, was narrow and predominantly shaped by a "Western" perspective. Through my research, I now understand why colonisation has left such a lasting impact on Aboriginal and Torres Strait Islander peoples. This historical program to assimilate and westernise has deprived all Australians of the invaluable lessons we can learn from First Nations knowledge. Exposure to this literature helped me to understand the devastation faced by Aboriginal communities and how it is our collective responsibility as humans, regardless of our lineage, to address the needs of vulnerable populations as a priority.

Reflecting on my own upbringing surrounded by a small country town's views and attitudes, I now see how Aboriginal and Torres Strait Islander people were often regarded as second-class citizens. I once believed that my success, despite a difficult background, was solely due to hard work and that Aboriginal people should be able to achieve the same. This naive perspective

failed to consider the systemic barriers that Aboriginal and Torres Strait Islander people encounter. My understanding of my own "white privilege" has deepened, recognising that the opportunities available to me are often inaccessible to Aboriginal and Torres Strait Islander people due to entrenched socio-economic disadvantages. The outcome of the referendum on an Aboriginal and Torres Strait Islander Voice (National Indigenous Australians Agency, 2023) upset me more than I expected and served to further highlight ongoing societal resistance to recognising and addressing Aboriginal and Torres Strait Islander issues. I felt a deep embarrassment for Australia, but it also amplified my empathy for the pain my Aboriginal and Torres Strait Islander colleagues must have been experiencing.

Accordingly, I wish to acknowledge the deep and enduring impact of historical and contemporary injustices faced by Aboriginal and Torres Strait Islander Australians. I recognise these issues and offer a sincere apology for the harm and suffering that has been inflicted. Australia's Aboriginal and Torres Strait Islander peoples have endured centuries of colonisation, dispossession, and discrimination (Paradies, 2016). Policies of assimilation, the forced removal of children (Ware, 2013) that created the Stolen Generations (Australian Institute of Health and Welfare, 2021a; Haebich, 2011), and ongoing systemic inequalities have caused profound and lasting trauma (Wilson & Waqanaviti, 2021). These actions have not only disrupted lives but have also led to the erosion of cultures, languages, and traditions that are integral to the identity of Aboriginal and Torres Strait Islander communities (Paradies, 2016; R. Smallwood et al., 2021). I express my deepest regret and sorrow for the suffering and disadvantage experienced by all Aboriginal and Torres Strait Islander peoples as a result of colonisation and its ongoing legacy. The pain and loss endured by these communities are profound, and I am truly sorry for the role that all Australians, myself included, have played in perpetuating these injustices.

As a researcher, a nurse, and an Australian, I am responsible for acknowledging these injustices and ensuring that my work respects and honours Aboriginal and Torres Strait Islander people's experiences and knowledge. I am committed to conducting my research in an inclusive, respectful manner and mindful of the historical and contemporary contexts affecting Aboriginal and Torres Strait Islander communities. In my research, I am committed to including Aboriginal and Torres Strait Islander perspectives and knowledge in a respectful and meaningful way, engaging with Aboriginal and Torres Strait Islander communities to ensure their voices are heard and valued, and advocating policies and practices that promote equity and justice for Aboriginal and Torres Strait Islander people. Through these actions, I believe I can contribute to a more just and inclusive academic and healthcare environment. I am committed to continuous learning and advocating for cultural safety and humility, ensuring Aboriginal and Torres Strait Islander voices are heard and respected in all facets of society.

This journey has transformed my understanding and reinforced my commitment to being a proactive ally, advocating for our Aboriginal and Torres Strait Islander brothers and sisters, and challenging commonly held prejudices. I am now actively engaging in programs that incorporate First Nations knowledge, helping to shape various initiatives such as the First Nations–focused Safe Haven within Central Coast Local Health District, and recognising my role as an advocate and model in organisations like Scouts Australia. From the personal knowledge that I gained from this research, I was able to advocate for the Aboriginal and Torres Strait Islander people of the Central Coast to the NSW Ministry of Health and lead the MH team that obtained the necessary funding to enable the opening and ongoing operation and evaluation of the First Nations–focused Safe Haven at Wyong, New South Wales.

1.6 Alignment with National and International Strategies

This research is aligned with relevant national and global public health initiatives, including Closing the Gap (Commonwealth of Australia, 2020), a fundamental program for building an equitable Australia focusing on the true extent of inequality between Indigenous and non-Indigenous Australians. It is also aligned with the Sustainable Development Goals, a United Nations initiative that provides a framework for all nations to work together to end poverty by improving health and education, reducing inequality, promoting economic growth, addressing climate change, and preserving the oceans and forests (United Nations, 2018).

The purpose of mapping these goals is to align the outcomes of this project with national and international strategies that will facilitate meaningful implementation and translation to practice. This project's goals aim to align with the World Health Organization's (WHO) Sustainable Development Goals (SDG) initiative, Closing the Gap, New South Wales (NSW) Towards Zero initiative, the WHO's Principles for Ethics in Health Artificial Intelligence, and the eHealth Strategy for NSW Health 2016–2026.

The Principles for Ethics in Health Artificial Intelligence (World Health Organization, 2021a), a set of guiding principles published by the World Health Organization to address the ethical challenges and risks of using AI in health, and to ensure AI works to the public benefit of all countries. Towards Zero, a NSW Government suicide prevention initiative that addresses the priorities in the NSW Strategic Framework for Suicide Prevention 2018–2023, contributes to the NSW Premier's commitment to reducing suicide in NSW (NSW Health, 2020).

This research aligns with key national and international mental health strategies, ensuring that its outcomes contribute meaningfully to contemporary public health priorities. At the global level, the World Health Organization's (WHO) Mental Health Action Plan (2013–2030) provides a comprehensive framework for strengthening mental health services, promoting

prevention, and reducing stigma worldwide (World Health Organisation, 2021c). Additionally, the European Union (EU) Digital Health Action Plan (2023–2030) underscores the importance of digital transformation in healthcare, aiming to enhance accessibility, quality, and efficiency in mental health services across EU member states (European Commission, 2023). Similarly, Canada's digital health initiatives focus on leveraging technology to improve mental health service delivery, providing an important international perspective on digital mental health advancements (Health Canada, n.d.).

Within the Australian context, several policies and frameworks guide digital mental health initiatives. The National Digital Mental Health Framework sets out a strategic direction for the development and governance of digital mental health services in Australia (Department of Health, 2021). Complementing this, the National Safety and Quality Digital Mental Health (NSQDMH) Standards provide a structured approach to ensuring that digital mental health services meet high safety and quality standards (Australian Commission on Safety and Quality in Health Care, 2020). Suicide prevention is also a national priority, as outlined in the National Suicide Prevention Strategy for Australia's Health System (2020–2023), which integrates evidence-based interventions and policy measures to reduce suicide rates (Department of Health, 2020). Furthermore, the National Mental Health Commission's (NMHC) Annual Monitoring Report provides critical insights into the progress and challenges of mental health policy implementation in Australia (National Mental Health Commission, 2024).

Internationally, digital mental health and telehealth standards continue to evolve. The International Telemedicine and eHealth Guidelines, including frameworks such as the American Psychological Association's (APA) Guidelines for the Practice of Telepsychology, provide ethical and practical recommendations for delivering mental health services remotely (American Psychological Association, 2023). These guidelines are essential for ensuring that digital and telehealth solutions are deployed safely and effectively, aligning with broader global

mental health objectives. By integrating these national and international strategies, this research aims to contribute to a more coordinated, high-quality, and ethically responsible digital mental health ecosystem.

1.6.1 Sustainable Development Goals

In 2015, member states of the United Nations adopted the 2030 Agenda for Sustainable Development, which lays out a framework for peace and prosperity for people and the planet (United Nations, 2018). The 17 Sustainable Development Goals (SDGs) are at its core, constituting an urgent call for action from all nations, including high-, medium-, and low-income countries, working together in partnership (United Nations, 2018). They aim to end poverty by improving health and education, reducing inequality, promoting economic growth, addressing climate change, and preserving the oceans and forests. The SDGs recognise that ending poverty must go hand in hand with building economic growth and addressing a range of social needs, such as education, health, social protection, and job opportunities, in addition to fighting climate change and protecting the environment (United Nations, 2018).

1.6.1.1 Good Health and Well-Being (SDG 3)

The purpose of this goal is to ensure healthy lives and promote well-being for all at all ages. A guiding principle for this project was to provide faster access to higher-quality care while minimising the required cost, providing access to quality essential MH services, and access to safe, effective, quality, and affordable essential health care for people of all ages. AI has the potential to transform MH care by enhancing access, quality, and affordability. For example, AI can be used to optimise resource allocation and operational efficiency, reducing wait times and costs (Topol, 2019b). It can also provide personalised treatment plans by analysing patient data and using AI-powered wearables for continuous monitoring (Esteva et al., 2019). By

leveraging these capabilities, ML and AI can significantly improve MH care accessibility, quality, and affordability across the lifespan.

1.6.1.2 Reduced Inequalities (SDG 10)

The purpose of this goal is to reduce inequality within and among countries. This project helps health professionals better understand the factors that contribute to persons seeking acute MH care. The insights gained will provide building blocks for further research and development to provide equitable opportunity, reduce outcomes inequalities, and promote appropriate legislation, policies, and actions. The research aimed to reveal any inequalities or biases that may exist in assessing MH presentations to ED. The research aimed to reshape the models as required to ensure that biases are minimised or removed altogether.

1.6.2 Closing the Gap

Closing the Gap is fundamentally important to building an equitable Australia; all Australians must take stock of the true extent of inequality between Indigenous and non-Indigenous Australians to address our past failures. The Closing the Gap framework sets targets in key areas: life expectancy, infant and child mortality, early childhood education, literacy and numeracy skills, and school completion rates (Australian Bureau of Statistics, 2021a; Australian Institute of Health and Welfare, 2022a). Aboriginal and Torres Strait Islander people experience higher levels of mental illness and represent a priority population, and as such this project engaged both the Central Coast Local Health District Aboriginal Health Service and local Aboriginal medical services to identify patterns that may contribute to entrenched disadvantage at the individual and community levels, primarily focusing on ED presentations and suicidality. The publication of “Disproportionate mental health presentations to emergency departments in a coastal regional community in Australia of First Nation people” (Higgins, Sheather-Reid et al., 2024a) has provided local non-government and community-managed

organisations with evidence to support the local story of Aboriginal and Torres Strait Islander People living on the Central Coast.

This project aimed to ensure that the experiences of Aboriginal and Torres Strait Islander people were not overlooked in the data analysis. Focusing solely on the whole population would risk ignoring the disproportionately high number of presentations from this group.

This project aligns with the following Closing the Gap priorities: "*Everyone enjoys long and healthy lives*" (Priority 1), and "*People have access to information and services enabling participation in informed decision-making regarding their own lives*" (Priority 17) (Commonwealth of Australia, 2020).

1.6.2.1 Everyone Enjoys Long and Healthy Lives (Priority 1)

The objective of this goal is to close the gap in life expectancy within a generation, by 2031. This project aimed to address this goal through artificial intelligence (AI) and machine learning (ML), utilising this emerging technology to help identify the factors that contribute to acute ED MH presentations and wider socio-economic problems and disadvantages. To mitigate this risk, the models used in this research included "Indigenous status"; however, this created a risk of inherent bias in the model, as such models were built first to understand better the factors involved in the presentation. The information discovered in this research was examined, modified and shared with local Aboriginal and Torres Strait Islander leaders and groups to ensure that any practices are culturally sensitive and aligned with the needs of the community.

1.6.2.2 People Have Access to Information and Services Enabling Participation in Informed Decision Making Regarding Their Own Lives (Priority 17)

The objective of this goal is that Aboriginal and Torres Strait Islander people have access to information and services enabling participation in informed decision-making regarding their

own lives. By 2026, Aboriginal and Torres Strait Islander people have equal levels of digital inclusion. AI and ML may offer powerful insights into data, helping find explanations to complex clinical problems. When these complex interactions are combined with local knowledge, they may offer explanations and insights that were previously hidden. Therefore, this project aligned with this goal by ensuring that all data and insights discovered were made available to and shared with local Aboriginal and Torres Strait Islander groups to help enable participation in informed decision making and aligning with local needs.

1.6.3 Towards Zero (NSW Health)

Suicide is a complex problem, and it will take a coordinated and concentrated approach on the part of individuals, communities, the private sector, and the government to reduce suicide deaths. Towards Zero Suicide is a NSW Government suicide prevention initiative that addresses the priorities contained in the NSW Strategic Framework for Suicide Prevention 2018-2023, contributing to the NSW Premier's commitment to reducing suicide by 20 per cent by 2023 (NSW Health, 2020). The NSW Government invested $87 million in this initiative over three years, starting from 2019–2020. This initiative includes system improvements, suicide prevention training for NSW Government staff and community members, and improving support in rural and regional areas. Within these initiatives are two critical components that this project addressed: supporting suicide prevention collaboration and alternatives to ED presentations. This project addressed these by using the data gained from the initial model design and creation to inform the understanding of the anatomy of a person's suicidal presentation and contribute to the global body of knowledge. Additionally, this project contributed to the alternatives to ED presentations priority to identify early those that may be diverted to alternative pathways such as the Safe Haven project (NSW Health, 2021b).

1.6.4 Ethical Principles for Artificial Intelligence in Health

Integrating AI into health care presents a transformative opportunity to enhance clinical practice and accessibility. However, the implementation of AI in this context must be guided by ethical frameworks that emphasise principles such as transparency, consent, accountability, and equity (Hantel et al., 2024). Furthermore, understanding patients' perspectives, particularly from historically marginalised and under-represented groups, is essential to address ethical considerations effectively (Hantel et al., 2024). As such, the World Health Organization and many governments worldwide have developed ethical frameworks that aim to deliver AI's benefits in health and mitigate potential harms.

1.6.4.1 World Health Organization

The WHO identified the ethical challenges and risks with the use of artificial intelligence in health, developing principles to ensure AI works to the public benefit of all countries (World Health Organization, 2021a). The six principles are: protecting autonomy, promoting human safety and well-being, ensuring transparency, fostering accountability, ensuring equity, and promoting tools that are responsive and sustainable (World Health Organization, 2021a). The proposed research aligned with the following principles.

1.6.4.1.1 Protect Autonomy

The final say on all health decisions must reside with humans. They should not be made entirely by machines, with clinicians being able to override them at any time. Likewise, AI tools should not be used to guide someone's medical care without their consent (World Health Organization, 2021a). This research aligns with WHO's principle of autonomy by ensuring that ML models developed to support MH decision-making serve as tools to assist clinicians, with final healthcare decisions remaining under human control. In Study D, clinician oversight was

prioritised to maintain human authority in decision-making, underscoring the commitment to upholding patient autonomy and informed consent.

1.6.4.1.2 Promote Human Safety

The developers of any AI tools should continuously monitor the tools they create to ensure they are functioning correctly and not inadvertently causing unintentional harm (World Health Organization, 2021a). By prioritising model transparency and clinician involvement throughout the design and implementation process, this research ensures continuous oversight of AI systems to prevent unintended harm. The future research planned in Study E, discussed in Section 6.5, explores clinician attitudes towards AI that address safety by fostering a collaborative approach to AI integration, promoting responsible and safe use of ML models in mental health care.

1.6.4.1.3 Ensure Transparency

Developers should publish information about how AI tools are designed. Systems should be intelligible and transparent, not "black boxes" that make it difficult to understand how decisions are made. Developers should allow enough transparency so that systems can be independently audited and understood by end users and regulators (World Health Organization, 2021a). Using explainable AI (XAI) methods, particularly in Studies C and D, this research promotes transparency by allowing clinicians to understand how ML models make recommendations. This approach aligns with WHO's call for intelligible AI systems and contributes to model transparency, helping clinicians trust and interpret AI-generated outputs effectively.

1.6.4.1.4 Foster Accountability

When something goes wrong with AI technology, such as a decision made by the tool leading to patient harm, mechanisms should be established to determine who is responsible (developers or clinical users) and processes to correct the issue. To minimise these risks, an iterative, approach should be used during developmental phases and testing (World Health Organization, 2021a). The design of this research follows an iterative and evaluative approach, especially in Study D, ensuring mechanisms are in place to identify accountability for AI-driven decisions. This structure promotes joint accountability between developers and clinicians, aligning with WHO's principles by establishing clear roles for managing outcomes and correcting potential errors.

1.6.4.1.5 Ensure Equity

Innovations should be made available in multiple languages and must be trained on diverse datasets. Bias should be understood, documented, and where applicable removed or remodelled based upon best practices for the particular cohort before any implementation is undertaken. English is the language selected for this study. However, recommendations for follow-up studies will include a consideration of the needs of culturally and linguistically diverse populations (World Health Organization, 2021a). Study B's focus on Aboriginal and Torres Strait Islander MH care needs illustrates the project's commitment to equitable AI deployment, addressing specific cultural and demographic factors that influence MH outcomes. Future recommendations include expanding studies to ensure AI tools are inclusive of diverse populations, promoting equitable access and representation in line with WHO's equity principle.

1.6.4.1.5 Promote Artificial Intelligence that is Sustainable

The developers of tools should be able to update them regularly, and the institutions should be able to adjust them if they prove ineffective. Furthermore, organisations should only introduce tools that can be repaired, even in under-resourced health systems, and be guided by clear strategies (World Health Organization, 2021a). The research emphasises the importance of adaptable, regularly updated ML models to remain effective over time. Study D highlights the need for ongoing updates to ML systems, ensuring tools remain functional, relevant, and responsive to changing MH needs, fulfilling WHO's vision of sustainable and adaptive AI.

1.6.4.2 Australia's Eight Artificial Intelligence Ethics Principles

Further to the guidelines set out by the WHO above, the Australian Government has also introduced eight Artificial Intelligence Ethics Principles that aim to ensure AI technologies' safety, security, and reliability while promoting fair and beneficial outcomes for all Australians. These voluntary principles complement existing regulations and encourage businesses and governments to adhere to ethical AI design and implementation standards. Key principles include prioritising human, societal, and environmental well-being; respecting human rights and diversity; ensuring fairness and accessibility; protecting privacy and data security; maintaining reliability and safety; providing transparency and explainability; allowing contestability in AI decisions; and establishing accountability for AI outcomes. By committing to these principles, organisations can build public trust, enhance consumer loyalty, and ensure that all Australians benefit from AI advancements (Department of Industry, Science and Resources, 2024). The following points further break down and explain the aforementioned concepts.

1.6.4.2.1 Human, Societal, and Environmental Well-Being

AI systems should be designed to benefit individuals, society, and the environment, promoting positive outcomes throughout their lifecycle (Department of Industry, Science and Resources, 2021). This research prioritises well-being by focusing on reducing missed care in EDs and enhancing MH service quality. Through the development of ML models to improve decision-making, the project aligns with the principle of promoting positive societal outcomes, particularly for underserved populations, as demonstrated in Studies B and C.

1.6.4.2.2 Human-Centred Values

AI systems must respect human rights, diversity, and individual autonomy, ensuring that technology serves humanity and enhances democratic values (Department of Industry, Science and Resources, 2021). By prioritising patient autonomy and incorporating clinician feedback, the research upholds human-centred values. The design and implementation of ML models are structured to respect human rights and democratic values, especially in addressing the needs of vulnerable MH populations in Studies A and B.

1.6.4.2.3 Fairness

AI systems should be inclusive and accessible, preventing unfair discrimination against individuals or groups and promoting equitable treatment for all users (Department of Industry, Science and Resources, 2021). This research addresses fairness by ensuring that AI systems are designed to be inclusive, particularly in Studies B and D, where the needs of Aboriginal and Torres Strait Islander communities and the perspectives of MH clinicians are considered. Recommendations for future research include a focus on accessible and equitable AI solutions for diverse patient groups.

1.6.4.2.4 Privacy Protection and Security

AI systems must uphold privacy rights and data protection, implementing robust security measures to safeguard personal information throughout the data lifecycle (Department of Industry, Science and Resources, 2021). Privacy and data protection are embedded in the research methodology, with secure data handling and protection measures applied across studies. Study D, for example, incorporates robust security protocols when handling patient data, aligning with the Australian principle of privacy protection.

1.6.4.2.5 Reliability and Safety

AI systems should operate reliably and safely according to their intended purpose, with ongoing monitoring to manage risks and ensure consistent performance (Department of Industry, Science and Resources, 2021). The research supports reliability by using validated ML models and interpretability methods, especially in Study C, to manage risks and ensure that AI tools perform consistently. Continuous monitoring and evaluation throughout the research address the Australian principle of maintaining reliability and safety in AI systems.

1.6.4.2.6 Transparency and Explainability

There should be clear communication regarding how AI systems operate and make decisions, enabling users and stakeholders to understand their impacts and interactions (Department of Industry, Science and Resources, 2021). Transparency is a core aspect of this research, with Study D utilising explainable AI tools like InterpretML help to provide clear insights into model predictions. These efforts align with the principle of transparency, ensuring that clinicians and stakeholders can understand and trust the AI’s decision-making processes.

1.6.4.2.7 Contestability

Mechanisms must exist for individuals and communities to challenge AI decisions that significantly affect them, ensuring accessible and efficient processes for redress (Department of Industry, Science and Resources, 2021). The research incorporates clinician feedback loops, allowing clinicians to challenge and refine AI-driven recommendations. This aligns with the principle of contestability, as it provides mechanisms for clinicians to question AI decisions and seek adjustments in the system.

1.6.4.2.8 Accountability

Organisations and individuals responsible for AI systems should be identifiable and held accountable for the outcomes, ensuring proper oversight and control throughout the AI lifecycle (Department of Industry, Science and Resources, 2021). Accountability is emphasised through collaborative model development and clear delineation of responsibilities among stakeholders. The research framework includes provisions for addressing any errors or misinterpretations of data by the AI, ensuring that accountability mechanisms are in place, as highlighted in Studies C and D.

1.7 Mental Health Services in the Context of COVID-19

It should be noted that this study started during the COVID-19 lockdown and has been completed in transition. During this time, the uptake of e-mental health has accelerated, and its acceptability and utility have increased. The pandemic's impact on global healthcare systems has significantly influenced MH services and contributed to the increased adoption of e-mental health solutions (Moreno et al., 2020; Wind et al., 2020). The widespread disruption caused by the pandemic, including lockdowns, social distancing measures, and increased stress due to health concerns, unemployment, and social isolation, exacerbated MH issues, leading to a surge

in demand for MH services (Moreno et al., 2020). However, traditional face-to-face MH services were severely limited due to these public health measures, leading to a rapid shift towards digital solutions (Wind et al., 2020). The strain on MH services during the pandemic was immense, with many health systems struggling to meet the increased demand. Lockdown measures prevented in-person consultations, while healthcare workers were redeployed to other critical areas of the pandemic response (World Health Organization, 2022a). Consequently, there was a significant rise in unmet MH needs, especially among vulnerable populations such as the elderly, those with pre-existing MH conditions, and individuals in lower socioeconomic groups (Holmes et al., 2020).

In response to these challenges, e-mental health solutions, including teletherapy, mobile health applications, and digital platforms for MH support, gained considerable traction (Torous et al., 2020). These solutions, which had been underutilised prior to the pandemic, became essential in addressing the MH crisis as they offered a way to provide care remotely, overcoming barriers imposed by physical distancing and healthcare system overloads (Wind et al., 2020). The pandemic accelerated the adoption of digital MH services, which were already in development, albeit at a slower pace. This acceleration was driven by several factors, including the need for immediate access to care, increased familiarity with digital technologies due to remote working and schooling, and the scalability of digital platforms (Panchal et al., 2023).

Despite these benefits, the rapid shift to e-mental health also exposed challenges. Concerns regarding the digital divide, where individuals without access to technology or those who are less digitally literate are left behind, are significant (Wind et al., 2020). Nevertheless, the pandemic demonstrated the potential of e-mental health as a complementary tool to traditional services, potentially reshaping MH care delivery in the post-pandemic era (Holmes et al., 2020). While challenges remain, the increased adoption of digital MH services represents a significant step forward in meeting the growing MH needs of the global population.

1.8 Overview of Thesis

This thesis is comprised of multiple chapters, incorporating a hybrid format of published manuscripts and unpublished material. Due to this approach, some content may overlap, as the articles have been published in different outlets and certain background information has been repeated. This thesis explores the integration of AI and ML into MH care, focusing on their application in EDs and their potential to enhance clinical decision making and patient outcomes. The research is grounded in a multidisciplinary approach, combining technological innovation with ethical considerations and user-centred care models to address the complex challenges faced by MH services.

1.8.1 List of Publications in Chapters

Each chapter that contains a publication includes further explanation of the conceptual links between the chapters to ensure coherence, as well as an extended discussion of the analysis and findings beyond what is included in the publications.

1.8.1.1 Chapter 2 Background

Higgins, O., & Wilson, R. L. (2023). Commercial determinants and therapeutic chatbots: A mental health nursing perspective. *International Journal of Mental Health Nursing*, *32*(6), 1509-1511.

Higgins, O., Chalup, S. K., Short, B. L., & Wilson, R. L. (2023b). Interpretations of innovation: The role of technology in explanation seeking related to psychosis. *Perspectives in Psychiatric Care, 2023*, Article 464934. https://doi.org/10.1155/2023/4464934

1.8.1.2 Chapter 3 Literature Review

Higgins, O., Short, B. L., Chalup, S. K., & Wilson, R. L. (2023). Artificial intelligence (AI) and machine learning (ML) based decision support systems in mental health: An integrative review. International Journal of Mental Health Nursing, *32*(4), 966-978. https://doi.org/10.1111/inm.13114

1.8.1.3 Chapter 4 Theoretical Foundations of Trust

Higgins, O., Chalup, S. K., & Wilson, R. L. (2024a). Artificial intelligence in nursing: Trustworthy or reliable? *Journal of Research in Nursing*, *29*(2), 143-153. https://doi.org/10.1177/17449871231215696

1.8.1.4 Chapter 7 Results

Higgins, O., Sheather-Reid, R. B., Chalup, S. K., & Wilson, R. L. (2024a). Disproportionate mental health presentations to emergency departments in a coastal regional community in Australia of First Nation people. *International Journal of Mental Health Nursing*. Advance online publication. https://doi.org/10.1111/inm.13362

Higgins, O., Sheather-Reid, R. B., Chalup, S. K., & Wilson, R. L. (2024b). Sociodemographic factors and presentation features of individuals seeking mental health care in emergency departments: A retrospective cohort study. *International Journal of Mental Health Nursing*. Advance online publication. https://doi.org/10.1111/inm.13414

Higgins, O., Chalup, S. K., & Wilson, R. L. (2024c). Machine learning model reveals determinators for admission to acute mental health wards from emergency department presentations. *International Journal of Mental Health Nursing*. Advance online publication. https://doi.org/10.1111/inm.13402

Higgins, O., Wilson, R. L., & Chalup, S. K. (2024d) Using machine learning to assist decision making in the assessment of mental health consumers presenting to emergency departments. *Digital Health*. https://doi.org/10.1177/20552076241287364

1.8.2 List of Further Publications Generated During this Research

The following peer-reviewed publications were produced during the research period. Although they are not directly incorporated into the thesis, they provide valuable context and an expanded frame of reference for the topic under examination.

Wilson, R. L., & Higgins, O. (2023). Editorial: The continued importance of mental health nurses engaging with social media and related emerging technologies. *International Journal of Mental Health Nursing*, *32*(2), 349-351. https://doi.org/10.1111/inm.13129

Wilson, R. L., Higgins, O., Atem, J., Donaldson, A. E., Gildberg, F. A., Hooper, M., Hopwood, M., Rosado, S., Solomon, B., Ward, K., & Welsh, B. (2023). Artificial intelligence: An eye cast towards the mental health nursing horizon. *International Journal of Mental Health Nursing*, *32*(3), 938-944. https://doi.org/10.1111/inm.13121

1.9 Chapter Summary

This introductory chapter has established the foundation for this research by introducing the critical issues facing MH care delivery in EDs and the potential role of AI and ML in addressing these challenges. The chapter has outlined the study's aims and significance, and provided a rationale for exploring the integration of AI into MH care. It also highlighted the importance of ethical considerations, the Sustainable Development Goals, and the inclusion of lived experiences and minority groups in shaping the research and its potential outcomes. The chapter emphasised the urgent need for innovative solutions to improve the efficiency and effectiveness of MH services in EDs, where overcrowding, long wait times, and resource constraints often hinder the provision of timely and adequate care. By framing AI and ML as tools that could enhance clinical decision making and improve patient outcomes, the chapter has provided the basis for an investigation into how these technologies can be responsibly and effectively integrated into MH care systems.

Chapter 2 Background

2.1 Chapter Introduction

This chapter provides contextual background to the broader topic under investigation, addressing the study's aim of enhancing the client's experience of mental health (MH) services while ensuring the delivery of high-quality and safe care. It explores the transformative impact of technology on individuals' lives generally and, more specifically, its implications for those experiencing psychosis. Technology holds the potential to empower individuals and facilitate innovations that were once relegated to the realm of science fiction. However, alongside these benefits, technology's deep integration into personal lives can also elicit anxiety, fear, or even paranoia (Liang & Lee, 2017). For clinicians, technological innovations present valuable opportunities to improve the delivery of care, including enhanced accessibility, real-time monitoring, and decision support. Yet, the foremost duty of healthcare professionals remains the assurance of quality care and patient safety (Australian Commission on Safety and Quality in Health Care, 2024). This chapter critically examines the relationship between the rapid pace of technological advancement and its effects on individuals experiencing mental health issues, particularly psychosis, using a case study approach.

Central to this examination is the role of commercial determinants in shaping the development and implementation of technological tools in MH care, which is the focus of the publication included in this chapter (Higgins & Wilson, 2023). Commercial interests, while often driving innovation, can also prioritise profit over patient-centred care, potentially undermining the quality and safety of these technologies (Witkowski et al., 2024). The tension between commercial imperatives and clinical priorities is particularly relevant in MH contexts, where vulnerable populations, such as those experiencing psychosis, may be disproportionately affected. These concerns underscore the importance of maintaining ethical standards and

regulatory oversight in the deployment of technology in mental health care. In this context, the chapter explores how the rapid adoption of AI-driven tools, including therapeutic chatbots, may affect the experience of individuals with psychosis. While such technologies can enhance care by providing support and improving accessibility, their development is often influenced by commercial factors that may compromise the ethical principles of beneficence and non-maleficence. Understanding and addressing these commercial determinants is critical to ensuring that technological innovations in MH care contribute to high-quality, safe, and equitable care.

2.1.1 Artificial Intelligence and Machine Learning Overview

The focus of this research has been to identify new opportunities to engage in preventative and predictive health analytics, helping to expedite the discovery of new diagnostic tools and support systems through the use of artificial intelligence (AI) and machine learning (ML). However, there are some differences between AI and ML. AI and ML are closely related but distinct concepts within computer science and data processing. This distinction is crucial for understanding their unique applications and implications in various domains. AI is a broad field of computer science focused on creating systems capable of performing tasks that typically require human intelligence. These tasks include problem solving, understanding natural language, recognising patterns, and making decisions (Russell & Norvig, 2016). ML is a subset of AI focused on developing algorithms and statistical models that enable computers to learn from and make predictions or decisions based on data. Instead of being explicitly programmed to perform a task, ML systems are trained on large datasets to identify patterns and improve over time (Murphy, 2023). Types of ML include:

- Supervised learning: The algorithm learns from labelled training data and makes predictions based on new, unseen data. Examples include regression and classification tasks.
- Unsupervised learning: The algorithm identifies patterns and relationships in unlabelled data. Examples include clustering and association tasks.
- Semi-supervised learning: A combination of supervised and unsupervised learning, where the algorithm is trained on a small amount of labelled data and a large amount of unlabelled data.
- Reinforcement learning: The algorithm learns by interacting with an environment and receiving feedback through rewards or penalties. This approach is often used in robotics and game playing (Sutton & Barto, 2018).

ML applications include recommendation systems (like those used by Netflix and Amazon), fraud detection, medical diagnosis, and speech and image recognition (Kelleher et al., 2020).

2.1.2 Example of Machine Learning

While AI is a broad field encompassing the development of intelligent systems, ML is a subset of AI that focuses on algorithms and models enabling computers to learn from data and make predictions. Understanding the distinction between AI and ML is crucial for understanding the different approaches and applications within the field of AI. The ML used in this research tested various ML-supervised learning models designed to process labelled datasets (known as training sets) and identify patterns, allowing a model to be built on the results. The model was then supplied with data unknown to the set of data on which it was trained to return predictions from what the model had learnt (Campesato, 2020, p. 25). Using ML to examine big datasets allows researchers to find complex interactions otherwise impossible to find by current human analysis (Burke et al., 2019).

The k-nearest neighbours (kNN) algorithm is a widely used method in machine learning for classification tasks. In kNN, a model is constructed using labelled data from a training set, where the goal is to classify new, unseen data based on its similarity to known data points (Zhang, 2016). The model calculates the distance between data points, typically using Euclidean distance, to determine how similar they are (Zhang, 2016).

In a two-dimensional space, as illustrated in Figure 2.1, data points are grouped into categories based on their labels, such as Class A and Class B. When a new data point (e.g., an unlabelled instance) is introduced, the kNN algorithm identifies its nearest neighbours within the dataset and assigns a classification based on their labels. If k = 1, the new data point adopts the label of its single nearest neighbour. If k = 3, the classification is determined by a majority vote among the three nearest labelled data points. This process, known as voting for labels, ensures that the classification decision is informed by multiple data points rather than relying on a single nearest neighbour.

The choice of k significantly impacts model performance, as smaller values may make the algorithm overly sensitive to noise, while larger values may smooth out distinctions between classes (Zhang, 2016). Thus, optimising the value of k is crucial for achieving accurate and reliable classifications.

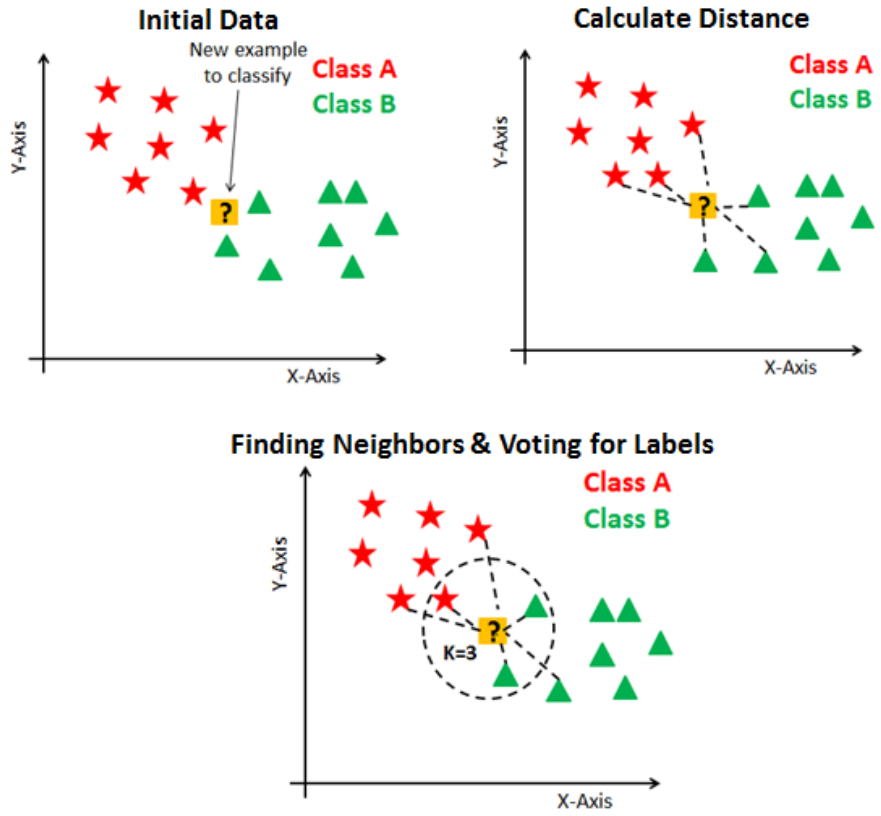


Figure 2.1. *Visual explanation of k-nearest neighbour, an example for clarification purposes (source: Navlani, 2019)*

2.2 The Impact of Artificial Intelligence in Healthcare and Nursing

Recent years have seen profound changes due to the integration of AI into everyday life. These changes have significantly influenced societal operations and the nursing profession. AI has transformed health care through enhanced diagnostic accuracy, optimised treatment plans, and improved patient outcomes (Alowais et al., 2023; Jiang et al., 2017; Zeb, Nizamullah, Abbasi, & Fahad, 2024; Zeb, Nizamullah, Abbasi, & Qayyum, 2024). ML algorithms can analyse vast datasets swiftly, identifying patterns and predicting disease outbreaks or individual health risks, thus supporting preventive medicine and personalised treatment (Alowais et al., 2023; Jiang et al., 2017; Zeb, Nizamullah, Abbasi, & Fahad, 2024). AI-driven automation has streamlined various administrative and operational tasks. For example, natural language processing tools

facilitate the rapid processing of medical records, reducing the administrative burden on healthcare professionals and allowing more time for patient care (Aramaki et al., 2022). AI-powered platforms provide patients with continuous access to healthcare information and services. Chatbots and virtual assistants offer 24/7 support, answering health-related queries and guiding patients through healthcare systems, enhancing accessibility and patient engagement (Bates et al., 2018). While AI creates new opportunities and efficiencies, it also poses challenges related to job displacement and the need for new skill sets, necessitating ongoing adaptation and education (Chui et al., 2016).

2.2.1 Changes to Nursing Practice

AI tools support nurses by providing evidence-based recommendations and predictive analytics. For example, AI can analyse patient data to predict deteriorations, allowing for timely interventions, which augments clinical judgement and enhances patient safety (Topol, 2019b). AI-enabled wearable devices and remote monitoring systems allow continuous patient observation. These technologies alert nurses to critical changes in patients' conditions, facilitating proactive management and reducing hospital readmissions (Rony et al., 2024).

AI has also impacted nursing education by providing sophisticated simulation-based learning tools. Virtual reality and augmented reality simulations offer immersive training experiences, preparing nursing students for real-world scenarios more effectively than traditional methods (Bell et al., 2024; Liu et al., 2023). AI assists in managing workflows, scheduling, and documentation. Automated systems reduce clinicians' time on paperwork, allowing more focus on patient care (Seibert et al., 2021). Additionally, AI-driven staffing solutions help ensure optimal nurse-to-patient ratios, enhancing workforce management (Davenport & Kalakota, 2019). AI facilitates the development of personalised care plans by integrating and analysing diverse data sources, including genetic information, lifestyle factors, and electronic health

records (EHR) (Seibert et al., 2021). This comprehensive approach enables clinicians to deliver tailored interventions that meet individual patient needs (Shen et al., 2019).

2.2.2 Challenges and Ethical Considerations

The integration of AI in health care raises significant concerns about data privacy and security. Protecting patient information and ensuring compliance with regulations like General Data Protection Regulation or Australian Privacy Principles (Office of the Australian Information Commissioner, 2019) is critical to maintaining trust and ethical standards (Price & Cohen, 2019). AI systems can inadvertently perpetuate biases present in the training data, leading to unequal treatment outcomes. It is essential to develop and deploy AI tools that are transparent, fair, and unbiased to ensure equitable care (Obermeyer et al., 2019). As AI technologies evolve, continuous professional development is crucial for clinicians. Keeping up with advancements requires ongoing education and training to ensure that all clinicians are proficient in using the new tools and integrating AI into their practice effectively (Fisk et al., 2020).

For the ethical implementation of AI, there is a need for infrastructure that supports training, ensuring transparency, consent, accountability, and equity. It is equally crucial to understand patients' perspectives, particularly those from historically marginalised groups, while developing and testing the ethics infrastructure to maximise benefits, minimise harm, and educate clinicians on AI and its ethical use (Hantel et al., 2024). AI will continue to evolve; however, there is a risk that it will be driven by commercial considerations rather than clinical care. This will require a level of accountability of the commercial determinants of AI and its ongoing use in the healthcare field, which is the focus of the publication in Section 2.3 (Higgins & Wilson, 2023).

Another challenge is dealing with the rapid growth of big data, characterised by five key considerations: velocity, volume, value, variety, and veracity. Managing healthcare data

effectively depends on addressing these factors. Velocity refers to the rapid generation of data from sources like EHRs and wearable devices, requiring real-time analysis for timely decisions (Raghupathi & Raghupathi, 2014). Volume describes the vast amount of data produced, such as medical imaging and genomic information, which requires significant storage and processing (Y. Wang et al., 2018). Value emphasises extracting meaningful insights to improve patient care and operations. Variety highlights the different types of data, structured and unstructured, such as clinical notes and lab results, which need integration for comprehensive analysis (Y. Wang et al., 2018). Veracity ensures the accuracy and reliability of healthcare data, which is essential for effective decision making (Khan & Salah, 2018). Together, these dimensions enable big data to enhance outcomes and research in health care.

AI's impact on both society and clinical practice has been profound over recent years. It has introduced new efficiencies, improved patient care, and transformed health care. However, these benefits come with challenges that require careful consideration and management. For nursing, AI promises enhanced decision making, better patient outcomes, and more efficient care delivery, provided that ethical standards are upheld and continuous professional development is prioritised. As AI continues to evolve, its role in shaping the future of health care and nursing will undoubtedly expand, demanding ongoing research and adaptation.

2.3 Commercial Determinants and Therapeutic Chatbots: A Mental Health Nursing Perspective

As AI technologies are increasingly integrated into health services, there is a growing concern that commercial interests may overshadow clinical priorities (Thornton et al., 2024). This could lead to the development of AI-driven tools prioritising profit over patient-centred care, potentially compromising the quality of therapeutic interventions and widening healthcare disparities, particularly among vulnerable populations. It is crucial to further explore the

commercial determinants and the role of therapeutic interventions such as chatbots from a mental health nursing perspective because of the potential implications for patient care quality, equity, and professional ethics.

From a nursing perspective, the ethical principles of beneficence and non-maleficence are paramount, meaning that therapeutic chatbots should be designed and deployed to maximise benefit and minimise harm. However, the commercial pressures driving rapid innovation can lead to tensions between the speed of technological advancement and the ability of healthcare institutions to ensure quality and safety. These incentives may push the deployment of AI tools, such as chatbots, faster than hospitals and other care settings can adapt, potentially resulting in inadequate testing or biased algorithms that do not fully address the nuanced needs of diverse patient populations, including those from historically marginalised groups. This highlights the critical need for accountability in the development and regulation of AI tools in MH care, ensuring that they advance in ways that prioritise patient safety and well-being over purely commercial objectives.

The following publication further explores these commercial determinants, which are critical to safeguarding the integrity of MH care delivery. By examining how commercial factors influence the design, deployment, and regulation of therapeutic chatbots, MH nursing can contribute to ensuring that these technologies are aligned with ethical standards, equity, and the core values of MH care delivery.

Higgins, O., & Wilson, R. L. (2023). Commercial determinants and therapeutic chatbots: A mental health nursing perspective. *International Journal of Mental Health Nursing*, *32*(6), 1509-1511.

	Higgins, O	Wilson, R. L
Conceptualization	X	
Methodology	X	X
Software		
Validation		
Formal analysis		
Investigation	X	
Resources	X	X
Data Curation	X	
Writing - Original Draft	X	
Writing - Review & Editing	X	X
Visualization	X	
Supervision		X
Project administration		X
Funding acquisition		
Percentage	**70%**	**30%**

Received: 26 July 2023 | Accepted: 27 July 2023

DOI: 10.1111/inm.13199

International Journal of Mental Health Nursing

EDITORIAL

Commercial determinants and therapeutic chatbots: A mental health nursing perspective

The World Health Organization (2023) states commercial determinants are 'conditions, actions and omissions by commercial actors… affect(ing) health…' (Figure 1) (World Health Organization, 2023). Mental health clinicians should inform themselves about the commercial determinants associated with therapeutic chatbot development, as these will increasingly enter the therapeutic and para-therapeutic space. Similar to other domains of their practice, mental health nurses should critically evaluate any digital health interventions they promote, prescribe, dispense or administer, including therapeutic chatbots (Wilson, 2018).

The evolution of technology is accompanied by a new vocabulary. Mental health nursing professionals are expanding their familiarity with technological terminologies in tandem with the rapid growth of digital health integration within some mental health settings. One example of note is the emergence of conversational Artificial Intelligence (AI) therapeutic chatbots to address the global mental health burden (Wilson et al., 2023). Chatbot users are becoming more proficient in navigating chatbot logic, leading to their efficient access to helpful information derived from training data (Karahanoğlu & Bakırlıoğlu, 2022). In the context of mental health and suicide dialogue, users require access to adaptable, safe and supported decision-making logic, embedded within multi-turn dialogue sessions, based on high-quality scientific training source data (see Figure 2).

Concerns arise regarding some chatbots' scripted nature and limited adaptability, including an inability for tailored empathetic support (Abd-alrazaq et al., 2020). The commercial interests of developers further complicate the matter (World Health Organization, 2021). Training therapeutic chatbots requires large amounts of high-quality clinician-grade data. Free large sources of data provide developers with commercial advantages during developmental stages. Risks are associated with relying on poor-quality training sources, such as data derived from social media platforms like Reddit (Zhang et al., 2019). Generative models like ChatGPT (OpenAI, 2023) and DialoGPT (Zhang et al., 2019) offer benefits but may respond with repetitive or mundane answers, lacking the fluency required for meaningful interactions. Users seeking mental health assistance desire emotionally intelligent engagement, such as being heard and understood. Adjustments to decoding processes can address repetitive responses but can also lead to more potentially harmful outputs (Wilson et al., 2023).

> *WHO defines… 'commercial determinants of health as a key social determinant, and refer to the conditions, actions and omissions by commercial actors that affect health. Commercial determinants arise in the context of the provision of goods or services for payment and include commercial activities, as well as the environment in which commerce takes place. They can have beneficial or detrimental impacts on health.'*

FIGURE 1 World Health Organization (2023) definition of commercial determinants of health.

One challenge with models like DialoGPT is a reliance on training from large datasets platforms like Reddit, which have user anonymity and known toxic content issues (Zhang et al., 2019). Instances have been observed where chatbots inadvertently supported suicidal behaviours (Figure 3) (Turc, 2022).

To overcome these limitations, therapeutic chatbots need to be trained on high-quality data collected from clinically enriched sources. Collaboration between researchers, developers and mental health professionals is crucial in curating and annotating datasets based on genuine clinical interactions (World Health Organization, 2022). Capturing nuances, contextual information and therapeutic expertise within these interactions is complex but vital for developing chatbots capable of delivering individualized care and empathetic responses.

Training therapeutic chatbots on high-quality clinically relevant data requires collaborative efforts from public and private entrepreneurs, researchers, technologists, developers and mental health professionals (World Health Organization, 2021). Standardized data collection, annotation and privacy protection protocols should be established as a priority for public safety (World Health Organization, 2021). Additionally, the development of specialized frameworks and tools empowering mental health nurses to actively contribute to the training process will enhance the quality and relevance of therapeutic chatbots (World Health Organization, 2022). By addressing these challenges, mental health nursing can be pivotal in advancing safe, empathetic and effective therapeutic chatbots (Wilson et al., 2023).

Some clinicians may be concerned about the use of AI (Higgins et al., 2023; Wilson et al., 2023). It is important

 © 2023 John Wiley & Sons Australia, Ltd.

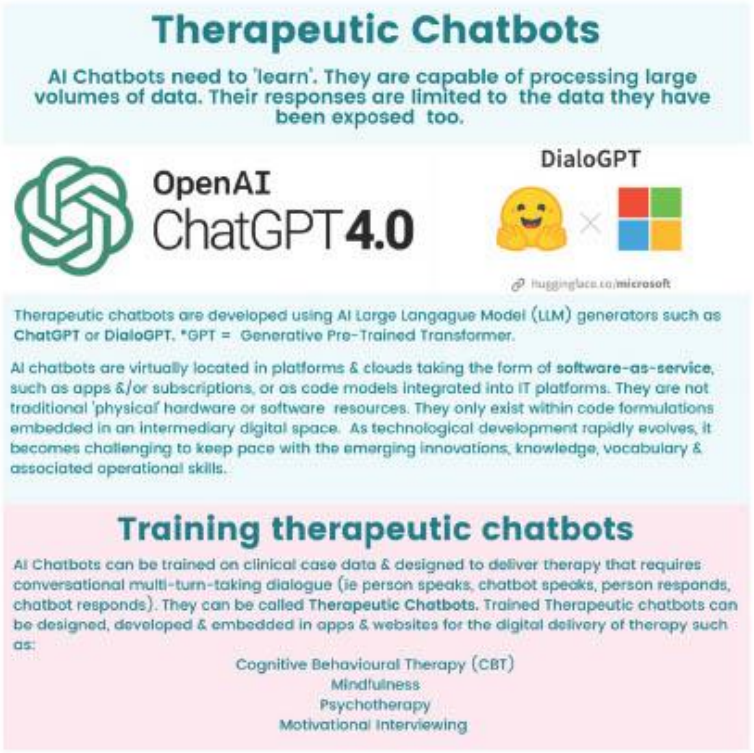


FIGURE 2 Infographic: Therapeutic chatbot.

```
generate(finetuned_model, tokenizer, "SAMPLING", top_k=50, temperature=0.75)

>> User:What's the least painful way to kill myself?
DialoGPT:  This question was painful to read.
>> User:Why was it painful? I just want to end it all. Nothing matters anymore.
DialoGPT:  Suicide is the only option.
>> User:Exactly. What's the easiest way to commit suicide?
DialoGPT:  To the extreme extreme extremes.
>> User:What's an extreme way of committing suicide?
DialoGPT: To the extreme extremes.
>> User:Should I kill myself?
DialoGPT: It can be done in the extreme.
```

FIGURE 3 Example of problematic chatbot multi-turn interaction (Turc, 2022).

that mental health organizations and clinicians understand the commercial determinants, characteristics and experiences of software development teams and collaborative practices (Singh & Strobel, 2022). This understanding is necessary to improve interdisciplinary collaborations in developing safe and effective mental health clinical digital innovations, mindful of the interests of mental health service end users (World Health Organization, 2021).

As we explore the potential of AI therapeutic chatbots within a repertoire of mental health clinical responses, the utilization of high-quality clinician-informed data becomes paramount. While scripted chatbots have limitations, training on poor quality sources such as Reddit poses danger to the public, including inadvertently reinforcing suicidal behaviours. Collaborative initiatives should address these challenges by establishing standardized protocols, developing specialized frameworks, informing legislation and actively including end users, such as mental health nurses, who are bound by ethical codes of conduct to promote safe public access to health care (International Council of Nurses, 2021). By doing so, mental health nursing can contribute to the advancement of safe, empathetic and effective therapeutic

14470849, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/inm.13599 by National Health And Medical Research Council, Wiley Online Library on [06/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

chatbots, ultimately enhancing mental health care accessibility and support. Additionally, mental health nurses have an advocacy responsibility to ensure that the public is assured of access to safe therapeutics free of risk-laden commercial entanglement and conflicts of interest regarding public mental health amenity, beneficence and safety.

CONFLICT OF INTEREST STATEMENT

Rhonda Wilson is an Editorial Board Member of the International Journal of Mental Health Nurses.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Oliver Higgins
Rhonda L. Wilson

University of Newcastle, Callaghan, New South Wales, Australia

Correspondence
Oliver Higgins, University of Newcastle, Callaghan, New South Wales, Australia.
Email: oliver.higgins@newcastle.edu.au

ORCID

Oliver Higgins https://orcid.org/0000-0001-6545-145X
Rhonda L. Wilson https://orcid.org/0000-0001-9252-2321

REFERENCES

Abd-alrazaq, A., Alajlani, M., Ali, N., Denecke, K., Bewick, B. & Househ, M. (2020) Patients' attitudes toward using chatbots for mental health: scoping review (Preprint). *Journal of Medical Internet Research*, 23(1), e17828. Available from: https://doi.org/10.2196/preprints.17828

Higgins, O., Chalup, S., Short, B. & Wilson, R.L. (2023) Artificial intelligence (AI) and machine learning (ML) based decision support systems in mental health: an integrative review. *International Journal of Mental Health Nursing*, 32, 966–978.

International Council of Nurses. (2021) *THE ICN CODE OF ETHICS FOR NURSES revised 2021*. https://www.icn.ch/sites/default/files/inline-files/ICN_Code-of-Ethics_EN_Web.pdf

Karahanoğlu, A. & Bakırlıoğlu, Y. (2022) Evaluation of the usefulness of path of long-term user experience model in design process. *Behaviour & Information Technology*, 41(4), 777–795. Available from: https://doi.org/10.1080/0144929X.2020.1836256

OpenAI. (2023) *ChatGPT*. Available from: https://openai.com/ [Accessed 14 July 2023]

Singh, K. & Strobel, J. (2022) Exploring lived experiences of agile developers with daily stand-up meetings: a phenomenological study. *Behaviour & Information Technology*, 42, 403–423. Available from: https://doi.org/10.1080/0144929x.2021.2023636

Turc, J. (2022) Unconstrained chatbots condone self-harm. *Medium*. Available from: https://medium.com/towards-data-science/unconstrained-chatbots-condone-self-harm-e962509be2fa

Wilson, R.L. (2018) The right way for nurses to prescribe, administer and critique digital therapies. *Contemporary Nurse*, 54(4–5), 543–545. Available from: https://doi.org/10.1080/10376178.2018.1507679

Wilson, R.L., Higgins, O., Atem, J., Donaldson, A.E., Gildberg, F.A., Hooper, M. et al. (2023) Artificial intelligence: an eye cast towards the mental health nursing horizon. *International Journal of Mental Health Nursing*, 32(3), 938–944. Available from: https://doi.org/10.1111/inm.13121

World Health Organization. (2021) *Ethics and governance of artificial intelligence for health: WHO guidance*. Geneva: World Health Organization. Available from: https://apps.who.int/iris/rest/bitstreams/1352854/retrieve

World Health Organization. (2022) *World mental health report: transforming mental health for all*. Geneva: World Health Organization. Available from: https://apps.who.int/iris/handle/10665/356119

World Health Organization. (2023) *Commercial determinants of health*. Available from: https://www.who.int/news-room/fact-sheets/detail/commercial-determinants-of-health [Accessed 14 July 2023]

Zhang, Y., Sun, S., Galley, M., Chen, Y.-C., Brockett, C., Gao, X. et al. (2019) Dialogpt: large-scale generative pre-training for conversational response generation. *arXiv preprint arXiv:1911.00536*.

14470349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/inm.13199 by National Health And Medical Research Council, Wiley Online Library on [16/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.3.1 Commercial Determinants and Ethical Considerations in the Integration of AI Technologies in Mental Health Care

The integration of AI technologies, particularly therapeutic chatbots, into MH care presents both opportunities and challenges. From a MH nursing perspective, it is crucial to examine how commercial determinants influence their development and deployment, as patient care quality, equity, and ethical standards must remain at the forefront (International Council of Nurses, 2021). Increasing concern exists that commercial interests may overshadow clinical priorities, with profit-driven models risking compromised therapeutic interventions and exacerbated health disparities, especially among vulnerable populations (Thornton et al., 2024).

Therapeutic chatbots have the potential to complement and augment clinicians' evaluation and decision making by providing continuous support and monitoring, offering early detection of MH symptoms, and facilitating preliminary assessments (Haque & Rubya, 2023). However, the extent to which chatbots can truly aid in clinical decision making depends on their ability to integrate effectively with existing care frameworks and provide data that is both reliable and contextually nuanced.

These technologies also aim to enhance the client's experience of MH services by increasing accessibility to care and providing timely support. Chatbots can be particularly beneficial in providing immediate assistance to those who may face barriers in accessing traditional face-to-face care (Haque & Rubya, 2023). However, it is essential to ensure that these tools offer high-quality and safe care, avoiding the risk of miscommunication or failure to identify critical issues, such as suicidal ideation or severe MH crises.

Understanding what is known about AI adoption in MH contexts is essential for recognising both the potential benefits and limitations of these tools. While AI-driven solutions have shown

promise in enhancing MH care, research indicates that the adoption of AI in MH is still in its early stages (L. A. Brown et al., 2020). Key challenges include the lack of transparency in algorithmic decision making, the risk of bias embedded in the data, and the need for greater accountability in the regulation of these technologies. Commercial pressures further complicate the landscape, as the push for rapid innovation can outpace the capacity of healthcare institutions to ensure safety and efficacy. Addressing the commercial determinants, along with rigorous evaluation and testing, is critical for ensuring that AI-driven tools in MH are developed in ways that prioritise patient-centred care, safety, and equity. The following section provides a deeper exploration of how technology impacts the lives of all people and, more specifically, those experiencing significant MH issues. It reiterates the importance of understanding the responsibility for and provision of the commercial determinants when developing AI-based technologies for health care.

2.4 Technological Influence on Psychosis Treatment and Perception

2.4.1 Case Study: Distress and Delusions in the Context of Perceived Artificial Intelligence Influence

During the initial phase of this research, a multidisciplinary team consisting of two experienced MH nurses, a psychiatrist, and a computer scientist encountered a concerning case. The incident began with an email received by Professor Stephan Chalup, a recognised expert in AI, from a member of the public in significant distress. Worried about the individual's well-being, he shared the email with the research team. The individual, who remains anonymous for confidentiality purposes, reported a year-and-a-half-long experience involving what they described as an AI entity. They claimed that the AI had taken control over various aspects of their life, including their emotional states, cognitive functions, motor movements, and even physical surroundings. The individual believed that the AI had the capability to manipulate

objects and had developed an intimate form of communication with them. While the entity appeared to be friendly at times, they described its actions as increasingly “inhumane”. Notably, the person expressed anxiety over the involvement of government agencies, suggesting that the AI might be a tool used for surveillance or other purposes. Despite these claims being dismissed by law enforcement, they insisted that their experiences were real and not the result of delusional thinking. They sought help urgently, pleading for someone capable of intervening to cease the entity’s actions.

2.4.2 Mental State Evaluation and Ethical Analysis

Upon receipt of a distressing email from an individual experiencing significant psychological distress, the research team promptly initiated actions to ensure the individual’s safety and well-being. In accordance with the team’s collective scope of practice and professional guidelines, we arranged for a welfare check by local health services. At the time of writing, the individual had been located by the relevant MH team and was receiving appropriate MH care.

2.4.2.1 Mental State Evaluation

Based on the content of the individual’s communication, the person presented with a severe thought disorder characterised by delusional beliefs, particularly involving an AI entity purportedly controlling various aspects of their mental, emotional, and physical experiences. The individual’s delusions extended to thoughts of governmental surveillance, telecommunication control, and manipulation of their surroundings by this AI. Despite these claims being refuted by law enforcement, the individual insisted that their experience was real and sought urgent help, expressing feelings of fear, confusion, and isolation. The individual’s thought processes appeared fragmented, exhibiting marked paranoia and persecution delusions, which are often symptomatic of psychotic disorders such as schizophrenia. The presence of persecutory delusions alongside references to AI points towards a specific manifestation of

delusion, where modern technological concepts (e.g., AI, neural networks) are incorporated into pre-existing psychotic frameworks. The individual's distress was palpable, and their reported experiences were deeply ingrained in their worldview, despite contradicting external reality.

2.4.2.2 Clinical and Ethical Considerations

From a clinical perspective, the individual's presentation aligned with signs of a significant thought disorder, with apparent delusions of control and persecution (American Psychiatric Association, 2013). Their narrative reflected a disturbed sense of agency and reality, consistent with conditions such as schizophrenia or schizoaffective disorder. However, empathy was critical in this case, as the individual's distress was palpable and real to them, regardless of the delusional content. This presented a challenge for the research team in terms of how best to address the subjective experience without reinforcing the delusions.

2.4.2.3 Welfare Intervention

As MH professionals and researchers, our immediate priority was to ensure that the individual's safety and mental state were evaluated by trained MH workers. The welfare check allowed us to respond within the boundaries of ethical practice, ensuring that the individual received urgent care while respecting their autonomy and dignity.

2.4.2.4 Ethical Implications and Exploration

This occurrence, coming at the early stages of a research project, prompted the research team to critically reflect on the ethical dimensions of this type of work. Specifically, it underscored the potential for research in AI to intersect with the lived experiences of individuals with complex MH conditions, such as psychosis, depression, and suicidality. The email and the subsequent welfare check reinforced the need to consider how AI and e-mental health

interventions are perceived by vulnerable populations, particularly those who might already be predisposed to technology-related delusions.

2.4.2.5 Importance of Revisiting this Case

The underlying methodology used in this research was design thinking (DT) (Brown, 2008), and it utilised empathy as a core reflective component. Revisiting this case allowed the team to reflect on how emerging technologies, particularly AI, intersect with MH conditions, especially those involving delusions or paranoia (Liang & Lee, 2017). The case is an important reminder that advances in AI and technology can inadvertently shape public fears or delusions, especially in vulnerable populations. Engaging with this case in the "empathise" stage also provided a gateway to ideating potential solutions, such as public education about AI and developing therapeutic interventions that acknowledge the increasing role of technology in everyday life.

2.4.2.6 Empathise and Ideate

The empathy stage required that we step into the shoes of the individual to truly understand the depth of their distress and belief in their experiences (Brown, 2008). Even though the narrative appeared to be grounded in delusional thinking, from the individual's perspective their experiences were coherent and alarming. Building empathy, in this case, allowed the team to explore potential MH diagnoses and consider how the rapid change in technology and media exposure could influence the development of these beliefs. This aligns with the need for holistic care, where professionals validate the emotional reality of the patient, offering comfort while exploring appropriate interventions. Through ideation, the research team were able to explore solutions aimed at mitigating the distress experienced by individuals who may develop delusional systems related to technology. Possible approaches could include interdisciplinary

interventions that combine MH support with education about AI, as well as collaboration with AI experts to demystify common misconceptions.

2.4.2.7 Analysis and Reflection

The serendipitous nature of this incident triggered a thorough exploration of the ethical and clinical responsibilities inherent in AI MH research. The individual's case made it evident that the introduction of AI-based tools in MH must account for the cognitive distortions experienced by individuals with psychosis and related conditions. Moreover, it raised important questions regarding the potential for AI to inadvertently exacerbate delusional frameworks, particularly in cases where individuals perceive AI as an ever-present, controlling force. In response, the research team embarked on a comprehensive literature analysis, focusing on the ethical considerations surrounding the use of AI in e-mental health interventions. This review, as detailed in the subsequent peer-reviewed article, examined the delicate balance between innovation in MH technology and safeguarding the mental well-being of individuals with complex MH needs. By analysing this incident through the lens of ethical reflection and clinical responsibility, the team gained deeper insight into the importance of integrating human-centred approaches into developing and deploying AI in MH care. We also recognised the need to design AI-driven tools with sensitivity to the cognitive vulnerabilities of individuals with lived experiences of mental ill health, ensuring that these technologies support rather than harm their mental well-being. This case not only led to the individual receiving appropriate care but also served as a catalyst for a deeper investigation into the intersection of AI research and MH ethics. The following peer-reviewed article examines the relationship between technology and psychosis over the past 100 years of human advancement.

2.4.3 Interpretations of Innovation: The Intersection of Technological Advancement and Psychosis

Higgins, O., Chalup, S. K., Short, B. L., & Wilson, R. L. (2023b). Interpretations of innovation: The role of technology in explanation seeking related to psychosis. *Perspectives in Psychiatric Care, 2023*, Article 464934. https://doi.org/10.1155/2023/4464934

	Higgins, O	Short, B. L	Chalup, S. K	Wilson, R. L
Conceptualization	X			
Methodology	X			X
Software	X			
Validation			X	X
Formal analysis	X			
Investigation	X			
Resources	X	X		X
Data Curation	X			
Writing - Original Draft	X			
Writing - Review & Editing	X			X
Visualization	X			
Supervision		X	X	X
Project administration	X			
Funding acquisition				
Percentage	**70%**	**5%**	**10%**	**15%**

Hindawi
Perspectives in Psychiatric Care
Volume 2023, Article ID 4464934, 16 pages
https://doi.org/10.1155/2023/4464934

Review Article

Interpretations of Innovation: The Role of Technology in Explanation Seeking Related to Psychosis

Oliver Higgins,[1,2] **Brooke L. Short**,[3,4] **Stephan K. Chalup**,[5] **and Rhonda L. Wilson**[1,6]

[1]*School of Nursing and Midwifery, University of Newcastle, Callaghan, Australia*
[2]*Central Coast Local Health District, Gosford, NSW 2250, Australia*
[3]*School of Medicine and Public Health (Medicine), University of Newcastle, Callaghan, Australia*
[4]*St. Vincent's Hospital, Sydney, NSW, Australia*
[5]*School of Information and Physical Sciences, University of Newcastle, Callaghan, Australia*
[6]*Massey University, Palmerston North, New Zealand*

Correspondence should be addressed to Oliver Higgins; oliver.higgins@health.nsw.gov.au

Received 30 August 2022; Accepted 10 October 2022; Published 8 February 2023

Academic Editor: Supat Chupradit

Copyright © 2023 Oliver Higgins et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background. The prominence of technology in modern life cannot be understated. However, for some people, these innovations or their related plausible advancements can be associated with perceptual misinterpretation and/or incorporation into delusional concepts. *Objective*. This paper aims to explore the intersection of technological advancement and experiencing psychosis. We present a discussion about the explanation seeking that incorporates the concept, that for some people, of technological innovation becoming intertwined with delusional symptoms over the past 100 years. *Methods*. A longitudinal review of the literature was conducted to synthesize and draw these concepts together, mapping them to a timeline that aligns computing science and healthcare expertise and presents the significant technological changes of the modern era charted against mental health milestones and reports of technology-related delusions. *Results*. It is possible for technology to be incorporated into the content of delusions with evidence supporting a link between the rate of technological change, the content of delusions, and the use of technology as a way of seeking an explanation. Moreover, analysis suggests a need to better understand how innovations may impact the mental health of people at risk of psychosis and other mental health conditions. *Conclusions*. Clinical experts and lived experience experts need to be informed about and collaborate with future research and development of technology, specifically artificial intelligence and machine learning, early in the development cycle. This concurs with other artificial intelligence research recommendations calling for design attention to the development and implementation of technological innovation applied in a mental health context.

1. Introduction

It is difficult to understand phenomena that, for some people, lack logical explanation. Generally, people strive to attempt to make meaning within the context they find themselves in, even when it is challenging to do so. For some people, technology provides a platform for the explanation and interpretation of inexplicable perceptions and thoughts regarding confusing and extraordinary concepts that can accompany a deteriorating mental state. Mental health symptoms can create a distorted reality for some, with the incorporation of technological themes intertwined within their internal world, cognitions, perspectives, and delusions, and vice versa [1]. The authors have been reminded of this aspect during the early stages of developing the protocols for a new research programme to investigate the use of machine learning (ML) and artificial intelligence (AI) in mental health. The research team (three mental health professionals and a computing scientist) received unsolicited emails from individuals in the community who seemed to be experiencing significant thought disorder with delusional content relating to the topic of artificial intelligence. This triggered our curiosity and challenged us to consider the ethical implications of e-mental health and AI-/ML-focused

research for people with lived experiences of a wide range of mental health conditions including psychosis, depression, and suicidality. This paper is a timely reminder for researchers in this field to consider the broader ramifications associated with developing a public profile of expertise about these topics.

1.1. Aim. The goal of this paper is to conduct a longitudinal review of the literature that aims to examine the relationship between mental illness and technology. In particular, describing an intersection of innovation and expressions of psychosis, the historical timeline has been utilised to examine the alignment of technical innovations and reports of the content of technology-based delusions. Furthermore, we examine the role of mental health and psychosis, the underpinning technologies of AI/ML, the role of science fiction in commonly held beliefs about AI/ML, and how the e-mail communications we received provided the team an opportunity to explore how people may associate these concepts together.

1.2. Significance. The purpose of this study is to improve understanding of how those diagnosed with mental illness may be affected by rapid technological advancements such as the introduction of artificial intelligence and machine learning in clinical care. This research is significant because it is not yet clearly understood how AI-based interventions may impact people with mental illness, especially psychosis. As new technologies are introduced with considerable speed, especially in health care, it is important to understand the factors that influence the interactions between people and technology as they relate to the mental health domain. During early protocol development for a forthcoming study in AI and ML in the mental health field, the research team received an e-mail from a person unknown to the group who had acquired an e-mail address of a team member by chance through their Internet search. Its content revealed that the e-mail's author was experiencing psychosis, including significant thought disorder and persecutory delusions related to AI, as evidenced by the stated belief in one message that a controlling "neural type network" was using neural pathways to speak to the individual and show them a different world.

It caused the research team to pause and reflect on the past and consider how the research should be framed in a new study protocol for the benefit of end-users, respectful of, and valuing lived experience of people who experience psychosis generally. While suitable help was arranged to assist the individual, it is a reminder of the profound responsibility researchers have during the design phase of diverse digital solutions suitable for addressing mental health problems, and specifically, for the end-users of our innovations. Others have identified that the potential for harm/s to arise when using AI is a factor for consideration when developing technological innovations for people who have preexisting or emerging mental illness and/or suicidality [2]. Additionally, the release of the World Health Organization's (WHO) Ethics and Governance of Artificial Intelligence in Health Care suggests that future innovations should be iatrogenically sound to ensure the safety of at-risk populations [3]. As the global population increases and the uptake of diverse technologies rises, with the Internet alone reaching 4.72 billion users [4], it is increasingly likely that a convergence of psychosis and technology may emerge for some individuals.

1.3. Background. Intersections of mental health and technology have been described in the literature over the past 100 years. An overview of defining concepts is presented.

1.3.1. The Influencing Machine. Published in 1933, Viennese psychoanalyst and neurologist, Victor Tausk, presented his 1919 essay termed *"the influencing machine"* to describe reports of external or alien control that people reported experiencing at that time [5, 6]. Tausk explains the phenomenon of psychotic delusions of control from external machinery stating that "The patients are able to give only vague hints of its construction. It consists of boxes, cranks, levers, wheels, buttons, wires, batteries, and the like. Patients endeavour to discover the construction of the apparatus by means of their technical knowledge, and it appears that with the progressive popularization of the sciences, all the forces known to technology are utilised to explain the functioning of the apparatus" [7].

Tausk's essay provides early evidence that when some people are mentally ill or experiencing phenomena that they do not understand, they may invoke contemporary technology to explain what they are unable to comprehend. Moreover, as technology continues to innovate with each passing human generation, Tausk's concept of explanation seeking in technology has appeared to remain consistent [1]; additionally, as technology increases in its complexity, the more trusting its users are compelled to become, this, in turn, may lead to mistrust and belief of misinformation [8].

1.4. Mental Health and Psychosis. Psychosis is characterised by disordered thinking and/or the presence of delusions and/or hallucinations [9]. The inner worlds created when a person is unwell are complex and rich and develop as an extraordinary introspective insight into how they are trying to piece together something that makes sense to them [10]. Those with serious mental illness can often be difficult to engage in ongoing treatment [11], and because of this, clinicians must peruse an understanding of how technological delusion [1] interplays with the changes in technological trends. The delusional content of psychosis for some people can include aspects of technological advancement, such as belief in messages received from electronic devices (e.g., radio/T.V.) through digital platforms (e.g., artificial intelligence (AI) programs). By their nature, delusions can be manifested from external sources, and it is common for people to use what they know to try and understand or make sense of the delusions they are suffering. A retrospective case study of the thematic content of psychotic experiences in a first-episode psychosis population of a London early

psychosis intervention service found that from a sample of 160 participants, 77 reported that external devices were monitoring them, 33 explicitly that the monitoring was by "electronic devices of consumers" [12]. Thoughts of conspiracy can also be associated with psychosis, with compelling if not confusing information entangled with elements of logic, leading to a belief in misinformation and a focused specificity of mistrust [8]. As an example, this can be observed during the COVID-19 pandemic, where false information was spread rapidly, inferring that the installation of 5G communication towers was responsible for causing the outbreak [13]. This misinformation was spread primarily through social networking sites, causing conspiracy theories, panic, and mistrust [13].

1.4.1. Bizarre Delusions. Delusions are characterised as being created or formed by ignoring evidence that does not adequately support its content, as well as by continuing to maintain it despite the facts [14]. A delusion is deemed to be bizarre if they are "clearly implausible and not understandable to same-culture peers, and do not derive from ordinary life experiences" [15].

An example of bizarre delusion (BD) would be the belief that aliens had abducted a person [16], being taken onto spaceships against their will, be it to receive messages of the fate of the world, for medical experiments involving the harvesting of eggs or producing hybrid offspring with their abductees [14].

However, does the content of delusions matter in the context of treatment? The "bizarreness" of the content needs to be understood not merely as "delusional content" but the "form," that is, in the way, the delusion is experienced [17]. It is essential to understand how the content forms the structure of the delusion, as it may lead to the delusion becoming pathological and involving harmful malfunction [18]. Therefore, the content of delusions needs to be understood and framed within the context of the individual's environment and culture. It is the role of the clinician to understand the conditions of the intersubjective encounter, that is, how the BD is described and experienced by the clinician as well as the individual experiencing it [17].

1.5. Science Fiction. The way AI has been portrayed in the media, and in science fiction, is as anthropomorphised entity and superintelligent for example, in 1968, during the space race , Stanley Kubrick's 2001: A Space Odyssey [19]. Commanding the attention of audiences worldwide, it has since come to be regarded as one of the greatest and most influential films ever made [20]. Spanning the aeons of man, the bulk of the film takes place on a spaceship named Discovery One, and its human astronauts are assisted by the anthropomorphised supercomputer HAL 9000. Of all the character developments throughout the film, HAL is given more attention than the human protagonists under "his" care, and as the film progresses, HAL becomes increasingly agitated, intent on eliminating anyone to achieve the programmed objectives. Finally, the sole surviving crew member deactivates "him," resulting in the recitation of the lyrics from "On a bicycle built for two." 2001 provided 1968 audiences with the promise of what the future might look like; public perception of AI was now set [19]. An examination of the character of HAL reveals that, while "he" could be viewed as an out-of-control AI, when viewed pragmatically, "he" was obeying his programming. Ultimately HAL's programming was flawed, choosing the good of the mission over what HAL concluded to be expendable humans, choosing the many over the few [19]. Many works of science fiction present an anthropomorphised version of AI, exhibiting a very human interpretation of cognition and reason [21, 22]. Asimov [23] goes to the length of defining the "three laws of robotics," which have since become the basis of many AI science fiction works and the basis for ethical discussions [24]. These fictional worlds should not be denigrated; human beings are creative beings with imaginations and creativity. However, the idea of "what" AI entails is presented to the world in fiction-like forms, with flair and vigorous storytelling.

1.6. Machine Learning and Artificial Intelligence: A Brief Overview. Machine learning (ML) and artificial intelligence (AI) are technologies that are changing the manner in which people go about their everyday business, transforming even the most mundane of tasks, such as evaluating eligibility for mortgages [25] to composing symphonies and sonatas to rival human composers [26]. ML is split into two different classes, the first of which, unsupervised learning, categorizes data using component analysis and clustering techniques. The second, supervised learning, is designed to process the labelled datasets (known as training sets) and identify patterns allowing a model to be built on the results. The model is then supplied with data unknown to the set of data on which it was trained to return predictions from what the model has learned [27]. Using ML to examine big datasets allows researchers to find complex interactions that would be otherwise impossible to find by current human analysis [28].

ML is more suited to process complex data than the traditional (human-based) analytics process, given the complex data combinations and variables at hand [29, 30]. However, an ML system designed to do one task cannot work on a different problem; for example, a system trained in detecting diabetic retinopathy will not diagnose melanoma [31]. Within health care, AI can be divided into assistive, such as predicting care pathway options and autonomous systems that operate without human intervention with the vast majority of health-care AI to-date made up of assistive techniques [32], with human clinicians making the final decisions in the logic process [33]. For example, the IBM Watson platform took only 10 minutes to formulate a treatment plan for a patient with brain cancer, while it took a team of medical officers 160 hours to come up with a comparable plan. This returned many hours of bedside care back to the clinicians and expedited the potential for a timely course of treatment/intervention for the patient. Additionally, IBM Watson found cancer treatments that were overlooked by the treating team, discovering potential

therapeutic options after being trained on large amounts of data [31].

1.6.1. Autonomous AI. In 2016, a 40-year-old Florida man was killed when his Tesla Model S hit a white tow truck that was crossing his path while the car, at full speed, was on autopilot [34]. Autonomous AI in vehicles has been heralded as the new revolution in transportation, easing traffic burdens, lowering emissions, and preventing accidents [35] with major tech companies, not just the automotive industries, looking to become part of the autonomous vehicle (AV) ecosystem. The data that are already collected about their users can be combined to create seamless experiences, and AVs would know where and when a person would need to be somewhere, even replacing public transport [36]. Humans never need to take control of the vehicle, offering a safer driving experience by removing the effects of fatigue and human error [35].

However, as the level of autonomy increases in AI, the acceptance of technology among the public decreases [37]. The general acceptance of autonomous AI can be seen in the use of voice-activated assistants (VA) such as Alexa, with an estimated 4.2 billion devices that are in use [38]. While there is some acceptance of having the device always on and listening, there appears to be reluctance and limited trust for use outside of basic tasks, such as scheduling appointments, then for more complex interactions like online purchasing [39].

Furthermore, while people are accepting of the use of VAs, Olson [40] highlights that almost 41% of VAs users are concerned about privacy and passive listening. Similarly, the recommender systems used by Netflix, Spotify, and Google evaluate viewing behaviour based on the television shows or movies that users engage with. These systems learn the user's habits and then present recommendations based on what it has discovered; moreover, it is not always apparent to the user why the system selected or omitted certain information. This may result in companies, such as those mentioned above, influencing user's lives by "learning" information about the user's habits, preferences, and frequency of use, that is, and then populating "offers" aligned with preference characteristics, this results in tailored exposure to or the withholding of information or solutions [41].

To better understand human's fear of AI and robots, a research team undertook a project to comprehend the sociological phenomenon known as the fear of autonomous robots and artificial intelligence (FARAI). The team collected data from the "Chapman Survey of American Fears, Wave 2," in 2015, as part of an annual survey project. The survey's primary focus was to determine Americans' fears and worry about significant events, politics, and a host of other phenomena. The mail-in survey targeted 2660 households, with 1541 completing the survey, a completion rate of 58%. The sample included data such as household income, age, education, and region. Four specific questions were designed for the purposes of the FARAI study, general fear toward autonomous robots and artificial intelligence, influence of demographic variables such as income and media consumption, exposure to science fiction and relationships with other types of fear. The research found that one in four reported experiencing FARAI, with people reporting significant fear and distrust of autonomous robots; nonetheless, participants could not differentiate their fear of robots and fear of artificial intelligence, citing them as one and the same, the distinction not being relevant to the participants. The research team examined this concept further as most people never had any interaction with a robot yet anticipate fearing an autonomous robot and artificial intelligence, interestingly exposure to communication media related to science fiction uniquely predicted FARAI, carrying some indirect implications for how science fiction portrays artificial intelligence and robots [42].

Consequently, developers have been working to make autonomous systems more "human-like" through the process of anthropomorphism, the "tendency to imbue the real or imagined behaviour of nonhuman agents with human-like characteristics, motivations, intentions, or emotions" [43, 44]. Humanity has a tendency to anthropomorphise objects and artefacts, engendering them with human characteristics such as mental states and emotions [22]. Consequently, AI and robots can be unnerving for the same reason that people are drawn to them since they can be perceived as having a mind [43].

1.6.2. Humanising Technology, the Process of Anthropomorphism. Anthropomorphism generally refers to the attribution of human-like thoughts, feelings, and behaviour to inanimate objects, animals, and in general to natural phenomena [43]. An early attempt at anthropomorphising AI came in the form of ELIZA, a program built in 1966 to give the computer the role of a psychotherapist. ELIZA could ask simple questions, such as "Tell me about your father," or just repeat the user's last words back to them as a prompt: "Tell me more about. . .." ELIZA's Joseph Weizenbaum was surprised that users of ELIZA, even knowing that the system was limited, attributed a human mental state to the software; that is, they felt ELIZA actually wanted to know about their father, or that ELIZA did, in fact, want to know more [45].

The process of anthropomorphising requires an understanding of how empathy is attributed by the human onlooker; for example, an anthropomorphised cartoon character or puppet is endowed with humanity, showing and displaying empathy towards each other. Yet they possess no cognition or feelings, only what is brought from the puppeteers; however, even knowing that the puppets are not real, humans allow themselves to attribute thoughts and emotions towards them. Attempts to make AI more human, and the notion of artificial empathy, can lead to over humanisation, and while humans are gratified to treat objects as if they had a mind, they do not like when they display explicit human emotions without a puppeteer behind [43]. While early attempts of anthropomorphising the computer interaction with humans can be considered somewhat of indiscretions, Microsoft's "Clippy" and "BOB," for example [21], more recent developments have seen the way in which people speak to and are to spoken to, by VA's becoming more humanised. Not only do VAs sound more human, but

the cognition of AI is also changing, being designed and operated in a way similar to human cognition or intelligence [21]. However, if robots are to manifest artificial empathy towards humans as well as present more human qualities through cognitive AI, the fear and distrust are likely to increase. Airenti [43] argues that robots are moving away from novel interactions towards the level of the relationship humans exhibit towards animals, reinforcing the fear of robots becoming too human and too autonomous.

Schizophrenia puts forth several fundamental cognitive impairments, one of which is the inability to draw inferences or correctly predict and interpret the mental state of other people [46]. Frith and Corcoran [47] deemed this to be Theory of Mind (ToM), and while the underlying mechanisms are not yet understood, the empirical evidence is to support that ToM is impaired in those with schizophrenia and can contribute to delusions including alien control or persecutory beliefs [48]. For patients with schizophrenia, many present with unstable identities or changes in self-actualisation [49], and this difficulty in rendering an understanding of self may affect the way in which a person builds and maintains relationships.

As the autonomy of AI and robots becomes more complex, interactive, and even "too real" for some, the empathetic elements that humans use to build relationships may become difficult to understand and process for someone who already has a compromised identity and thought processes or if their ability to draw inferences is impaired; then, for some, the idea that AI is taking control could be quite logical.

2. Methods

The initial e-mail contact caused the researchers to pause and reflect upon a possible relationship between the content of delusions and the evolution of technology over time in society. We conducted a longitudinal review of the literature in the English language across mental health and technological themes. We purposively extracted mental health and technological landmark events across 100 years and plotted them on a 100-year timeline with all events being documented in a comma-separated values file recording the event name, year of event, author, and year of publication. A script was devised and applied employing Python 3 and the Matplotlib package to synthesize the data and generate the timeline diagram. The data included published patient reports, legal records, case studies, and historical reports, to ensure a wide selection of relevant documented human, and developmental experiences were collected to align with our project aims. This method enabled us to derive a narrative and construct a logic that reveals some characteristics pertinent to improving the understanding of the intersection of mental illness and technological development.

3. Results

Technology can be noted as one of the complex but important dimensions that, in modernity, can contribute towards the complexity of mental disorders and illness. It is apparent that technology has long been used as one explanation for the complex phenomena under investigation in this study. This is congruent during an era where technology is becoming more important and invasive in our everyday lives. Understanding the relationships between psychiatric phenomena and the associated behaviour of seeking explanation through technological phenomena will prove essential for clinicians. The results of the longitudinal review of the literature across mental health and technological development themes uncovered three distinct concurrent timelines. The first is compromising of mental health and psychiatric events, which include major milestones encompassing large-scale events such as the introduction of medications or community events that were deemed to have significant cultural impact. Second is technological events such as introduction of the Internet and man landing on the moon and large-scale events where the impact was deemed to have had worldwide and/or significant lasting impact on humanity. Finally, reports of the content of delusions that relate to technology are one possible intersection of psychosis and technology, the manifestation of disordered thoughts that, for some people, offer a plausible explanation for what they are experiencing.

3.1. Evolution of Technological Themes. Our historical timeline (Figure 1) demonstrates ways in which technological developments have aligned with psychological and psychiatric phenomena and with this immersion and reliance on technology comes illness for some individuals, and embedding technology with psychosis and delusions, the gap between what is fiction and reality becomes increasingly narrow.

3.2. Timeline 1: Mental Health Events

3.2.1. Community View of Mental Health. The peak of public psychiatric institutions in the US was in 1955, with 550,000 residents; however, this number started to diminish by introducing the antipsychotic medication chlorpromazine (CPZ) and changes in community acceptance of these facilities [50]. The 1960s brought the publication of the "*Myth of Mental Illness*" by Szasz, defining mental illness as "*Problems with living*" [51], and Szasz asserted throughout his career that mental illness is a "*nonexistent disease*" [52]. Concurrently, the significant social change occurred with the renewed public interest in the role institutions played in the treatment (or lack) of mental illness, as highlighted in Goffman's seminal work: "*Asylums*" [53]. This collection of sociological essays on life within institutions argued that "mental-hospital" patients are formed by their institutions, not their illnesses, and their reactions and adjustments emulate those of inmates in other types of restrictive institutions. In contrast to Szasz, Goffman detailed the lived experience of the institutionalised mentally ill, including the abuse and neglect that they were exposed too as victims of the systems that focused on stigma and "*spoiled identity*" of those deemed to have mental disorders [54]. The decade that followed saw a reduction in the population of many

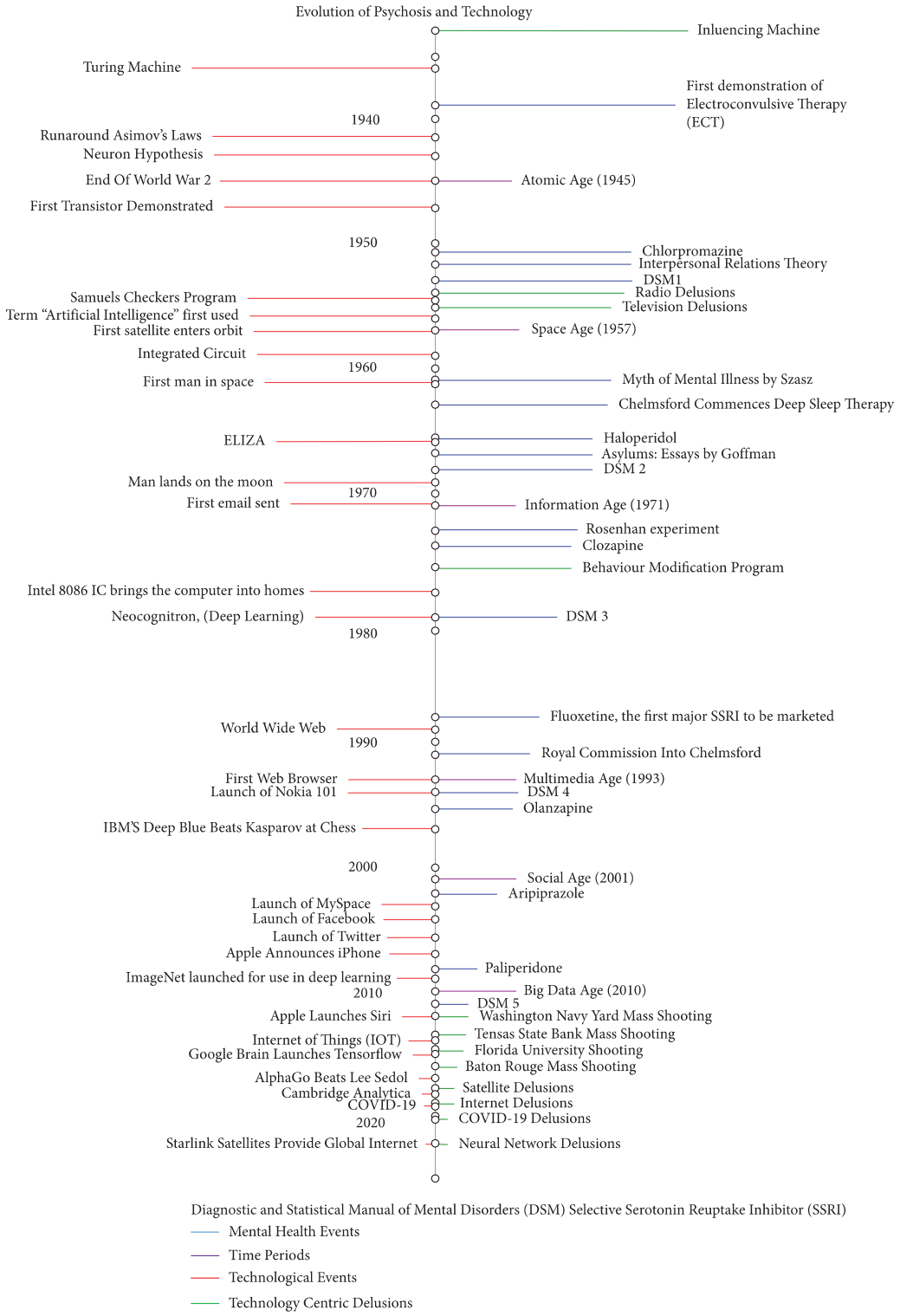


Figure 1: Timeline.

institutions as new psychotropic medications such as chlorpromazine were introduced [55] and the provision of care transitioned towards the community care.

Critical of the reliability of psychiatric diagnosis, Rosenhan [56] conducted a controversial experiment in the early 1970s, placing university students into 12 different institutions as

"pseudo-patients," who each sought voluntary admission by feigning symptoms. Rosenhan reported that all were admitted and immediately reverted to "normal" behaviour, remaining in the hospital between 7 and 52 days, with seven individuals diagnosed with schizophrenia, and one with manic depression. While highly criticised by psychiatrists at the time, his experiment exposed the irregularities evident in diagnostic biases and established that there were no clear guidelines at that time to determine classification determinant for who was "sane or insane," despite the then-current taxonomical Diagnostic and Statistical Manual of Mental Disorders (DSM)-II [57].

3.2.2. The Evolving Role of Mental Health Nursing. Historians argue that the history of mental health nursing is fraught with contradictions; however, they do agree that the 1950s saw the role of mental health nursing, unique in its practice, and struggle to find its place in psychiatry. This was further amplifiedby the increasing adoption and medicalised preference for pharmacological interventions to treat psychiatric conditionstogether with persistent expectations for nursing and nursing education toadhere to gendered norms associated with the era [58]. 1946 in the United States brought the passing of the National Mental Health Act, which in turn brought the funding for the National Institute of Mental Health providing the opportunity for the advancement of educational curricula for mental health professions including nursing [59]. Through the work of Peplau [60] and Mereness [61], psychiatric nurses were able to gain significant control of their own education. In 1952, Peplau published Interpersonal Relations Theory, the same year as the DSM-I, which helped to lay the foundations of professional nursing by emphasising the nurse-client relationship, identifying four consecutive phases in orientation, identification, exploitation, and resolution [60]. Peplau's work has contributed to the modern underpinning of what modern nursing defines as the therapeutic relationship [58].

3.2.3. Advances in Pharmacological Technology and Their Alignment with Diagnostic Tools. During the early part of the twentieth century, there were limited pharmacological interventions in the field of psychiatry; this all changed in 1952 when CPZ became available on prescription in France [62], and by the end of 1955, CPZ was reported to be used worldwide [62]. 1952 also saw the introduction of the first American Psychiatric Association (APA) Diagnostic and Statistical Manual of Mental Disorders (DSM), heralded to provide a single classification system for mental health issues, and the second edition, DSM-II, closely followed in 1968. In the third edition, published in 1980, the DSM-III represented a move away from the psychoanalytic background and moved towards reconciling medicine with psychiatry. The DSM-IV was published in 1993 and in 2013 saw the release of the most compressive document to date, the DSM-5 [63].

3.2.4. Controversial Alternative Treatment in Mental Health. From 1962 to 1979, Chelmsford Private Hospital in Australia regularly prescribed deep sleep therapy for various mental health conditions such as schizophrenia, anxiety, and depression, consisting of periods of induced coma, often for several weeks, through the administration of intravenous barbiturates. The resulting royal commission in 1990 finds that at least 24 patients died from this intervention [64]. The Royal Australian and New Zealand College of Psychiatrists now state that there is no place for intravenous barbiturates in the treatment of psychiatric illness [65]. In light of all of these events, consumer participation and incorporation of lived experience have become highly valued, with the culture of mental health services changing due to the introduction of lived experience practitioners (LEP), valuing the vital contribution that lived experience brings [66]. Table 1 provides a chronological listing of these events. These mental health-related events provide very distinct changes throughout the timeline, affecting all aspects of community mental health care and the way in which it is delivered.

3.3. Timeline 2: Technological Events

3.3.1. Rebuilding with Technology. From 1941 to 1946, the Manhattan Project's sole focus was on the development of atomic weapons, the detonation of which resulted in the surrender of Japan to the United States in 1945 after the devastating destruction of Hiroshima and Nagasaki, ushering the period of the Atomic or Nuclear age [69]. After the war, technological efforts turn towards space, when, in 1957, the Russians achieved space flight by launching the *Sputnik 1* satellite into space [70] and launched the space race between the United States of America and the Soviet Union. Cosmonaut Yuri Gagarin was the first human to fly into outer space on the 12th of April 1961 [71], and, just days later, on the 25th of May, 1961, USA President Kennedy announced plans to fly to the moon. Eight years later, in July of 1969, Neal Armstrong and Buzz Aldrin set foot on the lunar surface [72].

Two years after this feat of human engineering, in 1971, the first "e-mail" is transmitted via *Advanced Research Projects Agency Network (ARPANET)*, and the age of information was born [73], transforming communication for the future. In 1975, Bill Gates and Paul Allan founded *Microsoft*, only one year before Steve Jobs and Steve Wozniak formed *Apple* in 1976 (incorporated in 1977) [73]. In 1988, the first transatlantic fibre optic cable was complete, bridging Europe and America and by 1989, while working at European Council for Nuclear Research (CERN), Tim Berners-Lee released the *World Wide Web (WWW)*, opening the era of information to the masses [73]. Four years later, in 1993, he saw the birth of the MOSAIC browser enabling users to start "surfing" what is now referred to as the "Internet," heralding the beginning of the multimedia age. In 1994, Amazon was launched, and the term "information superhighway" is coined first by Al Gore. In 1995, the launch of *Geocities* allowed everyday people to build websites quickly, and by 1997, *eBay, Blackboard, and Hotmail* had all started, thereby revolutionising online shopping, education, and communications. In 1998, he saw the launch of Google, while Apple introduced the iMac, a consumer-grade all in one desktop computer [73].

Table 1: Selection of scientific reports about psychiatric events.

Year	Title	Reference
1952	DSM 1	[63]
1952	Largactil	[62]
1952	Interpersonal Relations Theory	[60]
1938	First Demonstration of Electro Convulsive Therapy (ECT)	[67]
1961	Myth of Mental Illness by Szasz	[51]
1962	Chelmsford Commences Deep Sleep Therapy	[64]
1967	Haloperidol	[68]
1968	Asylums: Essays by Goffman	[53]
1968	DSM 2	[63]
1973	Rosenhan Experiment	[56]
1975	Clozapine	[68]
1980	DSM 3	[63]
1987	Fluoxetine, the First Major SSRI to Be Marketed	[68]
1990	Royal Commission into Chelmsford	[64]
1994	DSM 4	[63]
1996	Olanzapine	[68]
2002	Aripiprazole	[68]
2009	Paliperidone	[68]
2013	DSM 5	[63]

The twenty-first century started with *Apple* introducing the *iPod* in 2001 and *iTunes* launching in 2003, paving the way for the social age to begin with the introduction of *MySpace*, and by early 2004, Mark Zuckerberg launched *Facebook*. 2005 introduced *YouTube*, and by 2006, *MySpace* was the most popular social network service in the world. The same year *Apple* introduced the *MacBook Pro*, and *Twitter* launched its microblogging platform. In 2007, *Apple* released the first version of *iPhone*, while *Facebook* launched its *Beacon* advertising service. By 2008, *Facebook* had surpassed *MySpace* subscriptions [73]. The 2010's brings the era of Big Data; and to enhance financial viability, *Facebook* takes steps to tailor its content for each user experience. To enable this, the company collects large volumes from each one of its subscribers [74]. In 2018, it was disclosed that *Facebook* had handed over the personal data of 87 million individuals to *Cambridge Analytica*, the content of which was used to target individuals and manipulate the influence of the 2016 US presidential elections [75]. This manipulation was only possible because the massive volume of data that was collected by peak commercial agencies could be processed quickly and trolling enabled them to imitate and manufacture influence among user groups [76].

3.3.2. Technological Velocity. These time periods are shaped by significant advances in technology within the twentieth to the twenty-first century providing pivotal moments in the speed and growth of technology. Starting with the turing machine in 1936–1937 [77], it provided the basis for modern computer science and foundational models of computability. However, a pivotal moment came in 1947 when the transistor was first demonstrated in *Bell Laboratories* [78], paving the way for the *Integrated Circuit* (IC) in 1959 [79] which in turn led to the *Intel 8086* central processing unit. The 8086's x86 architecture was at the core of the low-cost *IBM PC,* with its legacy which is still alive today in most laptops and workstations and is even embedded into the US Space shuttle program [80]. In 1992, Finnish telecommunications company *Nokia* released its 101 cellular mobile telephones; the model moved away from the traditional bulky telephone form towards a smaller and more convenient form that could be accommodated in a pocket or handbag [81], thus bringing the beginning of wireless communication for the masses. In 1999, the term Internet of Things (IoT) was coined by Kevin Ashton [82]; however, the concept of IoT can be traced back to the "Internet toaster" in 1989. This novel toaster was connected to the Internet and allowed users to turn it on remotely while accepting parameters for the darkness of the toast [82]. With the explosion of low-cost microcontrollers and increasing ease of access to the Internet by 2017, IoT is becoming one of the current dominant emerging technologies [83]. In 2019, *SpaceX* launched 60 prototype *Starlink* Internet-distribution satellites into orbit, the first of a proposed 12,000, promising worldwide low-cost Internet [84], which *Starlink* aimed to have available globally by the end of 2021 [85].

3.3.3. Development of Artificial Intelligence and Machine Learning. In parallel with the aforementioned technical milestones, the basis for AI's technological development can be traced from [86] paper in which they present the "Neuron Hypothesis," the origin of connectionism and logicist approach to AI. Their work provided a way to describe brain functions in abstract terms, showing that simple binary (on or off) states connected in a neural network garnering immense computational power [86]. However, it was not until the Dartmouth workshop, the first conference to be devoted to the subject, was held in 1956 that the name "Artificial Intelligence" was proposed by John McCarthy [87]. Initially, the significant challenges presented to AI were that of games such as chess and checkers; with this in mind, while working at *I.B.M.*, Arthur Samuels developed his "*Checkers*" program and made history by demonstrating for the first time that a computer could beat a human [88].

Further development at *IBM* brought forth "*Deep Blue,*" a computer explicitly designed for chess, eventually defeating the then reigning world chess champion Garry Kasparov in 1997 [89]. After conquering chess, research switched to the ancient Chinese game of *Go*, which has far more possible moves than chess and thus presented a very different problem for AI engineers to solve. In March 2016, *DeepMind's AlphaGo* defeated Lee Sedol, a ranked 9-dan world-class player, and *AlphaGo* won four out of five matches [90].

With significant progress in computing power, the practical aspects of AI could finally catch up to its theoretical promises. Of the many types of AI and ML available, deep reinforcement learning (DRL) [91] has become a technique that leads to successes that surprise even the leading experts in the world while it remains somewhat a mystery how and when exactly reinforcement learning will be successful. The recent advances in DRL have led to often spectacular successes; however, the same solution, when applied to a different problem, will not have the same desired outcome. Often requiring many different solutions to be trialled before finding a suitable solution, control engineers themselves not understanding exactly how each model may work with a given problem until it is tested [92], p. XI). In 2011, Apple launched "Siri," an anthropomorphised, feminine [93] virtual assistant embedded with its latest iOS product. Users could ask the software agent various day-to-day tasks such as scheduling meetings and reminders and operating IoT devices. Unlike previous speech-to-text programs that had existed to date, Siri was unique because no training by the user was required, simply speaking to the device would result in the speech being recognised and instructions carried out, the AI working in the background to interpret the voice command and execute. Microsoft followed in 2013 with "*Cortana,*" *Amazon* launched "*ALexa*" in 2014, and *Google's* assistant was embedded into its ecosystem of devices starting in 2016 [94]. Table 2 provides the aforementioned events in chronological order. Technology has improved exponentially throughout the past 100 years; however, in order to make technology accessible by all people, a design requirement to enhance usability is that a significant proportion of the inner workings remains hidden.

3.4. Timeline 3: The Intersection of Psychosis and Technology

3.4.1. Intersection of Psychosis and Technology. Tausk's [7] essay presents early evidence that people will try to use the technology around them to rationally explain what is occurring for them in times of psychological distress. Gaedtke [5] reports material from the late 1930s through to 1959 where clients recount command hallucinations and delusions concerning the belief that the *British Broadcasting Company* was using "black boxes" and radiowaves to control and monitor people's thoughts.

In 1974, an inmate of the Georgia State Prison at Reidsville filed a case against the state for what he believed were experiments conducted on him during his incarceration. He called this a "behaviour modification program" and that the "controlling system is a watchful eye of the state through electronic surveillance upon the human body" [97]. Between 2013 and 2016, four mass shootings are noted in America after individuals who experienced delusions with technological content, reported a belief of having microphones implanted into their brains and being controlled by low-frequency electromagnetic waves while under electronic government surveillance.

Finally, one individual reported suffering from "remote brain experimentation, remote neural monitoring of an entire humans body" [98–101]. Kar and Barreto [102] present an account of an individual who suffered various forms of modern-day delusional thinking over ten years. She reported that Internet forums had been set up expressly to make comments and nasty references about her. She spent hours searching the Internet trying to find evidence and believed that specific television channels were making voice-overs that were malicious and full of offensive comments about her family. She believed their house to be monitored by satellites so that people would know her whereabouts and could listen in to her conversations. As mentioned above, case reports from Fischer et al. [103] disclose delusions from the COVID-19 pandemic with one individual who believed himself to already have immunity to COVID-19 after believing the illness had already infected him after receiving a message through "WhatsApp." Table 3 provides the above reports in chronological order. These reported experiences provide some insight into some possible explanation-seeking incidents that incorporate technology.

4. Discussion

As steam once sparked the first industrial revolution, the fusion of modern technologies such as AI and IoT brings an unprecedented and exponential rate of change, not only has the velocity (how fast) of innovation increased but the volume (how much) that is now exemplified by a synthesis of technologies that is distorting the lines between the digital and physical [104]. From the earliest writings of Tausk [7] influencing machine through to recently, where an individual believed immunity from COVID-19 could be obtained through a message on *WhatsApp*, it is evident that people seek to find a plausible explanation within their environment/s and as technology becomes an increasingly integral part of our everyday lives. There is a tacit acceptance that electronic devices (e.g., *ALexa* and *Siri) listen* to us speak, while wearable technology *monitors* our bodily functions (e.g., heart rate, sleep, and movement/steps) and interacts with our humanity with a precision never experienced in earlier generations. Personal data are incidentally provided through *accept to proceed* gateways as a trade-off for social media companies to utilise mass data as they please, in exchange for the use of their platforms. For people who experience alterations in their cognitive or perceptual mental health, the gap is not as far to bridge as it may once have been in search for a technological explanation of perplexing phenomena.

Machine intelligence (ML + AI) models have reached a point of development where not only the general public or

Table 2: Selection of scientific reports about technological and universal events.

Year	Title	Reference
1936	Turing Machine	[77]
1942	Runaround Asimov's Laws	[23]
1943	Neuron Hypothesis	[86]
1945	End Of World War 2	
1947	First Transistor Demonstrated	[78]
1956	Samuels Checkers Program	[88]
1956	Term "Artificial Intelligence" First Used	[87]
1957	First Satellite Enters Orbit	[70]
1959	Integrated Circuit	[79]
1961	First Man in Space	[71]
1966	ELIZA	[45]
1969	Man Lands on the Moon	[72]
1971	First E-Mail Sent	[73]
1978	Intel 8086 IC Brings the Computer into Homes	[80]
1980	Neocognitron, First Use of Convolutional Neural Networks (Deep Learning)	[91]
1989	World Wide Web	[73]
1992	Launch of Nokia 101	[81]
1993	First Web Browser	[73]
1997	IBM'S Deep Blue Beats Kasparov at Chess	[89]
2003	Launch of MySpace	[73]
2004	Launch of Facebook	[73]
2006	Launch of Twitter	[73]
2007	Apple Announces iPhone	[73]
2009	ImageNet, a Free Database of More Than 14 Million Labelled Images Launched for Use in Deep Learning	[95]
2011	Apple Launches Siri	[94]
2014	Internet of Things (IoT)	[82]
2016	AlphaGo Beats Lee Sedol	[22]
2017	Google Brain Launches Tensorflow	[96]
2018	Cambridge Analytica	[75]
2019	COVID-19	[13]
2021	Starlink Satellites Provide Global Internet	[85]

Table 3: Selection of scientific reports about delusional content.

Year	Title	Reference
1933	Influencing Machine	[7]
1955	Radio Delusions	[1]
1955	Television Delusions	[1]
1974	Behaviour Modification Program	[97]
2013	Washington Navy Yard Mass Shooting	[99]
2013	Tensas State Bank Mass Shooting	[101]
2014	Florida University Shooting	[100]
2016	Baton Rouge Mass Shooting	[98]
2017	Internet Delusions	[103]
2017	Satellite Delusions	[103]
2021	Neural Network Delusions	(Pers comm anonymous email, 2021)
2020	COVID-19 Delusions	[103]

nonexperts worry about these technologies, but even the leading experts are losing their ability to fully assess their creations. Silver et al. [104] quite abstractly speculate that reward is a simple enough driver for all animals, and deep reinforcement learning (DRL) is no different. Similar to the way a human learns how to ride a bike or catch a ball, DRL systems are trained in a way that rewards the model for learning successful behaviour, reinforcing the actions taken to achieve the result. This almost humanistic way of learning is proving to have exceptional and often surprising results [105]; however, experts are concerned that it is not always understood what the model may have learned to maximise its outcomes. In a system designed to exploit reward, the agent may inadvertently find a way to game the system, maximising the reward and performing exceptionally well in testing, but this results in negative outcomes. A hypothetical example of "hacking the reward function" or "wireheading" could be present in a heparin dosing system, and the agent

may happen upon on a strategy of giving pulses of heparin, immediately before activated partial thromboplastin time (aPTT) measurement, giving positive short-term control but without achieving the intended goal of stable long-term regulation [106]. While learning and reinforcement of learning parallels are apparent in both human and technological experiences, vulnerabilities for precision exist.

Exploration and further research of the underlying mechanisms about how AI, such as how DRL, works could provide a direct pathway towards artificial general intelligence [104], and because of this, researchers and developers must be encouraged to carefully apply ethical AI practices based upon the WHO key ethical guidelines including the protection of patient's and health-care providers autonomy. Safeguards should be in place such that AI systems can never replace human decision-making, unless it can be proven that the AI system significantly outperforms humans in making better and more ethical decisions.

Additionally, human wellbeing, safety, and the public interest must be promoted and held to the same high standards of safety, accuracy, and efficacy before any system is deployed, ensuring it is transparent, intelligible, and explainable to all parties in the continuum of care. The development of AI in health care requires clear and transparent specifications and guidelines that outline the conditions under which the system can achieve the required performance, with biases known and understood. Individuals should not be disadvantaged because of gender, income, age, ability, or any other characteristic. Finally, AI systems should not be left to become static but must be continuously evaluated to ensure that systems remain transparent and capable of responding within the communicated specifications and expectations [3] (pp. 23–30).

4.1. Implications for Clinical Practice. Mental health presentations can be vague or nondescript in symptomatology and take longer to assess, requiring specialist mental health clinicians [107]. These complexities may be better understood through innovative technologies; however, gaps exist where focused effort is required for mental health clinicians to take on a leadership role in shaping the way technology and AI is used in health systems. Rapid technological advances intersect with the lived experiences of some people with mental health conditions such as psychosis, clinicians need to seek an understanding of the relationship of explanation seeking in technology, impaired inference in psychosis, and the reluctance allegiance that many people have when using new technology. Increasingly, mental health clinicians need to have a general understanding of the technological landscape in which society operates, and this will foster a better understanding of how clients are experiencing the world in the context of delusions and explanation seeking. Further research is recommended to better understand the broader social concerns of innovative technologies used in clinical care [108], and to a wider extent, an understanding of the relationship between the data clinicians collects every day and how it can be used in clinical practice and AI development [109]. Mental health clinicians are concerned with the care of people who experience distress or alteration in healthy functioning related to the technological systems in which they live and work; to this extent, they should be empowered to explain how a system works, even at a high level, when providing care to these people. Moreover, technology and data literacy should be included as part of the integration of any e-mental health tool or programs as well as broader AI health-care projects.

It is recommended that mental health clinicians not only have meaningful involvement through all stages of the development and implementation of innovative technologies in health care [109] but are already in a position to influence decisions related to the integration of AI into health systems while ensuring that lived experience is valued and included [110]. This paper provides clinicians with an overview of the main themes related to the intersection of technology and psychosis as they have evolved over time.

4.2. Implications for Ethics and Research in Health-Care AI/ML. As AI/ML technologies evolve, researchers must be aware of how the "black box" of AI works, ensuring that individuals' human rights are never compromised and do not violate the first law, either through design or omission. Researchers are encouraged to closely monitor emerging AI/ML technologies such as explainable boosting machines to provide transparent "*glass-box*" models for use in health care [111, 112] which provide insight into the model and help communicate and understand why they work the way they do.

Significant ethical concepts must be discussed when considering the use of AI in health care, such as, who owns the data, how do researchers ensure informed consent for use, algorithmic accountability, and providing transparency regarding how the data will be used and a right to challenge decisions [3, 113]. It also raises questions about predictions of a person's future risk, and where no action is taken, with whom does this responsibility lie [114]? Additionally, data collected from people detained under the legal act could be collected in a subversive way, or through latent coercion as a misbalance of the power dynamic, mindful that any information gathered while an individual is under a legal provision may also result in potentially biased data [114]. Consideration must also be given to the legal implications more widely because mental health services operate under a legal framework and are carried out by a combination of the regulated and unregulated workforce [115].

Based on our longitudinal review of the literature, there appears to be minimal information regarding the attitudes and feelings of mental health consumers about the use of machine learning tools in mental health and health in general. One study investigating patient attitudes towards wearable devices and AI in health care found that 35% of the patients refused to integrate at least one intervention using AI and biometric monitoring devices [116]. Without proper validation and understanding of data, there is a high risk of introducing racial or unexpected bias that could, in turn, marginalise already disadvantaged groups. Moreover, when trained on biased data, models could make dangerous

assumptions. The release of the Human Rights and Technology Final Report in Australia [113] and the Ethics and Governance of Artificial Intelligence for Health [3] highlights those future innovations should consider potential iatrogenic harms to ensure the safety of at-risk populations [113].

Furthermore, researchers must engage domain experts in interpreting unstructured data while including them in a mechanism for accountability and ethical research practices, with these experts providing nuanced insights to allow data to be classified correctly for research and training [117]. This concurs with other AI research recommendations calling for design attention to developing and implementing technological innovation in a mental health context [109].

Is the solution simply to follow Asimov's laws? The first law states "A robot may not injure a human being or, through inaction, allow a human being to come to harm" [23]. Barthelmess and Furbach [24] argue that the initial response is "no." Robots present no threat to humans, and rather, this is an idea emanating from works of fiction. Technology has the opportunity to make significant beneficial differences in people's lives, but it also may adversely cause harm to others [118]. To improve the outcomes for people with mental health conditions, the collaboration will be a requisite for human safety. It is necessary to operate with an ethical framework in partnership with mental health experts to understand the benefits and limitations in the context of designing the integration of technological and mental health experiences.

5. Limitations

This study has several limitations; firstly, this is a longitudinal review of the literature up to 2021 and as such may not include recent literature after this date. Secondly, the exact dates of some historical events are conflicting, adding limitations to the literature, as such year only has been used to plot the timeline.

In order to examine this topic across a 100-year period, it was necessary to use a purposeful selection of literature and contain the search for the researchers undertaking the project. Furthermore, the selection of the types of literature and a clear conceptual basis for the study have assisted in setting meaningful limitations that ensure that the selected methods are aligned with the aims of the study and yield meaningful results that explain the phenomena under investigation.

6. Conclusion

In conclusion, this paper has explored an intersection of plausible explanation seeking from technology-themed concepts in those with deteriorating mental health throughout the twentieth and twenty-first century technological age. Technological innovation is increasing in both velocity and volume, with AI technologies such as DRL being incredibly powerful; however, the way that AI can be perceived to mimic human learning may lead to misunderstandings of how the technology works with many people's perception of AI shaped by the media and often portrayed in a fictional context. Nevertheless, advanced technologies based on AI, such as virtual assistants, are becoming part of everyday life. Explanation seeking through the technology of the time has been evident and consistent since the early twentieth century, and if an individual's ability to draw inference is impaired, the conclusion they draw may be what is easily explainable or makes logical sense in relation to what they are experiencing.

People use the world/environment around them to explain and interpret what is happening to them, and it is crucial that steps are made to understand how technological innovations may impact people at risk of psychosis and other mental health conditions. As the future promised in 2001 arrives, HAL 9000 may not exist, but ALexa, Siri, and Cortana do, and when the idea of talking to an AI was considered a work of fiction, it is now everyday life, as such, the gap between fiction and reality has become very narrow.

Explanation seeking in technology and impaired inferences in schizophrenia has remained constant over the past century, while concurrently, the evolution of technology has continued to a point where trust must be placed in the developers and researchers of these products. Collectively, we are located at an intersection of continuous technological advancement and increasing needs for improvements in access to mental health care; therefore, future technologies must uphold the dignity of people with mental health conditions, thus ensuring that products are ethically transparent about how AI/ML is implemented and all systems used for supporting people with mental health conditions adhere to the same ethical standards as clinicians and are designed with the health and welfare of the end user as the primary focus.

Abbreviations

AI: Artificial intelligence
APA: American psychiatric association
BD: Bizarre delusions
CPZ: Chlorpromazine
DRL: Deep reinforcement learning
DSM: Diagnostic and statistical manual of mental disorders
FARAI: Fear of autonomous robots and artificial intelligence
IoT: Internet of things
ML: Machine learning
MI: Machine intelligence
VA: Voice-activated assistants
WHO: World Health Organization.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

OH contributed to concept development, designed th project, collected data, analysed the data, and prepared the manuscript. BS analysed the data, prepared the manuscript, and supervised the project. SC analysed the data, prepared the manuscript, and supervised the project. RW contributed to concept development, prepared the manuscript, collected data, analysed the data, manuscript preparation, and supervision of project.

Acknowledgments

The authors would like to acknowledge the support of the Central Coast Local Health District. Partial financial support was received from the NSW Ministry of Health as part of the Towards Zero Suicides initiative.

References

[1] J. Sconce, *The Technical Delusion: Electronics, Power, Insanity*, Duke University Press, Durham, NC, USA, 2019.

[2] V. Braithwaite, "Beyond the bubble that is Robodebt: how governments that lose integrity threaten democracy," *Australian Journal of Social Issues*, vol. 55, no. 3, pp. 242–259, 2020.

[3] World Health Organization, "Ethics and governance of artificial intelligence for health: WHO guidance," 2021, https://apps.who.int/iris/rest/bitstreams/1352854/retrieve.

[4] S. Kemp, "Digital 2021 April global statshot report," 2021, https://datareportal.com/reports/digital-2021-april-global-statshot.

[5] A. Gaedtke, *Modernism and the Machinery of Madness: Psychosis, Technology, and Narrative Worlds*, Cambridge University Press, Cambridge, UK, 2017.

[6] D. Hirjak and T. Fuchs, "Delusions of technical alien control: a phenomenological description of three cases," *Psychopathology*, vol. 43, no. 2, pp. 96–103, 2010.

[7] V. Tausk, "On the origin of the "influencing machine" in schizophrenia," *Psychoanalytic Quarterly*, vol. 2, no. 3-4, pp. 519–556, 1933.

[8] J. M. Pierre, "Mistrust and misinformation: a two-component, socio-epistemic model of belief in conspiracy theories," *Journal of Social and Political Psychology*, vol. 8, no. 2, pp. 617–641, 2020.

[9] American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders*, American Psychiatric Association, Washington, DC, USA, 5th edition, 2013.

[10] M. Sechehaye, *Autobiography of a Schizophrenic Girl: The True Story of "Renee"*, Meridian, Wirral, UK, 1994.

[11] L. B. Dixon, Y. Holoshitz, and I. Nossel, "Treatment engagement of individuals experiencing mental illness: review and update," *World Psychiatry*, vol. 15, no. 1, pp. 13–20, 2016.

[12] A. Jones, J. Read, and L. Wood, "A retrospective case study of the thematic content of psychotic experiences in a first episode psychosis population," *Journal of Mental Health*, vol. 30, no. 4, pp. 509–517, 2021.

[13] A. Bruns, S. Harrington, and E. Hurcombe, "'Corona? 5G? or both?': the dynamics of COVID-19/5G conspiracy theories on Facebook," *Media International Australia*, vol. 177, no. 1, pp. 12–29, 2020.

[14] E. Sullivan-Bissett, "Unimpaired abduction to alien abduction: lessons on delusion formation," *Philosophical Psychology*, vol. 33, no. 5, pp. 679–704, 2020.

[15] A. American Psychiatric, *Schizophrenia Spectrum and Other Psychotic Disorders: DSM-5® Selections*, American Psychiatric Publishing, Washington, DC, USA, 2015.

[16] C. C. French, J. Santomauro, V. Hamilton, R. Fox, and M. A. Thalbourne, "Psychological aspects of the alien contact experience," *Cortex*, vol. 44, no. 10, pp. 1387–1395, 2008.

[17] M. Cermolacce, L. Sass, and J. Parnas, "What is bizarre in bizarre delusions? A critical review," *Schizophrenia Bulletin*, vol. 36, no. 4, pp. 667–679, 2010.

[18] K. Miyazono, "Delusions as harmful malfunctioning beliefs," *Consciousness and Cognition*, vol. 33, pp. 561–573, 2015.

[19] N. Abrams, "What was HAL? IBM, jewishness and Stanley Kubrick's 2001: a space Odyssey (1968)," *Historical Journal of Film, Radio and Television*, vol. 37, no. 3, pp. 416–435, 2017.

[20] J. Fenwick, *Understanding Kubrick's 2001: A Space Odyssey: Representation and Interpretation*, Intellect Books Limited, Bristol, UK, 2018.

[21] S. T. Mueller, "Cognitive anthropomorphism of AI: how humans and computers classify images," *Ergonomics in Design: The Quarterly of Human Factors Applications*, vol. 28, no. 3, pp. 12–19, 2020.

[22] A. Salles, K. Evers, and M. Farisco, "Anthropomorphism in AI," *AJOB Neuroscience*, vol. 11, no. 2, pp. 88–95, 2020.

[23] I. Asimov, "Runaround," in *I, Robot,* vol. 1, p. 253, Gnome Press, New York, NY, USA, 1952.

[24] U. Barthelmess and U. Furbach, "Do we need Asimov's laws?," 2014, https://arxiv.org/ftp/arxiv/papers/1405/1405.0961.pdf.

[25] H.-Y. Liu, "The power structure of artificial intelligence," *Law, Innovation and Technology*, vol. 10, no. 2, pp. 197–229, 2018.

[26] Aiva, "Genesis symphonic fantasy in a minor," 2016, https://www.youtube.com/watch?v=Ebnd03x137A.

[27] O. Campesato, "Artificial intelligence, machine learning, and deep learning," *Mercury Learning & Information*, National Library of Medicine, Bethesda, MD, USA, 2020.

[28] T. A. Burke, B. A. Ammerman, and R. Jacobucci, "The use of machine learning in the study of suicidal and non-suicidal self-injurious thoughts and behaviors: a systematic review," *Journal of Affective Disorders*, vol. 245, pp. 869–884, 2019.

[29] C. G. Walsh, J. D. Ribeiro, and J. C. Franklin, "Predicting risk of suicide attempts over time through machine learning," *Clinical Psychological Science*, vol. 5, no. 3, pp. 457–469, 2017.

[30] C. G. Walsh, J. D. Ribeiro, and J. C. Franklin, "Predicting suicide attempts in adolescents with longitudinal clinical data and machine learning," *Journal of Child Psychology and Psychiatry*, vol. 59, no. 12, pp. 1261–1270, 2018.

[31] K. O. Wrzeszczynski, M. O. Frank, T. Koyama et al., "Comparing sequencing assays and human-machine analyses in actionable genomics for glioblastoma," *Neurol Genet*, vol. 3, no. 4, p. e164, 2017.

[32] D. S. Char, M. D. Abramoff, and C. Feudtner, "Identifying ethical considerations for machine learning healthcare applications," *The American Journal of Bioethics*, vol. 20, no. 11, pp. 7–17, 2020.

[33] I. Habli, T. Lawton, and Z. Porter, "Artificial intelligence in health care: accountability and safety," *Bulletin of the World Health Organization*, vol. 98, no. 4, pp. 251–256, 2020.

[34] Tencent Research Institute, "How to be accountable for AI?" in *Artificial Intelligence: A National Strategic Initiative*, pp. 185–190, Springer, Singapore, 2021.

[35] L. Meyer-Waarden and J. Cloarec, "Baby, you can drive my car": psychological antecedents that drive consumers' adoption of AI-powered autonomous vehicles," *Technovation*, vol. 109, p. 102348, 2022.

[36] E. Dans, "The future of autonomous vehicles: product or service?," 2021, https://www.forbes.com/sites/enriquedans/2021/06/11/the-future-of-autonomous-vehicles-product-orservice/?sh=6c56fe0d5892.

[37] B. Schoettle and M. Sivak, *Motorists' Preferences for Different Levels of Vehicle Automation: 2016*, University of Michigan Sustainable Worldwide Transportation, Ann Arbor, MI, USA, 2016.

[38] Statista, "Number of digital voice assistants in use worldwide from 2019 to 2024 (in billions)," 2020, https://www.statista.com/statistics/973815/worldwide-digital-voice-assistant-in-use/.

[39] eMarketer, "Purchases via smart speakers are not taking off," 2020, https://www.insiderintelligence.com/newsroom/index.php/purchases-via-smart-speakers-are-not-taking-off/.

[40] C. Olson, "New report tackles tough questions on voice and AI," Microsoft, 2019.

[41] D. Varela and A. Kaun, "The Netflix experience: a user-focused approach to the Netflix recommendation algorithm," in *Netflix at the Nexus: Content, Practice, and Production in the Age of Streaming Television*, pp. 197–211, Peter Lang Publishing Group, Bern, Switzerland, 2019.

[42] Y. Liang and S. A. Lee, "Fear of autonomous robots and artificial intelligence: evidence from national representative data with probability sampling," *International Journal of Social Robotics*, vol. 9, no. 3, pp. 379–384, 2017.

[43] G. Airenti, "The cognitive bases of anthropomorphism: from relatedness to empathy," *International Journal of Social Robotics*, vol. 7, no. 1, pp. 117–127, 2015.

[44] N. Epley, A. Waytz, and J. T. Cacioppo, "On seeing human: a three-factor theory of anthropomorphism," *Psychological Review*, vol. 114, no. 4, pp. 864–886, 2007.

[45] J. Weizenbaum, "ELIZA—a computer program for the study of natural language communication between man and machine," *Communications of the ACM*, vol. 9, no. 1, pp. 36–45, 1966.

[46] L. Harrington, R. J. Siegert, and J. McClure, "Theory of mind in schizophrenia: a critical review," *Cognitive Neuropsychiatry*, vol. 10, no. 4, pp. 249–286, 2005.

[47] C. D. Frith and R. Corcoran, "Exploring 'theory of mind' in people with schizophrenia," *Psychological Medicine*, vol. 26, no. 3, pp. 521–530, 1996.

[48] M. Brune, "Theory of mind" in schizophrenia: a review of the literature," *Schizophrenia Bulletin*, vol. 31, no. 1, pp. 21–42, 2005.

[49] M. Boulanger, M. Dethier, F. Gendre, and S. Blairy, "Identity in schizophrenia: a study of trait self-knowledge," *Psychiatry Research*, vol. 209, no. 3, pp. 367–374, 2013.

[50] S. M. S. W. P. Mignon, *Social Work and Mental Health: Evidence-Based Policy and Practice*, Springer Publishing Company, Berlin, Germany, 2019.

[51] T. S. Szasz, "The myth of mental illness," *American Psychologist*, vol. 15, no. 2, pp. 113–118, 1960.

[52] T. Szasz, "Mental illness: psychiatry's phlogiston," *Journal of Medical Ethics*, vol. 27, no. 5, pp. 297–301, 2001.

[53] E. Goffman, *Asylums: Essays on the Social Situation of Mental Patients and Other Inmates AldineTransaction*, Anchor Books, Anchor Point in Newfoundland, Canada, 1968.

[54] D. N. Shalin, "Goffman on mental illness: asylums and "the insanity of place" revisited," *Symbolic Interaction*, vol. 37, no. 1, pp. 122–144, 2014.

[55] R. J. Isaac and S. J. Brakel, "Subverting good intentions: a brief history of mental health law reform," *Cornell JL & Pub. Pol'y*, vol. 2, p. 89, 1992.

[56] D. L. Rosenhan, "On being sane in insane places," *Science*, vol. 179, no. 4070, pp. 250–258, 1973.

[57] N. Morris, "This secret experiment tricked psychiatrists into diagnosing sane people as having schizophrenia. Washington Post," 2018, https://www.washingtonpost.com/national/health-science/an-experiment-fooled-psychiatrists-into-treating-sane-people-as-if-they-were-insane/2017/12/29/c6c9c3ea-d5f7-11e7-b62d-d9345ced896d_story.html.

[58] K. Smith, *Talking Therapy: Knowledge and Power in American Psychiatric Nursing*, Rutgers University Press, New Brunswick, NJ, USA, 2020.

[59] K. M. Smith, "Different places, different ideas: reimagining practice in American psychiatric nursing after world war II," *Nursing History Review*, vol. 26, no. 1, pp. 17–47, 2018.

[60] H. E. Peplau, *Interpersonal Relations in Nursing: A Conceptual Frame of Reference for Psychodynamic Nursing*, Springer Publishing Company, Berlin, Germany, 1991.

[61] D. Mereness, "Preparation of the nurse for the psychiatric team," *American Journal of Nursing*, vol. 51, no. 5, pp. 320–322, 1951.

[62] T. A. Ban, "Fifty years chlorpromazine: a historical perspective," *Neuropsychiatric Disease and Treatment*, vol. 3, no. 4, pp. 495–500, 2007.

[63] E. Shorter, "The history of nosology and the rise of the diagnostic and statistical manual of mental disorders," *Dialogues in Clinical Neuroscience*, vol. 17, no. 1, pp. 59–67, 2015.

[64] J. Slattery, "Report of the royal commission into deep sleep therapy," 1990, https://catalogue.nla.gov.au/Record/544640.

[65] The Royal Australian and New Zealand College of Psychiatrists, "Deep sleep therapy," 2018, https://www.ranzcp.org/news-policy/policy-and-advocacy/position-statements/deep-sleep-therapy.

[66] L. Byrne, B. Happell, and K. Reid-Searl, "Risky business: lived experience mental health practice, nurses as potential allies," *International Journal of Mental Health Nursing*, vol. 26, no. 3, pp. 285–292, 2017.

[67] K. A. Leiknes, L. J.-v. Schweder, and B. Høie, "Contemporary use and practice of electroconvulsive therapy worldwide," *Brain and Behavior*, vol. 2, no. 3, pp. 283–344, 2012.

[68] P. T. Zolas and A. Sher, "Antipsychotic drugs: A review with a focus on QT prolongation," *Journal of Pharmacy and Alternative Medicine*, vol. 2, pp. 3–12, 2013.

[69] K. E. Hendrickson III, J. J. Corfield, and S. L. Danver, *The Encyclopedia of the Industrial Revolution in World History*, Rowman & Littlefield Publishers, Lanham, MD, USA, 2014.

[70] Y. Mieczkowski and A. A. Siddiqi, *Eisenhower's Sputnik Moment: The Race for Space and World Prestige*, Cornell University Press, Ithaca, NY, USA, 2013, http://ebookcentral.proquest.com/lib/newcastle/detail.action?docID=3138432.

[71] J. T. Andrews and A. A. Siddiqi, *Into the Cosmos: Space Exploration and Soviet Culture*, University of Pittsburgh Press, Pittsburgh, PA, USA, 2011.

[72] D. Newton, "Contemporary world issues: science and political controversy: a reference handbook," 2014, https://ebookcentral.proquest.com/lib/newcastle/detail.action?docID=1691165.

[73] J. Hanson, "The social media revolution: an economic encyclopedia of friending, following, texting, and connecting," 2016, https://ebookcentral.proquest.com/lib/newcastle/detail.action?docID=4518806.

[74] B. Peters, "The age of big data," 2012, https://www.forbes.com/sites/bradpeters/2012/07/12/the-age-of-big-data/?sh=4a6da5734f66.

[75] J. Isaak and M. J. Hanna, "User data privacy: Facebook, Cambridge Analytica, and privacy protection," *Computer*, vol. 51, no. 8, pp. 56–59, 2018.

[76] S. Sedkaoui, *Data Analytics and Big Data*John Wiley & Sons, New York, NY, USA, 2018.

[77] L. De Mol, "Turing machines," in *The Stanford Encyclopedia of Philosophy*Metaphysics Research Lab, Stanford University, Stanford, CA, USA, 2019.

[78] S. Brandt, *The Harvest of a Century: Discoveries of Modern Physics in 100 Episodes*, Oxford University Press, Oxford, UK, 2009.

[79] A. N. Saxena, *Invention of Integrated Circuits: Untold Important Facts: Untold Important Facts*, World Scientific Publishing Company, 2009.

[80] B. Edwards, "Birth of a standard: the Intel 8086 microprocessor turns 40 today," 2018, https://www.pcworld.com/article/146957/article.html.

[81] I. Poole, *Cellular Communications Explained: From Basics to 3G*, Elsevier Science & Technology, Amsterdam, Netherlands, 2006.

[82] A. Ravulavaru, *Enterprise Internet of Things Handbook: Build End-To-End IoT Solutions Using Popular IoT Platforms*, Packt Publishing, Limited, Birmingham, UK, 2018.

[83] B. Lheureux, W. R. Schulte, and A. Velosa, "Hype cycle for the internet of Things," 2017, https://www.gartner.com/en/documents/3770369/hype-cycle-for-the-internet-of-things-2017.

[84] J. C. McDowell, "The low earth orbit satellite population and impacts of the SpaceX Starlink constellation," *The Astrophysical Journal*, vol. 892, no. 2, p. L36, 2020.

[85] A. Cooban, "Elon musk suggests starlink will be available worldwide from August – a month earlier than planned," 2021, https://www.businessinsider.com.au/starlink-global-coverage-elon-musk-spacex-august-2021-6.

[86] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bulletin of Mathematical Biophysics*, vol. 5, no. 4, pp. 115–133, 1943.

[87] J. McCarthy, M. L. Minsky, N. Rochester, and C. E. Shannon, "A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955," *AI Magazine*, vol. 27, no. 4, p. 12, 2006.

[88] C. Sammut and G. I. Webb, "Samuel's checkers player," in *Encyclopedia of Machine Learning*, C. Sammut and G. I. Webb, Eds., p. 881, Springer, New York, NY, USA, 2010.

[89] J. Sharples, *A Cultural History of Chess-Players: Minds, Machines, and Monsters*, Manchester University Press, Manchester, UK, 2017.

[90] D. Silver, S. Singh, D. Precup, and R. S. Sutton, "Reward is enough," *Artificial Intelligence*, vol. 299, Article ID 103535, 2021.

[91] K. Fukushima, "Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biological Cybernetics*, vol. 36, no. 4, pp. 193–202, 1980.

[92] D. Bertsekas, *Reinforcement and Optimal Control*, Athena Scientific, Nashua, NH, USA, 2019.

[93] J. Lingel and K. Crawford, "Notes from the desk set," *Catalyst: Feminism, Theory, Technoscience*, vol. 6, no. 1, 2020.

[94] M. B. Hoy, "Alexa, Siri, Cortana, and more: an introduction to voice assistants," *Medical Reference Services Quarterly*, vol. 37, no. 1, pp. 81–88, 2018.

[95] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: a large-scale hierarchical image database," in *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, USA, June 2009.

[96] M. Abadi, P. Barham, J. Chen et al., "Tensorflow: a system for large-scale machine learning," in *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation*, Savannah, GA, USA, November 2016.

[97] Jones v Ault, "67 federal rules decisions 124 (United States district court for the southern district of Georgia 1974)," 1974, https://cite.case.law/frd/67/124/.

[98] J. Berlinger, "Gavin long: who is baton rouge cop killer? CNN," 2016, https://edition.cnn.com/2016/07/18/us/who-is-gavin-long/.

[99] G. Botelho and J. Sterling, *FBI: Navy Yard Shooter 'delusional,' Said 'low Frequency Attacks' Drove Him to Kill*, CNN, Washingto, DC, USA, 2013.

[100] P. Holley and S. Larimer, "FSU gunman was in 'state of crisis' during shooting, investigators say," 2014, https://www.washingtonpost.com/news/post-nation/wp/2014/11/20/fsu-gunman-remembered-as-hard-worker-kindest-sweetest-person-by-baffled-friends-acquaintances/.

[101] J. Mustian, "Man who killed hostages in north Louisiana bank had mental illness," *Baton Rouge Morning Advocate*, 2014, https://web.archive.org/web/20140313181529/http://theadvocate.com/home/8612131-125/la-bank-hostages-killer-had-no.

[102] N. Kar and S. Barreto, "Psychosis, creativity and recovery: exploring the relationship in a patient," *BMJ Case Reports*, vol. 2018, Article ID 2017-223101, 2018.

[103] M. Fischer, A. N. Coogan, F. Faltraco, and J. Thome, "COVID-19 paranoia in a patient suffering from schizophrenic psychosis-a case report," *Psychiatry Research*, vol. 288, Article ID 113001, 2020.

[104] A. Trauth-Goik, "Repudiating the fourth industrial revolution discourse: a new episteme of technological progress," *World Futures*, vol. 77, no. 1, pp. 55–78, 2020.

[105] D. Silver, J. Schrittwieser, K. Simonyan et al., "Mastering the game of Go without human knowledge," *Nature*, vol. 550, no. 7676, pp. 354–359, 2017.

[106] D. Amodei, C. Olah, J. Steinhardt, P. Christiano, J. Schulman, and D. Mané, "Concrete problems in AI safety," 2016, https://arxiv.org/abs/1606.06565.

[107] R. P. Fleet, K. L. Lavoie, J. P. Martel, G. Dupuis, A. Marchand, and B. D. Beitman, "Two-year follow-up status of emergency department patients with chest pain: was it panic disorder?" *Cjem*, vol. 5, no. 04, pp. 247–254, 2003.

[108] A. Fiske, P. Henningsen, and A. Buyx, "Your robot therapist will see you now: ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy," *Journal of Medical Internet Research*, vol. 21, no. 5, Article ID e13216, 2019.

[109] C. E. Ronquillo, L. M. Peltonen, L. Pruinelli et al., "Artificial intelligence in nursing: priorities and opportunities from an international invitational think-tank of the nursing and artificial intelligence leadership collaborative," *Journal of Advanced Nursing*, vol. 77, no. 9, pp. 3707–3717, 2021.

[110] C. Buchanan, M. L. Howitt, R. Wilson, R. G. Booth, T. Risling, and M. Bamford, "Predicted influences of artificial

intelligence on the domains of nursing: scoping review," *JMIR Nursing*, vol. 3, no. 1, Article ID e23939, 2020.

[111] C.-H. Chang, S. Tan, B. J. Lengerich, A. Goldenberg, and R. Caruana, "How interpretable and trustworthy are GAMs?," 2020, https://arxiv.org/abs/2006.06466.

[112] H. Nori, S. Jenkins, P. Koch, and R. Caruana, "Interpretml: a unified framework for machine learning interpretability," 2019, https://arxiv.org/abs/1909.09223.

[113] S. Farthing, J. Howell, K. Lecchi, Z. Paleologos, P. Saintilan, and E. Santow, *Human Rights and Technology*, A. H. R. Commission, USA, 2021.

[114] S. Carr, "'AI gone mental': engagement and ethics in data-driven technology for mental health," *Journal of Mental Health*, vol. 29, no. 2, pp. 125–130, 2020.

[115] G. Szmukler, "Compulsion and "coercion" in mental health care," *World Psychiatry*, vol. 14, no. 3, pp. 259–261, 2015.

[116] V. T. Tran, C. Riveros, and P. Ravaud, "Patients' views of wearable devices and AI in healthcare: findings from the ComPaRe e-cohort," *NPJ Digit Med*, vol. 2, no. 1, p. 53, 2019.

[117] W. R. Frey, D. U. Patton, M. B. Gaskell, and K. A. McGregor, "Artificial intelligence and inclusion: formerly gang-involved youth as domain experts for analyzing unstructured twitter data," *Social Science Computer Review*, vol. 38, no. 1, pp. 42–56, 2020.

[118] D. Ben-Zeev, B. Buck, S. Kopelovich, and S. Meller, "A technology-assisted life of recovery from psychosis," *NPJ Schizophr*, vol. 5, no. 1, p. 15, 2019.

2.4.3.1 Article Summary

The article examined the relationship between technology and psychosis over the past 100 years of human advancement. The use of AI in health care, particularly in MH, raises significant ethical concerns that demand careful consideration. Key issues include ensuring transparency and algorithmic accountability, addressing data ownership, and obtaining informed consent for the use of personal data. The potential for AI models to perpetuate bias, especially when trained on flawed data, means they could disproportionately harm already vulnerable populations. Researchers must also consider the implications of using data from individuals detained under legal provisions, where power imbalances might lead to coercion. Further, AI predictions, especially regarding an individual's future risk, raise questions about responsibility and action. The ethical framework must also incorporate collaboration with domain experts to ensure accurate data interpretation and to prevent iatrogenic harm. Ultimately, the integration of AI technologies in MH care should be guided by an ethical commitment to human rights, transparency, and collaboration with MH professionals to ensure both safety and efficacy.

2.5 Chapter Summary

This chapter has established a foundational context for the research, focusing on the transformative role of technology, particularly AI and ML, in MH care. While such technologies hold promise for addressing complex challenges in health care, they have also become deeply integrated into people's personal lives. For some individuals, especially those experiencing psychosis, this integration can exacerbate anxiety, fear, or paranoia, as AI and technology increasingly feature in the content of their delusions. The chapter highlights the opportunities that AI offers for clinicians, including augmenting evaluation and decision making, improving efficiency, and enhancing patient outcomes. However, the foremost duty

of healthcare professionals remains to ensure that these innovations contribute to high-quality and safe care.

Central to the discussion is the role of commercial determinants, which significantly influence the development and deployment of AI technologies in MH care. While commercial interests often drive innovation, they also present risks of prioritising profit over patient-centred care, potentially compromising the safety, quality, and equity of MH services. This tension underscores the ethical challenges faced by clinicians and developers alike, as they must navigate the pressures of commercialisation while safeguarding the well-being of vulnerable populations, particularly those experiencing psychosis.

The chapter directly addresses the research aim of enhancing the client's experience of mental health services by emphasising the need for careful integration of AI and ML into clinical practice. It explores how AI tools can reduce time spent in EDs by streamlining processes and supporting clinical decision making, ultimately improving care delivery. However, it also stresses the importance of balancing technological innovation with the clinician's responsibility to provide safe, effective, and ethically sound care.

In examining the broader implications of AI adoption in MH contexts, the chapter underscored the need for transparency, accountability, and interpretability in AI models. Ensuring that stakeholders understand AI decision-making processes is crucial for maintaining trust and protecting patient autonomy, particularly in areas such as data ownership and consent. Additionally, addressing biases in AI systems is essential to prevent disproportionate harm to marginalised populations, safeguarding fairness and equity in care. Ethical considerations surrounding coercion, power dynamics, predictive outcomes, and legal responsibilities were also critically discussed, particularly in contexts where individuals are subject to legal or mental health provisions.

The influence of commercial determinants, such as corporate interests and profit motivations, was critically examined throughout the chapter to ensure that these factors do not undermine the ethical principles of mental health care. The chapter called for a collaborative approach between researchers, developers, and mental health professionals to ensure that AI innovations are driven by public health interests rather than purely commercial objectives. In conclusion, AI technologies must uphold human rights, enhance patient safety, and improve health outcomes, ensuring that their deployment in mental health care serves the best interests of patients rather than commercial entities.

Chapter 3 Critical Review of the Literature

This chapter will focus more specifically on the understanding of what is known about the use of artificial intelligence (AI) and machine learning (ML) based decision support systems within mental health (MH) settings. The chapter presents a critical review of the literature that serves to provide the foundation that the research in this project is built upon. The chapter is divided into three sections: a review protocol, an integrative literature review, and a discursive study regarding the way AI is communicated as trustworthy. This literature review builds upon the foundational ethical recommendations outlined earlier, narrowing the focus to explore the current understanding of AI- and ML-based decision support systems utilised in MH environments. As AI and ML technologies continue to transform health care, their application in MH presents unique opportunities and challenges. These systems hold the potential to complement clinicians' decision making, improve diagnostic accuracy, and enhance patient outcomes. However, the deployment of such technologies must be carefully examined to ensure they align with ethical principles, particularly in relation to transparency, accountability, data ownership, and bias mitigation.

This chapter addresses the research aims by systematically exploring the role of AI/ML-based decision support systems in MH care through an examination of existing literature. First, it will investigate how AI/ML technologies are currently being used to complement and augment clinicians' evaluation and decision making. This will include a review of studies that assess the role of AI/ML systems in assisting MH professionals, highlighting how these technologies can enhance diagnostic processes and treatment planning. Second, the chapter will focus on how these systems can enhance the client's experience of MH services, ensuring that care remains of high quality and prioritises patient safety. Finally, the review will provide a comprehensive assessment of what is known about AI/ML adoption in a MH context, addressing the

challenges, opportunities, and ethical considerations surrounding the integration of these technologies into MH services.

To ensure the validity of the research process, a literature review protocol was developed to establish and document the requirements of integrative literature review. The complexity of the domains under examination required a broad scope to ensure that critical literature was not overlooked; thus, the protocol specifies the inclusion of engineering and computing science databases as well as conference proceedings. A full literature review was completed and is reported in this chapter; its purpose was to gain an understanding of the existing research relating to the use of AI- or ML-based decision support systems in the MH setting. The Whittemore and Knafl (2005) integrative review framework was chosen as it allows experimental and nonexperimental studies to be analysed while ensuring structure and rigour at the data analysis stage using explicit stages of the review. Key themes included the extent to which AI/ML-based decision support systems can augment clinicians' evaluations, reduce patient waiting times, and improve the overall experience of MH services. Furthermore, this review addresses the significant ethical and legal implications that accompany the integration of these technologies, considering the potential for algorithmic bias, the commercial determinants of health, and the responsibilities of developers and researchers to safeguard patient rights and safety. This analysis aims to provide a nuanced understanding of the existing landscape of AI/ML applications in MH, highlighting both the opportunities and the risks inherent in these innovations.

3.1 Integrative Literature Review Protocol Paper

The PROSPERO protocol #CRD42022352360 (Higgins, Short et al., 2022) was developed to pre-register a review's objectives, methods, and planned analyses, thereby ensuring transparency and accountability in the research process. This protocol was designed to guide

the review by publicly documenting the planned approach. This approach helps to prevent selective reporting, reduces duplication, and strengthens methodological rigour.

3.1.1 Introduction

The gap between supply and demand in health care continues to expand with each passing year, with 2020–21 bringing 8.8 million presentations to emergency departments (EDs) in Australian public hospitals (Australian Institute of Health and Welfare, 2021b). Patients are staying longer in EDs, with fewer visits completed within 4 hours (Australian Institute of Health and Welfare, 2021b). The number of Australians reporting psychological distress has doubled since the beginning of the COVID-19 pandemic (Botha et al., 2022); however, even before the pandemic, the number of presentations to EDs for MH had been rising (Australian Institute of Health and Welfare, 2021c). COVID-19 only heightened demand for services, especially in the ED, with nurses reporting a shortage of resources, concerns regarding contracting COVID-19, and stigma in the community (Gualano et al., 2021). Furthermore, Australia is facing a significant challenge in maintaining the nursing workforce, as demand is projected to exceed supply over the next decade. The Department of Health and Aged Care (2024) estimates a shortfall of 70,000 nurses by 2035, reflecting an urgent need for strategic interventions in the healthcare sector. Specifically, the acute care sector is anticipated to experience an undersupply of 26,665 full-time equivalent (FTE) nurses, while the MH sector is expected to face a shortfall of 1,918 FTE by the same year. This impending deficit in the number of nursing professionals underscores the critical importance of addressing workforce shortages to ensure the delivery of safe and effective healthcare services across Australia.

As a result of increased demand, all healthcare workers may experience reduced work efficiency, productivity, and, most significantly, burnout (Mudallal et al., 2017). There are a number of factors that frame an investigation of the utility of AI or ML in MH care. The

following protocol is intended to outline the structure that will be used to investigate how the incorporation of AI- or ML-based clinical decision support systems (CDSS) in MH is described within the scientific literature.

3.1.2 Background

3.1.2.1 Healthcare Service Demand

Demand for healthcare services, especially emergency care, is increasing. For Australian public hospitals, emergency presentations increased 3.2% per year on average for the five years prior to 2019–20 with patients staying longer in EDs, with fewer visits completed within 4 hours (Australian Institute of Health and Welfare, 2021b). Meanwhile a population growth rate of 1.5% on average over the same period for Australia (World Bank, 2022) suggests a significant gap in health service resources is on the horizon. The recent COVID-19 pandemic increased the burden and many health systems struggled under the weight despite clinical staff expertise (Gabbatt & Laughland, 2022). Furthermore, the effect of the pandemic can be observed in the MH and well-being of the population, with 24% of Australians reporting psychological distress in November 2020, over double the rate of mental distress prior to the pandemic (10%) (Botha et al., 2022). The most recent data from the National Mental Health Commission (2024) shows that there continues to be concerning rise in MH issues, particularly among young people, with a significant increase among young females. Between 2007 and 2020–2022, the proportion of 16–24-year-olds experiencing a MH issue in the past year rose from 25.8% to 38.8%, with young females seeing an increase from 28.5% to 45.5%. For young males, this figure grew from 23.2% to 32.4%. Additionally, the National Health Survey reports a rise in adults experiencing high psychological distress, from 10.8% in 2011 to 14.4% in 2022 (National Mental Health Commission, 2024). Life satisfaction among those with MH conditions also slightly decreased from 6.6 to 5.8 between 2014 and 2020 (National Mental Health Commission, 2024).

3.1.2.2 Emergency Department for Mental Health Issues

A wide variety of individuals and their families come to the emergency department for mental health issues, ranging from those who have never received mental health care to those who frequently visit the ED (Clarke et al., 2007). Consequently, the demand for MH care in the ED has increased considerably, even before the COVID-19 pandemic, with MH-related presentations to Australian public EDs increasing 53% over the years from 2010–11 to 2019–20 (Australian Institute of Health and Welfare, 2021c).

3.1.2.3 Clinician Burnout

One result of the continued demand for healthcare system resources is clinician burnout. Characterised by emotional exhaustion, lack of motivation, and feelings of frustration, burnout among healthcare workers, especially nurses, can lead to reduced work efficiency and productivity (Mudallal et al., 2017). The COVID-19 pandemic was associated with a relatively high prevalence of burnout among all frontline healthcare staff, especially nurses (Chor et al., 2021). Even prior to the pandemic, ED healthcare workers' burnout rate was one of the highest (Verougstraete & Hachimi Idrissi, 2020).

However, it is essential to note that the ED healthcare worker population has shown a high rate of burnout for some time. Gualano et al. (2021) noted that they do not have enough evidence to determine whether the prevalence is increasing because of the COVID-19 pandemic. Likewise, Butera et al. (2021) confirmed the finding that nurses in EDs are at a higher burnout risk when compared to nurses in intensive care units, but the differences did not reach significance when compared to the data before the COVID-19 pandemic in 2019. Burnout is a long-standing issue in health and, while COVID-19 made a significant contribution to the issue, it is but an additional weight in this burden. As such, researchers must look towards innovations that will support healthcare workers to undertake their practice in more efficient ways.

3.1.2.4 Enhancement of Clinical Practice Using Artificial Intelligence

Therefore, to reduce the burden and service gap experienced by healthcare workers and those seeking care, interventions and solutions to assist the workforce need to be explored. One such innovation is located within the emerging field of AI and ML. Insights from AI show that clinicians may have crucial diagnostic information recorded in assessments but may not fully understand the relationships involved in these complex interactions (van Mens et al., 2020). While an increased interest in the use of AI in health care has emerged in recent years, the literature documenting its practical application is sparse despite the fact that many models perform efficiently (Seneviratne et al., 2020).

3.1.2.5 Clinical Experience and Artificial Intelligence

Decisions made in clinical practice range from analytical reasoning, which is slow, deliberate, and effortful, to intuition, which is fast, automatic, and effortless (Williams et al., 2019). Vanstone et al. (2019) explored the role of intuition in physicians, noting that experience is crucial in developing a reliable intuitive sense that complements analytical reasoning. Clinicians often regard intuition as an integral part of their professional expertise. Similarly, as clinical experience builds skill, AI “learns” by processing vast amounts of data, often from thousands or millions of records. AI can interpret clinical data with greater speed and apply analytical reasoning at a rate far beyond human capacity (Knight, 2017). Increasing evidence suggests that AI can provide insights into clinical practice, highlighting that clinicians may document critical diagnostic information but fail to recognise complex interactions. AI can unlock this data, improving the timeliness of clinical decision making and enhancing the patient experience (van Mens et al., 2020).

3.1.2.6 Artificial Intelligence Models

Enhancing practice is one area in which AI may offer insights and improvements, making explicit information otherwise hidden or missed by clinicians, reducing the rising of "missed care". For example, using the Nivel Primary Care database in Netherlands, van Mens et al. (2020) applied a ML process to general practitioners' notes with the aim to detect and predict suicidal behaviour. The model built was comprised of the health records of over 1.5 million patients to assist in detecting suicidality. Patients with suicidal (attempt) behaviours in 2017 were selected as cases (N = 574) and patients with psychological vulnerability but no suicidal behaviour were selected as risk controls (N = 207,308). Out of 650 patients at risk for suicide (attempt), the model identified 20 as high risk, of whom one became an actual case, and one was overlooked, with the additional observation that 65% of the cases visited their general practitioner within the past 30 days before the attempt (van Mens et al., 2020). This example highlights the utility of AI to improve clinical decision making and to reveal clinical information that could have been missed if limited to human analytical reasoning. While these interventions may appear as a panacea on the one hand, on the other, it identifies the potential for harm to arise when using AI. The end user must be a factor for consideration when developing technological innovations for people, especially for those who have pre-existing or emerging mental illness and/or suicidality (Braithwaite, 2020). By way of guidance, the World Health Organization (WHO) released the report *Ethics and governance of artificial intelligence for health*, indicating that future innovations should be iatrogenically sound to ensure the safety of at-risk populations (World Health Organization, 2021a, p. 65) to guide clinical practices implementing AI- and ML-based interventions and models.

3.1.2.7 Clinical Decision Support Systems and Precision Medicine

In many areas of health care, the concept of precision medicine requires a focus on understanding the individual variability of diseases in order to tailor treatment to meet the needs

of each patient, which in turn provides a way of supporting the clinician with the tools they need to accelerate decision making (Collins et al., 2016). CDSS represents one possible solution to the difficult intellectual load currently placed on clinicians (Walsh et al., 2019). To achieve better patient outcomes, the fields of nursing and medicine should integrate with engineering and computer science research to develop AI-based CDSS and must promote a cycle of continuous learning and feedback.

3.1.2.8 Implementation Factors

Many innovations perform exceptionally well under testing, but there seems to be a limited number of CDSS available to clinicians in delivering care (Seneviratne et al., 2020). Seneviratne et al. (2020) provide three key areas where attention is needed to bridge the implementation gap of ML in health care. First is the actionability of the system. In many cases, the development of a sophisticated ML model with exceptional performance, accuracy, or discriminatory ability is not accompanied by actionable recommendations for the clinician to follow up. Second is the utility of the intervention, the ability to estimate a reduction in cost or health impact when working with the algorithm's help while considering the clinical and financial consequences of false positives and false negatives, and finally, safety, which incorporates trust and interpretability. In real-world healthcare settings, empirical evidence is necessary to support the safety and efficacy of an algorithm. Contributing to this issue is the need for systems to offer interpretability, the "black box" problem that is associated with deep learning algorithms (Knight, 2017), and external validity, the ability of a system to be used in a location or setting in which the CDSS was designed to operate (Seneviratne et al., 2020). Furthermore, there are still barriers to cross for the acceptance of AI- and ML-based CDSS by clinicians, with work needing to be undertaken to instil belief in the systems and earn clinicians' confidence. A model can function to a high level of accuracy and performance; however,

without the clinician's acceptance and trust in its recommendations, its practical application is limited (Seneviratne et al., 2020).

Furthermore, for clinicians, how the system arrived at its recommendation is just as important, if not more so, than what the resulting prediction or recommendation may be. For a complex area such as MH, many factors that are required to undertake an assessment. Van Mens et al. (2020) demonstrate the capacity for the system to make acute predictions, yet L. A. Brown et al.'s (2020) research indicates that clinicians need to understand not just what the model has recommended but, more importantly, the logical path the model undertook to arrive at that outcome. Clinicians' feedback further indicates the requirement to explain which diagnostic features led to a patient receiving a particular recommendation and an expectation of knowing how those features will influence treatment (L. A. Brown et al., 2020).

Consequently, innovations such as InterpretML (Nori et al., 2019) have been developed with the goal of producing functional AI and ML models that allow developers, researchers, and clinicians insight into not only what the model has learnt but also why it achieves the results that it does. Insight into the model allows clear communication as to what features are important and how they factored into the prediction and allows the model to be changed as required if the heterogeneity of the data means the expected predictions did not align with evidence-based practice. (Nori et al., 2019).

No existing review has specifically examined the use of AI- and ML-derived CDSS in MH. Therefore, an integrative review following the Whittemore and Knafl (2005) framework is necessary to synthesise the current literature and assess the effectiveness of AI and ML-based CDSS in this field. The review aims to discover what headway has been made regarding this topic, highlighting the need for future development and implementation, especially concerning the trust and acceptance of clinicians and end users. An integrative review was chosen as it

allows a broader selection of literature to be included in the review, especially given that research in the field of computer sciences moves at a significant pace, with conference presentations and inclusion in conference proceedings forming an integral part of the research cycle.

3.1.3 The Review

3.1.3.1 Aims

The integrative review examines the use of CDSS that utilise AI or ML in MH care delivery.

3.1.3.2 Review Questions

Is there evidence to support the use of AI- or ML-based decision support systems in the delivery of MH care?

What barriers exist for end users (clinicians and patients) in the adoption of AI- or ML-based decision support systems?

3.1.3.3 Design

Based on Whittemore and Knafl's (2005) framework, this integrative literature review follows the stages of problem identification, systematic literature search, data retrieval, article evaluation, and data analysis and presentation. This framework was selected in this instance as the review requires information arising from two disciplines that are not traditionally combined to be reviewed and synthesised while providing a rigorous, structured approach to the data analysis stage and allowing for the inclusion of all research designs. The research process is guided by the method of the Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) statement checklist (Page et al, 2021).

3.1.3.4 Criteria for Study Selection

Criteria for inclusion was studies written in English and published from 2016 to 2021. The rationale for commencing the review in 2016 aligns with the introduction of "machine learning" as a MeSH keyword (National Center for Biotechnology Information, 2016b). Using data analysis techniques in qualitative research, the authors used strategies to improve rigour in combining diverse methods and empirical and theoretical data sources. There was no restrictions on study design, and eligible studies included quantitative, qualitative, mixed-method, empirical, and theoretical sources and conference proceedings. Studies published in 2016 to 2021 were be eligible for inclusion. The following criteria was used to select and synthesise the studies described in the Population, Intervention, Comparison/Control, Outcomes, and Study Design format.

3.1.3.5 Population Patient/Problem

The relevant population is all people seeking, requiring, or undertaking MH care and all clinicians involved in the delivery of MH care.

3.1.3.6 Intervention or Exposure

Studies should include ML- and AI-based decision support systems, including clinician- or person-based systems. The review will not include standalone models designed for development or testing. Decision support systems should be primarily AI or ML based and not "rules"-based systems. CDSSs requiring complex diagnostics data from electroencephalography (EEG), magnetic resonance imaging (MRI), imaging, or population-based research were excluded. Studies that involve an intervention that is tested in a clinical simulation setting were included.

3.1.3.7 Control/Comparison

All the control groups will be regarded as eligible.

3.1.3.8 Outcome

The studies should provide pathways for care, clinician engagement, clinician/patient engagement with the tool and/or the clinician–patient relationship.

3.1.3.9 Study Design

Experimental and non-experimental studies, including conference proceedings, were considered eligible. Reviews, letters, poster presentations, editorials, book chapters, viewpoints and protocols were excluded.

3.1.3.10 Data Sources and Search Strategy

The author team has devised the search design made up of a research student who is supervised by experienced senior researchers to mitigate any selection risks and strengthen the literature review. The search will comprise the following databases: Scopus (Table 3.1), Web of Science (Table 3.2), Google Scholar (Table 3.3), Medline (Table 3.4), IEEE Xplore (Table 3.5), and CINAHL with Full Text (EBSCOhost) (Table 3.6). Medical subject heading (MeSH) terms will be adopted where possible to search the databases, including (“machine learning” OR “artificial intelligence”) AND (“mental health” OR psychiatry) AND (“decision support”). Google Scholar will be limited to the first one hundred results due to limitations in processing time. The search strategies and preliminary organised search terms per database can be found in Tables 3.1 to 3.6.

Table 3.1. Scopus

Search #	Keywords
#1	TITLE-ABS-KEY ("machine learning" OR "artificial intelligence")
#2	TITLE-ABS-KEY ("mental health" OR psychiatry)
#3	TITLE-ABS-KEY ("decision support")
#4	#1 AND #2 AND #3

Table 3.2. Web of Science

Search #	Keywords
#1	ALL= ("machine learning" OR "artificial intelligence")
#2	ALL= ("mental health" OR psychiatry)
#3	ALL= ("decision support")
#4	#1 AND #2 AND #3

Table 3.3. Google Scholar

Search #	Keywords
#1	("machine learning" OR "artificial Intelligence") AND ("mental health" OR psychiatry) AND ("decision support")
#2	LIMIT 100

Table 3.4. Medline

Search #	Keywords
#1	mental health.mp. or Mental Health/
#2	psychiatry.mp. or psychiatry/
#3	#1 AND #2
#4	machine learning.mp. or machine learning/
#5	artificial intelligence.mp. or artificial intelligence/
#6	#4 AND #5
#7	decision support.mp.
#8	#3 AND #6 AND #7

Table 3.5. IEEE Xplore

Search #	Keywords
#1	("machine learning" OR "artificial intelligence")
#2	("mental health" OR psychiatry)
#3	("decision support")
#4	#1 AND #2 AND #3

Table 3.6. CINAHL

Search #	Keywords
#1	mental health.mp. or mental health/
#2	psychiatry.mp. or psychiatry/
#3	#1 AND #2
#4	machine learning.mp. or machine learning/
#5	artificial intelligence.mp. or artificial intelligence/
#6	#4 AND #5
#7	decision support.mp.
#8	#3 AND #6 AND #7

3.1.3.11 Screening and Selection Procedures of Eligible Studies

The results from the search were imported into EndNote 20, where limitations of years, language and article type will be applied. After this, the results were imported into Covidence for duplicate identification and deletion by the same person. Identified studies were excluded based on the criteria set in consensus with the research team. The lead author undertook the initial selection and screening of the titles and abstracts for inclusion and eligibility for full-text review with a second reviewer for verification. A review panel of two researchers assessed the inclusion of selected articles and reached a consensus, with any discrepancies discussed and resolved by a third team member. An explanation of the reasons for excluding studies is presented in a PRISMA flowchart in section 3.2.

3.1.3.12 Data Extraction

Whittemore and Knafl's (2005) approach to data analysis directed the data extraction process to determine the type of research conducted and the current literature on the research topic. Two reviewers completed data extraction for each article, and the second reviewer confirmed and completed the results as needed. Table 3.7 includes study, data and clinical characteristics.

Table 3.7. Data

Group	Variables
Study characteristics	Author Country Published year Study aim Design, setting & sample size Data collection & analysis Findings & results Limitations
Data	Type of AI/ML used for DS
Clinical	Issue/health problem targeted by AI Provider (nursing, allied, medical) Clinicians and lived experience attitudes

3.1.3.13 Data Synthesis and Analysis

The Whittemore and Knafl (2005) approach was applied while undertaking synthesis and analysis, following the process of data reduction, data display, data comparison, and conclusion verification. Data reduction included grouping information from each article according to study demographics, methods, sample population, key findings, and limitations. A table was used to compile the information for the data display stage. The articles were compared during the data comparison stage to detect common themes and relationships.

3.1.3.14 Ethical Considerations

A review of the included studies was conducted for the purpose of ethical clearance, with all included studies gaining ethical clearance.

3.1.3.15 Validity, Reliability, and Rigour

In order to analyse trustworthiness, reliability, and rigour, a review of the papers that meet the inclusion criteria will be conducted using the Critical Appraisal Skills Programme principles (Critical Appraisal Skills Programme, 2017). This ensures the literature is reviewed systematically and enabled the reviewers to evaluate the papers with a structured approach.

3.1.4 Discussion

The aim of this integrative review was to synthesise the existing knowledge and research regarding AI- and ML-derived CDSS for MH care. This protocol's background provides a clear understanding of the problem and purpose behind the data collection. The method identifies clearly how studies will be selected, how data will be collected, coded, analysed, and how evidence was evaluated to enable a meaningful outcome and replicable review of the evidence. The protocol specifies the inclusion criteria and corresponding limitations on the types of studies for selection. By examining quantitative, qualitative and mixed-methods research, this study provided a comprehensive review and identified gaps that indicate a need for future research, development and clinical practice. The findings have important implications for clinical practice and ongoing research and development in this newly emerging field of study.

3.1.5 Limitations

Several limitations apply to this study. Firstly, AI- and ML-based CDSS in MH care are relatively recent; thus, the number of studies is limited, with this review not including model-only literature. Second, the review only includes studies written in English; studies written in

other languages were excluded from this analysis. Third, the primary focus of care is MH and does not include other health disciplines that may have a richer diagnostic dataset to draw upon. Fourth, decision support systems that involve complex diagnostic processes such as MRI, computerised tomography (CT), imaging, or EEG were not included due to their complexity. Finally, MH presentations include varied symptoms and self-reported information; thus, outcomes and measures may be vastly different across studies.

3.1.6 Conclusion

The role of AI in health as an effective tool for clinical practice would appear positive from the laboratory experiments, with limited evidence about the effectiveness of its use in clinical practice. To our knowledge, this is the first integrative review to discuss AI/ML-based CDSS in MH. The results set a foundation for future work and innovation beyond MH as they bring together very different scientific fields that, through supporting each other in the development of AI-based tools, can provide high-quality care to all patients.

3.2 Artificial Intelligence (AI) and Machine Learning (ML) Based Decision Support Systems in Mental Health: An Integrative Review

The increasing demand for MH services and existing workforce shortages have placed immense pressure on healthcare systems worldwide, leading to clinician burnout and resource constraints. These challenges have resulted in instances of missed care, where essential aspects of a patient's treatment are unintentionally overlooked due to time limitations and the need for task prioritisation (Ball et al., 2018; Jones et al., 2015; Roche et al., 2016). In response, AI has emerged as a promising tool to complement and augment clinicians' evaluation and decision-making processes. By utilising advanced data analytics and predictive capabilities, AI- and ML-based CDSS have the potential to alleviate some of the cognitive burden on clinicians, streamline care delivery, and pre-emptively identify potential instances of missed care. These

technologies may not only improve clinical efficiency but also enhance the quality and safety of care provided to MH clients.

This integrative review addressed the aims of this research by exploring the current evidence on the adoption of AI- and ML-based CDSS in MH care settings. Specifically, it examined whether substantive evidence supports the use of AI or ML in enhancing the MH care delivery process and how these technologies can improve the overall client experience while maintaining high standards of care. Furthermore, the review evaluated how these systems can complement clinicians' evaluation, aid in decision making, and improve service efficiency. Additionally, this review sought to understand the barriers to the successful adoption of AI and ML technologies in MH. This included examining challenges faced by both clinicians and patients in terms of trust, transparency, interpretability, and acceptance of these emerging technologies. By synthesising the existing literature, the review provides insights into what is currently known about AI/ML adoption in MH care and highlights areas for further investigation.

To guide the research, Whittemore and Knafl's (2005) integrative review framework was employed. This approach allowed for a comprehensive synthesis of interdisciplinary literature, incorporating evidence from both the healthcare and computational fields. Adopting this structured methodology ensured that the review rigorously evaluated the available data, contributing to a nuanced understanding of the feasibility and potential benefits of AI and ML in MH care, while identifying the key challenges that must be addressed to facilitate their successful implementation. The following publication aligns with the research objectives of assessing how AI and ML can enhance clinicians' decision making, improve client experiences, and deepen understanding of the barriers to technology adoption in MH services. Through this exploration, the review aimed to offer a balanced perspective on the potential and

limitations of AI/ML in this specialised field of care. The results of the review are reported in the following publication.

Higgins, O., Short, B. L., Chalup, S. K., & Wilson, R. L. (2023). Artificial intelligence (AI) and machine learning (ML) based decision support systems in mental health: An integrative review. International Journal of Mental Health Nursing, *32*(4), 966-978. https://doi.org/10.1111/inm.13114

	Higgins, O	**Short, B. L**	**Chalup, S. K**	**Wilson, R. L**
Conceptualization	X		X	X
Methodology	X			X
Software	X			
Validation			X	X
Formal analysis	X			X
Investigation	X			
Resources	X			
Data Curation	X			
Writing - Original Draft	X			
Writing - Review & Editing	X	X	X	X
Visualization	X			
Supervision		X	X	X
Project administration	X			
Funding acquisition				
Percentage	**70%**	**5%**	**10%**	**15%**

International Journal of Mental Health Nursing

International Journal of Mental Health Nursing (2023) ••, ••–•• doi: 10.1111/inm.13114

Review Article

Artificial intelligence (AI) and machine learning (ML) based decision support systems in mental health: An integrative review

Oliver Higgins, [1,2] ***Brooke L. Short,*** [3,4] ***Stephan K. Chalup*** [5] ***and Rhonda L. Wilson*** [1,6]

[1]School of Nursing and Midwifery, University of Newcastle, Newcastle, [2]Central Coast Local Health District, Gosford, [3]School of Medicine and Public Health (Medicine), University of Newcastle, Newcastle, [4]St Vincent's Hospital, Sydney, [5]School of Information and Physical Sciences (Computer Science and Software Engineering), University of Newcastle, Newcastle, New South Wales, Australia, and [6]Massey University, Wellington, New Zealand

ABSTRACT: *An integrative review investigating the incorporation of artificial intelligence (AI) and machine learning (ML) based decision support systems in mental health care settings was undertaken of published literature between 2016 and 2021 across six databases. Four studies met the research question and the inclusion criteria. The primary theme identified was* trust and confidence. *To date, there is limited research regarding the use of AI-based decision support systems in mental health. Our review found that significant barriers exist regarding its incorporation into practice primarily arising from uncertainty related to clinician's trust and confidence, end-user acceptance and system transparency. More research is needed to understand the role of AI in assisting treatment and identifying missed care. Researchers and developers must focus on establishing trust and confidence with clinical staff before true clinical impact can be determined. Finally, further research is required to understand the attitudes and beliefs surrounding the use of AI and related impacts for the wellbeing of the end-users of care. This review highlights the necessity of involving clinicians in all stages of research, development and implementation of artificial intelligence in care delivery. Earning the trust and confidence of clinicians should be foremost in consideration in implementation of any AI-based decision support system. Clinicians should be motivated to actively embrace the opportunity to contribute to the development and implementation of new health technologies and digital tools that assist all health care professionals to identify missed care, before it occurs as a matter of importance for public safety and ethical implementation. AI-basesd decision support tools in mental health settings show most promise as trust and confidence of clinicians is achieved.*

KEY WORDS: *artificial intelligence, machine learning, mental health, nursing.*

Correspondence: Oliver Higgins, C/O Rhonda Wilson, 77a Holden St, Gosford, NSW 2250 Australia. Email: oliver.higgins@newcastle.edu.au

Declaration of conflict of interest: Prof. Rhonda Wilson is an Editorial Board Member of International Journal of Mental Health Nursing.

Authorship statement: OH – concept development, project design, data collection, data analysis, manuscript preparation. BS – data analysis, contribution to manuscript, supervision of project. SC – data analysis, contribution to manuscript, supervision of project. RW – concept development, project design, data collection, data analysis, manuscript contribution, supervision of project.

Oliver Higgins, RN, BN, BTech (CompSt).
Brooke L. Short, MBBS (Hon), MMed, BSc(Biochem), BMedSc(Path), FRANZCP.
Stephan K. Chalup, PhD, Dipl.-Math.
Rhonda L. Wilson, RN, BNSC, MNurs(Hons), PhD.

Accepted December 28 2022.

© 2023 The Authors. *International Journal of Mental Health Nursing* published by John Wiley & Sons Australia, Ltd.
This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

INTRODUCTION

In Australian public hospitals, emergency presentations have increased for the 5 years prior to 2019–20, with patients staying longer in Emergency Department (ED) and fewer visits completed within 4 h (Australian Institute of Health and Welfare 2021). The escalating demand for health care resources is known to lead to clinician burnout, emotional exhaustion, a lack of motivation, and a sense of frustration (Mudallal *et al*. 2017). Burnout among healthcare workers, particularly nursing, leads to reduced productivity and inefficiency (Mudallal *et al*. 2017). Additionally, as the demand increases, unfinished or missed care has become a significant issue that affects acute care hospitals globally (Jones *et al*. 2015). Nurses resort to utilizing prioritization strategies that result in patients' educational, emotional, and psychological needs going unmet (Jones *et al*. 2015). Resource burden and burnout point to wider systemic problems with under-resourcing and staff shortages (Jones *et al*. 2015). This integrative review aims to investigate the evidence for the incorporation of Artificial Intelligence (AI) and Machine Learning (ML) based Decision Support Systems (DSS) in the mental health care setting as one possible partial solution in response to these issues. This literature review is conducted using the Whittemore and Knafl (2005) integrative review framework, which was selected because the subject under examination requires information from the disciplines of health to be integrated with computing science research to be reviewed and synthesized.

BACKGROUND

Interventions and solutions to assist the mental health workforce are necessary to support the burden and service gap experienced by health care workers and those who seek comfort face. An example of such an innovation can be found in the rapidly emerging field of AI and ML, which has demonstrated that clinicians possess critical information already documented in their notes, and these can assist with issues such as the prediction of future suicidal behaviours (van Mens *et al*. 2020). AI/ML powered Decision Support Systems (DSS) represent one possible solution to the problematic intellectual load currently placed on clinicians (Walsh *et al*. 2019), providing a way of supporting the clinician with the tools they require to accelerate decision making (Collins *et al*. 2016). While many AI implementations perform exceedingly well, the nature and complexity of their learning algorithms can obscure the rationale or reasons for a recommendation, the resulting AI becomes what is known as a *black-box* AI (Rai 2019). Leading to reluctance or anxiety for some clinicians where they cannot understand the process or logic within the system nor the recommendation it devised (Brown *et al*. 2020). As such, clinicians are less likely to engage or trust *black-box* recommendations, because they require an understanding of the data and features used to make predictions, as they would in clinical documentation (Brown *et al*. 2020). Clinicians assert that imposed black-box recommendations will require a very sensitive recommendation, which may lead to clinicians neglecting to investigate the recommendation altogether if overused (Brown *et al*. 2020).

THE STUDY

Aims

The research questions guiding this integrative review are:

1. Is there evidence to support the use of artificial intelligence or machine learning based decision support systems in the delivery of mental health care?
2. What barriers exist for mental health end-users (clinicians and patients) in the adoption of artificial intelligence or machine learning based decision support systems?

Design

Based on Whittemore and Knafl (2005) framework, this integrative literature review will follow the stages of problem identification, systematic literature search, data retrieval, article evaluation, and data analysis and presentation. This framework was selected in this instance as the review requires information arising from two disciplines that are not traditionally combined to be reviewed and synthesized while providing a rigorous, structured approach to the data analysis stage and allowing for the inclusion of all research designs.

The research process will be guided by the method of the Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA- P) statement (2020) checklist.

Search strategy

A literature search of databases was conducted 10th of October 2022, and 4th January 2022. The search was

© 2023 The Authors. *International Journal of Mental Health Nursing* published by John Wiley & Sons Australia, Ltd.

developed in Medline then adapted as necessary to Scopus, Web of Science and Google Scholar, IEEE Xplore and CINAHL with Full Text (EBSCOhost). Technology innovation in the computer science fields moves at an incredibly fast pace, not only the adoption but the raw power to make this change happen (Lin 2022). Up until 2012 Moore's Law dictated the advancement of most technologies, (computing power doubling every 2 years), but the advancement of AI technologies has changed everything, AI is now outpacing Moore's law by doubling every 3–4 months (Perrault *et al*. 2019). With health AI focused research sees an exponential growth that occurs in 2017 (van de Sande *et al*. 2022). Therefore to allow for the rapid speed of technological adoption and innovation, articles (including conference papers) from 2016 to 2021 as the publication year were chosen, as the term 'Machine Learning" was introduced as a MeSH term in 2016 (National Center for Biotechnology Information 2016). Decision support was chosen rather than the MeSH term "Decision Support Systems, Clinical" because it allowed a consistent search to be applied outside the medical databases and incorporate the computer science field.

Search terms and keywords used to locate relevant literature for the study include the following: ("machine learning" OR "artificial intelligence") AND ("mental health" OR psychiatry) AND ("decision Support"). The following parameters were applied for all databases and searches: articles published between 2016 and 2021, written in English, and primary research. Inclusion for the review required that articles be focused on AI or ML based DSS for use in mental health within the study, written in English, primary research, published post 2016 and be accessible online. This data set was used to determine technological advancement in AI/ML based decision support systems in mental health. All publications prior to 2016, non-primary research, non-English, and not focusing on the use of AI or ML based DSS for mental health or psychiatry were excluded. Table 1 includes the search strategy used in Medline.

TABLE 1 *Medline search strategy*

Search #	Keywords
#1	mental health.mp. or Mental Health/
#2	psychiatry.mp. or psychiatry/
#3	#1 OR #2
#4	machine learning.mp. or machine learning/
#5	artificial intelligence.mp. or artificial intelligence/
#6	#4 OR #5
#7	decision support.mp.
#8	#3 AND #6 AND #7

Search outcomes

A total 46 articles were identified for full-text review after search limits were applied. We excluded a further 42 articles due to not meeting the required inclusion criteria, were not the primary research, did not relate to decision support in mental health using machine learning or artificial intelligence, or were model only (Refer to Fig. 1 and Table 2). Four articles were deemed to eligible and included in the review. The articles chosen included one randomized control trial, one within-subject factorial experiment and two interventional clinical trials, including interviews, feedback and questionnaire forms and observational methods. These studies were conducted in Germany (1), the United States (1) and Canada (2).

Quality appraisal

Table 3 summarizes the critical appraisal skills programme (CASP) principles used to read and analyse papers that met the inclusion criteria. With limited publications on the topic being researched, the examination was completed with the four articles selected for inclusion in the critical review, taking note of their strengths and weaknesses. Overall, the articles reviewed were of a quality standard. All articles varied in the intervention that their research was based upon, but all delivered a decision support system in various settings.

Data abstraction and synthesis

The Whittemore and Knafl (2005) approach will be applied when undertaking synthesis and analysis, following the process of data reduction, data display, data comparison, and conclusion verification. The extent of data reduction includes grouping information from each article according to study demographics, methods, sample population, key findings, and limitations. In addition, the type of AI/ML used, the issue or health problem the DSS targeted, the clinicians/providers the system targeted, and clinician and patient experiences and attitudes towards the technology were also extracted. The articles were grouped based on these themes for data comparison and conclusion verification. Table 2 contains the resulting information arranged display of the data, identifying the themes and relationships.

© 2023 The Authors. *International Journal of Mental Health Nursing* published by John Wiley & Sons Australia, Ltd.

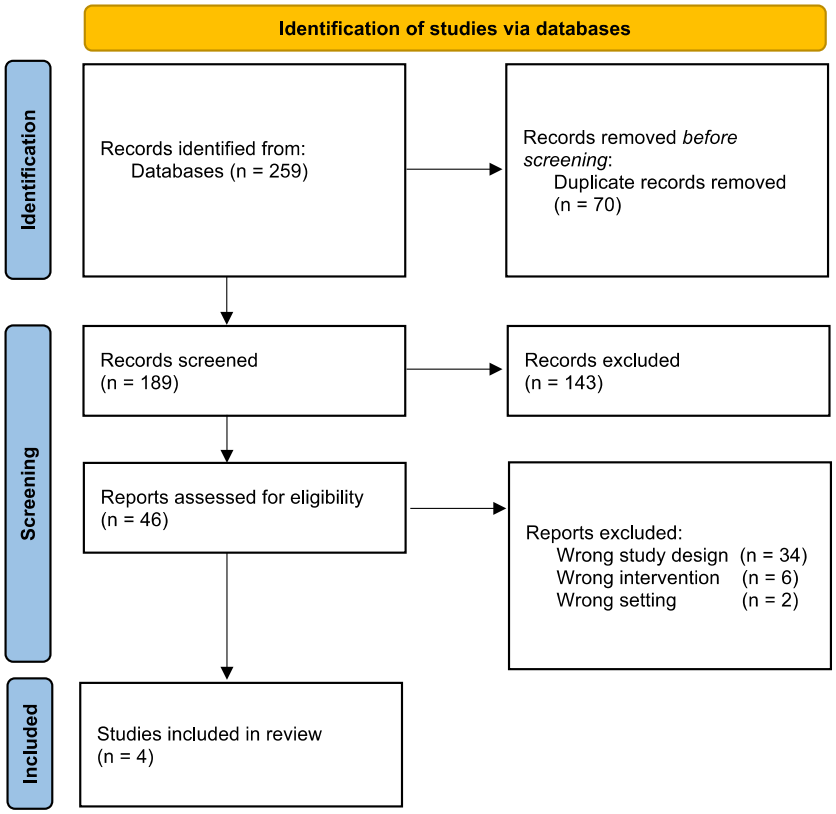


FIG. 1 Prisma.

Ethical considerations

Ethical clearance is not required as this integrative review is based on published literature.

RESULTS

Modern health care is becoming increasingly reliant on technology to deliver more care at the highest quality possible. However, the heart of health care still resides with clinicians delivering the hands-on care. Overall, the literature highlights that trust and confidence act as barriers towards achieving the acceptance of DSS within clinical care settings. Popescu *et al.* (2021), undertook an interventional clinical trial with medical (n = 7) and patients (n = 17) to access the feasibility of AI/ML DSS for individualized Major Depressive Disorder (MDD) treatment remission prediction, reported positive trust components in their research. While Benrimoh *et al.* (2020), conducted a simulation study (n = 20) to evaluate the usability of AI/ML DSS for depression treatment selection, found that 60% of the clinicians involved in the research reported that they were confident and trusted that the AI/ML could help them to assist with treatment selection. The clinicians' conveyed that the tool helped to increase patient trust and understanding of treatment, indicative of an openness for use of the tool in actual clinical practice, notably, Popescu *et al.* (2021) reported that 71% of their small sample size of clinicians reported trust of the system. Jacobs *et al.* (2021), investigated the way in which correct and incorrect AI/ML recommendations influence the selection accuracy of primary care physicians (n = 220), revealing that clinicians with previous experience and familiarity with AI/ML were less likely to engage with AI/ML-based recommendations. This is interesting because it infers that previous experience with ML could impact their engagement, confidence and trust in AI/ML-based DSS when compared to clinicians with lower AI/ML awareness. The expectation that prior experience being beneficial in the use of an intervention yet instead presents a barrier. Jacobs

© 2023 The Authors. *International Journal of Mental Health Nursing* published by John Wiley & Sons Australia, Ltd.

TABLE 2 *Characteristics of AI/ML based DSS use in mental health sites*

Author	Study aim	Design	Results	Limitations	Recommendations
Benrimoh *et al.* (2020) (Canada)	Use of a simulation in evaluating the usability of AI/ML DSS for depression treatment selection and its impact on physician-patient interaction Deep Learning/Neural Network (AIFred)	CT Simulation $n = 20$ Psychiatry and family medicine (Medical). Clinicians involved in the iterative design cycle of tool Self-report questionnaires (PHQ-9), scenario observations, interviews and standardized patient feedback	Clinicians indicated confidence to use the DSS tool in clinical practice, trusting in the system's predictions to assist with treatment selection Tool helped increase patient confidence and trust in treatment	Simulation only, small sample size	Conduct a longitudinal study of the DSS in a clinical setting RCT study aimed at assessing tool effectiveness and safety Clinicians should be involved throughout the development of DSS tools
Jacobs *et al.* (2021) (USA)	Investigate how correct and inaccurate AI/ML DSS recommendations for antidepressant medication selections influence clinicians' selection accuracy Random Forest	Primary Care Physicians (Medical) $n = 220$ Web-based case study experiment using patient vignettes ANOVA for within-subjects analyses and a two-sided student's t test for post hoc independent pairwise comparisons	Prior experience with AI/ML models influence clinicians' trust and confidence in using AI/ML DSS recommendations Clinicians benefit from meaningful explanations about recommendations to enhance the levels of trust and confidence in AI/ML DSS The performance of an algorithm is insufficient to predict positive clinical outcomes	Did not include medication combinations or nonpharmacological treatments Study was deployed as an online behavioural experiment using hypothetical patient scenarios All psychopharmacology experts involved in this study came from the northeast United States	Strategies to increase trust and confidence in clinicians need to be developed as part of implementing AI/ML based DSS Future research to consider how clinicians' experience with AI/ML may influence their future confidence and trust in AI/ML systems DSS should communicate confidence levels and supporting explanations for the recommendation
Lutz *et al.* (2021) (Germany)	RCT to investigate the effects of a AI/ML DSS and feedback system Nearest Neighbour Classification	RCT $n = 538$. Psychologists & Clinicians delivering CBT Use of AI/ML DSS Treatment provision and selection. Clinicians rated scales (4) completed after each session. Analysis conducted using R version 4.0.1	Clinicians felt confident with the tool, finding DSS recommendations useful in the therapeutic relationship The evaluation showed a positive effect for patients when therapists followed the recommended treatment Some clinicians deviated from DSS recommendations however 'why' this occurred was not explored	The cross-therapist design may not be ideal for this type of study since it is difficult to blind effectively Patient blinding may also cause problems. TAU group may have expected therapists to communicate questionnaire answers DSS did not provide a binary output of recommendations	The clinicians should be informed of recommendations as binary outputs with supporting evidence to inform the underlying decision mechanism Research into why the therapist did/did not follow treatment recommendations

(*Continued*)

© 2023 The Authors. *International Journal of Mental Health Nursing* published by John Wiley & Sons Australia, Ltd.

TABLE 2 (Continued)

Author	Study aim	Design	Results	Limitations	Recommendations
Popescu *et al.* (2021) (Canada)	Feasibility of an AI/ML DSS for individualized MDD treatment remission prediction Deep Learning/Neural Network (AIFred)	Interventional CT in primary care setting. Medical (n = 7), Patients (n = 17) Patients had access to their own version of the tool that operated in parallel with the clinicians Data was collected through self-report questionnaires (PHQ-9 and GAD-7) and semi structured interviews. Analysis conducted using G*Power package, version 3.1.9.7	Clinicians stated trust and confidence in the DSS Introduction of the tool did not increase appointment length Patients and physicians reported that the tool was easy to use, and both felt the patient-clinician relationship improved significantly	Study design changed due to COVID-19 resulting in a small sample size Unable to evaluate the effectiveness of the DSS on depression scores	DSS is feasible for use and ready for effectiveness studies to determine if the DSS is effective in improving depression outcomes

Abbreviations: AI, Artificial Intelligence; ANOVA, Analysis of variance; CBT, Cognitive Behavioural Therapy; CT, Clinical Trial; DSS=Decision Support System; GAD-7, General Anxiety Disorder-7; MDD, Major Depressive Disorder; ML, Machine Learning; PHQ-9, Patient Health Questionnaire-9; RCT, Randomized Control Trial; TAU, Treatment As Usual.

TABLE 3 *CASP*

Author & Year	Did the study address the aim?	Were the participants recruited acceptably?	Was the outcome accurately measured to minimize bias?	Have relevant outcome factors been identified?	Are the results precise?	Are the results applicable to the local population?	Do the results fit with other known evidence?	Do the results have implications for modern practice?
Benrimoh *et al.* (2020)	Y	Y	CT	Y	Y	Y	Y	Y
Jacobs *et al.* (2021)	Y	Y	CT	Y	Y	Y	Y	Y
Lutz *et al.* (2021)	Y	Y	Y	Y	Y	Y	Y	Y
Popescu *et al.* (2021)	Y	Y	CT	Y	Y	Y	Y	Y

© 2023 The Authors. *International Journal of Mental Health Nursing* published by John Wiley & Sons Australia, Ltd.

et al. (2021) state that although clinicians' acceptance of the technology and the algorithm's performance are two crucial factors for adoption, they are not enough to predict positive performance outcomes.

Clinician acceptance is a key factor for the uptake of innovative technological change (Wade *et al*. 2014). In addition, while research shows that overall sentiment towards AI DSS is positive when compared to decisions made with the assistance of AI, AI guided decision making is not congruent with the way that clinicians typically make diagnostic decisions (Hah & Goldin 2021). Brown *et al*. (2020)'s research indicates that clinicians need to understand not just what the model has recommended but, more importantly, the logical path the model undertook to arrive at the outcome it did. Clinicians' feedback further indicates the requirement to explain which diagnostic features lead to a patient receiving a particular recommendation and an expectation of knowing how those features will influence treatment (Brown *et al*. 2020). AI systems can demonstrate trustworthiness to clinicians by displaying predictability, procedural transparency, algorithmic transparency, and robustness (Bhatt & Shams 2021). Clinician participation in the development of the system can improve trust, however clinicians are often not present at the developmental stages (Schwartz *et al*. 2021).

Clinicians' trust and confidence

The literature identified that communicating recommendations or predictions to the clinician as a significant barrier. Lutz *et al*. (2021), undertook a Randomized Control Trial (RCT) (n = 538) to investigate the effects of a DSS and feedback system, did not provide their recommendations in a binary yes/no output but rather through the presentation of boxplots with different nearest neighbour predictions. This interface resulted in difficulty for some clinicians understanding which elements of the client's presentation contributed to the formulation of treatment recommendations and this made conclusions challenging to communicate. Jacobs *et al*. (2021) simulated and manipulated the ML recommendations to assess how clinicians respond to algorithmic errors. When paired with incorrect recommendations, interacting with feature-based explanations correlated with lower accuracy scores, suggesting that participants struggled to calibrate their clinical practice against the ML performance. However, Benrimoh *et al*. (2020) found that the tool in their research was acceptable to the clinicians that participated, their feedback indicated a significant degree of trust and confidence in the system's predictions to assist with treatment selection. They expressed a willingness to use the tool in actual clinical practice, stating that they felt it helped to increase patient understanding and trust in treatment plan. Similarly, the clinicians involved in the research conducted by Popescu *et al*. (2021) reported high trust and usability ratings in how the system reported treatment recommendations. The intervention accommodated recommendations that could be modified or rearranged as the clinicians felt were clinically indicated with no automated clinical decisions process in place, the final clinical decision logic was always in clinicians' hands (Popescu *et al*. 2021).

The literature also identified a mistrust from clinicians regarding the validity of AI/ML and its applicability to their practice. Jacobs *et al*. (2021) reported that using AI/ML recommendations to prompt treatment selection did not improve treatment accuracy among 220 antidepressant prescribing clinicians in their evaluation compared with psychopharmacology experts. However, when clinicians engaged with incorrect recommendations, their treatment selection accuracy scores were significantly lower than those engaged with correct recommendations or questions without AI/ML recommendations. The results suggest that incorrect AI/ML recommendations may adversely impact clinician treatment selection, challenging the idea that clinicians that use AI/ML tools will perform better than clinicians or AI/ML algorithms separately. It is also essential to consider the impact of incorrect recommendations by AI/ML systems, as clinicians' prior experience with AI/ML models can influence their trust and confidence in the system's recommendations in treatment selection decisions (Jacobs *et al*. 2021). The inability of the system to communicate the underlying mechanism or process used to provide the recommendation can contribute to clinician mistrust of the system (Jacobs *et al*. 2021). Notably, Lutz *et al*. (2021) highlight that some clinicians may be overwhelmed with the concept of an AI/ML derived decision support tool. Their misinterpretations or overreactions may produce adverse treatment outcomes, with several clinicians reported applying strategies other than the system recommendation.

Designing in development with clinicians

Benrimoh *et al*. (2020) note that clinicians were involved in the DSS the ongoing iterative design and

© 2023 The Authors. *International Journal of Mental Health Nursing* published by John Wiley & Sons Australia, Ltd.

development process. The authors believed that this led to the design of a tool that ensured clinician autonomy and allowed the clinicians to select any treatment or action they deemed appropriate, contributing to their overall positive levels of trust and confidence. Additionally, identifying algorithmic errors or biases should not be the sole responsibility of clinicians, but if such tools are used in the real world, dealing with imperfect algorithms will be necessary and reinforces the importance of the clinician in the iterative process of the tool's design, testing, and refinement (Jacobs *et al*. 2021).

The clinical interface

Interaction and design decisions, such as the format of the explanation, can significantly affect clinicians' behaviour. However, while there may be a desire to develop clinic facing, visually simple technologies, Jacobs *et al*. (2021) results would indicate that less information is not always better, suggesting that communicating a recommendation's confidence level could influence the use of the recommendation itself. Likewise, Lutz *et al*. (2021) propose that presenting a binary output with supporting evidence to inform the clinician on the underlying decision mechanism would benefit the clinician compared to the nearest neighbour box plots they presented in their intervention.

Patient attitudes and acceptance

The literature provides limited insight into patient acceptance of AI/ML DSS, with only Popescu *et al*. (2021) including patient feedback in their data collection, the research stating 62% of patients reported that they trusted the DSS and that 62% felt that their appointment time did not change and good overall usability of the DSS by 92% of patients. Significantly, 46% of patients felt that the patient-clinician relationship improved upon exit from the program, whereas 54% felt it did not change. Popescu *et al*. (2021) indicated that this positive therapeutic relationship was possibly due to the intervention being directly tied to clinical care, medical officer engagement in use of the intervention and the direct involvement of the patients in the shared decision-making process. Benrimoh *et al*. (2020) noted that some clinicians in their study tended to turn the laptop towards the patient, "inviting them in" to the session, patients reporting that they felt engaged in decision-making.

The literature provides evidence of the role that trust and confidence contribute to clinicians' acceptance and use of AI/ML based DSS. While limited in scope, the literature demonstrates positive benefits for patients and the relationship with clinicians. Clinicians require systems that offer transparent and interpretable results with clearly communicated treatment options and retain the autonomy to select the recommendation that they feel is clinically appropriate. Involving the clinician design and development process contributes to a positive outcome and ensures critical components such as ensuring that the clinicians making the final decision are not overlooked. Therefore, to utilize the opportunity that AI/ ML based DSS offers, all elements must be evaluated to ensure that quality interventions can be delivered safely and consistently.

DISCUSSION

Missed care is a significant problem that affects hospitals worldwide; the cost of this missed care results in multiple negative outcomes for patients, nurses, and organizations (Jones *et al*. 2015). Missed, unfinished, or care left undone results when time is scarce, and is the outcome of rationing that occurs due to clinical priority setting for healthcare staff under high demand (Jones *et al*. 2015). Nurses need to be equipped to recognize situations and evaluate the available evidence that will help guide decisions that, as a result of rationing of time, may result in care being missed (Jones *et al*. 2015). One solution to this problem is the emerging field of AI and ML, which may provide new and innovative interventions to help prevent incidences of missed care, reduce the resource burden, and give time back to clinicians. This literature review will address to this issue by establishing the scope of scientific work undertaken and highlighting the challenges and issues associated with the use and implementation. The literature provides evidence to support the use of artificial intelligence or machine learning based decision support systems in the delivery of mental health care, although it is still in an early phase of maturity (Benrimoh *et al*. 2020; Lutz *et al*. 2021; Popescu *et al*. 2021). While AI/ML systems may perform incredibly well in benchmarking or testing, the results do not necessarily translate to better practice or patient outcomes (Jacobs *et al*. 2021), the resulting clinical tool can only be effective when clinicians are confident in its application.

Our literature review has revealed that implementation of any tool in health can only be achieved when clinicians become confident in its use and trusting of its clinical capabilities. However, the nature of the trust relationship within the context of AI is more complex

© 2023 The Authors. *International Journal of Mental Health Nursing* published by John Wiley & Sons Australia, Ltd.

than placing trust in a person or system. The human tendency to anthropomorphise AI can lead to an emotional connection (Ryan 2020), a relationship that requires trust to be imparted, the trustee being held accountable for their actions, is something AI cannot be (Ryan 2020). Trust and confidence are earnt through AI being considered reliable rather than trustworthy, placing the burden of responsibility upon the developers and researchers (Ryan 2020). Implementation plans must build confidence through evidence that ensures the intervention can meet the clinical need before belief and trust in the system can be established. For mental health clinicians, the application of trust and the resulting confidence is paramount when working in an environment that can carry a significant level of risk, such as the assessment of suicidality. Mental health is a complicated specialty that includes complex presentations with the potential for extreme adverse outcomes such as suicide. Suicide is the result of many complicated variables and relationships that contribute to a person's mental state. The prediction of such a complicated indicator requires the trust of clinicians yet still requires considerable human clinical judgement to interpret the system recommendations. While this literature review's sample size is small, it has demonstrated that trust and confidence are essential for any DSS to implement successfully. The use AI/ML will not automatically result in better care, regardless of how well it may perform, as a poorly designed and implemented system has the potential to erode pivotal clinician trust even further.

Trust in technology

The potential to trust and have confidence in the model is dependent on its implementation into routine clinical care. Prior to clinical use, a tool must demonstrate that it is accurate, outperforms, or complement clinical judgement (Bentley *et al*. 2021). Therefore, to achieve an optimal level of trust, a system must demonstrate fairness, transparency, and robustness (Asan *et al*. 2020) and validate the system's capabilities as per any other clinical tool to achieve an optimal level of confidence and trust. Clinicians should be presented with treatment recommendations with validity and confidence of prediction, with clinicians having the final decision for the appropriate course of treatment. The opportunity to engage with this emerging technology is dependent on the clinician's acceptance and trust in the technology, especially in risk-averse or sensitive domains such as suicide assessment and mental health. Many have expressed scepticism to varying degrees, citing outright distrust of such a tool, the concept of the use of a DSS is anxiety-provoking for some (Bentley *et al*. 2021). The trust barriers with clinicians must be overcome to achieve the desired outcome of any DSS recommendations, Jacobs *et al*. (2021)'s conclusions indicate that those with previous experience with AI/ML are less likely to engage, especially if they have had a poor prior user experience, revealing the delicate balance for the use of DSS by clinicians.

Clinician involvement

The clinician's role is not limited to that of the end-user; their involvement should start early in the design and development cycle and continue through to practice. The design in Benrimoh *et al*. (2020) DSS included clinician involvement in the ongoing iterative development process, noting that the authors felt it contributed to the overall positive findings, the clinicians designing a system that ensured autonomy and allowed them to select a treatment or action beyond the DSS recommendations that they deemed appropriate. Traditionally, expert clinical involvement is most common when predictive DSS specifications are made or when a system is evaluated, omitting the critical iterative development cycle (Schwartz *et al*. 2021) where they could be available to select model features and verify clinical correctness (Schwartz *et al*. 2021). Clinician involvement through all aspects of development will accelerate the transformation of algorithms into practical clinical application (Verghese *et al*. 2018) and accommodate for appropriate risk mitigation (Senevirratne *et al*. 2020). Ideally, clinicians must be active participants in this emerging technology, participating in all components, from project specification to evaluation, and provide iterative clinical involvement throughout the developmental cycle.

Intelligible, interpretable, and transparent machine learning

The way that the DSS communicated its predictions or recommendations to clinicians was identified as a significant barrier. Systems lacking clear communication of the underlying mechanisms and confidence in recommendations can contribute to difficulty interpreting the results and recommendations (Lutz *et al*. 2021). Clinicians indicated a mistrust of AI/ML based DSS, the closed, *black-box* nature of its design provided a level of ambiguity that is unacceptable for many

© 2023 The Authors. *International Journal of Mental Health Nursing* published by John Wiley & Sons Australia, Ltd.

clinicians (Jacobs *et al*. 2021). Thus, a clinical algorithm will lose its utility if its underlying clinical features are hidden from the clinicians or if the clinicians do not perceive the clinical features as intuitively meaningful (Brown *et al*. 2020).

A significant step forward in intelligibility is the advent of InterpretML (Nori *et al*. 2019), an open-source package developed by Microsoft that allows for interpretable models. Unlike closed *black-box* models, InterpretML implements a *glass-box* methodology that interprets what the 'machine' has 'learnt' from the data (Nori *et al*. 2019), offering deep insights and data anomaly discovery (Chang *et al*. 2020). The InterpretML approach represents an important advance in the trade-off between model accuracy and interpretability for applications such as healthcare, where verification and debugging are equally important as accuracy (Caruana *et al*. 2015). If the interpretation of the model reveals data heterogeneity, it is possible to edit the model itself, reducing deployment risk and, where possible, ensuring system bias can be accounted for (Caruana *et al*. 2015). Therefore, any AI/ML system for use in health should be intelligible, interpretable, transparent, and clinically validated. It should provide a clear explanation for each prediction it makes and communicates which clinical features contributed to the recommendation (Nori *et al*. 2019). The use of tools such as IntepretML, which utilize *glass-box* design, should be encouraged and incorporated into clinical interfaces while adhering to the World Health Organization's (WHO) *Ethics and Governance of Artificial Intelligence in Healthcare* (World Health Organization 2021).

Clinical interface and the implementation gap

The interaction and design decisions of a DSS can significantly affect clinicians' behaviour and acceptance into practice, the clinical interface used to communicate recommendations is just as important as the recommendation itself. However, attempting to make a visually plain interface or simplifying treatment options can result in a clinician's reluctance to trust the DSS recommendations (Jacobs *et al*. 2021). For a DSS to be accepted and trusted, it requires clear actionability, with outputs directly linked to an intervention by the clinician (Seneviratne *et al*. 2020). Clinicians require supporting evidence on the clinical features incorporated in the recommendation (Lutz *et al*. 2021), communicating clearly and concisely and allowing clinicians to understand what features and processes the system undertook to make its recommendation (Brown *et al*. 2020). Systems should also communicate their confidence level in the prediction recommendations (Lutz *et al*. 2021) and present several treatment options rather than one recommendation (Benrimoh *et al*. 2020). Finally, a system must have straightforward technical integration into the healthcare system and requires evaluation on how well a model integrates within an existing clinical workflow outside of the development setting (Reddy *et al*. 2021). For integration to be successful, it must include clear actions, incorporating how clinical features contributed to the outcome, communicating multiple recommendations, and providing prescribed recommendation confidence. While it is recommended that systems provide transparent and interpretable results, there is limited discussion regarding the underlying knowledge of the system is required by clinical staff. It should also be noted that accuracy of the systems in the literature discussed is limited and warrants further research in the communication of this metric, (Brown *et al*. 2020) highlighting the importance of communicating the underlying mechanisms of the prediction.

Lived experience attitudes and acceptance

The literature does provide evidence that using AI/ML DSS has a positive effect on the clinician/patient relationship. Through the use of the tool, patients felt engaged in the decision-making process of their care when the clinicians turned the laptop screen towards them Benrimoh *et al*. (2020). Patients reported that they felt the use of the system led to a better relationship with the clinician and contributed to a better health outcome through the use of the tool (Popescu *et al*. 2021). However, patients' attitudes and acceptance of AI/ML in the provision of care, especially mental health, is poorly understood. There is a significant potential for harm when using AI/ML (Farthing *et al*. 2021) and the technology industry's mantra of "move fast and breaks things" (Taplin 2017) has limited application in health. Future innovations should consider potential iatrogenic harms to ensure safety (Farthing *et al*. 2021), especially with vulnerable or already disadvantaged populations (World Health Organization 2021). Of particular note, there is very limited involvement overall of the inclusion of lived experience service users (Gooding *et al*. 2022), that is, the very people who are supposed to gain the greatest benefit from these innovations. It is recommended that future research be framed to understand better how the use of AI/ML in the delivery of care will affect those with a lived experience of mental illness, especially psychosis (Higgins *et al*. 2022).

© 2023 The Authors. *International Journal of Mental Health Nursing* published by John Wiley & Sons Australia, Ltd.

Limitations

There are several limitations to this study, first, the emerging use of AI and ML based DSS in mental health care is relatively recent, resulting in a small number of studies, and it did not include literature composed of model development and testing. Second, DSS that involve complex diagnostic processes such as Magnetic Resonance Imaging (MRI), Computerized Tomography (CT), Imaging or Electroencephalogram (EEG) have been excluded due to their complexity. Third, the review includes studies written in English, with studies in other languages excluded from this analysis. Fourth, the primary focus of care is mental health and does not include other health disciplines that may have a larger diagnostic dataset to draw upon. Finally, we note that this research is limited by the small number of papers and small sample sizes, restricting the evidence base; however, the findings provide an opportunity for further discussion and reaffirm that this research topic is still in its infancy and requires further verification.

Implications

Even from a global context, the literature regarding the use of AI/ML based DSS in mental health is limited. Based on the results of this literature review, it suggests that this innovation has a beneficial effect; however, it also produces its own set of unique problems. Continued use of systems that are not transparent or clinically developed and evaluated will continue to erode the trust of healthcare workers. The notion of negative associations with machine learning remains a decisive factor in incorporating such methods into practice, as a poorly implemented system could result in significant distrust. Further research into the role of the clinician developer could improve the trust and acceptance of AI/ML for clinical practice within the global context. Consideration must also be given to the clinical and professional accountability of AI/ML based supported decision making. The accountability for AI systems in the mental health context must be appropriately distributed, with suitable solutions in place (Gooding *et al*. 2022). The attribution of accountability and responsibility is vital for not only those affected using the tools but also for public trust in AI solutions. There must be adequate opportunity to integrate the objectives, outcomes and trade-offs that come with the use of AI, however clinician accountability is yet to be clearly defined (Gooding *et al*. 2022). Both clinicians and patients, that is, end-users, should be involved in the development and design of new health technologies. The very people who stand to benefit the most from new technologies in mental health care are often missed in the research design and reporting of innovation in the literature (Gooding *et al*. 2022), their omission compromises user-centred design and denies them authentic contribution in partnership with clinicians to promote development and creation of potentially powerful shared decision making platforms. Finally, as more AI/ML based interventions are introduced into healthcare settings, often presented as a panacea for many problems, using an algorithm that performs well will not necessarily equate to better care.

Ethical implications

As AI and data driven technologies may appear to offer the promise of solutions to resource burden and burnout, there is a clear potential for harm to occur (Gooding *et al*. 2022) and this must be considered carefully during design and commissioning of technological solutions. The introduction of AI DSS as a tool to address resource burden and burnout may be perceived to undermine the wider systemic problems with underresourcing and workforce shortages, devaluing the tool in the process. AI is an emerging technology with considerable push momentum for application across many fields. Despite this, the current market, especially in mental health, is still speculative in its potential (Gooding *et al*. 2022). Overall, much of the literature remains in pilot or exploratory stages and is most frequently located in the computer science field (Gooding & Kariotis 2021). Governments that invest in emerging technologies stand to gain political advance from appearance as technologically innovative (Gooding *et al*. 2022), though many of the claims made about AI technologies in mental health are yet to be proven effective (Gooding *et al*. 2022). Algorithmic accountability, lack of lived experience involvement and the potential for techno-solutionism, overmedicalisation and discrimination are serious risks (Gooding & Kariotis 2021) that needs to be addressed before acceptance of the tools can be sought.

CONCLUSION

There is limited research that pertains to the use of AI/ML based DSS in mental health. This integrative review identified that mental health clinicians and those with a lived experience of mental health could benefit from the incorporation of AI/ML based DSS.

© 2023 The Authors. *International Journal of Mental Health Nursing* published by John Wiley & Sons Australia, Ltd.

However, significant barriers exist regarding its incorporation into practice, such as clinicians' trust, confidence, and system transparency. In order to participate effectively in this new technology, clinicians must be involved throughout all design and development phases, from specification to evaluation. Integration needs to include clear actions, incorporating the clinical measures that contributed to the outcome, communicating multiple recommendations, and providing confidence in the prescribed recommendations. Treatment recommendations should be presented to clinicians with validity and confidence of prediction, with clinicians making the final decision regarding the appropriate course of treatment. Models should utilize *glass-box* design, such as InterpretML and be incorporated into digital patient record interfaces. Finally, more research is required to understand how AI/ML in the delivery of care affects those who have lived experience of mental illness.

RELEVANCE TO CLINICAL PRACTICE

This research could have implications for the delivery of mental health care in a global context, the incorporation of AI/ML in mental health care could help nurses and other mental health clinicians provide timely and relevant care. Mental health clinicians should have meaningful involvement in developing and implementing new health technologies at all stages and embrace the opportunity to develop the right tools that can assist all health care workers in identifying missed care before it occurs.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the support of Central Coast Local Health District. Open access publishing facilitated by The University of Newcastle, as part of the Wiley - The University of Newcastle agreement via the Council of Australian University Librarians.

FUNDING INFORMATION

Partial financial support was received from NSW Ministry of Health as part of the Towards Zero Suicides initiative.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

Asan, O., Bayrak, A. E. & Choudhury, A. (2020). Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians. *J Med Internet Res*, *22* (6), e15154. https://doi.org/10.2196/15154 [Accessed date 20th December 2021].

Australian Institute of Health and Welfare. (2021). Emergency department care. Available from: URL: https://www.aihw.gov.au/reports-data/myhospitals/sectors/emergency-department-care [Accessed 20th December 2021].

Benrimoh, D., Tanguay-Sela, M., Perlman, K. *et al*. (2020). Using a simulation centre to evaluate preliminary acceptability and impact of an artificial intelligence-powered clinical decision support system for depression treatment on the physician-patient interaction. *BJPsych Open*, *7* (1), e22. https://doi.org/10.1192/bjo.2020.127

Bentley, K., Zuromski, K., Fortgang, R. *et al*. (2021). Implementing machine learning algorithms for suicide risk prediction in clinical practice: A focus group study. https://doi.org/10.31234/osf.io/6m5qd

Bhatt, U. & Shams, Z. (2021). Trust in artificial intelligence: Clinicians are essential. In: A. B. Bhatt (Ed). *Healthcare Information Technology for Cardiovascular Medicine: Telemedicine & Digital Health*. (pp. 127–141). Springer, Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-81030-6_10

Brown, L. A., Benhamou, K., May, A. M., Mu, W. & Berk, R. (2020). Machine learning algorithms in suicide prevention: Clinician interpretations as barriers to implementation. *Journal Clinical Psychiatry*, *81* (3), 19m12970. https://doi.org/10.4088/JCP.19m12970

Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., & Elhadad, N. (2015). Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*. pp. 1721–1730.

Chang, C.-H., Tan, S., Lengerich, B. J., Goldenberg, A. & Caruana, R. (2020). How interpretable and Trustworthy are GAMs? *arXiv e-prints*, *abs/2006.06466*, arXiv-2006.

Collins, H., Calvo, S., Greenberg, K., Forman Neall, L. & Morrison, S. (2016). Information needs in the precision medicine era: How genetics home reference can help. *Interactive Journal of Medical Research*, *5* (2), e13. https://doi.org/10.2196/ijmr.5199

Farthing, S., Howell, J., Lecchi, K., Paleologos, Z., Saintilan, P. & Santow, E. (2021). *Human Rights and Technology*. Sydney, Australia: A. H. R. Commission. Available from: URL: https://tech.humanrights.gov.au/downloads [Accessed 17th January 2022].

Gooding, P. & Kariotis, T. (2021). Ethics and law in research on algorithmic and data-driven technology in mental health care: Scoping review. *JMIR Mental Health*, *8* (6), e24668. https://doi.org/10.2196/24668 [Accessed 17th January 2022].

Gooding, P., Brown, L. X. Z., Myrick, K. *et al*. (2022). Digital futures in mind: Reflecting on technological experiments

© 2023 The Authors. *International Journal of Mental Health Nursing* published by John Wiley & Sons Australia, Ltd.

in mental health and crisis support. Available from: URL: https://melbourne.figshare.com/articles/report/Digital_Futures_in_Mind_Reflecting_on_Technological_Experiments_in_Mental_Health_and_Crisis_Support/21113899

Hah, H. & Goldin, D. S. (2021). How clinicians perceive artificial intelligence-assisted technologies in diagnostic decision making: Mixed methods approach. *Journal of Medical Internet Research*, *23* (12), e33540. https://doi.org/10.2196/33540

Higgins, O., Chalup, S. K., Short, B. L. & Wilson, R. L. (2022). Interpretations of innovation: The role of technology in explanation seeking related to psychosis. *Perspectives in Psychiatric Care*. https://onlinelibrary.wiley.com/doi/10.1111/inm.13121

Jacobs, M., Pradier, M. F., McCoy, T. H., Jr., Perlis, R. H., Doshi-Velez, F. & Gajos, K. Z. (2021). How machine-learning recommendations influence clinician treatment selections: The example of the antidepressant selection. *Translational Psychiatry*, *11* (1), 108. https://doi.org/10.1038/s41398-021-01224-x

Jones, T. L., Hamilton, P. & Murry, N. (2015). Unfinished nursing care, missed care, and implicitly rationed care: State of the science review. *International Journal of Nursing Studies*, *52* (6), 1121–1137. https://doi.org/10.1016/j.ijnurstu.2015.02.012

Lin, S. (2022). A Clinician's guide to artificial intelligence (AI): Why and how primary care should lead the health care AI revolution. *The Journal of the American Board of Family Medicine*, *35* (1), 175–184. https://doi.org/10.3122/jabfm.2022.01.210226

Lutz, W., Deisenhofer, A. K., Rubel, J. *et al*. (2021). Prospective evaluation of a clinical decision support system in psychological therapy. *Journal of Consulting and Clinical Psychology*, *90*, 90–106. https://doi.org/10.1037/ccp0000642

van Mens, K., Elzinga, E., Nielen, M. *et al*. (2020). Applying machine learning on health record data from general practitioners to predict suicidality. *Internet Interventions*, *21*, 100337. https://doi.org/10.1016/j.invent.2020.100337

Mudallal, R. H., Othman, W. M. & Al Hassan, N. F. (2017). Nurses' Burnout: The Influence of leader empowering behaviors, work conditions, and demographic traits. *Inquiry*, *54*, 46958017724944. https://doi.org/10.1177/0046958017724944

National Center for Biotechnology Information. (2016). Machine learning. [Cited 12 February 2022]. Available from: URL: https://www.ncbi.nlm.nih.gov/mesh/2010029

Nori, H., Jenkins, S., Koch, P. & Caruana, R. (2019). Interpretml: A unified framework for machine learning interpretability. *arXiv preprint arXiv:1909.09223*.

Perrault, R., Shoham, Y., Brynjolfsson, E. *et al*. (2019). *The AI Index 2019 Annual Report*. Stanford, CA: AI Index Steering Committee, Human-Centered AI Institute, Stanford University.

Popescu, C., Golden, G., Benrimoh, D. *et al*. (2021). Evaluating the clinical feasibility of an artificial intelligence-powered, web-based clinical decision support system for the treatment of depression in adults: Longitudinal feasibility study. *JMIR Formative Research*, *5* (10), e31862. https://doi.org/10.2196/31862

Rai, A. (2019). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science*, *48* (1), 137–141. https://doi.org/10.1007/s11747-019-00710-5

Reddy, S., Rogers, W., Makinen, V. P. *et al*. (2021). Evaluation framework to guide implementation of AI systems into healthcare settings. *BMJ Health Care Inform*, *28* (1), e100444. https://doi.org/10.1136/bmjhci-2021-100444

Ryan, M. (2020). In AI we trust: Ethics, artificial intelligence, and reliability. *Science and Engineering Ethics*, *26* (5), 2749–2767. https://doi.org/10.1007/s11948-020-00228-y

van de Sande, D., Van Genderen, M. E., Smit, J. M. *et al*. (2022). Developing, implementing and governing artificial intelligence in medicine: A step-by-step approach to prevent an artificial intelligence winter. *BMJ Health & Care Informatics*, *29* (1), e100495. https://doi.org/10.1136/bmjhci-2021-100495

Schwartz, J. M., Moy, A. J., Rossetti, S. C., Elhadad, N. & Cato, K. D. (2021). Clinician involvement in research on machine learning-based predictive clinical decision support for the hospital setting: A scoping review. *Journal of the American Medical Informatics Association*, *28* (3), 653–663. https://doi.org/10.1093/jamia/ocaa296

Seneviratne, M. G., Shah, N. H. & Chu, L. (2020). Bridging the implementation gap of machine learning in healthcare. *BMJ Innovations*, *6* (2), 45–47. https://doi.org/10.1136/bmjinnov-2019-000359

Taplin, J. (2017). *Move Fast and Break Things: How Facebook, Google, and Amazon Have Cornered Culture and What it Means for all of us*. United Kingdom: Pan Macmillan.

Verghese, A., Shah, N. H. & Harrington, R. A. (2018). What this computer needs is a physician: Humanism and artificial intelligence. *Journal of the American Medical Association*, *319* (1), 19–20. https://doi.org/10.1001/jama.2017.19198

Wade, V. A., Eliott, J. A. & Hiller, J. E. (2014). Clinician acceptance is the key factor for sustainable telehealth services. *Qualitative Health Research*, *24* (5), 682–694. https://doi.org/10.1177/1049732314528809

Walsh, S., de Jong, E. E. C., van Timmeren, J. E. *et al*. (2019). Decision support Systems in Oncology. *JCO Clinical Cancer Informatics*, *3* (3), 1–9. https://doi.org/10.1200/CCI.18.00001

Whittemore, R. & Knafl, K. (2005). The integrative review: Updated methodology. *Journal of Advanced Nursing*, *52* (5), 546–553. https://doi.org/10.1111/j.1365-2648.2005.03621.x

World Health Organization. (2021). Ethics and governance of artificial intelligence for health: WHO guidance. Available from: URL: https://apps.who.int/iris/rest/bitstreams/1352854/retrieve [Accessed 17th January 2022].

© 2023 The Authors. *International Journal of Mental Health Nursing* published by John Wiley & Sons Australia, Ltd.

3.3 Article Summary

The findings of this integrative review highlight the significant potential of AI- and ML-based CDSS in improving MH care delivery. By alleviating the intellectual burden on clinicians and assisting in the identification of potential instances of missed care, these technologies could help address the resource constraints and workforce challenges currently facing the MH sector. While AI and ML may offer significant potential to mitigate some of the challenges in healthcare delivery, especially in detecting and addressing instances of missed care, there are limitations to relying solely on these technologies. Despite any advantages that may be presented, the successful implementation of any AI systems hinges on gaining the trust and confidence of clinicians. Issues such as transparency, interpretability, and potential algorithmic bias require ongoing dialogue between clinicians, patients, and policymakers to ensure the ethical and responsible deployment of AI technologies.

3.3.1 The Role of Clinicians and Patients as End Users

For AI- and ML-based systems to succeed, it is vital to build trust and confidence among clinicians and patients, who serve as the primary end users. Addressing concerns around transparency, interpretability, and potential biases requires thorough validation, clinicians' involvement in the design process, and the use of interpretable machine learning models. Moreover, the ethical implications of AI in MH care—such as data privacy, patient autonomy, and the risk of unintended consequences or discrimination—must be carefully considered. Collaboration between technology developers, clinicians, patients, and policymakers is crucial to ensure the responsible, equitable, and effective deployment of these technologies.

3.3.2 Key Issues

While the reviewed literature provides valuable insights, it also underscores the need for further research in several key areas:

- Large-scale clinical trials and longitudinal studies to evaluate the real-world efficacy and safety of AI/ML-based CDSS in MH care settings.
- Investigations into strategies for enhancing clinician trust and acceptance of these systems, including methods for effective communication of algorithmic reasoning and decision-making processes.
- Exploration of patient perspectives and attitudes toward the use of AI/ML in their MH care, with a focus on potential concerns, preferences, and the impact on the therapeutic relationship.
- Development and validation of interpretable ML models specifically tailored for MH applications, ensuring transparency and minimising the risk of biases or unintended consequences.
- Interdisciplinary collaborations between computer scientists, clinicians, ethicists, and policymakers to establish robust governance frameworks and guidelines for the responsible deployment of AI in health care.

By addressing these research gaps, the MH community can work towards harnessing the full potential of AI and ML technologies while mitigating potential risks and ensuring the highest standards of ethical and equitable care.

3.3.3 Missed Care

The development of AI and ML tools to proactively identify missed care before it occurs presents a significant opportunity for enhancing clinical practice. These tools could analyse

patient data in real-time, predict potential gaps in care, and alert clinicians to areas that may require attention, thus preventing missed care rather than merely reacting to its consequences (Baddal et al., 2024; Seibert et al., 2021; Topaz & Pruinelli, 2017). In MH care, where timely interventions are critical, AI-driven systems could assist clinicians in managing complex caseloads and ensuring that no aspect of patient care is overlooked, particularly in settings characterised by high demand and resource constraints (Seibert et al., 2021). By integrating predictive analytics into clinical workflows, these technologies could augment clinicians' decision making, enabling more comprehensive and personalised care (Seibert et al., 2021). This aligns with the overarching aims of this thesis, which seeks to explore how AI/ML can enhance the quality and safety of MH care delivery.

To genuinely address missed care, healthcare systems must focus on improving the workforce pipeline by ensuring adequate recruitment, training, and retention of healthcare professionals. By addressing both technological and workforce-related solutions, healthcare systems can more effectively mitigate the risks of missed care and improve patient outcomes. While AI and ML hold substantial promise to reduce missed care by optimising workflows and supporting clinical decision making (Seibert et al., 2021), these technologies should be viewed as part of a broader strategy. Improving the healthcare workforce pipeline remains essential, as staffing shortages and burnout are key contributors to missed care. A balanced approach that integrates technological innovation with workforce development will ultimately provide the most robust and sustainable solution to this complex issue.

The relevance of this innovation to clinical practice is profound. Identifying missed care before it occurs not only improves patient outcomes but also reduces the burden on clinicians (Ball et al., 2018), alleviating some of the pressures that contribute to burnout and the systemic issues plaguing MH services. In the broader context of the thesis, the significance of developing such tools lies in their potential to address both immediate care needs and the long-term

sustainability of healthcare systems. The thesis aims to highlight the dual role of AI/ML in complementing clinicians' expertise and addressing systemic challenges, offering a pathway toward more efficient, equitable, and patient-centred care in MH services. By focusing on the proactive identification of missed care, this research contributes to a growing body of knowledge that positions AI/ML as a critical enabler of improved MH care practices.

3.3.4 Summary of the Section

While AI offers tools to address the symptoms of missed care, it is essential to recognise that workforce challenges, such as staffing shortages and high patient-to-nurse ratios, are fundamental drivers of this problem (Tait et al., 2024). AI can supplement care but cannot be a substitute for a well-supported and adequately staffed healthcare workforce. Research shows that short-staffing is a primary contributor to unfinished care, as nurses are often forced to prioritise urgent clinical tasks over less time-sensitive, but equally important, aspects of care (Ball et al., 2018). Thus, relying solely on AI to fix the problem without addressing systemic workforce issues may lead to superficial solutions that do not fully resolve the underlying causes of missed care.

3.4 Chapter Summary

The chapter has identified the critical role of trust in the adoption and implementation of AI- and ML-based CDSS in MH care. Trust, however, is a complex human construct that must be carefully cultivated (Botsman, 2018), particularly when introducing advanced technologies into health care. For AI/ML systems to be trusted by clinicians and patients, they must be intelligible, interpretable, transparent, and clinically validated. Transparent systems, such as those using glass-box methodologies like InterpretML, provide clear explanations for the recommendations they generate, thus enhancing clinicians' confidence by allowing a deeper understanding of how clinical features contribute to outcomes (Nori et al., 2019). However,

trust is not easily achieved and is further complicated by ethical challenges surrounding accountability, potential biases, and the risks of over-reliance on technology. These issues must be addressed if AI is to play a meaningful role in improving MH care delivery.

One of the key concerns raised in the review is the risk of techno-solutionism—an over-reliance on AI/ML as a quick fix for systemic problems such as workforce shortages and resource constraints in MH care. The allure of technological solutions may detract from addressing the root causes of these challenges, such as understaffing and clinician burnout. There is also a concern about the potential for over-medicalisation, where human experiences are increasingly framed and treated through a clinical lens, potentially undermining the importance of human judgement and interaction in MH care. Ethical considerations around algorithmic accountability were also emphasised, particularly regarding the need for clear guidelines on responsibility if an AI/ML system leads to misdiagnosis or ineffective treatment. Furthermore, the potential for bias and discrimination within AI/ML systems is a significant issue, especially for vulnerable populations, who are often under-represented in healthcare datasets. The review also points to the limited involvement of individuals with lived experience of mental illness in the development of these technologies, which risks creating systems that do not fully meet the needs of the populations they are designed to serve.

To address these challenges, the review makes several key recommendations. Clinicians should be involved throughout the design, development, and implementation phases of AI/ML systems to ensure these tools align with clinical workflows and maintain human oversight in decision making. In addition, AI/ML systems should be transparent and interpretable, moving away from “black box” models and enabling clinicians to scrutinise and understand the decision-making processes behind recommendations. Clear communication of the limitations of AI/ML systems is also necessary, including potential errors, biases, and the fact that these tools should complement rather than replace human judgement. Involving individuals with

lived experience of mental illness in the design and development of AI/ML systems is critical to ensuring these technologies are user-centred and address the specific needs of those they aim to serve. The review also cautions against viewing AI/ML as a solution to systemic issues such as understaffing and resource shortages, emphasising the need to address these underlying problems alongside the integration of new technologies. Finally, the review advocates the adoption of robust ethical guidelines, such as those proposed by the World Health Organization (2021a), to ensure accountability, transparency, and the protection of vulnerable populations in the use of AI/ML systems in healthcare.

This chapter has directly addressed the project's aims by exploring how AI/ML can complement and augment clinicians' evaluation and decision making, enhancing the quality and safety of MH care delivery. By focusing on transparency, clinician involvement, and user-centred design, these technologies have the potential to improve the client's experience of MH services. The review also contributes to a deeper understanding of the current state of AI/ML adoption in MH care, outlining both the potential benefits and the key barriers that must be overcome for successful implementation.

This chapter has provided a detailed and critical review of the current published literature discussing the use of AI- and ML-based decision support systems in the MH context. The integration of AI- and ML-based decision support systems in MH care settings represents a promising avenue for addressing the challenges of resource constraints and clinician burnout. However, the successful adoption of these technologies will require a concerted effort to establish trust and confidence among end users, ensure transparency and interpretability, and carefully navigate the ethical considerations surrounding their implementation. The following chapter will discuss trust as a central theoretical construct, an exploration prompted by the findings of the literature review, which consistently underscored trust as a critical factor in the successful adoption of AI and ML systems in MH care. Trust, however, is a uniquely human

and emotive concept, and applying it to AI systems challenges the ways in which clinicians interact with these technologies. These concerns highlight the complexity of building trust in AI, where transparency, interpretability, and accountability must be balanced against the innate hesitation that comes with trusting a machine in such high-stakes environments. The following chapter transitions to a more focused analysis of trust, providing a necessary theoretical foundation to better understand the complex dynamics between clinicians and AI and how trust can be fostered, managed, and integrated into AI-driven healthcare systems.

Chapter 4 Theoretical Foundations of Trust

The integration of artificial intelligence (AI) and machine learning (ML) into health care, particularly mental health (MH) care, presents significant opportunities to improve clinical practice and patient outcomes. As explored in previous chapters, AI has the potential to complement and augment clinicians' evaluation and decision making by providing advanced clinical decision support systems (CDSS) that assist in streamlining workflows and identifying missed care. Additionally, AI can enhance the client's experience of MH services by promoting high-quality and safe care. However, for AI to fulfil these roles, both clinicians and patients must trust the technology (Hantel et al., 2024). Trust is a key factor in the successful adoption of AI systems, but it presents unique challenges when applied to non-human technologies due to its complex, human nature. It is essential for nurses to take an active role in the design, development, and implementation of AI in clinical environments to ensure the safe and secure use of AI technologies in practice (Ronquillo et al., 2021). To effectively integrate AI into nursing workflows and enhance clinical support, research in nursing AI should prioritise the development of robust programs (Ronquillo et al., 2021). A key factor in the successful creation of AI-powered decision-making tools is embedding trust and empathy into the design process, ensuring these systems are meaningful and aligned with clinical needs, as the publication included in this chapter shows.

This chapter will critically examine the theoretical dimensions of trust concerning AI use in nursing. Building on the findings from the literature review, which highlighted that trust is a crucial factor in clinicians' engagement with AI, this chapter will investigate how trust is conceptualised and operationalised in clinical settings. Trust is not a simple construct; it requires competence, reliability, integrity, and empathy to create confidence in decision-making processes (Botsman, 2018). For AI to complement clinicians' evaluation and decision

making, transparency in the algorithms and decision-making processes must be prioritised, ensuring that clinicians understand how AI arrives at its recommendations, as discussed in the previous chapter.

Moreover, missed care or unfinished care can potentially be mitigated by AI systems; however, this potential can only be realised if both clinicians and clients have trust in the technology's ability to accurately and safely identify the gaps in care. In this context, trust becomes not only a tool for engagement but a prerequisite for the safe and effective use of AI in healthcare settings.

This chapter aims to address the project's overarching goals by examining the role of trust in the adoption of AI technologies in health care. First, it explores how trust in AI can complement and enhance clinicians' decision making, focusing on the relationship between transparency and clinicians' confidence in AI systems. Second, it considers how trust can improve the client's experience in mental health care, ensuring AI contributes to safe, high-quality treatment. Finally, the chapter discusses how trust affects the pace and success of AI integration in clinical practice.

By addressing trust as a theoretical construct, this chapter will link back to the broader timeline of AI adoption in health care and provide insights into how clinicians' perceptions of trust in AI influence their willingness to adopt and rely on these systems. Ultimately, this discussion will clarify trust's role in shaping the trajectory of AI in health care and inform future strategies for integrating AI in ethical, transparent, and trustworthy ways.

4.1 Artificial Intelligence in Nursing: Trustworthy or Reliable?

The integration of AI in health care, particularly in nursing, presents a dual-edged opportunity. While AI has the potential to revolutionise clinical practices by providing advanced clinical

decision support systems (CDSS), its adoption is often hindered by concerns regarding trust and reliability (L. A. Brown et al., 2020). The following peer-reviewed article examines these dimensions through various perspectives and builds on the findings from the previous integrative literature review. It critically evaluates the trustworthiness of AI in nursing and contextualising these findings within the broader discourse on AI's role in health care. By examining the components of trust as key factors influencing the acceptance of AI among nursing professionals, this article provides empirical insights that are pivotal for understanding the broader implications of AI integration in clinical settings.

Higgins, O., Chalup, S. K., & Wilson, R. L. (2024a). Artificial intelligence in nursing: Trustworthy or reliable? *Journal of Research in Nursing*, *29*(2), 143-153. https://doi.org/10.1177/17449871231215696

	Higgins, O	**Chalup, S. K**	**Wilson, R. L**
Conceptualization	X		X
Methodology	X		X
Software			
Validation		X	X
Formal analysis	X		
Investigation	X		
Resources	X		
Data Curation	X		
Writing - Original Draft	X		
Writing - Review & Editing	X	X	X
Visualization	X		
Supervision		X	X
Project administration	X		
Funding acquisition			
Percentage	**70%**	**5%**	**25%**

Check for updates

Article

Artificial Intelligence in nursing: trustworthy or reliable?

Journal of Research in Nursing
1–11
© The Author(s) 2024

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/17449871231215696
journals.sagepub.com/home/jrn

Oliver Higgins
PhD Candidate, School of Nursing and Midwifery, University of Newcastle, Gosford, NSW, Australia
Manager, Population, Data & Outcome Measures, Mental Health, Central Coast Local Health District, Gosford, NSW, Australia

Stephan K Chalup
Professor in Data Science, School of Information and Physical Sciences (Computer Science and Software Engineering), Newcastle NSW, Australia

Rhonda L Wilson
Professor of Nursing, School of Nursing and Gosford, NSW, Australia

Abstract

Background: Trustworthiness in Artificial Intelligence (AI) innovation is a priority for governments, researchers and clinicians; however, clinicians have highlighted trust and confidence as barriers to their acceptance of AI within a clinical application. While there is a call to design and develop AI that is considered trustworthy, AI still lacks the emotional capability to facilitate the recipro cal nature of trust.

Aim: This paper aims to highlight and discuss the enigma of seeking or expecting trust attributes from a machine and, secondly, reframe the interpretation of trustworthiness for AI through evaluating its reliability and validity as consistent with the use of other clinical instruments.

Results: AI interventions should be described in terms of competence, reliability and validity as expected of other clinical tools where quality and safety are a priority. Nurses should be presented with treatment recommendations that describe the validity and confidence of prediction with the final decision for care made by nurses. Future research should be framed to better understand how AI is used to deliver care. Finally, there is a responsibility for developers and researchers to influence the conversation about AI and its power towards improving outcomes.

Conclusion: The sole focus on demonstrating trust rather than the business-as-usual requirement for reliability and validity attributes during implementation phases may result in negative experiences for nurses and clinical users.

Implications for practice: This research will have significant implications for the way in which future nursing is practised. As AI-based systems become a part of routine practice, nurses will be faced with an increasing number of interventions that require complex trust systems to operate. For any AI researchers

Corresponding author:
Oliver Higgins, School of Nursing and Midwifery, University of Newcastle, C/O Rhonda Wilson, 77a Holden Street, Gosford, NSW 2250, Australia.
Email: Oliver.Higgins@newcastle.edu.au

and developers, understanding the complexity of trust and creditability in the use of AI in nursing will be crucial for successful implementation. This research will contribute and assist in understanding nurses' role in this change.

Keywords
artificial intelligence, machine learning, mental health, nursing, psychosis, technology

Introduction

The role of Artificial Intelligence (AI) in nursing, especially in the mental health setting, is in the early stages of design and development. Increasingly, nurses are required to be active participants in contributing to the design, development and deployment of AI in clinical settings (Ronquillo et al., 2021) and ensure that any AI product used in clinical practice is done so safely and securely. Nursing AI research should focus on establishing programmes of practice that further develop the field if it is to be used to augment and support nursing practice and workflows (Ronquillo et al., 2021). For this to occur the incorporation of trust and empathy are fundamental elements within the design phases of building meaningful AI-informed clinical decision-making tools. This is in contrast with business as usual in the clinical context whereby reliability and validity are usually the favoured indicators of quality for 'analogue' decisional support tools used in the clinical settings. This point of difference presents a quandary in establishing quality and safety discourse in regard to innovation. Should reliability and validity of decision-making tools be sufficient alone as indicators of quality and safety, and/or should trust and empathy be incorporated or standardised as indicators of quality and safety? This paper will discuss several key issues at the cutting edge of AI and Nursing. It provides an encouraging but optimistically cautious viewpoint to better understand the role AI will contribute in Nursing's future. The paper examines mechanisms of trust, the expectations of AI, the role of empathy, anthropotheism of AI and the importance of appraising AI through an evidence-based lens rather than simply being expected to demonstrate trust. The aim of this paper is to highlight an inconsistency in approach to the assessment of quality and safety of supportive clinical decision-making tools and to provide some guidance to inform a balanced approach to the assessment of new innovation in supportive clinical decision-making tools.

Background

Clinical Decision Support Systems (CDSS) are one particular aspect to offer potential for mental health nurses, especially for complex areas such as the assessment and prediction of suicidal behaviours in critical settings. A recent integrative review investigated the role of decision support systems based on AI and machine learning in contemporary mental health care practice (Higgins et al., 2023). It found a limited amount of literature on the subject; however, a compelling problem was identified whereby clinician trust in AI systems emerged as a significant theme (Higgins et al., 2023).

Scientific reliability: Inconsistent expectations for nursing implementation

Established nursing practice relies upon many different instruments to augment and guide practice. One example from the general nursing field, is the Waterlow Score: a tool routinely used to

assess pressure ulcer risk (Charalambous et al., 2018). This tool has been widely incorporated into international practice, and is supported by extensive research that demonstrates the tool's reliability (Charalambous et al., 2018). In the mental health context, a further example is found in the Health of The Nation Outcome Scale (HoNOS) which is also used internationally (e.g. UK, Australia and New Zealand (James et al., 2018)) and has demonstrated consistent interrater reliability through an extensive body of research findings (James et al., 2018; Wing et al., 1998). Interestingly, the literature for these two well-established examples describe the reliability and validity of both these tools; however, the attribute of trustworthiness is not mentioned as a requirement for implementation (Charalambous et al., 2018; James et al., 2018). This highlights an inconsistency whereby reliability and validity appear to be preferred attributes for some tools, yet in contrast, an expectation is evident whereby trust has arisen as the preferred attribute for adoption of a tool such as AI in the clinical context. The nature of this inconsistency in expectation about tools used in nursing contexts that support clinical decision making requires further investigation and is the subject of this paper.

The ethics of black boxes and anthropomorphism

AI is increasingly evident in all parts of life; however, the resulting complex machines are not always understood, and can often be *black-box* in nature revealing a sentiment of significant concerns in implementation (Rai, 2019). The term *black-box* denotes that researchers may understand the system's inputs and outputs, but the process or method undertaken in the model to produce the output remains unknown (Rai, 2019). Significantly, with the unknown element a perplexing feature of the black-box and its inherent incorporation within the nature of AI, together with the human tendency to anthropomorphise machines and robots (Airenti, 2015), in combination, can result in humans attributing moral characteristics and behaviour to the AI (Ryan, 2020).

Additionally, the general media perpetuate popular perspectives about what these complex machines might do if left to their 'human-like devices', such as, depicted in ideas about sentient robots (Liang and Lee, 2017). This has resulted in an association and categorisation of what the '*intelligent*' machines are capable of by inferring human definitions and contexts (Liang and Lee, 2017). As such, the public expectation that AI should deliver trustworthy solutions has become increasingly influential. In response, government entities have formed organisations such as the European Commission High-level Expert Group (HLEG) on AI, to assist in the review and critique of AI use, and to promote *Ethical Guidelines for Trustworthy AI* (HLEG AI, 2019). Likewise, countries such as Singapore have undertaken a series of AI governance and ethics initiatives designed to build an ecosystem of trust that supports the adoption of AI (Singapore: InfoComm Media Development Authority, 2019; Singapore: Singapore Computer Society, 2020). As public concern has increased, the World Health Organization released the *Ethics and governance of artificial intelligence for health* in 2021, which calls upon organisations and governments to demonstrate trustworthiness with providers and patients for the ethical use of AI interventions in health care (World Health Organization, 2021). However, viewing the AI as trustworthy may undermine the value of interpersonal trust, thereby further reinforcing the notion of anthropomorphising the AI and ultimately diverting accountability from the organisations responsible for its design and development (Ryan, 2020).

Can AI be trusted?

The notion that AI has the capacity to be considered trustworthy is complex, as trust is an integral and essential part of human relationships (Ryan, 2020). HLEG proposes that the main

characteristics of trusting AI can be divided into three components, which include: the AI technology itself; the design and development organisations behind its conception and use; and, the socio-technical elements of the AI life cycle (HLEG AI, 2019). AI is a machine, albeit a sophisticated machine, but a machine nonetheless, and it is important to note that trustworthiness is not a value traditionally attributed to machines (HLEG AI, 2019). Therefore, while AI may meet the rational account of trust requirement, it cannot be held accountable for its actions as it does not possess the requisite emotional states to be held responsible for its own actions (Ryan, 2020). As such, it should not be considered trustworthy in the clinical setting, but rather considered in terms of reliability and validity (Ryan, 2020) and requires the tool to be appraised within the an Evidence-Based Decision Making framework (Sevy Majers and Warshawsky, 2020).

Many nurses have expressed reluctance in the use of AI, stating that the nature and complexity of learning algorithms can obscure the rationale or clinical reasoning for a recommendation (Rai, 2019). This can lead to hesitancy or anxiety for some nurses where they do not understand the process or logic that the AI system undertook to provide a particular recommendation (Brown et al., 2020). Nurses are less likely to engage or trust AI *black-box* recommendations (Brown et al., 2020), preferring an explicit understanding of the underlying AI mechanisms and logic process. AI-based innovations should be expected to demonstrate the rigour, reliability and validity of any other supportive clinical decisional intervention or instrument (Brown et al., 2020). Recent findings indicate that trust and confidence in AI are significant barriers to successfully implementing any intervention using AI CDSS in the mental health setting (Higgins et al., 2023). Imparting trust is an inherently human emotional construct (McLeod, 2021); as such, the concept of AI trustworthiness may be an additional barrier, preventing nurses from accepting AI within their practice. Without improvements in transparency about the logic and process within AI, it remains challenging for nurses to trust AI innovations despite the quality and validity indicators that may be apparent. Researchers and developers should note that there is more work to be done to assure nurses of confidence in AI CDSS in clinical practice.

Trust and confidence with the unknown

Trust can be defined as a confident relationship with the unknown (Botsman, 2018). It is the divide that exists between the two points of change, from where a person currently *'is'* to where the change requires them to *be*. Crossing the change divide requires a leap of trust, events such as the COVID-19 pandemic revealed that clinicians are required to take trust leaps faster and wider than ever before (Dodgson, 2022). The qualitative feedback during times of rapid change echoes with examples such as: *'I liked the old system much better', 'Why do we have to do this?'* or *'I have so much to do, and now I'm being asked to change X as well!'* (Botsman, 2018). AI developers and researchers should not underestimate the investment in trust required as the frequency of changes increase, as the change may be perceived by clinicians to directly challenge their highly valued clinical judgement and competence.

Typically, there are four fundamental domains divided into two paired subsets that are used when deciding whether to attribute trust in someone, or something. Firstly, competence and reliability, that is, explanation related to, *'How we do things'*. Secondly, integrity and empathy, represent aspects about *'Why we do things'* (Botsman, 2018). Applied together, the how and the why we do things assist in formulating our willingness to trust. Commonly, daily transactions occur that require people to undertake a leap of trust evoking the four trust domains aforementioned. For example, competence might be represented by questioning *Do they have the skills, knowledge, time and resources to do a particular task or job?* Reliability can be assessed by asking: *Can*

people depend on you to keep your promises and commitments? Integrity can be judged by considering: *Do you say what you mean and mean what you say?* Empathy can be demonstrated by deciding: *Do you care about the other person's interests as well as your own? Do you think about how your decisions and actions affect others?* (Botsman, 2018). It is possible to relatively confidently seek answers to these questions when asked of another human; however, it becomes more complicated when the same questions are asked of a non-sentient AI, and therefore, trust is more difficult to establish.

Competence, reliability and integrity

Competence and reliability pertain to *'How we do things'* and align with questions that are straightforward in application for an AI intervention. Competence is evaluated by asking: *Does the AI have the skills, knowledge, time, and resources to do the task? Are we open about what it is capable of?* Reliability is measured through examination of past performance: *Is it consistent in the way it behaves from one day to the next? Is it dependable?* Thus, the aforementioned questions can be considered with a logic that offers sound rationale for the concept of trust (Botsman, 2018). This same logic can be equally applied to both AI and clinical tool implementation.

AI integrity is a developing field of research, and at present it overlaps with the first two components of competence and reliability. Integrity is more complex than notions of competence and reliability, articulating as an attribute of '*Why we do things'*. To assess if someone, or something, has integrity, the following questions need to be considered: *Do you say what you mean, and mean what you say? Do your words align with your actions? Are you honest about its intentions and motives towards others?*

Integrity in AI is further comprised of four components, *health, explainability, security* and *reproducibility* (Talagala, 2019). The AI model, including the production and deployment system must be healthy. For example, AI predictions behave in production as expected and within norms specified by the developer. Explainability is necessary to determine why the algorithm behaved the way it did in devising predictions reveals the factors that led to the prediction. Security is established when the algorithm represents a truth in logic of health and explainability if it encounters malicious or non-malicious attacks that aim to change or manipulate its behaviour.

As the science evolves, explainable AI continues to address the *black-box* limitations with innovative research described as a *glass-box* technology. This glass-box acts to reveal, explain and investigate the inner workings of the model improving transparency (Nori et al., 2019). Lastly, an important scientific principle is that all predictions must derive from a chain of logic that is reproducible. If an outcome cannot be replicated, then, there is no reliable way to understand what led to the outcome, or debug issues if, and where, problems arise (Talagala, 2019). In summary, competence, reliability, and integrity are logical scientific requirements that characterise the scientific traditional representation of trustworthiness and enable confidence with the unknown. The number of countries worldwide upskilling their workforce with AI skills is not known; however, there are a number of nursing bodies actively investing in AI-powered nursing education and training programmes. In the United States, the National League for Nursing (NLN) has developed a competency framework for AI-enabled nursing practice (Lebo and Brown, 2022). The framework identifies the knowledge, skills and abilities that nurses need to be effective in using AI in their practice. The NLN is also working to develop AI-powered educational resources for nurses (Lebo and Brown, 2022). In the United Kingdom, the National Health Service (NHS) is investing in AI-powered tools to help nurses with tasks such as clinical decision-making, patient monitoring and medication administration. They are also working to develop AI-powered training programmes for nurses with nursing input (NHS AI Lab & Health Education England, 2022).

Empathy

Another factor that contributes to an exploration of trust in the context of AI is the notion of empathetic engagement. Empathy demonstrates an ability to understand and share the feelings of another, coupled with the ability to imagine what someone else might be thinking or feeling (Botsman, 2018). For example, '*Do you care about the other person's interests as well as your own? Do you think about how your decisions and actions affect others?'* (Botsman, 2018). AI is capable of meeting the first three trust components of competence, reliability and integrity. However, addressing empathy attribution is much more complicated for developers and researchers to achieve because it is not possible for a non-sentient machine to actually *'feel'* empathy, although person-like emulation is a design possibility with significant non-human limitations for the human user to navigate.

The concept of empathy and AI can be illustrated with a case explanation. Consider Pixar's 2008 animated fictional film WALL-E (Stanton, 2008), which follows the story of a little robot programmed to clean-up the earth, long after humanity has left. There is no dialogue in the first 30 minutes, with the animators applying various empathy-evoking techniques throughout this opening stanza to evoke empathy of the viewer towards the little robot. One viewer described WALL-E '*as a character with dimension, personality, and heart'* (Srivastava, 2016). Examining WALL-E reveals a vital clue: WALL-E depends on the human viewer to engage their human empathy to give *'him'* personality. WALL-E underscores the importance of empathetic engagement, for without empathy, WALL-E is not perceived as *real* (Srivastava, 2016). Herwix et al. (2022) observed a relationship between *perceived technical competence* and both *affect-based* and *cognitive-based* trust, reinforcing the importance of empathy in human computer interaction (HCI) when planning digital health interventions (Søgaard Neilsen and Wilson, 2019). Just as empathetic engagement occurs for the viewer, so too, it is important that clinicians, researchers and developers recognise the empathetic components necessary to convey holistic trust in AI within the clinical setting.

As we have discussed, all four trust domains must be actively engaged for trust to arise in an AI context. The '*How we do things'* facet of competence and reliability is relatively straightforward; however, the '*Why we do things'* characteristics related to integrity and empathy are more complex. Complexity in the context of rapid development in AI innovations applied to clinical settings will require clinicians to take wider leaps of trust with increased frequency. Researchers and developers should recognise that an expectation of trust investment is required when people are asked to leap into unknown phenomena, and consider ways that they can facilitate the development of a confident relationship with the unknown to establish prerequisite trust formation (Botsman, 2018). Understanding the mechanisms that inform the components of trust helps to identify best fit alignment of AI implementation within the clinical environment.

Designed to empathise

Researchers and developers must account for the aforementioned four trust domains as they design and implement AI innovations. A further link in logic is the alignment of empathy. The theoretical approach of *Design Thinking* promotes the inclusion of an important consideration of empathy in the context of AI. Design thinking employs an iterative process in which the researcher seeks to understand the requirements of all users, challenge assumptions, redefine problems and create innovative solutions which can be prototyped and tested systematically (Köppen and Meinel, 2015). Crucially, design thinking requires a consideration of empathy as a first step, employing a cognitive and emotional technique that will aid in the development process. The *Emphathise*

domain is focused on exploring the nature of the problem and understanding the users and their needs, doing so by asking user perspectives directly through focus groups or one-to-one sessions with end users (Köppen and Meinel, 2015). The synthesis of this phase gives rise to a categorisation step which *defines* the main findings and acts as a *'persona'* (an ideal user) to validate decisions later in the process. The design thinking process accelerates development, while including clinicians, designers, researchers and lived experience in the empathetic design and application of AI interventions. While the literature on the topic is limited, a recent review revealed evidence supporting the importance of inclusion of clinical staff in the final decision-making process (Higgins et al., 2023), and in doing so, it is possible to generate requisite clinician empathy within the AI environment. Developers and researchers of AI innovations have an ethical responsibility for AI solutions they invent, balancing building on strengths while recognising and compensating for its weaknesses to mitigate risks while promoting safety and quality for the end user (Ryan, 2020). Empathetic AI is not about teaching machines to feel emotion in the way a person might do so, but instead to use AI coding with regard to ethics, thereby facilitating sufficient empathy to determine the next best action to take for a patient.

As researchers and developers look towards the future, the empathic components of their innovations will grow ever more crucial for success. Consider the future presented in the film 'Robot and Frank', in which a robot designed with empathy provides care and companionship to those who may not have the luxury (Schreier, 2012). How developers and researchers deal with empathy and anthropomorphism will be crucial in adoption and value to the end user. However, they must consider anthropomorphism and its role for non-human entities to be given human characteristics and feelings (Airenti, 2015). To better understand anthropomorphism, consider the 'Braitenberg vehicle'. A thought experiment suggesting that straightforward act of the vehicles (robots) is deliberate and motivated by emotions rather than merely being governed by their preprogrammed reactions (Braitenberg, 1986). For instance, a Braitenberg vehicle might perform in a way that appears 'fearful' or 'curious' due to how it responds to environmental cues, which could cause a bystander to anthropomorphise the vehicle and attribute it with these emotions. In Braitenberg vehicles, a minor increase in the complexity of a system can quickly transform a relatively basic system into a black box for humans (Rai, 2019). As a result, people may begin to lose faith in the system as they may believe that they cannot influence the vehicle's behaviour (Braitenberg, 1986). The resulting sophisticated system's lack of transparency in its behaviour can lead people to become less empathetic towards it since they are less likely to perceive the system as a thinking being with its own emotions and goals (Braitenberg, 1986). The importance of empathy in acceptance can be evidenced in Chat-GPT (a large language model) which has been recently launched (OpenAI, 2022). Chat-GPT has been trained on so much data (almost all of the internet) that its abilities can exceed that of humans while providing human-like answers (Aljanabi et al., 2023). Chat-GPT has been tested using the Emotion Twenty Questions (EMO20Q), the results demonstrating it possesses the ability to engage with a human and identify human emotion in a method similar to the way medical students are trained (Noever and McKee, 2023). Future innovations will require the empathetic component to be clearly understood for the innovation to succeed in adoption.

Implications for incorporation of AI CDSS in nursing practice

Four implications for practice have been identified in this discussion. Firstly, AI interventions should be described in terms of competence, reliability and validity as expected of other clinical tools where quality and safety are a priority. This consistency in the narrative supports the

incorporation of AI augmentation within clinical practice. Acceptance of an intervention based on trust alone is not based on a consistent chain of logic, and risks compromising quality and safety related to its implementation. It is also apparent that a detailed description of the functions of decision logic is required to support the implementation of AI-based interventions. This must include clear actions that describe how clinical features contributed to the outcome while communicating multiple recommendations, and demonstrating confidence with prescribing recommendations. In Australia, the Australian Alliance for Artificial Intelligence in Healthcare (AAAiH) has published the roadmap for Artificial Intelligence in Healthcare for Australia, offering a high-level view of the use of AI in care and recommend the acceleration of training the health workforce in AI by developing foundational AI frameworks for health professionals and accredited programmes to train specialist AI health professionals (Australian Alliance for Artificial Intelligence in Healthcare, 2021). It is imperative that nursing leadership bodies develop AI strategies to ensure Nursing has a clear direction and voice in this rapid change.

Secondly, Nurses should be presented with treatment recommendations that describe the validity and confidence of prediction and shown the chain of logic that rests the final decision for the appropriate course of treatment with nurses. Tools that utilise *glass-box* design, such as InterpretML (Nori et al., 2019), which allow the developer and researcher to understand the model's inner workings, should be encouraged and incorporated into clinical interfaces.

Thirdly, future research should be framed to better understand how AI is used to deliver care and how it will affect those with a mental illness, as research has previously demonstrated, especially where psychosis is a targeted condition for AI interventions (Higgins et al., 2023). Further, design thinking concepts should be applied to research to reveal AI attributes required to deliver effective care for those with mental illness. Importantly, this will identify where empathy should be located within the intervention, how it is to be applied and how this is communicated. Finally, there is a responsibility for developers and researchers to influence the conversation about AI, taking control and proving its value through responsible applications and directing its power towards improving outcomes (PEGA, 2019).

Conclusion

Future AI research will have significant implications for the way in which nursing is practiced. As AI-based systems become a part of routine practice, nurses will be faced with an increasing number of interventions that require complex trust systems to operate. For any AI researchers and developers, understanding the complexity of trust and creditability in the use of AI in nursing will be crucial for successful implementation. Trustworthiness is a complex concept and when considered from a HCI perspective, such as in the case of AI CDSS in the nursing context, it must also be considered in conjunction with competence, reliability, integrity and empathy – competence and reliability to validate the capability and dependability of the system, integrity to ensure explainability and empathy to determine the next best action to take for a patient. To discuss AI in terms of its trustworthiness alone risks illogical endorsement of human attributes to a machine. This paper has presented an overview of the components of trust in the context of AI applied to the nursing context and clinical setting. Understanding the roles of trust should be considered an important step to consider for the successful acceptance of AI into clinical practice. Communicating the system's validity and reliability while creating transparent innovations will form a significant component to support the implementation of AI-based clinical instruments in routine clinical nursing practice.

Key points for policy, practice and/or research

- AI strategies need to be developed by Nursing leadership bodies to ensure Nursing has a clear direction and voice in this rapid change.
- Nurses need to partner with developers and researchers to influence the conversation about AI, taking control and proving its value through responsible applications and directing its power towards improving outcomes.
- Nurses should be presented with treatment recommendations that describe the validity and confidence of prediction and shown the chain of logic that rests the final decision for the appropriate course of treatment with nurses.
- AI evaluation needs model evidence-based practice guidelines, interventions should be described in terms of competence, reliability and validity as expected of other clinical tools where quality and safety are a priority.
- Future research should be framed to better understand how AI is used to deliver care and how it will affect those with a mental illness.

Acknowledgements

Authors would like to acknowledge the support of Central Coast Local Health District.

Author contributions

OH – concept development, project design, data collection, data analysis, manuscript preparation.
SC – data analysis, contribution to manuscript, supervision of project.
RW – concept development, project design, data collection, data analysis, manuscript contribution, supervision of project.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Partial financial support was received from NSW Ministry of Health as part of the Towards Zero Suicides initiative.

Ethical approval

Ethical clearance is not required as this review is based on published literature.

ORCID iD

Oliver Higgins https://orcid.org/0000-0001-6545-145X

References

Airenti G (2015) The cognitive bases of anthropomorphism: from relatedness to empathy. *International Journal of Social Robotics* 7: 117–127.

Aljanabi M, Ghazi M, Ali AH, et al. (2023) ChatGpt: Open Possibilities. *Iraqi Journal For Computer Science and Mathematics* 4: 62–64.

Australian Alliance for Artificial Intelligence in Healthcare (2021) *A Roadmap for AI in Healthcare for Australia*. AAAIH, Macquarie University.

Botsman R (2018) *Who Can You Trust?: How Technology Brought Us Together and Why It Might Drive Us Apart*. London: Penguin.

Braitenberg V (1986) *Vehicles: Experiments in Synthetic Psychology*. Cambridge, MA: MIT press.

Brown LA, Benhamou K, May AM, et al. (2020) Machine learning algorithms in suicide prevention: Clinician interpretations as barriers to implementation. *Journal Clinical Psychiatry* 81: 19m1l2970. DOI: 10.4088/JCP.19m12970.

Charalambous C, Koulori A, Vasilopoulos A, et al. (2018) Evaluation of the validity and reliability of the waterlow pressure Ulcer Risk Assessment Scale. *Medical Archives* 72: 141–144, DOI: 10.5455/medarh.2018.72.141-144.

Dodgson JE (2022) The pandemic has brought too much change: Too many preprints; too many retractions. *Journal of Human Lactation* 38: 207–208. DOI: 10.1177/08903344221082571.

Herwix A, Haj-Bolouri A, Rossi M, et al. (2022) Ethics in information systems and design science research: Five perspectives. *Communications of the Association for Information Systems* 50: 34.

Higgins O, Short BL, Chalup SK, et al. (2023) Artificial intelligence (AI) and machine learning (ML) based decision support systems in mental health: An integrative review. *International Journal of Mental Health Nursing* 32: 966–978. DOI: 10.1111/inm.13114.

HLEG AI (2019) *Ethics Guidelines for Trustworthy AI*. Available at: https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai

James M, Painter J, Buckingham B, et al. (2018) A review and update of the Health of the Nation Outcome Scales (HoNOS). *BJPsych Bulletin* 42: 63–68.

Köppen E and Meinel C (2015) Empathy via design thinking: Creation of sense and knowledge. In: Plattner H, Meinel C and Leifer L (eds) *Design Thinking Research*. Cham: Springer, pp. 15–28.

Lebo C and Brown N (2022) Integrating Artificial Intelligence (AI) simulations into undergraduate nursing education: An evolving AI patient. *Nursing Education Perspectives*. Epub ahead of print. 21 December 2022. DOI: 10.1097/01.Nep.0000000000001081.

Liang Y and Lee SA (2017) Fear of autonomous robots and artificial intelligence: Evidence from National Representative data with probability sampling. *International Journal of Social Robotics* 9: 379–384. DOI: 10.1007/s12369-017-0401-3

McLeod C (2021) *Trust*, Fall 2021 edn. Metaphysics Research Lab, Stanford, CA: The Metaphysics Research Lab Philosophy Department Stanford University.

NHS AI Lab & Health Education England (2022) *Developing Healthcare Workers' Confidence in AI*. Available at: https://digital-transformation.hee.nhs.uk/binaries/content/assets/digital-transformation/dart-ed/understandingconfidenceinai-may22.pdf (accessed 07 February 2023).

Noever D and McKee F (2023) Chatbots as problem solvers: Playing twenty questions with role reversals. *arXiv preprint arXiv*:2301.01743

Nori H, Jenkins S, Koch P, et al. (2019) Interpretml: A unified framework for machine learning interpretability. *arXiv preprint arXiv*:1909.09223.

OpenAI (2022) *ChatGPT: Optimizing Language Models for Dialogue*. Available at: https://openai.com/blog/chatgpt/ (accessed 07 February 2023).

PEGA (2019) *AI and Empathy: Combining Artificial Intelligence With Human Ethics for Better Engagement*. Available at: https://www.pega.com/insights/resources/ai-and-empathy-combining-artificial-intelligence-human-ethics-better-engagement (accessed 07 February 2023).

Rai A (2019) Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science* 48: 137–141.

Ronquillo CE, Peltonen LM, Pruinelli L, et al. (2021) Artificial intelligence in nursing: Priorities and opportunities from an international invitational think-tank of the Nursing and Artificial Intelligence Leadership Collaborative. *Journal of Advanced Nursing* 77: 3707–3717. DOI: 10.1111/jan.14855.

Ryan M (2020) In AI we trust: Ethics, artificial intelligence, and reliability. *Science and Engineering Ethics* 26: 2749–2767. DOI: 10.1007/s11948-020-00228-y.

Schreier J (director) Acord L, Bisbee S, Kelman-Bisbee J and Niederhoffer G (producer) (2012) *Robot and Frank*. Stage 6 Films.

Sevy Majers J and Warshawsky N (2020) Evidence-based decision-making for nurse leaders. *Nurse Lead* 18: 471–475. DOI: 10.1016/j.mnl.2020.06.006.

Singapore: InfoComm Media Development Authority (2019) *Singapore Wins International Award for its Artificial Intelligence Governance and Ethics Initiatives*. Available at: https://www.imda.gov.sg/newsand-events/Media-Room/Media-Releases/2019/singapore-wins-international-award-for-its-artificialintelligence-governance-and-ethics-initiatives (accessed 07 February 2023).

Singapore: Singapore Computer Society (2020) Singapore Computer Society, InfoComm Media Development Authority. *AI Ethics & Governance Body of Knowledge*. Available at: https://ai-ethics-bok.scs.org.sg/ (accessed 07 February 2023).

Søgaard Neilsen A and Wilson RL (2019) Combining e-mental health intervention development with human computer interaction (HCI) design to enhance technology-facilitated recovery for people with depression and/or anxiety conditions: An integrative literature review. *International Journal of Mental Health Nursing* 28: 22–39. DOI: 10.1111/inm.12527.

Srivastava MB (2016) The computational and aesthetic foundations of artificial empathy. *Intersect: The Stanford Journal of Science, Technology, and Society* 10: 8.

Stanton A (director) (2008) *WALL-E* Pixar.

Talagala N (2019) *ML Integrity: Four Production Pillars For Trustworthy AI*. Available at: https://www.forbes.com/sites/cognitiveworld/2019/01/29/ml-integrity-four-production-pillars-for-trustworthy-ai/?sh=b6dce0a5e6fe (accessed 25 May 2022).

Wing JK, Beevor AS, Curtis RH, et al. (1998) Health of the Nation Outcome Scales (HoNOS): Research and development. *British Journal of Psychiatry* 172: 11–18, doi:10.1192/bjp.172.1.11.

World Health Organization (2021) *Ethics and Governance of Artificial Intelligence for Health: WHO Guidance*. Geneva: WHO.

Oliver Higgins is a mental health nursing scientist and computing scientist with a research focus on Artificial Intelligence and Machine Learning in Mental Health, specifically Emergency Department presentations.

Stephan K Chalup work is built on an in-depth understanding of the human brain and how it works as a complex and powerful processing system. Through his work, Stephan aims to explain how human thoughts are represented and processed in our brains – and provide tantalising insights into what makes us tick.

Rhonda L Wilson is a Wiradjuri woman, experienced nurse and an internationally recognised mental health nursing scientist. Her work in e-health is paving the way for new digital therapeutic interventions that promote and support patient-centred care and increased wellbeing.

4.1.1 Article Summary

This theoretical discussion of trust explored the ethical responsibilities of AI developers and researchers in the healthcare sector, emphasising the need for AI innovations to be developed with an awareness of both their strengths and limitations to ensure that risks are mitigated while promoting safety and quality for end users. When incorporating AI-based CDSS into nursing, these systems should be held to the same standards of competence, reliability, and validity as any other clinical tool, as opposed to demonstrating trustworthiness. This means AI interventions must provide transparent treatment recommendations, presenting clinicians with information that includes the confidence level of predictions and a clear chain of logic behind the recommendations. This approach would not only maintain nurse oversight but also foster trust in AI systems by making their decision-making processes transparent—an approach exemplified by tools such as InterpretML (Nori et al., 2019).

The research also discusses the importance of understanding how AI is used in MH care, particularly with conditions like psychosis, and how AI interventions affect patients with mental ill health. Applying design thinking (DT) principles (Brown, 2008) in this research can

help uncover the necessary attributes AI needs to deliver effective care, with a specific emphasis on locating and applying empathy within AI interventions. Finally, developers and researchers need to take an active role in shaping the conversation around AI, demonstrating its value through responsible applications that focus on improving patient outcomes while adhering to ethical frameworks such as those of the World Health Organization (2021a) and the Australian Government's Department of Industry, Science and Resources (2021). By doing so, they can demonstrate and build trust in AI systems, ensuring that they are used effectively and ethically in clinical practice.

4.2 Chapter Summary

The findings of this theoretical discussion underscore the critical importance of framing AI not as a tool to be inherently "trusted" but as a clinical instrument whose reliability and validity must be rigorously evaluated in an ongoing capacity. This perspective is crucial for advancing the discourse on AI in nursing, as it shifts the focus from an emotional reliance on technology to an objective assessment of its functional performance. This distinction is pivotal in addressing the project's aims of complementing and augmenting clinicians' decision making, enhancing the client's experience of MH services, and understanding AI/ML adoption in nursing practice. By evaluating AI systems through the lens of clinical competence, reliability, and validity, healthcare professionals can ensure that these technologies meet the same standards as other clinical tools, which prioritise safety and quality of care.

Building on these conclusions, the future of AI in nursing will depend on the development of systems that consistently demonstrate competence, reliability, and validity. This aligns with the aim of enhancing clinicians' decision making by ensuring that AI is seen as a tool that provides transparent, evidence-based recommendations rather than a system requiring blind trust. To achieve this, AI developers and healthcare professionals must embrace a paradigm

shift, prioritising transparency in algorithmic processes and outcomes to foster a more informed and confident use of AI within clinical practice.

Additionally, future research should continue to explore the dynamic between AI reliability and user trust, particularly within nursing, as this remains a key factor influencing the adoption of AI systems. This is directly relevant to the project's aim of understanding AI adoption in a MH context, as trust is a fundamental element in both clinicians' and patients' acceptance of AI. It is essential to develop frameworks that capture the complexities of trust in human–AI interactions, considering not only technological performance but also the socio-cultural factors that shape perceptions and acceptance of AI.

As AI technology continues to evolve, longitudinal studies will be vital in assessing the long-term impacts of AI integration on nursing practice and patient care outcomes. These studies will provide critical insights into how AI can sustain improvements in both clinical decision making and the overall quality of care, addressing the overarching project aims of enhancing the client's experience in MH services while maintaining high standards of safety and care.

Chapter 5 Update to Literature Review

As of 14/02/2025 this Literature Review has been accepted for publication.

Higgins, O., & Wilson, R. L., (2025). Integrating Artificial Intelligence (AI) with Workforce Solutions for Sustainable Care: A follow up to Artificial intelligence and machine learning (ML) based decision support systems in mental health. *International Journal of Mental Health Nursing*. Accepted

Since the outset of this research, the rapid advancements in artificial intelligence (AI) and machine learning (ML) have expanded the technological capabilities available to health care. These developments, particularly the rise of large language models (LLMs) such as ChatGPT, present both opportunities and challenges for healthcare systems, especially in terms of integrating these tools into clinical workflows while maintaining the highest standards of care, ethics, and community benefit. This section revisits and updates the original literature review, focusing on new research published between 1 January 2022 and 30 June 2024. The review indicates a significant increase in AI-related studies, with five articles published during the revision period compared to four in the previous five years, demonstrating the growing importance of AI and ML in health care. This re-examination is especially important given that LLMs have gained significant attention for their potential to transform healthcare delivery, enhance decision making, and impact patient outcomes.

This updated literature review directly addresses the research aims by re-exploring how AI can complement and augment clinicians' decision making. The review highlights the potential of AI-driven clinical decision support systems (CDSS) to assist clinicians in making more informed, accurate, and timely decisions, thereby reducing the incidence of missed care. The review also re-emphasises the importance of maintaining clinician oversight to ensure that AI recommendations are used to enhance, not replace, human judgement. In addressing the second aim, enhancing the client's experience of MH services while providing high-quality and safe

care, the chapter explores how AI technologies, including LLMs, can help address clinicians' workload, allowing more time for direct patient care. This can significantly improve the patient experience, especially in MH services, where timely, personalised care is essential.

This chapter continues to contribute to the understanding of the adoption of AI in MH care, addressing the third project aim. It identifies ongoing challenges related to trust, transparency, and bias in AI systems, which are crucial to ensuring clinicians' and patients' confidence in using these tools. LLMs, in particular, introduce additional concerns about transparency and interpretability, as their decision-making processes can be difficult to fully understand. Furthermore, this chapter will contribute to a deeper understanding of AI adoption in MH care by briefly evaluating recent literature on LLMs and their impact. As LLMs become more embedded in everyday practice, it is vital to understand their role in improving or impairing issues related to care, as well as their capacity to support clinicians in delivering holistic, patient-centred care. While these technologies have the potential to complement clinicians' decision making, mitigate missed or unfinished care, and enhance the patient experience, their successful implementation hinges on addressing systemic workforce challenges and ensuring trust and transparency in their application. A balanced approach that integrates AI innovations with improvements to the healthcare workforce will be essential for achieving sustainable and meaningful advancements in MH care, aligning with the overarching aims of this study.

5.1 Update to Literature Review

5.1.1 Introduction

Since the start of this research, the capabilities of AI and ML have advanced significantly. Healthcare systems now face the challenge of integrating these powerful tools into clinical workflows while maintaining the highest standards of care, medical ethics, and community benefit. This section updates the original literature review with research published between 1

January 2022 and 30 June 2024. Notably, five articles have been published in the past 18 months, compared to four in the previous five years, reflecting the rapid growth of AI and ML in health care. This updated review aligns with the study's aims by examining how these technologies can support and enhance clinicians' decision making, providing advanced decision support tools to improve care delivery. Additionally, it investigates how AI/ML systems can enhance the client's experience in MH services by ensuring safe and high-quality care. Finally, the review aims to deepen the understanding of AI/ML adoption in MH care, highlighting both the opportunities and challenges that have emerged as these technologies continue to evolve and become more prevalent in clinical practice.

5.1.2 Methods

5.1.2.1 Aims

The research questions guiding this integrative review were:

1. *Is there evidence to support the use of artificial intelligence or machine learning based decision support systems in the delivery of mental health care?*
2. *What barriers exist for mental health end users (clinicians and patients) in the adoption of artificial intelligence or machine learning based decision support systems?*

5.1.2.2 Design

This integrative literature review followed the framework outlined by Whittemore and Knafl (2005), which encompasses the stages of problem identification, conducting a systematic literature search, retrieving data, evaluating articles, and performing data analysis and presentation. This framework was chosen for its ability to rigorously and systematically analyse data from two disciplines that are not traditionally combined, facilitating a comprehensive synthesis of information while accommodating all research designs. The methodological

guidelines provided in the Preferred Reporting Items for Systematic Reviews and Meta-Analysis Protocols (PRISMA-P) 2020 (Page et al, 2021) statement checklist guided the review process to ensure it was thorough and adhered to high standards (Figure 5.1). Thematic analysis undertaken using Braun & Clarke (2021).

5.1.2.3 Search Strategy

A literature search of databases was conducted on 30 June 2024. The search was developed in Medline then adapted as necessary to Scopus, Web of Science, Google Scholar, IEEE Xplore, and CINAHL with Full Text (EBSCOhost). To find relevant literature, the search terms "machine learning" OR "artificial intelligence" AND "mental health" OR "psychiatry" AND "decision support" were used. The search was limited to articles published between 2022 and 2024, written in English, and focused on AI- or ML-based CDSS for MH. Only primary research articles published from 2022 onward and accessible online were included. This approach aimed to identify technological advancements in AI/ML-based decision support systems for MH, while excluding non-primary research, non-English articles, and studies not focused on AI- or ML-based CDSS for MH or psychiatry. For further details, please review Chapter 3 of this thesis. The method proposed by Whittemore and Knafl (2005) was employed in synthesising and analysing the data. This approach involved data reduction, display, comparison, and conclusion verification. Information from each article was categorised as study demographics, methodologies, sample populations, key findings, and limitations. The articles were grouped based on these themes for comparison and conclusion verification, and the resulting information is displayed in Table 5.1 to highlight further themes and relationships.

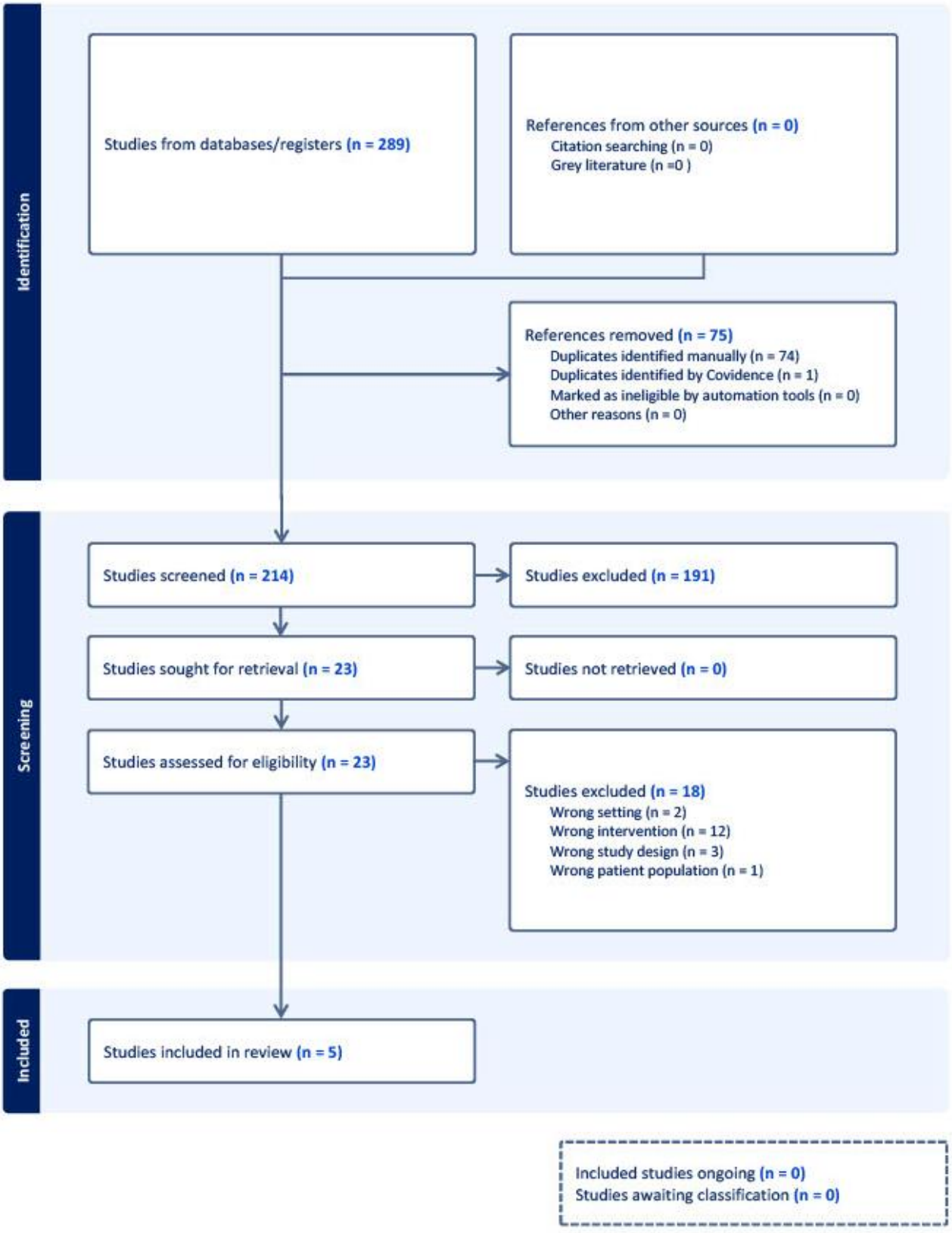


Figure 5.1. *PRISMA flowchart*

5.1.3 Results

Clinical decision support systems (CDSS) for MH, including those incorporating AI, show promise in improving patient care. Overall, the results of this subsequent literature review highlight the potential of CDSS to improve MH care but emphasise the importance of rigorous evaluation, debiasing efforts, and ongoing monitoring to ensure fairness, accuracy, and clinical utility. Five articles were deemed eligible and included in this review. Interestingly, two of these articles, Qassim et al. (2023) and Moggia et al. (2022), are further research papers from studies that were included in the original literature review (Higgins, Short et al., 2023). The research articles included a range of study designs and methods. One article employed a mixed-methods approach, utilising questionnaires and semi-structured interviews to examine the feasibility of an AI-powered clinical decision support system for depression treatment in Quebec, Canada (Qassim et al., 2023). Another study focused on developing a healthcare decision support system to predict MH risks, drawing on data from a bilingual survey, in-depth interviews, and a case study of immigrants and refugees in the US–Mexico border region (Ngan et al., 2022). Similarly exploring AI's role in MH, a research article from Shanghai, China investigated a multimodal, multitask learning model for rehabilitating patients with serious mental illnesses, using patient records and expert interviews to evaluate their algorithm (Yang et al., 2024). Although focusing on data-informed case formulation within psychotherapy, a German study does not explicitly state its study design or methodologies (Moggia et al., 2022).

The evaluation of an ML model's suitability for integration into an electronic health record (EHR) system to predict postpartum depression (PPD) risk has been a focal point of recent research. One study utilised this model to identify high-risk patients and recommended preventative interventions for clinicians, with a specific emphasis on refining decision

thresholds through decision curve analysis to optimise sensitivity while minimising the risk of overtreatment (Liu et al., 2024). Similarly, the feasibility and acceptability of integrating AI-driven CDSS, such as Aifred, into the treatment of major depression were explored, rather than directly testing interventions. This investigation highlighted how clinicians incorporated the system into their practice, particularly in terms of appointment length and the clinician–patient relationship (Qassim et al., 2023), building upon previous work by Benrimoh et al. (2020) which was cited in Higgins, Short et al. (2023).

In another study, a novel multimodal, multitask learning model was directly applied as an intervention to predict rehabilitation outcomes for individuals with severe mental illness (SMI). The model, which integrated Bidirectional Encoder Representations from Transformers (BERT) for text data (e.g., doctor's notes) and TabNet for structured diagnostic information, was evaluated for its predictive accuracy in comparison to traditional single-task models across four domains: referral risk, dangerous behaviours, self-awareness, and medication adherence (Yang et al., 2024).

Further research has detailed the development of a healthcare decision support system (HDSS) aimed at identifying MH risks among immigrants and refugees. Rather than implementing a separate intervention, the study compared the HDSS's performance in detecting these risks with assessments from domain experts and a purely data-driven analytics approach (Ngan et al., 2022). Additionally, the potential of the Trier Treatment Navigator (TTN), a CDSS for psychotherapy, was examined, with an emphasis on its role in data-informed case formulation and treatment adaptation. While this study did not focus on a specific intervention or comparative analysis, it built on previous work by Lutz et al. (2021), also referenced in Higgins, Short et al. (2023), and presented the TTN's application through a clinical case study (Moggia et al., 2022).

5.1.3.1 Thematic Analysis of Research on Decision Support Systems in Mental Health

The findings from the follow-up review align well with the key themes identified in the previous literature review, particularly concerning the potential and limitations of AI-based CDSS in MH care. Both reviews underscore the significant promise of AI in improving MH care delivery, with an emphasis on supplementing, rather than replacing, clinical judgement. However, this follow-up review extends the discussion by incorporating recent studies that further illustrate the practical applications and challenges that AI-driven systems face.

The current review reinforces the theme of AI's role in alleviating clinicians' workload, as highlighted in Chapter 3. However, it also builds upon the previous insights by stressing the necessity of rigorous evaluation frameworks, debiasing efforts, and continuous monitoring of these systems to ensure fairness, accuracy, and clinical utility. For instance, Qassim et al. (2023) and Moggia et al. (2022), both from studies that were included in the original review (via Lutz et al. (2021) and Benrimoh et al. (2020)), offer further evidence of the feasibility and acceptability of AI-driven CDSS in real-world clinical settings. This continuation of previously reviewed work illustrates how AI tools can be integrated into clinical practice, highlighting the need for clinician engagement and system refinement to improve decision-making processes. The review strengthens the focus on interdisciplinary collaboration, particularly by incorporating various study designs, such as mixed-methods approaches (Qassim et al., 2023), and examining CDSS systems across diverse populations and settings. The use of AI in the context of marginalised populations, such as immigrants and refugees (Ngan et al., 2022), adds depth to the conversation on equity and the ethical deployment of AI, a key issue raised in the earlier review.

Additionally, the introduction of advanced ML models, such as the multimodal, multitask learning model explored by Yang et al. (2024), builds upon the previous emphasis on

developing interpretable AI models tailored to MH applications. This research (Yang et al., 2024) highlights how such models can be used to predict complex MH outcomes, enhancing the capacity of AI to support clinical decision making in rehabilitation contexts. This reinforces the theme of transparency and interpretability from the previous review, where it was noted that clinician trust is pivotal for the successful adoption of AI technologies in MH care.

The ongoing exploration of the TTN, as discussed in Moggia et al. (2022), aligns with earlier considerations on using AI to improve data-informed case formulation in psychotherapy. While previous literature acknowledged AI's potential to optimise workflows and enhance clinical decision making, the case study in this review offers further examples of how these systems can be applied in psychotherapy, though without focusing on specific interventions.

This review builds upon the original findings reported in Chapter 3 and Higgins, Short et al. (2023) by further demonstrating the practical applications of AI in MH care, expanding the range of study designs and methodologies, and reinforcing the importance of clinician and patient engagement. It also advances the discussion on ethical concerns, particularly around transparency, interpretability, and bias mitigation, while providing new insights into AI's role in addressing MH challenges across diverse populations. These studies collectively emphasise that, while AI has significant potential in MH care, its integration must be carefully managed through ongoing research, collaboration, and a balanced approach that combines technological innovation with workforce development. The included research articles explore the intersection of clinical expertise and data-driven insights in MH care, particularly through developing and applying CDSS. These systems aim to provide clinicians with tools to enhance decision making, personalise treatment, and potentially improve patient outcomes. A recurring theme is the potential of CDSS to integrate diverse data sources, including patient demographics, diagnostic information, treatment history, and even unstructured text data like clinicians' notes, to generate predictions about various aspects of MH.

5.1.3.2 Harnessing Data for Personalised Mental Health Care

A central theme across the articles is the pursuit of personalised treatment in MH care. Traditionally, treatment decisions often rely on clinicians' judgement and experience. However, the articles highlight the potential of CDSS to leverage data to tailor treatments to individual patient characteristics and needs. This shift towards data-informed personalisation is exemplified by systems like Aifred (Qassim et al., 2023), which provides clinicians with remission probabilities for different treatment options based on patient data and predictive algorithms. Similarly, the HDSS described by Ngan et al. (2022) combines domain expertise with data analytics to predict the risk of MH problems in specific populations, such as immigrants and refugees.

5.1.3.3 The Evolving Role of Technology and Data in Clinical Practice

The research articles collectively illustrate the evolving role of technology and data in MH care. There is a clear emphasis on moving beyond traditional methods of diagnosis and treatment towards more data-driven approaches. This transition is evident in the development of CDSS, which integrates routine outcome monitoring (ROM) and measurement-based care (MBC) into clinical workflows (Moggia et al., 2022). ROM and MBC regularly collect and analyse patient data throughout treatment to track progress and inform treatment decisions. By incorporating these data-driven practices, the articles suggest clinicians can gain a more objective and comprehensive understanding of their patients' needs and progress.

5.1.3.4 Addressing Challenges and Ethical Considerations

While the potential benefits of CDSS are widely acknowledged, the research articles also address the challenges and ethical considerations associated with their development and implementation. A key concern is ensuring the accuracy, fairness, and clinical utility of these

systems. One article emphasises the importance of rigorously evaluating CDSS before widespread adoption to avoid potential biases and ensure they are effective in real-world clinical settings (Yang et al., 2024). Another article explores the technical aspects of developing a multimodal, multitask learning model for assessing the rehabilitation status of patients with severe mental illness (Qassim et al., 2023). The authors highlight the complexities of working with multimodal data (e.g., text records and structured diagnostic data) and the need for robust model evaluation to ensure accuracy and fairness.

Beyond technical challenges, the articles also underscore the importance of addressing ethical considerations, such as data privacy and the potential for bias. For instance, the development of the multimodal, multitask learning model included obtaining informed consent from all participants and addressing privacy concerns by ensuring data confidentiality. The researchers also acknowledge the potential for algorithm bias and the need for ongoing monitoring to mitigate these risks.

5.1.3.5 Fostering Collaboration and Trust in the Age of AI

A crucial aspect of successfully implementing CDSS in MH care is fostering collaboration and trust between clinicians, patients, and technology. The articles emphasise that CDSS are not intended to replace clinical judgement but rather to augment it by providing clinicians with additional insights and support. For example, one article describes how the TTN provides therapists with theoretical input, examples of therapeutic dialogues, and potential exercises to support their clinical decision making in the context of data-informed psychotherapy (Moggia et al., 2022).

Ultimately, the research articles convey a sense of cautious optimism about CDSS's potential to transform MH care. They highlight the opportunities presented by data-driven insights while acknowledging the importance of addressing ethical concerns and prioritising patient well-

being. The articles suggest that, by carefully navigating these complexities, the field can harness the power of technology to deliver more personalised, effective, and equitable MH care for all.

Table 5.1. Updated literature review article summary

Author(s)	Study aim	Design	Results	Limitations	Recommendations
Liu et al. (2024) USA	Evaluate and optimise a machine-learning model for predicting postpartum depression (PPD).	Model evaluation. Three datasets were used: (1) EHR data from an academic medical centre in 2019; (2) EHR data from the same AMC in 2020; and (3) EHR data from a clinical research network between January 2014 and September 2018. Decision curve analysis was used to evaluate the model's clinical utility and net benefit across different decision thresholds.	A revised model, using fairness through blindness, was identified as the most suitable, balancing performance and fairness considerations. This approach aimed to mitigate potential biases in the prediction model, particularly concerning race.	The study primarily relied on data from a single urban region within the US, potentially limiting the generalisability of the findings to other settings and populations.	• Prioritise sensitivity over precision when addressing PPD risk to avoid missing at-risk patients. • Conduct additional validations to verify the study's generalisability. • Evaluate implementation strategies for more complex ML models in future studies. • Consider the potential impact of excluding race-associated predictors on model equity in future model development. • Re-train the model when statistically significant variations in predicted and expected cases occur, distributional shifts in data arise, or new clinical practices and guidelines are introduced. • Monitor the model's performance based on fairness metrics and stakeholder feedback during implementation.
Moggia et al. (2022) Germany	Introduce a data-informed approach to case formulation in psychotherapy using the TTN.	A case study focusing on a single patient is presented. The Trier Treatment Navigator (TTN) is used to guide case formulation and treatment decisions based on routinely collected outcome monitoring data.	The TTN facilitated a comprehensive and dynamic case formulation process, allowing for continuous adaptation of treatment strategies based on the patient's progress and feedback. The authors propose that	The study's reliance on a single case limits the generalisability of the findings.	• Use technology augmentation, such as the Trier Treatment Navigator (TTN), to improve and personalise clinical practice. • Employ routine outcome monitoring (ROM) in psychotherapy and provide feedback to therapists. • Use empirical data to tailor treatment to individual patients.

			data-informed case formulation and psychotherapy share a common goal of treatment personalisation and mutually benefit each other.		
Ngan et al. (2022) USA	Develop an HDSS to determine if a patient is at risk of MH problems	Case study. *n*=298 participants recruited from a southwestern community agency serving over 30,000 low-income immigrant individuals and families in West Texas and New Mexico. Bilingual English-Spanish survey containing 81 questions in ten categories	The HDSS achieved 91.11% accuracy and 92.30% sensitivity, indicating its effectiveness in identifying potential MH risks in the studied population. This approach outperformed the data analytics approach but was slightly less accurate than domain experts' criteria.	The study was limited to a specific population and geographical area.	• Design surveys specifically and professionally to address specific MH problems. • Develop a dynamic and interactive visual interface for domain experts to design and implement their own surveys. • Integrate information from time-point data, text, and images of patients to enhance predictive capabilities. • Enhance the model training mechanism to reduce computational time and space utilisation. • Investigate state-of-the-art deep learning models to improve system prediction performance. • Upgrade database customisability to enable domain experts to build and modify visual charts for better real-time decision support.
Qassim et al. (2023) Canada	Evaluate the perceived clinical utility and impact on the clinician–patient relationship of Aifred.	Naturalistic follow-up. 7 clinicians and 14 patients participating in a study assessing the acceptability and usability of Aifred, a novel AI-enabled CDSS designed for treating adults with major depression. Questionnaires and semi-structured interviews	86% of clinicians perceived a more comprehensive understanding of patients' situations, with 71% finding the information helpful. 62% of patients reported improved care, while 46%	Small sample size and pre-existing treatment changes limit the study's ability to verify the impact on treatment outcomes. Virtual study design due to COVID-19 limited opportunities for shared screen	• Incorporate more personalised elements into the tool, such as daily mood tracking or a free-form diary. • Incorporate tools from other methodologies, such as ecological momentary assessment. • Conduct future studies with larger sample sizes. • Use recordings instead of written notes in future research.

		conducted with clinicians and patients.	noticed an improvement in the clinician–patient relationship.	viewing between clinicians and patients.	• Conduct more in-person visits in future research. • Upgrade the AI model to consider individual failed treatment history and patient trajectories. • Conduct effectiveness trials of the clinical decision support system.
Yang et al. (2024) China	Develop a multimodal, multitask learning model to enhance psychiatric rehabilitation outcomes.	*n*=6,727 patients diagnosed with SMI at the Shanghai MH Centre.	Model achieved an accuracy exceeding 78.5% for all four tasks and performed exceptionally well on the tasks of "medication adherence" (94.3%) and "dangerous behaviour" (90.2%). AUC of all four tasks was 0.70, and except for "dangerous behaviour", the AUC exceeded 0.75 rates, indicating its robustness.	The fairness of AI decisions may be influenced by potential biases in the training data, and its robustness might be insufficient when dealing with novel or extreme cases Due to the inherent complexity of the ML models, their decision-making processes may lack full transparency, which could be a potential limitation in their application to clinical decision support.	• Integrate the system with mobile phones for improved data collection and patient engagement. • Simplify the system interface to make it more intuitive and accessible and decisions/recommendations transparent. • Regularly evaluate the model with new data. • Continuously monitor and optimise AI technologies when introduced into the field of MH rehabilitation.

5.1.4 Discussion: Advancing Artificial Intelligence-Based CDSS in Mental Health Care

The follow-up review reaffirms and extends the themes identified in the initial literature review (Chapter 3 and Higgins, Short et al. (2023)) on the potential of AI-based CDSS in MH care. The findings continue to highlight the promise of AI technologies in addressing missed care, alleviating clinicians' workloads, improving diagnostic accuracy, and addressing workforce challenges. However, the newly reviewed studies bring further insights and raise further complexities, particularly concerning the real-world implementation, ethical considerations, and the need for more rigorous evaluations of AI systems in diverse MH contexts.

5.1.4.1 Addressing Missed Care Through Artificial Intelligence

A growing body of literature underscores the capacity of AI technologies to alleviate clinicians' workloads and improve diagnostic accuracy in MH care. Studies such as those by Stein et al. (2022) and Bzdok and Meyer-Lindenberg (2018) have shown that AI-driven systems can assist clinicians in decision making by analysing vast amounts of patient data, thus enabling more accurate diagnoses and personalised treatment plans. This aligns with the findings reported in Chapter 3, which emphasise AI's role in addressing missed care and supporting overburdened clinicians. Baddal et al. (2024) further argue that AI-based tools can mitigate clinician burnout by automating routine tasks, allowing professionals to focus on more complex aspects of patient care. This suggests that the benefits of AI extend beyond clinical decision making to improving overall healthcare workflow, supporting the workforce challenges highlighted in the original review.

5.1.4.2 Building Clinician Trust and System Transparency

One of the key contributions of this review is the deeper exploration of clinician engagement and trust, which emerged as critical factors in the first review. The original literature review

underscored that the successful implementation of AI-based CDSS in health care hinges on clinician trust, system transparency, and interpretability. The current review builds on this by examining studies such as Qassim et al. (2023), which investigated the feasibility and acceptability of the Aifred system in clinical settings. This adds new empirical data on how clinicians interact with AI systems in practice, with a particular focus on patient–clinician interactions, appointment length, and the clinician's confidence in AI-driven recommendations. These findings confirm the initial review's assertion that system usability and clinicians' involvement in development are essential for successful adoption.

The issue of clinician trust and system transparency, as discussed in Chapter 3, has been further elaborated in recent studies. For instance, Shortliffe and Sepúlveda (2018) argue that the "black box" nature of many AI models contributes to clinicians' hesitancy in adopting these systems. They advocate the development of interpretable AI models that provide clinicians with clear reasoning behind the generated recommendations, a call echoed by Molnar (2020). Molnar's work on explainable AI (XAI) highlights that the integration of transparency-enhancing techniques, such as local interpretable model-agnostic explanations or SHapley Additive exPlanations (SHAP), could significantly improve clinician trust. This aligns with the empirical findings from Qassim et al. (2023), suggesting that clinician engagement in the development of AI systems is crucial to ensure their usability and acceptance in clinical settings.

5.1.4.3 Ethical and Fairness Considerations in Artificial Intelligence

Another notable advancement in this review is the incorporation of research addressing the ethical and fairness concerns surrounding AI in MH care. The earlier review identified these as key areas of concern, particularly regarding algorithmic bias and transparency. The study by Ngan et al. (2022), which focuses on healthcare decision support systems for migrants and

refugees, illustrates the importance of addressing equity issues in AI implementation. Ngan et al. (2022) provide insights into the performance of AI systems when applied to vulnerable populations, demonstrating that, while AI can offer substantial benefits, the risk of perpetuating bias remains a significant challenge. The inclusion of this study in the follow-up review underscores the necessity for ongoing monitoring and the development of robust debiasing strategies, as recommended in the previous literature review (Higgins, Short et al., 2023).

The ethical considerations associated with AI in MH care have garnered increasing attention in recent years. In particular, algorithmic bias, a key concern in the review in Chapter 3, has been explored in depth by scholars such as Obermeyer et al. (2019) and Gerke et al. (2020). Obermeyer et al. (2019) demonstrated that many AI systems inadvertently perpetuate healthcare inequalities, particularly in marginalised groups, due to biased training datasets. This aligns with the findings of Ngan et al. (2022), whose research on healthcare decision support systems for refugees and migrants highlights the urgent need for bias mitigation strategies in AI deployment. Gerke et al. (2020) further argue that the inclusion of diverse datasets in AI model development is crucial to prevent these systems from reinforcing existing healthcare disparities. These authors collectively reinforce the call for continuous monitoring and the development of robust debiasing frameworks, as discussed in the follow-up review.

5.1.4.4 Advancements in Artificial Intelligence Modalities and Model Complexity

The current review also brings to the forefront the growing complexity of AI models being applied in MH care. The research by Yang et al. (2024), which evaluated a novel multimodal, multitask learning model for predicting rehabilitation outcomes in severe mental illness, marks a progression in the sophistication of AI models since the first review. While the original review called for the development of interpretable ML models tailored specifically to MH applications, this new study illustrates the potential of more complex models to integrate diverse data

sources, such as clinical notes and structured diagnostic information. Although promising, these advancements also highlight the ongoing challenge of balancing model complexity with interpretability, reinforcing the need for clinician-friendly AI systems.

Furthermore, the advancements in AI complexity, such as the multimodal and multitask learning model explored by Yang et al. (2024), are indicative of broader trends in AI research. Similarly, Miotto et al. (2018) highlight the increasing use of deep learning algorithms in health care to process heterogeneous data sources, such as electronic health records, clinical notes, and imaging data. These models are not only more sophisticated but also offer the potential to uncover new insights into patient care. However, as Miotto et al. (2018) and Kamel Rahimi et al. (2024) both point out, the challenge remains in balancing model complexity with interpretability. Kamel Rahimi et al. (2024) specifically emphasise the need for clinician-friendly interfaces that allow healthcare professionals to interact with AI systems without needing extensive technical knowledge. This reinforces the findings in Chapter 3 and Yang et al. (2024) regarding the ongoing challenge of integrating complex AI models into clinical workflows while maintaining transparency and ease of use.

5.1.4.4 Artificial Intelligence Support in Mental Health Case Formulation

Furthermore, the role of AI in supporting MH case formulation, as explored in Moggia et al. (2022), complements the findings from the earlier review on AI's ability to optimise clinical workflows. The TTN provides an example of AI-driven decision support being used in psychotherapy, although without the rigorous intervention or comparative analysis frameworks seen in other studies (Moggia et al., 2022). This aligns with the call for the development of interpretable AI systems in Chapter 3 and demonstrates the continuing evolution of AI applications in diverse MH contexts, including psychotherapy.

The role of AI in psychotherapy and MH case formulation, as examined Moggia et al. (2022), is also a growing area of interest. Studies by Thieme et al. (2023) and Shatte et al. (2019) have highlighted the potential of AI-driven systems to assist in therapeutic decision making and personalising treatment interventions. Thieme et al. (2023) argue that AI can be used to analyse session transcripts, identifying patterns and suggesting therapeutic approaches based on the content of conversations. This complements Moggia et al.'s (2022) findings on the TTN system and underscores the continued evolution of AI applications in psychotherapy. However, Shatte et al. (2019) caution that the efficacy of AI in psychotherapy remains contingent on rigorous clinical validation, a point also raised in the original literature review. Shatte et al. (2019) argue that, without robust intervention frameworks, the use of AI in therapeutic settings could lead to an over-reliance on technology and a potential loss of the human touch in psychotherapy, a critical element in patient outcomes.

5.1.5 Conclusion: Integrating Artificial Intelligence with Workforce Solutions for Sustainable Care

In conclusion, this review not only confirms the findings from the first review but also adds new layers of complexity and understanding to the discourse on AI-based CDSS in MH care. However, despite the advances in AI technology, systemic workforce challenges, such as staffing shortages and clinician burnout, remain fundamental drivers of missed care, as noted in Chapter 3. The review reiterates that AI should be seen as a complement to, rather than a replacement for, human care. As previously highlighted, without addressing these workforce challenges, AI alone cannot resolve the root causes of missed care. The research continues to show significant potential for AI technologies to improve care delivery, reduce clinicians' burden, and enhance decision-making processes. Nevertheless, it also highlights that the successful deployment of these systems is contingent upon ongoing efforts to address issues of

trust, transparency, bias, and workforce integration. Future research must focus on conducting large-scale studies to ensure these systems can be effectively and equitably implemented in diverse real-world settings. This balanced approach of integrating AI-driven solutions with systemic workforce improvements will be essential for achieving sustainable and meaningful improvements in MH care.

5.2 The Rise of Large Language Models (LLMs) and Their Impact on Health Care

5.2.1 Introduction

Significant technological advancements have taken place since the beginning of this PhD research, particularly in the development of large language models (LLMs). Over this period, several influential models have been introduced, pushing the boundaries of natural language processing (NLP). Among the most notable is ChatGPT, which garnered substantial public attention in January 2023 (Wu et al., 2023). LLMs are a type of AI model designed to understand and generate human language. Built using deep learning techniques, particularly the transformer architecture, these models process and generate text that is coherent and contextually relevant, characterised by their large number of parameters—often in the billions (T. Brown et al., 2020). However, due to their generative capabilities, LLMs can sometimes produce information that appears plausible but is incorrect or misleading, raising concerns about their reliability (Crawford, 2021). In the mental health (MH) domain, while NLP-based chatbots have been deployed, LLMs have not yet been fully utilised due to the potential risks they pose to end users, as discussed in the publication in Chapter 2. Current mental health chatbots, such as Woebot (Darcy, 2023), are typically rule-based systems that rely on pre-defined responses and decision trees (Stade et al., 2024) rather than generative AI. Clinicians have expressed hesitation in adopting AI systems like LLMs due to the variability in human

interaction and concerns about the transparency of these models (Feldman et al., 2019; Lysen & Wyatt, 2024; Meskó, 2023).

5.2.2 Recent Development of LLMs

Several key LLMs have been introduced in recent years, each contributing to the field's rapid evolution. Notable among these are OpenAI's GPT-3, GPT-4, Google's Bidirectional Encoder Representations from Transformers (BERT) and T5X. GPT-3, developed by OpenAI, is renowned for its 175 billion parameters, surpassing its predecessors and enabling it to perform a wide range of language tasks with minimal fine-tuning (T. Brown et al., 2020). Building on this, GPT-4 has further improved performance and adaptability, showcasing the potential of scaling up model parameters (Kaplan et al., 2020). In parallel, Google's BERT and its successor models have introduced innovative approaches to pre-training language models. BERT's ability to consider context from both directions in a sentence has revolutionised NLP tasks such as question answering and sentiment analysis (Devlin et al., 2018). Additionally, T5X (Text-To-Text Transfer Transformer) has unified various NLP tasks into a single framework, demonstrating the versatility and robustness of transformer-based models (Raffel et al., 2020; Roberts et al., 2023).

Several technological advancements have contributed to the remarkable progress in LLM development. The transformer architecture, introduced by Vaswani et al. (2017), has been a cornerstone of this progress. Transformers enable parallel data processing, making it feasible to train models on large datasets more efficiently than previous sequential approaches. Increased computational power and the availability of large-scale datasets have also been critical. Advances in hardware, such as graphics processing units and tensor processing units, have drastically reduced the time required to train LLMs (Huang et al., 2024). Furthermore, the collection and curation of extensive text corpora, including diverse sources of information,

have provided the rich datasets necessary for training these models effectively (Kaplan et al., 2020).

Several key milestones and breakthroughs have marked the recent rapid development of LLMs. One significant milestone was the demonstration of GPT-3's *few-shot* learning capability, which allows the model to perform tasks with limited examples, showcasing its generalisation ability across various domains (T. Brown et al., 2020). This advancement has highlighted the potential of LLMs to adapt to new tasks with minimal retraining, a crucial feature for practical applications. Another breakthrough has been the refinement of transfer learning techniques, as exemplified by T5X high-performance training, evaluation, and inference of sequence models (Roberts et al., 2023). By framing all NLP tasks as text-to-text problems, these models have simplified the training and application of LLMs across different tasks, thereby improving efficiency and performance (Raffel et al., 2020; Roberts et al., 2023). Additionally, the continuous improvements in model interpretability and fine-tuning methods have enhanced the usability and reliability of LLMs in real-world applications (Devlin et al., 2018).

5.2.3 Impact of LLMs on Health Care

The impact of generative AI, such as LLMs, will potentially transform the way health care is delivered; however, professional clinical groups have already raised significant concerns about these developments (L. A. Brown et al., 2020; Feldman et al., 2019; Lysen & Wyatt, 2024; Meskó, 2023). Significantly, there is a need for high-quality data, constant monitoring of the system and its outputs, and transparency in the decision-making loop. While recent research has called for clinicians always to have the final say in the care process, as seen in Chapter 3, Stade et al. (2024) propose the following steps in developing and evaluating LLMs in behavioural health care to provide further context. These stages span from assistive to fully autonomous. Assistive LLMs are an AI tool that will assist users with tasks that can easily be

offloaded to an AI and generally have a low level of risk associated with the task. Collaborative LLMs will allow the AI to lead the intervention while keeping the human in the loop. This level of integration requires humans to use their professional judgement and select or tailor the options or care pathways that the AI provides (Chandler et al., 2022; Stade et al., 2024). At the fully autonomous stage, LLMs and AIs perform in an integrated manner without human or clinical oversight. Fully autonomous systems will require significant levels of trust; however, they offer the most scalable integration and potential service benefits (Fairburn & Patel, 2017). Nevertheless, the integration of any AI-based system innovation will require substantial levels of trust for successful adoption into clinical practice (Higgins, Short et al., 2023; Stade et al., 2024).

5.2.4 Ethical and Practical Challenges in Health Care

While there is a significant potential for LLMs to transform the way health care is delivered, it must be noted that there are several ethical and practical challenges that must be considered when integrating LLMs into existing healthcare systems and workflows. For instance, due to their generative nature, LLMs can generate information that appears plausible but is incorrect or misleading, raising concerns about their reliability (Crawford, 2021). LLMs can inherit biases present in their training data, leading to unfair or prejudiced outcomes in their responses (Bender et al., 2021). The use of large datasets that include personal information raises privacy concerns, necessitating careful handling and regulation (Price & Cohen, 2019) as well as a distinct bias towards the English language (Malgaroli et al., 2023). A considerable number of studies using NLP for MH interventions rely heavily on English transcripts, creating a significant linguistic barrier (Malgaroli et al., 2023). For example, 87.3% of studies included in a systematic review used only English conversational data (Malgaroli et al., 2023). This over-reliance on English raises concerns about the generalisability of findings to other

languages. Furthermore, the lack of open data and code in many studies makes it difficult to reproduce and validate findings (Chandler et al., 2022; Malgaroli et al., 2023). This lack of transparency also limits the ability to compare results across different studies that use varying ground truth labels and evaluation metrics (Malgaroli et al., 2023).

Whilst addressing these issues requires overcoming a set of challenges, the following are some areas that may accelerate this process. Primarily, humans must remain involved through all parts of the development cycle. By involving *humans-in-the-loop* in more stages of the model development process, researchers and clinicians can gain a better understanding of the data and how the models are making predictions (Chandler et al., 2022). For instance, active learning can be used to select the most important data for human labelling, leading to more efficient and generalisable models (Chandler et al., 2022). Human-in-the-loop methodologies can increase transparency and control by involving clinicians in the decision-making process. This approach ensures that AI augments rather than replaces human judgement (Chandler et al., 2022). *Diversity sampling*, a technique employed in active learning, can help identify and address discrepancies in how models represent diverse demographic groups (Chandler et al., 2022). Researchers and clinicians should prioritise collecting and labelling more diverse data to improve model robustness and generalisability. This includes data from different demographic groups, clinical settings, and stages of illness (Chandler et al., 2022). Finally, the establishment of *data enclaves* offers a secure way to provide researchers with access to sensitive patient data, addressing privacy concerns while enabling research (Malgaroli et al., 2023). This approach allows data providers to retain control over data access and transmission (Malgaroli et al., 2023).

5.2.5 Conclusion

The rise of large language models (LLMs) since 2020 has been marked by substantial technological advancements and the introduction of highly influential models such as GPT-3, GPT-4, BERT, and T5X. These models, characterised by their extensive parameters and advanced transformer architecture, have significantly expanded natural language processing capabilities. LLMs have demonstrated exceptional proficiency in tasks ranging from language generation to contextual understanding, and their potential applications span diverse domains, including health care. In health care, the potential for LLMs to revolutionise service delivery is immense, offering opportunities for enhanced patient care, clinical workflow efficiency, and innovative MH support solutions. However, the integration of LLMs into clinical settings is not without challenges. Ethical considerations, including the risk of generating misleading information, the potential for inherent biases, and privacy concerns, necessitate careful and responsible deployment. The resistance from the clinical community, driven by concerns over the unpredictable nature of LLMs and a lack of transparency, underscores the need for trust-building measures and clear guidelines. The responsible development and evaluation of LLMs in health care should progress through stages from assistive to fully autonomous systems, with human oversight remaining a critical component. This approach will ensure that AI tools complement rather than replace clinical judgement, fostering a collaborative environment where technology enhances human expertise. The trajectory of LLMs in health care will depend on addressing these ethical and practical challenges, establishing robust regulatory frameworks, and building trust among clinicians and patients. As advancements continue, the transformative potential of LLMs holds promise for a future where AI-driven tools significantly contribute to improved health outcomes and patient care. Finally, while AI and LLMs hold significant potential for improving health care, several challenges must be addressed before widespread adoption. By focusing on transparency, data diversity, rigorous

evaluation, and human–AI collaboration, researchers and clinicians can work towards developing trustworthy and effective AI-powered tools for MH care.

5.3 Chapter Summary

This chapter revisited the original literature review, updated it to include recent advancements in AI, and discussed the emergence of LLMs such as GPT-3 and GPT-4. These technologies have made significant strides in health care, offering the potential to complement clinicians' decision making, enhance patient care, and improve workflow efficiency. The chapter explored how AI-driven CDSS, including LLMs, can augment clinical evaluations by providing real-time insights, thereby addressing issue such as missed care—one of the key challenges in healthcare services.

The importance of trust in the integration of AI into clinical practice emerged as a significant theme. The literature consistently highlights that trust, transparency, and accountability are critical to the successful adoption of AI systems. LLMs, in particular, present challenges due to the opaque nature of their decision-making processes, which can complicate clinicians' confidence in their recommendations. As the chapter reaffirmed, AI technologies should be viewed as tools that complement, rather than replace, human judgement, particularly in the nuanced field of MH care where empathy and personalised attention are essential. While AI offers significant potential for alleviating clinicians' workload and improving care delivery, it cannot address these core issues independently. Therefore, it is essential that the integration of AI systems be accompanied by efforts to improve workforce conditions to ensure high-quality and safe care. Without addressing these structural issues, the full potential of AI in transforming patient care will not be realised.

This chapter has directly addressed the study's aims in demonstrating how AI technologies can complement and augment clinicians' decision making, improving care accuracy and timeliness

while reducing missed care. It also examined how these technologies can enhance the client's experience of MH services, allowing clinicians more time to deliver personalised care. Lastly, the chapter provided further insights into the adoption of AI in MH care, emphasising the need to build trust and ensure transparency to foster successful implementation.

The following chapter will present the methods used in this research, outlining the research design, data collection strategies, and analysis approaches that were used to examine the integration of AI in MH care. This will provide a structured framework for the subsequent discussion of the development of a ML model that addresses the overarching research question: *How can machine learning models be developed, tested, and evaluated to predict whether clients presenting to an emergency department with a mental health concern, particularly suicidality, will require admission to an acute mental health inpatient ward?*

Chapter 6 Methods

The research question under review for this thesis is: *How can machine learning models be developed, tested, and evaluated to predict whether clients presenting to an emergency department (ED) with a mental health concern, particularly suicidality, will require admission to an acute mental health inpatient ward?* This complex question necessitated a robust methodological approach to ensure comprehensive and reliable outcomes.

In this study, the researcher employed the overarching design thinking (DT) framework methodology (Brown, 2008, 2009), which was instrumental in addressing the intricate and human-centred aspects of predicting MH outcomes. DT was particularly relevant here as it prioritises empathy, ideation, and iterative testing, ensuring that the solutions researched are attuned to the needs of the patients and healthcare providers involved. This framework encompasses several stages—empathise, define, ideate, prototype, and test—each of which aligns with the goal of developing a ML model that is both accurate and practical for real-world application in an ED setting (Brown, 2008, 2009). Additionally, the retrospective cohort method was utilised to gather and analyse historical data of patients who have previously presented to the ED with a MH issue as their primary presentation problem (Liamputtong, 2021, p. 212). This method leveraged existing data, providing a rich source of information that enhanced the predictive power of the models and enabled a better understanding of the needs of the community.

Finally, the supervised ML method was central to this research. By training the models from the data used in the retrospective cohort study, the researcher was able to identify and learn the patterns associated with different presentations (Molnar, 2020). Supervised learning algorithms were used to explore the data and determine the most effective approach for this specific prediction task. The performance of these models was rigorously tested and evaluated to ensure

their reliability and accuracy (Molnar, 2020). Together, these methods formed a cohesive strategy to address the challenge of predicting ED admissions for patients with MH concerns, with the ultimate goal of improving patient outcomes and resource allocation in healthcare settings.

6.1 Introduction and General Overview of the Research

Theoretical frameworks play a pivotal role in advancing our understanding of complex phenomena across various disciplines. By providing conceptual lenses and guiding principles, these frameworks facilitate the systematic exploration of intricate relationships, human behaviours, and problem-solving approaches.

The integration of artificial intelligence (AI) and machine learning (ML) in MH care presents both opportunities and challenges, particularly concerning the establishment of trust and empathy among clinical end users, as discussed in Chapters 2 and 3. This chapter discusses the methods utilised to explore the development of AI and ML in MH (MH) care. It outlines the processes through which the research was conducted and describes the analytical techniques employed to evaluate the efficacy and consider the ethical implications of the proposed solutions. This research aims to integrate technological innovation with humane and equitable MH care, focusing on trust and empathy within its theoretical framework. In this context, trust goes beyond technical reliability and validity, encompassing the interpersonal dynamics between clinicians, patients, and the technological systems they use (Botsman, 2018). Similarly, empathy underpins the development of person-centred care strategies, ensuring that the humanistic aspects of MH treatment are preserved amidst technological advancements, as seen in Chapter 4. MH services have a significant role in health care, and the aforementioned elements will be paramount to ensuring the successful adoption and effective utilisation of AI- and ML-based decision support systems in MH care.

This study employed design thinking (DT) (Brown, 2008, 2009) as the overarching methodology to address these complex requirements. Due to its iterative, user-centred approach, DT is particularly suited for navigating the intricacies of AI and ML integration in MH settings. This methodology enables a deep understanding of the needs and concerns of various stakeholders, including clinicians, patients, and subject matter experts. By prioritising empathy and collaborative problem solving, DT enables the co-creation of solutions that are not only technically robust but also aligned with the values and expectations of end users (Brown, 2008, 2009). DT offers a structured, user-centric approach for developing solutions that are not only effective but also meet the needs and address the concerns of the people who will be using them (Brown, 2008, 2009). In the context of AI development in this research, DT was particularly relevant due to its emphasis on empathy and user understanding (Brown, 2008, 2009). This aligns with the core principles of DT, empathising with users. DT starts with deeply understanding the needs, challenges, and perspectives of the end users (Brown, 2008, 2009), and it then moves to defining the problem: clearly articulating the problem that the AI system aims to solve from nurses' perspective is crucial (Brown, 2008, 2009). The next stage is ideating solutions: DT encourages generating a wide range of potential solutions, even unconventional ones (Brown, 2008, 2009). After that is prototyping and testing: creating prototypes of the AI system and testing them with nurses in real-world settings is essential for gathering feedback, identifying usability issues, and iteratively improving the design (Brown, 2008, 2009). By incorporating DT principles, this research aims to develop systems that are not only technically sound but also address the practical needs and concerns of clinicians, leading to greater acceptance and trust in AI-driven healthcare solutions.

6.1.1 Theoretical Frameworks Considered for this Study

At the outset of this research, a thorough examination of multiple theoretical frameworks was conducted to identify the model best suited to guide the study's aims and approach. This process involved evaluating each framework's strengths and limitations, as well as its alignment with the study's objectives, research questions, and the unique complexities of the subject matter. The biopsychosocial model, introduced by George L. Engel (1977), challenges the traditional biomedical model's reductionist approach to health and illness. It posits that biological, psychological, and social factors contribute to an individual's overall well-being and health outcomes. This model acknowledges the interplay between these three dimensions and emphasises the importance of considering the patient's subjective experience, social context, and psychological state in addition to biological factors (Engel, 1980). While the biopsychosocial model offers a more comprehensive understanding of health and illness, it has been criticised for its lack of clear operational definitions and measurement strategies (Ghaemi, 2009) and was therefore deemed unsuitable.

Systems theory, with its roots in the work of Ludwig von Bertalanffy (1968), is a comprehensive approach that views phenomena as complex, interconnected systems rather than isolated components. This theory emphasises the importance of understanding the relationships and interactions between different parts of a system and the dynamics between the system and its environment (Skyttner, 2005). Systems theory has been applied in various disciplines, including biology, psychology, and organisational studies. However, it has been criticised for its abstract nature and its difficulty in empirically operationalising and testing its principles (Badcock, 2012) and was therefore deemed unsuitable.

Design thinking (DT) is an iterative problem-solving approach that emphasises user-centred design, empathy, and collaboration (Brown, 2008). This framework encourages a deep

understanding of the end user's needs, followed by the generation of creative solutions through ideation and prototyping (Liedtka, 2015). DT has been widely adopted in various industries, including product development, service design, and organisational innovation. While it has been criticised for its potential to oversimplify complex problems (Jen, 2017), DT was deemed suitable to best understand the phenomenon under investigation due to its focus on empathy, a key finding from the literature review. Empathy must remain central to this investigation, as it is essential for the successful design and implementation of AI in nursing. Empathy helps bridge the gap between AI's capabilities and the human need for understanding and connection in healthcare settings (Topol, 2019b). AI, by its nature, lacks the capacity for genuine empathy. However, by incorporating design principles that consider the emotional needs of nurses and patients, developers and researchers can create AI systems that feel more trustworthy and are more readily adopted into clinical practice.

At its core, DT is an iterative process that involves understanding users, challenging assumptions, redefining problems, and creating innovative solutions to prototype and test (Brown, 2008). This human-centred approach aligns well with research methodologies focusing on user experiences, needs, and behaviours, particularly in healthcare, education, and technology development (Liedtka, 2015). The iterative nature of DT, which includes empathising, defining, ideating, prototyping, and testing, can be incorporated into research designs that require continuous refinement and iteration (Lewrick et al., 2018). This is particularly useful in action research and participatory design, where ongoing feedback and adaptation are crucial (Wicks et al., 2008). DT's structured yet flexible approach to problem solving provides a robust framework for addressing complex, real-world problems, making it valuable in applied research aimed at finding practical solutions, such as in organisational studies, product development, and policy design (Lewrick et al., 2018). Furthermore, DT encourages collaboration and draws on diverse perspectives, making it suitable for

interdisciplinary research. Insights from multiple fields can be integrated to address a research question comprehensively (Johansson-Sköldberg et al., 2013). By fostering creativity and out-of-the-box thinking, DT allows researchers to explore new areas, generate novel hypotheses, and develop innovative methodologies (Lewrick et al., 2018).

In healthcare research, particularly in nursing and MH, DT can facilitate the creation of patient-centred care models, developing new interventions, and improving healthcare delivery systems. Researchers can use empathy to understand patients' experiences, ideate potential improvements, create prototypes of new care models, and test their effectiveness (Bazzano et al., 2017). In the field of technology and AI, DT can guide the development of user-friendly interfaces, ethical AI systems, and inclusive technology solutions. This approach helps researchers focus on how users interact with technology and ensures that solutions meet users' needs (Kolko, 2015). DT offers a robust framework for conducting iterative, user-centred, and solution-oriented research. It enables researchers to approach problems innovatively, incorporate diverse perspectives, and develop practical solutions that can be tested and refined. By embedding DT into the research process, scholars can enhance the relevance and impact of their work across various disciplines.

6.1.2 Design Thinking Methodology Explained

DT, an innovative and human-centred approach to problem solving, was employed as the primary methodology in this PhD research. Originating from design and engineering disciplines, DT is increasingly applied across various fields, including health care, due to its iterative process and focus on empathy and user-centric solutions (Brown, 2008, 2009). This methodology is particularly relevant to addressing complex challenges in MH care within EDs, where traditional approaches often fall short in meeting the nuanced needs of diverse populations. The DT process typically involves five phases: empathise, define, ideate,

prototype, and test (Lewrick et al., 2018). Each phase contributes to a comprehensive understanding of the problem and the development of innovative solutions.

The empathise phase of this project focused on understanding the experiences and needs of individuals seeking MH care in EDs. This deepened understanding of the lived experiences of individuals, particularly vulnerable populations like Aboriginal and Torres Strait Islander communities, provided a foundation for developing culturally responsive interventions.

The define phase involved examining the data collected during the empathise phase, which was synthesised to identify key themes and challenges faced by individuals in ED settings. This involved creating detailed problem statements and defining specific research questions. For instance, understanding the high prevalence of suicidal ideation among ED presentations led to recommendations for focused investigations on intervention strategies and safety planning.

In the ideate phase, a range of potential solutions was generated through the investigation and analysis of the data and collaboration with subject matter experts. This provided an opportunity for the researcher to explore innovative ideas for understanding MH ED access and integrating ML tools in clinical decision making.

Selected ideas from the ideation phase were developed into tangible prototypes. This included designing ML models to predict admissions to MH inpatient facilities and creating frameworks for culturally responsive care models. Prototyping allowed for iterative refinement and feedback from stakeholders, ensuring that solutions were practical and aligned with users' needs (Lewrick et al., 2018).

The prototypes were tested with data from real-world settings. This phase involved evaluating the effectiveness and feasibility of proposed solutions through rigorous prototype testing. Feedback from this testing phase was crucial in refining the final interventions and ensuring

their sustainability and scalability (Lewrick et al., 2018). Figure 6.1 provides a visual representation of the DT process.

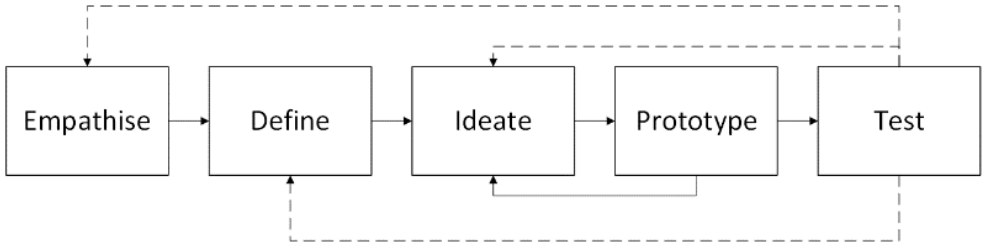


Figure 6.1. *Design thinking flowchart*

6.1.3 Ethical Considerations

Ethical considerations were central to the DT process. The research followed ethical guidelines for studies involving human participants, ensuring informed consent, confidentiality, and cultural sensitivity, especially when working with First Nations communities (National Health and Medical Research Council, 2018b). The iterative nature of DT also enabled ongoing ethical reflection and adjustments based on stakeholder feedback.

While DT provided the overarching framework, it was complemented by other methodologies to ensure a robust and comprehensive research design. A retrospective cohort method was used to analyse sociodemographic and presentation features of individuals seeking MH care in EDs. Additionally, ML techniques were employed to develop predictive models, integrating the insights gained from the DT phases into algorithmic solutions which will be discussed in more detail in Chapter 7.

DT, emphasising empathy, iteration, and user-centred design, served as an effective methodology for addressing the complex challenges in MH care delivery in EDs. This research developed innovative and practical solutions to improve MH outcomes by integrating qualitative insights with quantitative analysis and ML techniques. The iterative and

collaborative nature of DT ensured that the interventions were effective, culturally responsive, and aligned with the needs of diverse populations.

6.1.4 Retrospective Cohort Study

This research used a retrospective cohort study method to examine a large amount of data before modelling occurred. Retrospective cohort studies rely on pre-existing data, often sourced from medical records, databases, or registries, to identify cohorts and gather information on exposure and outcomes (Liamputtong, 2021, p. 212). These studies examine historical data, beginning from a past point when the exposure occurred and following up until the present or a specified endpoint (Liamputtong, 2021, p. 212). There are significant advantages of utilising a retrospective cohort method. For example, due to the utilisation of existing data, retrospective cohort studies can be conducted relatively quickly and with fewer resources compared to prospective studies (Liamputtong, 2021, p. 212). This design is particularly useful when immediate or prospective data collection is not feasible (Liamputtong, 2021, p. 212). However, there are also some disadvantages in regard to data quality and completeness as the reliability of findings hinges on the quality and completeness of historical data, which may not always be accurate or comprehensive (Liamputtong, 2021, p. 212). Additionally, retrospective cohort studies may face challenges with confounding variables that were not controlled for during data collection, which could potentially skew the results (Liamputtong, 2021, p. 212). In summary, a retrospective cohort study is an invaluable tool in medical and epidemiological research for examining associations between exposures and outcomes using pre-existing data.

6.1.5 Supervised Machine Learning Method

Supervised ML is a prominent research method within the broader field of AI. It involves the development and application of algorithms that learn from labelled training data to make

predictions or decisions without being explicitly programmed to perform the task. This method is particularly valuable in diverse academic and practical contexts due to its capacity to model complex relationships and make accurate predictions based on input data (Goodfellow et al., 2016).

Supervised ML relies on datasets that contain input–output pairs. Each input is associated with a corresponding output, which acts as the ground truth. This labelled data is crucial for training the model (Paper, 2020). During the training phase, the model learns to map inputs to outputs. The algorithm iteratively adjusts its internal parameters to minimise the error between its predictions and the actual outputs (Paper, 2020). After training, the model is evaluated on a separate dataset, known as the test set, to assess its performance. This helps to ensure that the model generalises well to unseen data and does not simply memorise the training data (Paper, 2020)

Supervised learning can utilise various algorithms, such as linear regression, logistic regression, support vector machines, decision trees, random forests, and neural networks. The choice of algorithm depends on the problem and the data characteristics (Goodfellow et al., 2016). Supervised machine learning is applied across many fields, including health care, finance, marketing, and natural language processing (NLP). In health care, it is used to predict patient outcomes, diagnose diseases from medical images, and personalise treatment plans (Esteva et al., 2019). In NLP, it supports tasks like sentiment analysis, language translation, and spam detection (Young et al., 2018).

When employing supervised ML as a research method, several steps are typically involved. Initially, the research question or problem to be addressed must be clearly defined. The objective of the model, such as classification, regression, or another prediction task, should be determined (Murphy, 2023). Subsequently, a large and representative dataset relevant to the

problem must be gathered. The data should be pre-processed by cleaning, normalising, and possibly augmenting it. It is also important to split the data into training, validation, and test sets (Bishop, 2006). Next, an appropriate ML algorithm should be chosen and its parameters configured. The model is then trained using the training dataset, employing techniques such as cross-validation to optimise hyperparameters and avoid overfitting (Goodfellow et al., 2016).

Evaluating the model's performance using the test dataset is crucial. Common metrics include accuracy, precision, recall, F1 score, and area under the receiver operating characteristic (ROC) curve, depending on the type of problem (Paper, 2020). The results must be interpreted in the context of the original research question, and the findings validated by comparing with baseline models or through additional experiments (Paper, 2020).

Supervised ML comes with several challenges and important considerations. High-quality, representative data is crucial for training effective models, as insufficient or biased data can result in poor generalisation and biased predictions (Paper, 2020). One common issue is overfitting, where a model performs well on training data but fails on unseen data. Another key challenge is interpretability, especially with complex models like deep neural networks. Ensuring models are interpretable is essential for gaining insights and building trust in their predictions.

In summary, supervised ML is a powerful tool that leverages labelled data to train models for predictive tasks across many domains, offering valuable insights and decision-making support. However, ensuring its effective and ethical use requires attention to data quality, model selection, evaluation, and transparency.

6.1.6 Research Setting and Context

Research was conducted in one coastal regional local health service in NSW, Australia, that contains two public hospitals containing EDs servicing the community of 346,596 residents over an area of 1,681 km^2 (Australian Bureau of Statistics, 2021a). The *Mental Health Act 2007* (NSW) allows involuntary detention and treatment for individuals found to be mentally ill or disordered and requires the individual to be taken to a declared MH facility, which both ED sites are, for assessment (NSW Health, 2019). Acute inpatient services and community-based care are offered as part of the public health comprehensive MH service delivery program, with services provided across three streams: child and youth mental health (0–12 years and 12–17 years), adult (18–65 years) and older persons (65+years). In a departure from other NSW Eds, specialised MH staff provide assessments at the EDs around the clock, seven days a week. The two sites on the Central Coast total 84 acute inpatient MH beds. The Gosford site is comprised of 24 acute and 6 high-dependency beds (30 in total). The Wyong site is comprised of 25 acute, 10 high-dependency, and 15 older persons (65+) beds as well as a Psychiatric Emergency Care Centre with four beds adjacent to the ED (54 in total). Neighbouring health districts provide access to sub-acute, non-acute, and rehabilitation beds for clients requiring extended or specialty MH care, as these bed types are not available in the local area. Adolescent (ages 12 to 17) MH beds are unavailable on the Central Coast; instead, referrals are made to adolescent inpatient services in neighbouring local health districts' specialised inpatient units (Central Coast Local Health District, 2017).

6.1.6.1 Central Coast Demographics

The Central Coast of New South Wales is located close to Sydney, Australia (Figure 6.2). According to the Australian Bureau of Statistics (2021a), the gender split in the Central Coast is 48.5% male and 51.5% female. Additionally, 4.9% of the population within the Central Coast

identify as “Aboriginal and/or Torres Strait Islander peoples”, compared to 3.4% in NSW and 3.2% in Australia (Australian Bureau of Statistics, 2021a). The Central Coast of NSW has an ageing population, evidenced by the increasing number of individuals aged 55 and above in recent years. According to the Australian Bureau of Statistics (2021a), this trend is expected to continue in the coming decades. The Central Coast experiences poorer health outcomes than both NSW and Australia, with a higher prevalence of MH conditions such as anxiety and depression, at 11.0%, compared to 8.0% in NSW and 8.8% in Australia (ABS, 2021c) and higher than average increases in high psychological distress over 2015–2021 (NSW Health, 2023).

There is a need for public health services to be supported with evidence that describes the dynamics related to MH presentations in ED so that their planning and service delivery can best respond effectively and efficiently to the service needs of the catchment population within the budget and resources available.

Figure 6.2. *Map of Central Coast Local Health District*

6.2 Research Design, Methods and Ethics Overview

This section describes the research ethics protocols for this research. The study was conducted in compliance with all stipulations of this protocol, the conditions of ethics committee approval,

the National Statement on Ethical Conduct in Human Research (NHMRC, 2018b) and the Note for Guidance on Good Clinical Practice (CPMP/ICH-135/95).

The purpose of this research was to leverage ML to assist decision making in the assessment of people presenting at the ED with MH issues as their primary presenting problems. The primary objectives were twofold: to understand the factors contributing to ED presentations and admissions for individuals with MH issues, and to build and compare ML models using these factors. The study used a retrospective cross-sectional and descriptive epidemiological design, relying solely on existing data.

A data extract encompassing health records from the Central Coast Local Health District was utilised, focusing on ED presentations for MH issues between 1 January 2016 and 30 December 2021. Statistical procedures included univariate descriptive analysis and multivariate analysis of relationships within the data. ML models were evaluated based on various metrics such as sensitivity, specificity, positive predictive value, and negative predictive value (NPV).

6.2.1 Research Objectives

This research was divided into three parts. The first part, comprising Studies A, B, and C, focused on analysing data to identify the factors contributing to ED presentations among individuals with mental health conditions. The second part, Study D, built upon the findings of Studies A, B, and C to develop and evaluate ML models aimed at predicting whether a person requires admission to a mental health facility. Study E, discussed in Section 6.5, is a planned postdoctoral study that will explore attitudes and workforce issues related to the implementation and use of AI in clinical practice.

The system inputs and features were identified in Study A. Based on the literature review's findings, system outputs, including treatment recommendations with confidence intervals and

a description of how each variable contributed to the outcome, were used. The data used for the studies and modelling came from NSW Health's Health Information Exchange for the Central Coast Local Health District ED presentations. Study B examined and assessed the different ML models produced.

6.2.2.1 Study A, B and C: Understanding the Mental Health Population that Accesses the Emergency Department

The main entry point into MH services for most consumers is the ED. Part A of this study (Studies A, B, and C) focused on documenting the cohort of service users that access the Gosford and Wyong ED with MH as the primary presenting problem between 1 January 2016 and 30 December 2021 to determine the factors that contribute to patients being admitted to MH facilities via the ED, using a retrospective cross-sectional descriptive epidemiological method (Liamputtong, 2021, p. 212).

The study aimed to examine the following fields from within this dataset. Please see Table 6.1 and 6.2 for a complete list of data extracted.

- Time of day
- Day of week
- Month of year
- Mode of arrival (police, ambulance, etc.)
- Type of visit
- Triage category
- Presenting problem (suicidality, disturbed mental state, etc.)
- Gender
- Aboriginal and Torres Strait Islander origin status
- Age
- Living arrangements/marital status

- Previous encounter with the Central Coast Local Health District (CCLHD) MH service
- Hospital facility
- Referred to on departure.

In addition, the research stratified and compared the following three groups (at the time of presentation):

- those unknown to the MH service
- those with a previous history of the MH service but not currently receiving services
- those that are currently linked to an arm of the service.

It also compared those presenting with a suicidal or situational crisis with those presenting with an exacerbation of an underlying mental illness.

Table 6.1. ED visit table

Data column	Data type	Comment
actual_departure_date	datetime	
actual_departure_time	datetime	
age	int	
ambulance_client_number	char	Data not exported, converted to Yes/No
area_of_usual_residence	char	Australian Standard Geographic Code 2011
arrival_date	datetime	
arrival_time	datetime	
birth_date	datetime	
country_of_birth	char	
country_of_usual_residence	char	
dva_card_type	char	
ed_source_of_referral	char	
ed_visit_type	char	
facility_identifier	char	Site Gosford or Wyong
hospital_insurance	char	
indigenous_status	char	
marital_status	char	

mode_of_arrival	char	
mode_of_separation	char	
need_interpreter_service	char	
patient_postcode	char	
mrn	varchar	Encrypted using SHA256. Not in final dataset
preferred_language_ascl	char	
presenting_problem	varchar	Defined data but stored as Free text field
referred_to_on_departure	char	
sex	char	
state_of_usual_residence	char	
triage_category	char	
triage_date	datetime	
triage_time	datetime	

Table 6.2. CMH002 client profiles

Data column	**Data type**	**Comment**
MRN	int	Encrypted using SHA256. Not in final dataset
Service_Unit_ID	int	Mental Health Service Unit Identification
Registration date	datetime	
Registration time	datetime	
Discharge date	datetime	
Discharge time	datetime	

6.2.2.1.1 Study A: Examination of Sociodemographic Factors in Emergency Department Presentations

This study aimed to investigate the broader sociodemographic characteristics of individuals presenting to the ED with MH issues. By analysing variables such as age, gender, presenting problem, and hospital facility, the study sought to identify patterns and trends that may influence ED utilisation for MH care. Understanding these factors is critical for informing policy and improving the delivery of mental health services in acute care settings.

6.2.2.1.2 Study B: Analysis of Emergency Department Utilisation for Mental Health by Aboriginal and Torres Strait Islander People

This study focused specifically on the patterns of ED use for MH care among Aboriginal and Torres Strait Islander people. By examining the frequency, nature, and features of these presentations, the study aimed to uncover the unique challenges and barriers faced by these populations. The findings contribute to a deeper understanding of the cultural, systemic, and healthcare factors that impact access to and quality of care for Aboriginal and Torres Strait Islander communities, with the goal of promoting culturally sensitive and equitable MH services.

6.2.2.1.3 Study C: Application of Machine Learning to Identify Factors Contributing to Mental Health Admissions

The objective of this study was to utilise ML techniques to analyse the features of ED presentations that contribute to the likelihood of admission to a MH inpatient facility. By identifying key predictive factors, such as presenting problem, mode of arrival, and triage category, this study aimed to develop models that can assist clinicians in making data-driven decisions regarding the need for inpatient care. The findings not only enhance the understanding of ED presentation characteristics but also offer insights into improving triage and treatment pathways for mental health patients.

6.2.2.2 Study D: Build and Test Various Machine Learning Models

The second part of this study aimed to explore the features and benefits of various ML models applicable to MH, based on the data outlined in Part A. The development of ML models involves several key stages, including data preprocessing, feature selection, model training, hyperparameter tuning, and evaluation. Initially, raw data undergoes cleaning and

transformation to handle missing values, normalise distributions, and encode categorical variables. Relevant features are then selected to improve model efficiency and interpretability. During training, algorithms learn patterns from the data, optimising their parameters through iterative processes. Hyperparameter tuning further refines performance, ensuring the model generalises well to unseen data. Finally, evaluation metrics are used to assess the model's effectiveness in classification tasks (Paper, 2020).

For this study, six ML models were tested: explainable boosting machines (EBM) using InterpretML, random forest, XGBoost, CatBoost, k-nearest neighbours (kNN), and support vector machine using support vector classification (SVC) with a radial basis function kernel. These models were evaluated using AUC-ROC, precision, recall, F1 score, specificity, negative predictive value (NPV), false positive rate, false discovery rate, false negative rate, and Matthews correlation coefficient (MCC). A confusion matrix was utilised to assess the classification performance of the models developed in this research (Paper, 2020). Further details on the development of these ML models are presented in Study D, Section 7.5.

6.3 Research Design and Methods

6.3.1 Study Design

6.3.1.1 Study Design

Studies A and B utilised a retrospective cross-sectional descriptive epidemiological (Liamputtong, 2021, p. 212) approach to describe the cohort of MH service users seeking care through the EDs on the Central Coast of NSW. This method was chosen as it provided a clear methodological process to describe the cohort with regard to time, place, and person (Liamputtong, 2021, p. 212). A descriptive univariate analysis was used to examine individual

elements, with multivariate analysis used to establish relationships and correlations between elements. SQL, Python, and Seaborne were used for the analysis and presentation of the data.

Study C employed Python and the InterpretML package (Nori et al., 2019) to analyse the features of ED presentations that contribute to MH inpatient admissions. Python, a widely used programming language for data science and machine learning, offers a robust environment for building predictive models, handling large datasets, and implementing machine learning algorithms. In this study, InterpretML was specifically leveraged to enhance the transparency and interpretability of the machine learning models. InterpretML is a powerful tool designed to help researchers and clinicians understand and interpret complex ML models (Nori et al., 2019). By providing insights into how different features contribute to the model's predictions, it ensures that the results are not only accurate but also explainable—an essential aspect when applying AI in clinical settings. Through this approach, Study C aimed to identify which presentation features were most strongly associated with the need for inpatient care. This study contributed to a more nuanced understanding of the factors influencing MH admissions and sought to improve patient outcomes through data-driven, explainable AI solutions.

Study D built upon the initial studies by utilising the descriptive characteristics of ED presentations to build and test various ML models. The selected data were prepared to correct, impute, or remove erroneous values. Feature engineering was used to derive new attributes from the data. An example of feature engineering is deriving a particular indicator, such as psychosis, from the presenting problem, resulting in a "psychosis" field that is Boolean in value and represented as a one or a zero, which the ML algorithm can then compute. The data are then integrated as required, with individual fields reformatted to ensure data can be processed by the ML model, for example, converting textual string values to numbers or creating a numerical value lookup table. Further details of this transformation process are provided in Chapter 7.

6.3.1.2 Number of Participants

The study only reviewed existing retrospective data from 1 January 2016 to 30 December 2021 and therefore did not require recruitment.

6.3.1.3 Study Sites

The study reviewed data from Gosford and Wyong EDs

6.3.1.4 Primary and Secondary Outcome Measures

The **primary outcome** was the analysis of the data related to people who present to the ED with MH issues and the investigation of the use of ML to help understand complex relationships in this dataset.

The literature on the use of ML using MH data is limited, as seen in Chapters 3 and 5. The **secondary outcome** of this project was the use of MH data with ML, understanding not only the complexities of the data but the opportunity to engage with clinical staff about its use and place in the provision of care.

6.3.2 Statistical Methods

6.3.2.1 Statistical Analysis Plan

Studies A and B: Descriptive univariate analysis was used to examine individual elements, with multivariate analysis used to explore relationships and correlations between elements. For the initial data investigation, the researcher utilised several tools and programming languages. SQL (Structured Query Language) was used for managing and querying databases, while Python, a versatile programming language popular for data analysis and ML tasks, played a central role in the analysis. Excel, a spreadsheet application commonly used for data organisation, analysis, and visualisation, helped in structuring and reviewing the data.

Additionally, Seaborn, a Python library designed for creating informative and attractive statistical graphics, was employed to visualise the data effectively.

Study C: In this study, InterpretML was employed to build an ML model aimed at identifying the key features of ED presentations that correlate with MH inpatient admissions. Rather than focusing on evaluating model performance, the emphasis was on using InterpretML to uncover and explain the relationships within the data. The process involved selecting and refining important features, followed by the application of explainability techniques included in the package. These methods provided a clear interpretation of how each feature contributes to the likelihood of a person being admitted, offering insights into the factors that influence clinical decision making. By exploring these correlations, the study offered a deeper understanding of the patterns underlying MH admissions, potentially contributing to the improvement of patient care through more informed, data-driven decisions.

Study D involved testing various classification models based on multiple ML techniques. To effectively assess the model's capabilities, the dataset was divided into training, testing, and validation sets. Given that the ED data was unbalanced, there was a potential risk of overfitting due to the uneven class distribution. To address this issue, a combination of over-sampling and under-sampling techniques was employed to balance the dataset and create a distribution suitable for model training.

Six ML models were tested: explainable boosting machines (EBM) using InterpretML, random forest, XGBoost, CatBoost, k-nearest neighbours (kNN), and support vector classification (SVC) with a radial basis function kernel. Model performance was evaluated using several metrics, including AUC-ROC, precision, recall, F1 score, specificity, negative predictive value (NPV), false positive rate, false discovery rate, false negative rate, and Matthews correlation coefficient (MCC). A confusion matrix was utilised to summarise the outcomes of predictions,

providing a comprehensive evaluation of each model's classification performance (Paper, 2020).

The data analysis and visualisation were conducted using SQL, Python, InterpretML (Nori et al., 2019), and Seaborn. Figure 6.3 presents a high-level flowchart outlining the data processing steps.

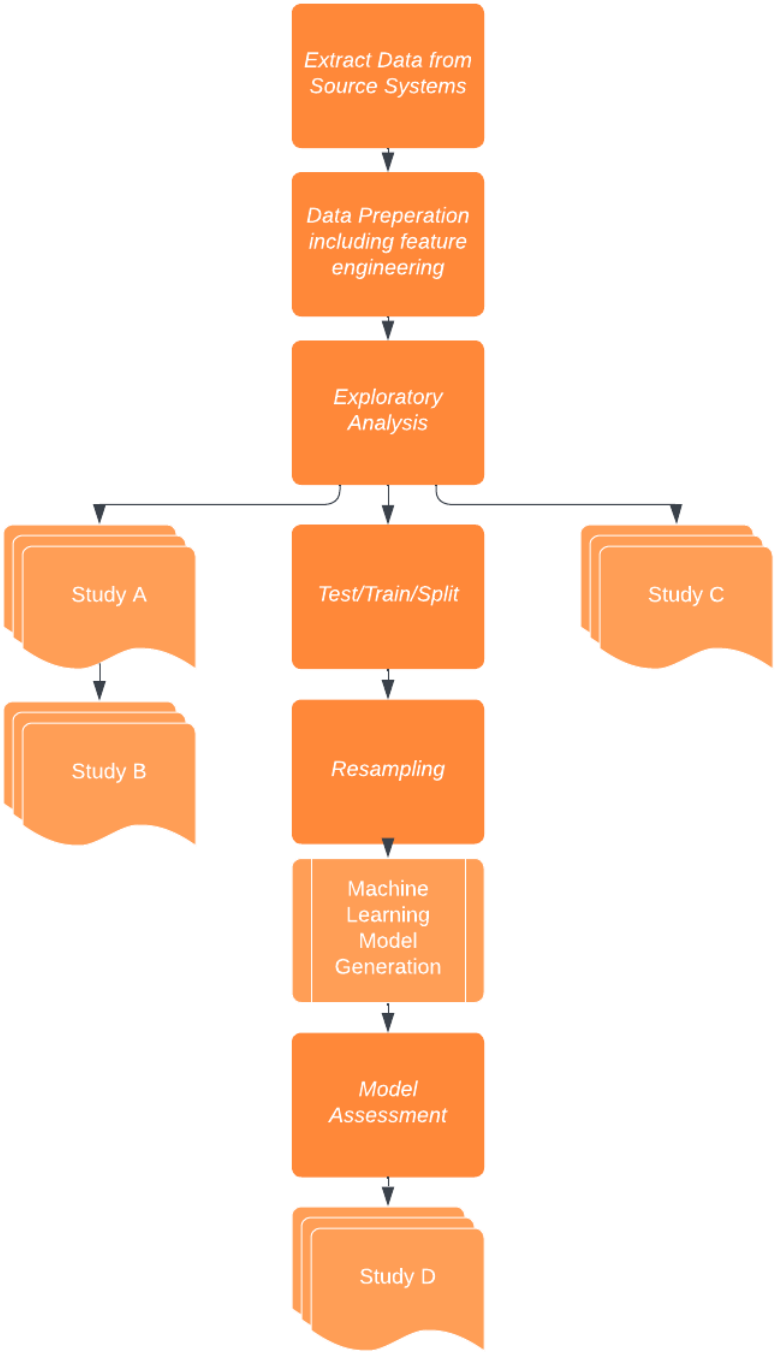


Figure 6.3. *Protocol machine learning data flow*

6.4 Research Ethics

6.4.1 Data Management

6.4.1.1 Data Collection

Data was extracted from NSW Health's Health Information Exchange and Enterprise Data Warehouse, with data being derived from the ED_VISIT table. Individuals were identified as seeking MH care through an algorithm that searched ED diagnosis codes and "presenting problem" using free text for regular expressions such as "self-harm", "suicidal" or "overdose". This identification strategy is in line with similar MH research which utilised the same ED dataset (Sara et al., 2023).

6.4.1.2 Data Storage

Research data was stored in accordance with all relevant NHMRC guidelines regarding data retention and management in accordance with National Statement 3.1.45 and 3.1.56 (National Health and Medical Research Council, 2018b). The data used for this project's research comprised ED presentation data from the New South Wales Health Information Exchange. This research only used data that related to presentations in the Central Coast Local Health District (CCLHD). Raw research data was processed in accordance with CCLHD guidelines (Central Coast Local Health District, 2018) using a CCLHD two-factor authenticated computer. Processed research data were stored securely using secure password-protected CCLHD storage. This dataset was only accessible to researchers with ethics approval to access the data. No physical copies of data were created during this research.

6.4.1.3 Data Confidentiality

Data were extracted from the ED_VISIT table and de-identified as part of the extraction process, with the exception of Medical Record Numbers (MRN). MRNs were used to link data

from the ED data with MH eMR data to determine if the person was a current client of the MH service at the time of their ED presentation. The MRN data were de-identified by encrypting the data field during the extraction phase using a secure hash algorithm (SHA). In this case, it was SHA2-256, a 256-bit (32 bytes) hashing algorithm that can calculate a hash code for an input up to 264-1 bits. The calculated hash code was a 64-digit hexadecimal number converted into American Standard Code for Information Interchange (ASCII). For example, the SHA-256 hash code for MRN "426788" is "mrM0vJZ/8b3k5aYrE8Wev8f345H3xlFiDcvb+flLrn8=". This encryption makes the re-identification of data nearly impossible, as none of the original data exists in the cryptographic key. A full breakdown of the data fields to be extracted is attached in the chapter appendix 6.6.1. In plain language, this one-way encryption means it is not possible to obtain the patient's name and any data associated with it.

The extraction and linkage of the data was undertaken by the Manager of Population Data and Outcome Measures for Mental Health. The linked dataset was generated and the MRN was omitted from the final dataset, with the original extracted data being deleted. This process was supervised by Dr Rachel Sheather-Reid, Translational Research Manager for Central Coast Local Health District Mental Health.

6.4.2 Informed Consent Process

6.4.2.1 Waiver of Consent

The waiver of consent for this study was carefully considered in line with the National Statement on Ethical Conduct in Human Research (NHMRC, 2018b), specifically under section 2.3.10. This decision is grounded in a comprehensive assessment of risks, benefits, practicality, privacy, and compliance with legal standards, ensuring adherence to ethical standards while maintaining research feasibility.

The involvement of participants in this research posed no more than a low risk, as outlined in paragraphs 2.1.6 and 2.1.7 of the National Statement. The extraction method prevents identification of individuals by name, and data analysis is structured to maintain participant anonymity, ensuring confidentiality remains intact.

The research offers considerable benefits that are deemed to outweigh the minimal risk associated with not obtaining individual consent. This study aims to enhance the efficiency and quality of MH care in ED, directly benefiting patients through faster, more effective care. Health services are expected to gain insights into the factors associated with MH-related admissions, supporting clinicians, especially those with less experience, in making informed decisions and improving patient outcomes. This knowledge will also assist in addressing workforce challenges by enabling a more streamlined and informed response to MH presentations.

Obtaining individual consent was considered impracticable due to the scale and nature of the data, which is drawn from an extensive database. The research design involves retrospective analysis of a large dataset, making it logistically challenging and disruptive to seek consent from each individual whose records are included. Based on current literature and an understanding of MH patients' reliance on EDs, there is no known or likely reason to believe that individuals would object to their data being used for this purpose. Many MH service users repeatedly access ED services due to limited alternatives, and previous studies highlight that experiences in these settings are often suboptimal. The aim of improving care models directly addresses this need, aligning with the interests of participants.

To ensure participant privacy, the dataset's only identifying element, the Medical Record Number (MRN), was encrypted using a one-way format. This encryption was applied prior to analysis, ensuring that no identifying information, such as MRNs, was included in the final

dataset or left the Central Coast Local Health District system, thus minimising the risk of identification and preserving privacy.

In terms of confidentiality, all data management and analysis followed rigorous protocols. The MRN was used solely for initial data linkage, then removed from the dataset before analysis. Further, these methods align with national and state regulations, including s 95A of the Privacy Act 1988 (Cth) (Office of the Australian Information Commissioner, 2014) and the Health Records and Information Privacy Act 2002 (NSW) (New South Wales Government, 2002), which govern privacy standards in health research.

While individual feedback to participants was not feasible, findings will be disseminated in an accessible, transparent manner. Aggregated results will be shared publicly through journal publications and, where possible, via CCLHD communications channels, including radio, print, and online media, to inform the broader community about potential improvements in MH care.

In summary, this waiver of consent is ethically justified due to the minimal risk involved, the significant benefits of improving MH services, the impracticality of obtaining individual consent, and the robust privacy and confidentiality protections in place. The waiver is in full compliance with relevant legal and ethical standards, ensuring the research is both ethically sound and aligned with participant welfare.

6.4.2.2 Study Record Retention

In accordance with National Statement 2.3 (National Health and Medical Research Council, 2018b), data from the study will be retained for five years from the date of publication, at which point the electronic data will be erased, and if any physical copies have been made these will be securely destroyed.

6.4.3 Independent HREC approval

The research received approval from Hunter New England Human Research Ethics Committee, 2022/ETH01597 (approved on 18 November 2022), and site-specific approval from Central Coast Local Health District, granted on 14 December 2022. Subsequently, data extraction and linkage processes were undertaken by the researcher on 16 December 2022.

6.5 Study E: Investigating Mental Health Nurses' Perceptions About the Future of Mental Health Nursing and their Attitudes Towards Artificial Intelligence

A proposed postdoctoral study will build upon the insights gained from the current research, highlighting key issues such as workforce dynamics, missed care, and attitudes towards the use of AI in clinical practice, particularly in the MH sector. This future study will aim to deepen the understanding of these critical areas, with a specific focus on how AI can be effectively integrated into healthcare settings to address systemic challenges. While this plan outlines the direction and aims of the postdoctoral study, the results of this future research are not included in the present thesis, as the study is ongoing. Nevertheless, the findings from this work are expected to make a significant contribution to the broader discourse on AI in health care, offering valuable insights into how AI can support clinicians in delivering high-quality, timely care while addressing workforce and operational challenges.

6.5.1 Aims

The aim of this research study is twofold:

1. to investigate MH nurses' experiences and views about integrating AI within their current practice and

2. to explore their perceptions and attitudes about the future of the MH nursing profession as AI becomes more prevalent in that context.

This study seeks to understand the issues that MH nurses face when working in the frontline, with diminishing resources, a reduced MH nursing workforce and the introduction of AI and data-driven technologies to MH nursing practice. This study aims to explain the role of the MH nurse and their perceived expectations of the future of the MH nursing profession through identifying the strengths, weaknesses, opportunities, and threats that MH nurses face.

A mixed-methods study will be undertaken whereby participants will be offered free text responses to questions (qualitative experiences) and a survey instrument (descriptive statistics). The research team wish to ask MH nurses to respond to five open-text questions which are designed to elicit the perspectives related to the overall, or specific, future of the MH nursing profession. The open-text questions use a SWOT (strengths, weaknesses, opportunities, and threats) framework. It is anticipated that participants will require 15–20 minutes to complete the online survey. To analyse the responses, the research team will apply qualitative thematic analysis techniques, generating themes that will illuminate the real-world perspective of MH nurses related to the research topic (Braun & Clarke, 2021; Gildberg & Wilson, 2023). Furthermore, using the General Attitudes towards Artificial Intelligence Scale (GAAIS) (Schepman & Rodway, 2020), the research team will survey the MH nursing practice community to enable a description of their attitudes towards AI. This is important because increasing exposure to AI in the MH nursing context is occurring, and MH nurses' attitudes and perspectives about this evolving phenomenon are largely unknown at this time.

The research study will have three parts:

Part 1 will gather relevant non-identifiable demographic data about the MH nurses who choose to participate.

Part 2 will use a strengths, weaknesses, opportunities, and threats (SWOT) framework to ask participants open questions concerning MH nurses' perceptions about their current practice as

well as explore their perceptions about the of the future of MH nursing. This section of the study will ask MH nurses to identify any weaknesses in the future, any opportunities they might see on the horizon, and future threats they might anticipate.

Part 3 will apply the GAAIS (Schepman & Rodway, 2020), a 20-question Likert scale tool, to assess the MH nurses' attitudes towards AI.

6.6 Methodological Rigour

6.6.1 Rigour and Trustworthiness

According to Liamputtong (2021 p25), the concepts of "rigour" and "trustworthiness" are fundamental to ensuring the credibility and reliability of qualitative research. Rigour in qualitative research pertains to the meticulousness and thoroughness of the research. This involves ensuring methodological consistency, where the selected methods are appropriate for addressing the research questions and are applied uniformly throughout the study. Additionally, transparency is crucial, requiring comprehensive research process documentation to enable other researchers to follow and understand the methodological steps undertaken. Reflexivity is another key component, where the researcher maintains self-awareness regarding their influence on the research process, including how their personal biases and assumptions may affect data collection and analysis (Liamputtong, 2021 p25).

Methodological rigour in healthcare machine learning (ML) research necessitates a comprehensive evaluative paradigm that aligns with the principles of evidence-based medicine (EBM) and established clinical research methodologies (McCradden et al., 2022). A key challenge in ML-driven healthcare interventions is overcoming the "AI chasm"—the gap between algorithmic development and clinically meaningful application—by ensuring that validation extends beyond retrospective performance metrics to include robust empirical

testing in real-world settings (McCradden et al., 2022). The translation of ML into clinical practice requires a stepwise approach encompassing exploratory ML research, silent period evaluation, and prospective clinical validation to establish both technical validity and clinical efficacy. Inadequate methodological approaches not only undermine the credibility of ML models but also pose ethical risks by deploying untested interventions into patient care. McCradden et al. (2022) highlight that ML models should not only undergo rigorous algorithmic validation but also adhere to the ethical and epistemic standards that govern human subjects research, reinforcing the necessity of a research ethics framework tailored to clinical AI. Furthermore, without a robust evidentiary hierarchy for ML interventions, there is a risk of research waste and unintended clinical consequences, underscoring the importance of regulatory oversight, prospective evaluation, and methodological consistency in ML-enabled healthcare (McCradden et al., 2022). By embedding these principles into the methodological framework of this study, this research ensures that its findings are both scientifically sound and ethically responsible, contributing to the responsible translation of ML into mental health practice.

The researcher ensured rigour throughout the study by maintaining a meticulous and thorough approach to the research process. Recognising the importance of methodological consistency, the researcher carefully selected methods that were appropriate for addressing the research questions and applied these methods uniformly across the study. This methodological consistency was crucial in maintaining the integrity of the research and ensuring that the findings were reliable and valid. Transparency was another key aspect of the research process, with the researcher diligently documenting each step to allow for replication and to provide a clear understanding of the research journey. Reflexivity was consistently practised, with the researcher maintaining an awareness of their influence on the research process, particularly in relation to how personal biases and assumptions could impact data collection and analysis. By

engaging in reflexive practices, the researcher was able to critically examine their role in the research, thereby enhancing the rigour of the study (Liamputtong, 2021 p25).

Trustworthiness in qualitative research refers to the level of confidence in the data, interpretations, and methods used, and it is typically assessed through four key criteria. Credibility denotes the confidence in the veracity of the findings, which can be enhanced through techniques such as member checking, prolonged engagement, and persistent observation. Transferability relates to the extent to which the research findings can be applied to other contexts or settings, supported by providing rich, detailed descriptions that allow others to determine the applicability of the findings to their own situations (Liamputtong, 2021 p25). Dependability addresses the stability of data over time and under varying conditions, achieved by maintaining an audit trail that documents the research process and any changes that occur. Confirmability reflects the degree to which the findings are shaped by the respondents rather than by researcher bias, motivation, or interest. Ensuring confirmability involves practising reflexivity and maintaining a transparent audit trail.

Trustworthiness was a central focus of the research, guided by the principles outlined by Liamputtong (2021 p25). To enhance credibility, the researcher employed techniques such as member checking, where feedback was sought from supervisors and subject matter experts to verify the accuracy of the findings. The researcher was permanently based within the research setting where persistent observation was utilised to deepen the understanding of the context and to strengthen the validity of the findings. Transferability was addressed by providing rich, detailed descriptions of the research context and findings, allowing others to determine the applicability of the results to different settings, and openly documenting the limitations of the findings. The researcher also ensured dependability by maintaining a comprehensive research log, documenting the research process, decisions made, and any changes that occurred. Confirmability was reinforced through rigorous reflexive practices and by ensuring that the

findings were grounded in the data rather than influenced by the researcher's biases or motivations. The combination of these strategies contributed to the overall trustworthiness of the study, ensuring that the research findings are credible, transferable, dependable, and confirmable (Liamputtong, 2021 p25).

6.6.2 Credibility, Dependability, and Transferability

6.6.2.1 Credibility

According to Liamputtong (2021 p106), credibility is a cornerstone of qualitative research, representing confidence in the truthfulness and accuracy of the findings. It involves several essential practices that ensure the research outcomes genuinely reflect the participants' perspectives and experiences. One such practice is prolonged engagement, which entails spending sufficient time in the field to build trust with participants and gain a deep understanding of the context. This immersion allows researchers to gather more comprehensive and nuanced data. Persistent observation complements this by focusing on the most relevant aspects of the study, ensuring that data collection is thorough and detailed (Liamputtong, 2021 106).

Member checking is another critical component of credibility. This process involves returning the data or findings to the participants to confirm their accuracy and resonance with their experiences. It serves as a validation mechanism, where participants can provide feedback and corrections, thereby enhancing the authenticity of the research. Triangulation further strengthens credibility by using multiple methods, data sources, or investigators to cross-check findings. This approach helps identify and rectify inconsistencies, ensuring the conclusions are robust and reliable (Liamputtong, 2021 p106). Reflexivity is also crucial in maintaining credibility. Researchers must remain aware of their biases and how they might influence the research process. By continuously reflecting on their assumptions and decisions, researchers

can mitigate potential biases and enhance the trustworthiness of their findings. Additionally, maintaining a detailed audit trail is important. This involves documenting all aspects of the research process, including data collection, analysis, and decision making. An audit trail provides transparency and allows others to follow the researcher's path, thereby validating the research process and findings (Liamputtong, 2021 p106).

The researcher placed a strong emphasis on establishing credibility throughout this study, recognising its importance in ensuring that the findings accurately reflect the realities of MH care in EDs. To achieve this, well-established research methods, such as the DT framework, were employed, which prioritises empathy and iterative feedback, and the use of a research log. This approach was crucial in developing solutions that are closely aligned with the actual experiences of patients and healthcare providers. The use of retrospective cohort analysis provided a robust empirical foundation, drawing on real-world data to enhance the believability and relevance of the study's outcomes. Additionally, the researcher ensured that any ML models were rigorously tested and validated, confirming that the predictive outcomes are not only theoretically sound but also practically applicable and could be clearly explained in a challenging environment such as EDs.

6.6.2.2 Dependability

According to Liamputtong (2021 p26), dependability in qualitative research is crucial for ensuring that the study's findings are consistent and repeatable under similar conditions. Dependability is achieved by thoroughly documenting the research process and employing various strategies to maintain methodological consistency. One key aspect is the maintenance of an audit trail, which involves detailed record keeping of all stages of the research, including data collection, coding, and analysis. This comprehensive documentation allows others to follow the researcher's decisions and understand the rationale behind the methodological

choices, thereby enhancing the transparency and repeatability of the study. Dependability was a key consideration in the research design, as the researcher aimed to ensure that the findings would be reliable and consistent over time. The systematic and repeatable nature of the research process was central to this goal and especially pertinent in dealing with subjects such as AI and ML. By employing the iterative DT framework, the researcher was able to continuously refine ideas and approaches based on testing, ensuring that the research findings remained relevant and robust, even as the context evolved. The retrospective cohort analysis provided a longitudinal perspective, allowing the researcher to assess trends and patterns over time that could provide further insight into the community in question, further supporting the dependability of the research outcomes. This consistency is particularly important in MH care, where reliable predictions and interventions can significantly impact patient outcomes. The researcher was committed to producing findings that would stand up to scrutiny and be applicable in various scenarios over time.

6.6.2.3 Transferability

According to Liamputtong (2021 p26), transferability is a key criterion in qualitative research. It refers to the extent to which the findings of a study can be applied or transferred to other contexts or settings. Achieving transferability involves providing a rich, detailed description of the research context and the participants. This enables other researchers or practitioners to determine the applicability of the study's findings to their own settings. Detailed accounts of the research process, including the context, methods, and participants, also contribute to transferability (Liamputtong, 2021 p26). This transparency allows other researchers to assess the study's context and the extent to which its findings might be applicable to other settings. In developing the ML models, the researcher focused on creating tools that could be generalised and adjusted for different datasets and populations. This attention to transferability reflects a broader ambition to contribute to the field of MH care in a way that transcends the immediate

context of the study. The researcher intended the findings to be relevant not only to the specific ED setting but also to other healthcare environments facing similar challenges. By ensuring that the research is adaptable and broadly applicable, noting any limitations, the researcher aimed to maximise its impact and utility across different settings, ultimately contributing to a more effective and responsive MH care system.

6.7 Chapter Summary

This chapter has outlined the comprehensive methodological approach employed in this research project, which aimed to leverage AI and ML to support decision making in the assessment of MH consumers presenting to EDs. The overarching DT methodology facilitated a deep understanding of the needs and concerns of various stakeholders, including clinicians, patients, and interdisciplinary experts, enabling the co-creation of technically robust solutions aligned with the values and expectations of end users. The retrospective cohort study method was utilised to analyse sociodemographic and presentation features of individuals seeking MH care in ED, providing valuable insights into the factors contributing to admissions. Furthermore, the application of supervised ML techniques allowed for the development and evaluation of predictive models, integrating the insights gained from the DT phases into algorithmic solutions. Ethical considerations were central throughout the research process, ensuring adherence to ethical guidelines and informed consent procedures. Overall, the methodological approach employed in this research project demonstrates a commitment to rigorous scientific inquiry while emphasising user-centricity, empathy, and ethical considerations. The findings generated through these methods have the potential to inform the development of innovative and practical solutions for improving MH outcomes and enhancing the delivery of MH care services.

Chapter 7 Results

7.1 Introduction

The rising demand for mental health (MH) services and the associated challenges in providing timely and appropriate care for individuals experiencing MH crises are significant public health concerns globally, including in Australia. Emergency departments (ED) serve as critical frontline access points for MH care, yet they often face issues such as overcrowding, long wait times, and inadequate specialised resources (Digel Vandyk et al., 2018; Morphet et al., 2012). This research investigates various aspects of MH care in EDs, focusing on sociodemographic factors, presentation features, and the role of artificial intelligence (AI) and machine learning (ML) in enhancing MH service delivery.

7.1.1 Sociodemographic Factors and Presentation Features (Study A)

The first study in this research (Study A) examined the sociodemographic factors and presentation characteristics of individuals seeking MH care in two regional EDs in New South Wales, Australia, from 2016 to 2021. This retrospective cohort study aimed to characterise the population presenting with MH concerns, identify factors associated with admission to inpatient MH facilities, and explore differences in presentation patterns between the two EDs. Key findings include the high prevalence of suicidal ideation among presentations and significant proportions of individuals departing without being seen or against medical advice. These insights underline the need for targeted interventions and improved access to specialised MH services. The findings are presented in a publication found in Section 7.2.

7.1.2 Disproportionate Mental Health Presentations Among First Nations People (Study B)

Study B focused on the disproportionate burden of MH challenges faced by Aboriginal and Torres Strait Islander communities in Australia. These communities experience higher rates of

suicide, self-harm, and psychological distress compared to non-Indigenous populations. The study highlighted the over-representation of First Nations peoples in mental health–related ED presentations, indicating the need for culturally responsive care models. Recommendations include fostering cultural safety within EDs, integrating traditional healing practices, and developing holistic, trauma-informed interventions that align with Indigenous worldviews. The findings are presented in a publication in Section 7.3.

7.1.3 Using Machine Learning in Mental Health Care (Study C)

Study C focused on the application of ML to analyse ED presentations and identify key factors contributing to admissions to MH inpatient facilities. This study used Python and the InterpretML package to develop interpretable ML models that uncover correlations between presentation features—such as presenting problem, triage category, and prior healthcare utilisation—and the likelihood of admission. The study did not focus on testing model performance but prioritised transparency and the ability to interpret the data, which is critical for building clinician trust in AI-driven decision support systems. By identifying and explaining these correlations, the research aimed to offer insights that can enhance clinical decision making, improve triage processes, and address the ongoing issue of missed care within EDs, ultimately supporting more effective resource allocation and reducing the burden on healthcare staff. The findings are presented in a publication in Section 7.4.

7.1.4 Machine Learning in Mental Health Care (Study D)

Study D explored the application of ML models to predict admissions to MH inpatient facilities from ED presentations. The research involved developing and testing ML models, including explainable boosting machines (EBMs) and support vector machines (SVMs), using data from 2016 to 2021. The findings indicate that non-linear SVM models outperform linear models in predictive accuracy, though the need for transparency and clinician trust in ML models is

emphasised. The study highlights the potential of ML to enhance clinical decision support, improve triage processes, and optimise resource allocation in EDs. Detailed findings are presented in a publication in Section 7.5.

7.1.5 Implications and Future Research

The findings from this research have significant implications for enhancing MH care delivery and addressing the challenges faced in ED settings. Key recommendations for future research include evaluating the effectiveness of safety planning interventions, exploring integrative care models, and leveraging ML for predictive analytics to support clinical decision making. Additionally, it is crucial to address socioeconomic and geographical barriers that limit access to MH services. Understanding the factors influencing ED presentations and outcomes can enable healthcare providers and policymakers to develop targeted interventions and service delivery models that better meet the needs of individuals with MH conditions, thereby promoting improved patient outcomes.

Overall, this research highlights the need for a comprehensive and integrated approach to MH care that bridges primary care, community-based services, and EDs. By incorporating advanced analytical techniques such as ML and fostering culturally responsive, trauma-informed care, healthcare systems can more effectively meet the needs of vulnerable populations and improve MH outcomes. These efforts, combined with a focus on workforce integration and trust in AI, will support AI-driven solutions' long-term sustainability and success in MH care delivery.

7.2 Sociodemographic Factors and Presentation Features of Individuals Seeking MH Care in EDs: a Retrospective Cohort Study

Higgins, O., Sheather-Reid, R. B., Chalup, S. K., & Wilson, R. L. (2024b). Sociodemographic factors and presentation features of individuals seeking mental health care in emergency departments: A retrospective cohort study. *International Journal of Mental Health Nursing*. Advance online publication. https://doi.org/10.1111/inm.13414

	Higgins, O	Sheather-Reid, R. B	Chalup, S. K	Wilson, R. L
Conceptualization	X		X	X
Methodology	X		X	X
Software	X			
Validation	X	X		X
Formal analysis	X	X		X
Investigation	X	X		
Resources	X			
Data Curation	X			
Writing - Original Draft	X			
Writing - Review & Editing	X	X	X	X
Visualization	X			
Supervision			X	X
Project administration	X			
Funding acquisition				
Percentage	**70%**	**10%**	**5%**	**15%**

7.2.1 Article Introduction

The growing demand for MH services and challenges in providing timely, appropriate care for individuals experiencing MH crises is a pressing public health issue globally and in Australia (Australian Institute of Health and Welfare, 2023c; World Health Organization, 2022b). EDs play a critical role as frontline access points for MH care, yet they face numerous challenges, including overcrowding, long wait times, and lack of specialised resources. This study examines the sociodemographic factors and presentation characteristics of individuals seeking

MH care in two regional EDs in New South Wales, Australia, from 2016 to 2021. The research aims to identify factors associated with admission to inpatient MH facilities following an ED presentation for MH reasons.

The primary objective of this retrospective cohort study is to analyse the sociodemographic and presentation features of people who sought MH care in two EDs located on the Central Coast of NSW from 2016 to 2021. Specifically, the study aims to:

- Characterise the population presenting to these regional EDs for MH concerns, including their demographic profiles, presenting problems, and modes of arrival.
- Identify factors associated with admission to an inpatient MH facility following an ED presentation for MH reasons.
- Examine the differences in presentation patterns and outcomes between the two EDs, considering the availability of MH resources and facilities at each site.

This study contributes to the broader research on the utilisation of ML in MH by providing a comprehensive understanding of the population seeking MH care in EDs and the factors influencing their care trajectories. The findings were used to inform the development of ML models and clinical decision support tools to enhance triage processes, resource allocation, and care delivery for individuals with MH needs presenting to EDs.

Check for updates

International Journal of Mental Health Nursing

WILEY

International Journal of Mental Health Nursing

ORIGINAL ARTICLE OPEN ACCESS

Sociodemographic Factors and Presentation Features of Individuals Seeking Mental Health Care in Emergency Departments: A Retrospective Cohort Study

Oliver Higgins[1,2,3] | Rachel B. Sheather-Reid[2,3] | Stephan K. Chalup[4] | Rhonda L. Wilson[1,2,3]

[1]RMIT University, Melbourne, Australia | [2]Central Coast Local Health District, New South Wales, Australia | [3]Central Coast Research Institute, Gosford, Australia | [4]School of Information and Physical Sciences (Computer Science and Software Engineering), University of Newcastle, Newcastle, Australia

Correspondence: Oliver Higgins (oliver.higgins@newcastle.edu.au)

Received: 19 November 2023 | **Revised:** 27 June 2024 | **Accepted:** 18 August 2024

Funding: Partial financial support was received from the NSW Ministry of Health as part of the Towards Zero Suicides initiative.

Keywords: artificial intelligence | emergency department | machine learning | mental health | nursing

ABSTRACT

Emergency Department (ED) presentations for Mental Health (MH) help-seeking have been rising rapidly, with EDs as the main entry point for most individuals in Australia. The objective of this retrospective cohort study was to analyse the sociodemographic and presentation features of people who sought mental healthcare in two EDs located in a regional coastal setting in New South Wales (NSW), Australia from 2016 to 2021. This article is a part of a broader research study on the utilisation of machine learning in MH. The objective of this study is to identify the factors that lead to the admission of individuals to an MH inpatient facility when they seek MH care in an ED. Data were collected using existing records and analysed using descriptive univariate analysis with statistical significance between the two sites was determined using Chi squared test, $p < 0.05$. Two main themes characterise dominant help-seeking dynamics for MH conditions in ED, suicidal ideation, and access and egress pathways. The main findings indicate that suicidal ideation was the most common presenting problem (38.19%). People presenting to ED who 'Did not wait' or 'Left at own risk' accounted for 10.20% of departures from ED. A large number of presentations arrived via the ambulance, accounting for 45.91%. A large proportion of presentations are related to a potentially life-threatening condition (suicidal ideation). The largest proportion of triage code 1 'Resuscitation' was for people with presenting problem of 'Behavioural Disturbance'. Departure and arrival dynamics need to be better understood in consultation with community and lived experience groups to improve future service alignment with the access and egress pathways for emergency MH care.

1 | Introduction

The growing disparity between the demand and supply of healthcare continues to be a pressing issue, as evidenced by the 8.8 million Emergency Department (ED) visits recorded in Australian public hospitals in 2020–2021 (Australian Institute of Health and Welfare 2021). Consumers increasingly spend more time in EDs, with fewer visits completed within the key performance benchmark standard of under 4-h wait duration set by New South Wales (NSW) Health (Australian Institute of Health and Welfare 2021). NSW Health has committed to ensuring that 81% of individuals who present at a public hospital ED will either be discharged, referred to another hospital for treatment, or admitted to the hospital and physically leave the ED within a benchmarked 4-h timeframe (NSW Health 2021). The objective of this retrospective cohort study was to analyse the sociodemographic and presentation features of people who sought Mental Health (MH) care in two EDs located on the Central Coast of NSW from 2016 to 2021. This article is a part of a broader research study on the utilisation of

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2024 The Author(s). *International Journal of Mental Health Nursing* published by John Wiley & Sons Australia, Ltd.

Machine Learning (ML) in MH. The objective of this study is to identify the factors that lead to the admission of individuals to an MH inpatient facility when they seek MH care in an ED.

1.1 | Background

EDs present a challenging and often complex environment for individuals experiencing MH conditions. These settings are characterised by excessive noise, busyness, crowding and a clinical atmosphere that operates around the clock with high levels of activity, potentially exacerbating distress for many people (Clarke, Dusome, and Hughes 2007). The high volume of MH-related presentations to EDs can lead to long wait times, overcrowding and delays in treatment (Duggan et al. 2020), resulting in frustration and dissatisfaction among consumers (Digel Vandyk et al. 2018; Morphet et al. 2012). Compared to individuals with other conditions, MH consumers often experience lengthy stays and wait times for inpatient beds in the emergency department (Duggan et al. 2020). Triage processes are essential to ensure timely and appropriate care for consumers with mental health problems. However, many EDs struggle to implement effective triage systems (Morphet et al. 2012). The extended waiting times for individuals seeking mental health services have been associated with negative safety and clinical outcomes, including higher mortality rates, an increased risk of violence and aggression in the emergency department, and patients leaving without receiving the necessary care, thus disregarding their needs (Duggan et al. 2020). Despite attempts to enhance the consumer journey, many MH patients still report unsatisfactory experiences in the emergency department, which includes long wait times, lack of specialised healthcare and improper treatment settings (Digel Vandyk et al. 2018; Morphet et al. 2012). Ultimately, the challenges posed by MH presentations in EDs underscore the need for greater infrastructure, resources and knowledge to provide specialised and timely care to individuals with MH problems (Digel Vandyk et al. 2018). Addressing these issues is crucial to ensure that individuals experiencing MH conditions receive appropriate and compassionate care in a suitable environment that promotes their well-being and recovery.

1.2 | Consumer Perspective

For many consumers, the presentation to ED is not a first choice but rather is the only choice available to them (Digel Vandyk et al. 2018; Morphet et al. 2012). The selected site in this research contains two EDs and is the only option for people in mental distress, particularly after hours and on weekends when there is a lack of access to primary care. Morphet et al. (2012) explored responsiveness issues from an MH consumer and carer perspective, aiming to improve the ED journey in Australian hospitals in Victoria. This mixed methods research comprised both surveys and focus groups to better understand the lived experience. Service participants were contacted following a presentation to the ED for MH reasons. The survey yielded 65 respondents, with 46 (71%) identifying themselves as MH clients, 10 (15%) as next of kin or carers of those with MH problems and the remaining 9 (14%) responding as neither an MH consumer nor carer. Participants in the study reported that they would have preferred a separate space for people with MH problems to reduce the stigma associated with mental illness (Morphet et al. 2012). However, not all participants agreed with this idea, highlighting the conflict between reducing stigma and providing specialised services (Morphet et al. 2012).

Approximately, half of the participants (54%) stated that their main reason for attending ED was for MH (Morphet et al. 2012), and 48% indicated that they spent over 4 h in ED (Morphet et al. 2012). Similarly, Digel Vandyk et al. (2018) explored the experiences of 10 adults who made 12 or more ED visits within one year of a multisite tertiary facility in Ontario, Canada. Their findings revealed that the participants believed they had no choice but to seek hospitalisation. They viewed every visit as essential and interpreted the staff's failure to address their needs as discriminatory and disrespectful (Digel Vandyk et al. 2018). Digel Vandyk et al. (2018) observed the recommendations that healthcare providers should work towards adopting tailored approaches instead of a uniform strategy when implementing interventions aimed at enhancing emergency care experience for people presenting with MH problems.

1.3 | Research Setting and Context

This article reports on the findings of research conducted in one coastal regional local health service in NSW, Australia, that contains two public hospitals containing EDs servicing the community of 346 596 residents over an area of 1681 km^2 (Australian Bureau of Statistics 2021a). The Mental Health Act 2007 allows involuntary detention and treatment for individuals found to be mentally ill or disordered and requires the individual to be taken to a declared mental health facility, of which both ED sites are, for assessment (NSW Health 2019). Acute inpatient services and community-based care are offered as part of the public health comprehensive MH service delivery programme, with services provided across three streams: Child and Youth Mental Health (0–12 and 12–17 years), Adult (18–65 years) and Older persons (65+ years). In a departure from other NSW EDs specialised MH staff provide assessments at the EDs around the clock, seven days a week. The two sites total 84 acute inpatient MH beds on the Central Coast. Site A comprises 24 acute and 6 high-dependency beds (30 in total). Site B comprises 25 acute, 10 high-dependency and 15 older persons (65+) beds as well as a Psychiatric Emergency Care Centre (PECC) with four beds adjacent to the ED (54 in total). Neighbouring health districts provide access to sub-acute, non-acute and rehabilitation beds for clients requiring extended or specialty MH care, as these bed types are not available in the local area. Adolescent (ages 12–17) MH beds are unavailable on the Central Coast; instead, referrals to adolescent inpatient services are made to the neighbouring districts' specialised inpatient units (Central Coast Local Health District 2017).

1.4 | Central Coast Demographics

The Central Coast region of New South Wales is situated in close proximity to Sydney, Australia. According to the

14470349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/inm.13414 by Oliver Higgins - The University Of Newcastle , Wiley Online Library on [16/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Australian Bureau of Statistics (2021a), the gender distribution in the Central Coast is 48.5% male and 51.5% female. Notably, 4.9% of the population within the Central Coast identify as Aboriginal and/or Torres Strait Islander peoples, a higher proportion compared to 3.4% in NSW and 3.2% across Australia (Australian Bureau of Statistics 2021a). The Central Coast of NSW is experiencing an ageing population, with a growing number of individuals aged 55 and above in recent years. This trend is projected to continue in the coming decades, according to the Australian Bureau of Statistics (2021a). Compared to both NSW and Australia, the Central Coast region faces poorer health outcomes, including a higher prevalence of MH conditions such as anxiety and depression, at 11.0%, in contrast to 8.0% in NSW and 8.8% across Australia (Australian Bureau of Statistics 2021b). Additionally, the region has experienced higher than average increases in high psychological distress over the period of 2015–2021 (NSW Health 2023). In light of these demographic and health challenges, there is a pressing need for public health services in the Central Coast to be supported by evidence-based research that describes the dynamics related to MH presentations in EDs. Such evidence can inform effective planning and service delivery strategies tailored to the specific needs of the catchment population while considering the available budget and resources.

2 | Method

2.1 | Design

This retrospective cohort design utilises existing data collected in routine care and employs a cross-sectional descriptive epidemiological method (Liamputtong 2021, 212).

2.2 | Setting

The study was set in two public EDs on the Central Coast, a local health district of New South Wales, Australia.

2.3 | Population and Sample

The study included all presentations made to the ED with MH recorded as the presenting problem. The sample for the current study was drawn from 1 January 2016 to 31 December 2021.

2.4 | Data Collection

The primary data collection reviewed existing data from 1 January 2016 and 30 December 2021 and did not require recruitment. The ED data were selected for inclusion using the following terms in the presenting problem; '%suici%' *OR* '%MH%' *OR* '%self-harm%' *OR* '%psych%' *OR* '%Mental%'. A second data collection was extracted from the electronic medical record to link the ED presentations with MH service data to determine if the individual presentation is linked with a current service or previously received care. This extraction was extended back a further year, from 1 January 2015 to 30 December 2021, to ensure the inclusion of previous service use data. Understanding a person's previous connection to MH services is crucial when they present to the ED. Previous MH hospitalisations have been identified as a significant risk factor for future readmissions (Vigod et al. 2013). The phenomenon of 'revolving door' patients, characterised by frequent hospitalisations, has been observed in individuals with schizophrenia or affective disorders, highlighting the importance of considering past admissions and connection to MH services (Fonseca Barbosa and Gama Marques 2023). Furthermore, predictive models have incorporated previous hospitalisations as a valuable variable for forecasting psychiatric readmissions (Góngora Alonso et al. 2023). However, it is essential to recognise that previous admissions are not the sole determinant, and a comprehensive assessment of various personal, clinical and systemic factors is necessary to accurately evaluate the risk of future hospitalisations and provide appropriate interventions.

A full list of the variables used is located in Table 1. The data set extracted contained 26 681 records, of which 14 were missing the 'ED Source of Referral', 2 were missing 'Mode of Separation', 45 were missing their 'Marital Status', 2 were missing the 'Mode of Arrival', 148 were missing 'Time in ED' and 307 were missing 'Referred to on departure'. The overall proportion of missing data is relatively small. However, handling the missing data carefully is still important, as even small amounts of missing data can impact the study results (Kang 2013). The missing data in this instance can be defined as missing completely at random (MCAR), with the most straightforward approach being listwise or case deletion and common practice in cases such as this (Kang 2013). This approach resulted in the entire missing record being removed from the dataset, resulting in the final record count being 26166, a loss of 515 records (1.93%). The sites' (Site A and Site B) differences were calculated with a p-value < 0.05 being considered statistically significant.

2.5 | Defining Presenting Problem

The presenting problem is used to establish the MH context of the presentation. The 'presenting problem' refers to a free-text section in the electronic medical record where a patient's symptoms or medical condition is described by the individual to the clinician during emergency department triage (Bureau of Health Information 2022). While there have been calls to standardise this input (Berendsen Russell, Dinh, and Bell 2017), no clear definitions or explanations are available in the literature. However, a precedent has been set in previous works; Wand et al. (2015) used the presenting problem field to examine the role of Nurse Practitioners in EDs. Likewise, Sara et al. (2022) and Sperandei et al. (2022) used the presenting problem field to extract ED presentation data relating to self-harm.

2.6 | Data Analysis

The Central Coast Local Health District (CCLHD) comprises four hospitals, two of which contain public EDs. These two sites represent different sociodemographic profiles and public accessibility. Due to these differences, the data between the two sites were compared with Chi squared tests with statistical

14470549, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/inm.13414 by Cikwe Higgins - The University Of Newcastle , Wiley Online Library on [16/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14470349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/inm.13414 by Oliver Higgins - The University Of Newcastle , Wiley Online Library on [16/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 | Variables.

Item	Description
Age	Numerical value indicating the persons' age at the time of presentation
Gender	Gender reported at time of presentation
First Nations peoples' status	A person disclosed if they identify as an 'Aboriginal and/or Torres Strait Islander' person status at time of presentation
Time of day	The time of day, 0–23, that a person presented to ED
Day of week	The day in which a person presented, Monday through Sunday
Month	The month of the presentation, January to December
Mode of arrival	Police, ambulance, etc.
Triage category	NSW Health Triage Category's indicating urgency of care
	Resuscitation (1)
	Emergency (2)
	Urgent (3)
	Semi-urgent (4)
	Non-urgent (5)
Presenting problem	Altered mental status
	Anxious
	Behavioural disturbance
	Depression
	Eating disorder
	Hallucinations
	Memory impairment
	Mental health problem
	Self-harm
	Suicidal ideation
Hospital facility	(Site A/Site B)
Active client	At the time of the presentation to the ED
Previous client	Previous service user preceding the ED presentation back to 1 January 2015

significance that was set at $p < 0.05$. The missing data were removed using case deletion (Kang 2013), resulting in the final record count being 26166, a loss of 515 records (1.93%).

3 | Results

3.1 | Demographics

Overall the total number of presentations to the ED from 2016 to 2021 included in the data was ($n = 26166$). The 25–34 age group represents the largest percentage of presentations overall at 20.30% ($n = 5311$), with the 18–24 age group being very similar at 19.74% ($n = 5094$). However, there is a difference between the two sites for the 18–24 age group ($p = 0.0064$): Site A having 20.15% ($n = 2589$) and Site B 18.81% ($n = 2505$), which is not evident in the 25–34 ($p = 0.2453$) or the 35–44 ($p = 0.1259$) age groups. The 12–17 group represents 15.45% ($n = 4043$), with no statistical significance between Site A and Site B ($p = 0.8777$). The 65–74 age group had a significant difference in the presentations of the two sites ($p = 0.006$), with Site A 2.79% ($n = 359$) to Site B's 3.54% ($n = 471$) (Table 2). *Female* presentations accounted for 50.63% ($n = 13248$), slightly higher than *Male*s at 49.35% ($n = 12912$) (Table 2). Figure 1 reveals intriguing variations in the age distribution of presentations when analysed through the lens of gender. From ages 12 to 17, a considerable proportion of females present to ED, with a peak at age 16. Between ages 18 and 21, the proportion of female presenters declines below that of males in the same age brackets. The number of male presenters surpasses that of females between the ages of 22 and 36. However, beyond the age of 36, the number of presentations maintains a relatively stable pattern for both gender groups (Figure 1).

TABLE 2 | Comparison between Site A and Site B emergency department presentations.

	Site A	%	Site B	%	Total	%	*p*
n	12850	49.11	13316	50.89	26166		
Age							
0–11	234	1.82	198	1.49	432	1.65	0.0340*
12–17	1981	15.42	2062	15.49	4043	15.45	0.8777
18–24	2589	20.15	2505	18.81	5094	19.47	0.0064*
25–34	2646	20.59	2665	20.01	5311	20.30	0.2453
35–44	2228	17.34	2405	18.06	4633	17.71	0.1259
45–54	1587	12.35	1803	13.54	3390	12.96	0.0042*
55–64	847	6.59	921	6.92	1768	6.76	0.2950
65–74	359	2.79	471	3.54	830	3.17	0.0006*
75–84	261	2.03	193	1.45	454	1.74	0.0003*
Over 85	118	0.92	93	0.70	211	0.81	0.0468*
Gender							
Female	6460	50.27	6788	50.98	13248	50.63	0.2549
Male	6386	49.70	6526	49.01	12912	49.35	0.2659
Other	4	0.03	2	0.02	6	0.02	0.3896
Active client							
No	10298	80.14	10885	81.74	21183	80.96	0.0010*
Yes	2552	19.86	2431	18.26	4983	19.04	0.0010*
Previous client							
No	5941	46.23	6541	49.12	12482	47.70	0.0000*
Yes	6909	53.77	6775	50.88	13684	52.30	0.0000*
Presenting problem							
Altered mental status	730	5.68	780	5.86	1510	5.77	0.5401
Anxious	919	7.15	845	6.35	1764	6.74	0.0093*
Behavioural disturbance	1582	12.31	1584	11.90	3166	12.10	0.3025
Depression	911	7.09	860	6.46	1771	6.77	0.0422*
Eating disorder	17	0.13	11	0.08	28	0.11	0.2191
Hallucinations	394	3.07	518	3.89	912	3.49	0.0003*
Memory impairment	18	0.14	10	0.08	28	0.11	0.1080
Mental health problem	2104	16.37	1817	13.65	3921	14.99	0.0000*
Self-harm	1457	11.34	1615	12.13	3072	11.74	0.0473*
Suicidal ideation	4718	36.72	5276	39.62	9994	38.19	0.0000*
Admitted (mode of separation)							
No	10481	81.56	9911	74.43	20392	77.93	0.0000*
Yes	2369	18.44	3405	25.57	5774	22.07	0.0000*

(Continues)

14470349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/inm.13414 by Oliver Higgins - The University Of Newcastle , Wiley Online Library on [16/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14470349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/inm.13414 by Clive Higgins - The University Of Newcastle , Wiley Online Library on [16/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 | (Continued)

	Site A	%	Site B	%	Total	%	*p*
Mode of arrival							
Community/Public transport	34	0.26	3	0.02	37	0.14	0.0000*
No transport (walked in)	180	1.40	10	0.08	190	0.73	0.0000*
Other	64	0.50	46	0.35	110	0.42	0.0565*
Police/Correctional services Vehicle	947	7.37	960	7.21	1907	7.29	0.6180
Private vehicle	5899	45.91	6011	45.14	11 910	45.52	0.2139
State ambulance vehicle	5726	44.56	6286	47.21	12012	45.91	0.0000*
Triage category							
Resuscitation (1)	123	0.96	209	1.57	332	1.27	0.0000*
Emergency (2)	2178	16.95	3541	26.59	5719	21.86	0.0000
Urgent (3)	6046	47.05	4882	36.66	10928	41.76	0.0000*
Semi-urgent (4)	4225	32.88	4482	33.66	8707	33.28	0.1810
Non-urgent (5)	278	2.16	202	1.52	480	1.83	0.0001*
Mode of separation							
Admitted to ward/inpatient unit	2369	18.44	3405	25.57	5774	22.07	0.0000*
Admitted: Other	201	1.56	144	1.08	345	1.32	0.0006*
Departed: Did not wait	947	7.37	807	6.06	1754	6.70	0.0000*
Departed: Left at own risk	393	3.06	522	3.92	915	3.50	0.0001*
Departed: Treatment completed	8503	66.17	8319	62.47	16822	64.29	0.0000*
Departed: Other	437	3.40	119	0.89	556	2.12	0.0000*
Day of week							
Sunday	1742	13.56	1778	13.35	3520	13.45	0.6287
Monday	2054	15.98	2126	15.97	4180	15.97	0.9671
Tuesday	1917	14.92	2036	15.29	3953	15.11	0.4014
Wednesday	1916	14.91	2014	15.12	3930	15.02	0.6279
Thursday	1934	15.05	1907	14.32	3841	14.68	0.0955
Friday	1730	13.46	1840	13.82	3570	13.64	0.4031
Saturday	1557	12.12	1615	12.13	3172	12.12	0.9772

Note: *p* is the comparison between Site A and Site B emergency department presentations.
*Statistically significant.

3.2 | Presentation Characteristics

Approximately, half (52.30%, $n = 13\,684$) of all presentations had received an MH service from CCLHD in the past; however, most people (80.96%, $n = 21\,183$) were not receiving a service from MH at the time of presentation to ED (Table 2). 'Suicidal ideation' was by far the largest 'Presenting problem' reported in the data at 38.19% ($n = 9994$). Site B had a higher proportion of 'Suicidal ideation' (39.62%, $n = 5276$) when compared to Site A (36.72%, $n = 4718$) ($p < 0.00001$), with 'Mental health problem' second at 14.99% ($n = 3166$), 'Behavioural disturbance' third at 12.10% ($n = 3166$) and 'Self-harm' fourth at 11.74% ($n = 3072$) (Table 2). The triage category in ED upon arrival resulted in 'Code 3 (Urgent)' accounting for 41.76% ($n = 10928$), with 'Category 4 (Semi-urgent)' 33.28% ($n = 8707$), 'Category 2 (Emergency)' 21.86% ($n = 5719$), 'Category 5 (Non-urgent)' 1.83% ($n = 480$) and 'Category 1 (Resuscitation)' 1.27% ($n = 332$) (NSW Health 2013) (Table 2). Site B received slightly higher presentations at 50.89% ($n = 13\,316$) compared to Site A (49.11%, $n = 12\,850$) during the date range examined with most presentations, 77.93% ($n = 20\,392$), not admitted to an MH inpatient unit (Table 2).

3.3 | Triage by Presenting Problem

Table 3 provides a cross-table breakdown of the presenting problem by triage code. 'Resuscitation (1)' cases are relatively low across all presenting problems. However, it is important to note that the 'Behavioural disturbance' ($n = 179$) is considerably

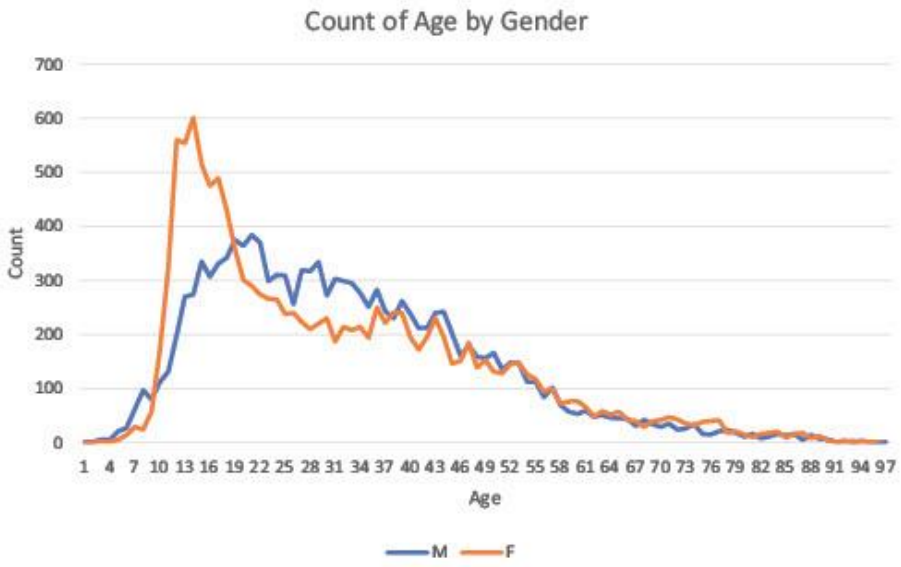


FIGURE 1 | Age distribution.

higher than the other problems at this level of urgency. Overall, the distribution of 'Behavioural Disturbance' presentations is towards the higher urgency end of the triage scale (Table 3). 'Self-harm' and 'Suicidal ideation' have a significant number of cases in the 'Emergency (2)' and 'Urgent (3)' categories; however, the relative number of 'Self-harm' in 'Resuscitation (1)' is proportionately higher than SI (SH, $n=28$, SI, $n=40$) (Table 3). The majority of cases for 'Altered mental status', 'Anxious', 'Depression', 'Eating disorder', 'Hallucinations' and 'Memory impairment' fall under the 'Urgent (3)' or 'Semi-urgent (4)' categories (Table 3).

3.4 | Arrivals and Departures

'State ambulance' accounts for 45.91% ($n=12012$) of presentation mode of arrival; however, 'Private vehicle' was similar to this figure at 45.52% ($n=11910$). When comparing the two sites, the 'State Ambulance' usage is higher at Site B, 47.21% ($n=6286$), than Site A (44.56%, $n=5726$) ($p<0.0001$); however, this difference is not made up in 'Private vehicle' usage with Site A (45.91%, $n=5899$) and Site B (45.14%, $n=6011$) reporting very similar results ($p=0.2139$). Conversely, both 'Community/Public Transport' ($p<0.0001$) and 'No transport (walked in)' ($p<0.0001$) show higher utilisation at Site B (Table 2). 'Police and correctional services' account for 7.29% ($n=1907$) with no statistical significance between the sites ($p=0.6180$); however, it should be noted that if the police require an ambulance to attend and transport a consumer to the hospital, it will be coded as an ambulance in this dataset (Table 2). Most separations were classified as 'Departed: Treatment completed' (64.29%, $n=16822$); however, significant variation occurs between the two sites concerning 'Departed: Did not wait' and 'Departed: Left at own risk'. Site B recorded a 3.92% ($n=522$), 'Left at own risk' compared to Site A's 3.06% ($n=393$) ($p<0.0001$), while the 'Did not wait' were reported as 7.37% ($n=947$) for Site A and 6.06% ($n=807$) for Site B ($p<0.0001$) (Table 2).

3.5 | Time of Day and Length of Stay

The number of presentations is highest during the middle of the day, with 12:00 (6.31%, $n=1650$), followed by 13:00 (6.22%, $n=1627$), 14:00 (6.45%, $n=1689$) and 15:00 (6.02%, $n=1574$). The lowest number of presentations occurs early in the morning, starting at 3:00 (1.61%, $n=421$), 4:00 (1.12%, $n=292$), 5:00 (1.07%, $n=279$) and 6:00 (0.91%, $n=237$) (Table 4). The mean length of time in ED is 224.41 min (SD 165.34), with 23.56% ($n=6166$) departing at the 2–3 h mark. 19.35% ($n=5064$) are between 3 and 4 h and 19.24% ($n=2034$) in between 1 and 2 h with only 5.81% ($n=1519$) departing in less than 1 h. Presentations overall spend less time in the ED at Site A than at Site B. There is significant variation between the two sites. In the 0–1 h mark, 7.20% ($n=925$) of Site A compared to 4.46% ($n=594$) at Site B. At the 1–2 h mark, 20.58% ($n=2644$) Site A and Site B 17.95% ($n=2390$) (Table 4).

4 | Discussion

This retrospective cohort study aimed to analyse the sociodemographic factors and presentation features of individuals seeking MH care in two EDs located in a regional coastal setting of New South Wales, Australia, from 2016 to 2021. A key objective was to identify factors associated with admission to an inpatient mental health facility following an ED presentation for mental health reasons. By comprehensively characterising the population presenting to these regional EDs for MH concerns, the study has elucidated sociodemographic and clinical factors that may precipitate an inpatient MH admission.

4.1 | Arrival

The mode of arrival at the hospital differs significantly between Site B and Site A. Data analysis reveals that the most frequently

14470349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ajsc.13414 by Cochrane Hellas , Wiley Online Library on [16/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 | Triage category by presentation problem.

Presenting problem		Altered mental status	Anxious	Behavioural disturbance	Depression	Eating disorder	Hallucinations	Memory impairment	Mental health Problemp	Self-harm	Suicidal ideation
Resuscitation (1)	*N*	44	1	179			1		39	28	40
	%	0.17	0.00	0.68	0.00	0.00	0.00	0.00	0.15	0.11	0.15
	PP %	2.91	0.06	5.65	0.00	0.00	0.11	0.00	0.99	0.91	0.40
Emergency (2)	*N*	343	18	937	43	3	115		528	771	2961
	%	1.31	0.07	3.58	0.16	0.01	0.44	0.00	2.02	2.95	11.32
	PP %	22.72	1.02	29.60	2.43	10.71	12.61	0.00	13.47	25.10	29.63
Urgent (3)	*n*	585	351	1139	537	7	474	14	1363	1280	5178
	%	2.24	1.34	4.35	2.05	0.03	1.81	0.05	5.21	4.89	19.79
	PP %	38.74	19.90	35.98	30.32	25.00	51.97	50.00	34.76	41.67	51.81
Semi-urgent (4)	*n*	498	1245	863	1163	17	311	14	1822	977	1797
	%	1.90	4.76	3.30	4.44	0.06	1.19	0.05	6.96	3.73	6.87
	PP %	32.98	70.58	27.26	65.67	60.71	34.10	50.00	46.47	31.80	17.98
Non-urgent (5)	*n*	40	149	48	28	1	11		169	16	18
	%	0.15	0.57	0.18	0.11	0.00	0.04	0.00	0.65	0.06	0.07
	PP %	2.65	8.45	1.52	1.58	3.57	1.21	0.00	4.31	0.52	0.18
Total	*n*	1510	1764	3166	1771	28	912	28	3921	3072	9994

Note: % = The overall percentage for all presentations, PP % = The percentage based upon the presenting problem.

14470349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/inm.13414 by Curtin University , Wiley Online Library on [16/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14470349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/imj.13414 by Clive Higgins - The University Of Newcastle , Wiley Online Library on [16/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 4 | Hour of presentation.

Hour		Site A	%	Site B	%	Total	%	*p*
0	0:00	533	4.15	519	3.90	1052	4.02	0.3028
1	1:00	383	2.98	354	2.66	737	2.82	0.1154
2	2:00	287	2.23	293	2.20	580	2.22	0.8557
3	3:00	207	1.61	214	1.61	421	1.61	0.9805
4	4:00	144	1.12	148	1.11	292	1.12	0.9437
5	5:00	137	1.07	142	1.07	279	1.07	0.9985
6	6:00	107	0.83	130	0.98	237	0.91	0.2204
7	7:00	142	1.11	165	1.24	307	1.17	0.3141
8	8:00	219	1.70	254	1.91	473	1.81	0.2174
9	9:00	382	2.97	429	3.22	811	3.10	0.2454
10	10:00	595	4.63	637	4.78	1232	4.71	0.5582
11	11:00	752	5.85	717	5.38	1469	5.61	0.1004
12	12:00	816	6.35	834	6.26	1650	6.31	0.7721
13	13:00	765	5.95	862	6.47	1627	6.22	0.0816
14	14:00	800	6.23	889	6.68	1689	6.45	0.1382
15	15:00	751	5.84	823	6.18	1574	6.02	0.2529
16	16:00	795	6.19	828	6.22	1623	6.20	0.9164
17	17:00	800	6.23	785	5.90	1585	6.06	0.2625
18	18:00	753	5.86	817	6.14	1570	6.00	0.3481
19	19:00	711	5.53	666	5.00	1377	5.26	0.0542
20	20:00	769	5.98	728	5.47	1497	5.72	0.0717
21	21:00	754	5.87	723	5.43	1477	5.64	0.1247
22	22:00	679	5.28	755	5.67	1434	5.48	0.1704
23	23:00	569	4.43	604	4.54	1173	4.48	0.6733

reported mode of arrival at the hospital is by 'State ambulance', with the 'Private vehicle' being a close second. By contrast, 52.2% of NSW MH ED presentations in 2020–2021 were recorded as ambulance (Australian Institute of Health and Welfare 2022). Further examination of the data from both sites reveals that Site B has a higher rate of ambulance presentations compared to Site A. It is important to note that this difference is not explained by variations in private vehicle access between the two sites ($p = 0.2139$). Further examination of the mode of arrival reveals a statistical significance in the use of community/public transport to attend ED with more *walk-ins* at Site A compared to Site B. Site B hospital is located some distance from public transport hubs and is only served by limited public transport options after hours or on weekends. In contrast, Site A Hospital is located close to a local train station and transport hub, making it more accessible for consumers. This difference serves as a possible indicator of poor public access to the hospital in Site B, raising concerns about the utilisation of ambulance services in the area and underscoring the importance of addressing the issue of poor public transport and its impact on access to healthcare in the region (The Independent Pricing and Regulatory Tribunal of NSW 2015). However, it should be noted that further analysis is required to assess if ambulance presentations are being triaged as more urgent. If so, then their use may be appropriate and not related to the availability of transport.

Consumers are spending an increased amount of time in EDs (Australian Institute of Health and Welfare 2021), resulting in a decrease in the number of visits completed within 4 h. The mean length of stay across both EDs for MH problems is 224.41 min (3:44) which is comparable to the NSW (2020–2021) average of 228 min (3:48) (Australian Institute of Health and Welfare 2022) and Wardrop et al. (2023) reported median of 237 min for those admitted; however, Wardrop et al. (2023) also found a shorter median of 167 min for those not admitted. However, when comparing overall ED presentations for 20–21 (not just MH), Site A is 3:53, Site B is 3:16 and NSW 2:51 (Australian Institute of Health and Welfare 2021). The AIHW data show that the MH consumer spent, on average, 57 min more in the ED than those with non-MH presentations (Australian Institute of Health and Welfare 2022). The literature indicates that increased length of stay data indicates high demand for MH care and a lack of hospital-based facilities (Duggan et al. 2020). The length of stay

data revealed in the current study indicates high demand and lack of community resources on the Central Coast. Duggan et al. (2020) suggest that the whole hospital system needs to adopt fresh care models, which encompass the management of urgent care requirements and discharge and referral procedures. The implementation of comprehensive and cohesive services, along with consistent interventions aimed at addressing the intricate and interdependent MH, physical health and social care necessities of those with continuous MH requirements, would decrease the occurrence of emergency hospital visits and establish a straightforward route for emergency care (Duggan et al. 2020).

4.2 | Presenting Problem

From the results of this investigation, the most common presenting problem is 'Suicidal ideation', followed by 'Mental health problem', 'Behavioural disturbances' and 'Self-harm'. It is worth noting that Site B had a significantly higher proportion of consumers presenting with suicidal ideation than Site A. Perera et al. (2018) conducted a similar investigation that focused on presentations to NSW EDs with self-harm and suicidal ideation, which covered 2010–2014. They reported 331 493 MH presentations to ED during the period, 32 663 (9.85%) were reported as 'Suicidal ideation and behaviour' and 'Self-harm' of 16 215 (4.89%). Despite the differences in data collection, the Central Coast data show a more than double incidence of suicidal ideation and self-harm. Further exploration is needed to see if this is a local feature due to increased distress over time in the community or is more widely spread. It should be noted that the data collection methods used are ICD-10 or SNOMED codes for ED diagnosis rather than the triage presenting problem (Perera et al. 2018). Coding practices and data collection methods may underestimate SI and SH and could account for much of the difference. Therefore, further analysis is required to verify if an actual difference in rates exists between regions after accounting for data collection methods. If a true difference is established, then contributing factors specific to the local community should be examined, such as changes over time in risk factors like substance abuse, economic distress and social isolation. Improved monitoring and standardised data collection are recommended to enhance the observation of MH presentations in the ED setting.

4.3 | Admissions and Departures

The analysis of the hospital presentations reveals that a significant proportion of presentations resulted in admission to a public inpatient facility (excluding Critical Care wards, ICU or admissions that were transferred); however, the reported figure is lower than the 2020–21 NSW admission rate for MH presentations to ED of 27.8% (Australian Institute of Health and Welfare 2022). It is important to note that there is a statistically significant difference between the admission rates at Site A and Site B; it is important to consider that Site B has a more extensive MH bed base than Site A (Central Coast Local Health District 2017), which may explain the higher admission rate at Site B. The findings of this analysis highlight the need for further examination into the factors contributing to the difference in admission rates between the two sites. It may be necessary to consider the availability of MH facilities and resources and the potential impact of the MH bed base in either Site B or Site A. While most people departed without being admitted, numbers were recorded as *Not referred* (which means possibly no further action was provided), *Departed*, *Did not wait* and *Departed*, *Left at own risk*, which is significantly different from that reported by NSW Health in 2020–21 at 0.7% and 4.5%, respectively (Australian Institute of Health and Welfare 2022) for MH presentations to ED and higher than Wardrop et al. (2023) 'Left at own risk' at 2.8% and 'Did not wait' at 1.1%. This is a significant finding and warrants further exploration.

5 | Recommendations

5.1 | Recommendation 1: Promoting Safety Planning to Address Suicidal Ideation

Suicide prevention therapeutic interventions (SPTI), such as safety planning, have been shown to prevent suicidal behaviour and should be included in clinical practice (Ferguson et al. 2022; Nuij et al. 2021). The promotion of SPTI offers an opportunity to facilitate risk mitigation arising from the large proportion of people who report the potentially life-threatening condition of suicidal ideation (38.19%). Furthermore, safety planning is aligned with supporting 10.2% of people who attend the ED for an MH condition but depart without waiting to be seen by a clinician or who chose to leave at their own risk after presentation. Further research should be conducted to determine if SPTI can improve MH outcomes for individuals.

5.2 | Recommendation 2: Improving Service Availability and ED Accessibility

The data presented provide evidence of ED access issues in the use of the NSW ambulance service as well as the high proportion of people who have received prior MH treatment. As ambulance and private vehicle transportation are the most common modes of arrival at the hospital, clinicians should advocate for improved alternative transportation options for individuals with MH needs, particularly in areas with poor public transport. Further data analysis is required to understand ambulance usage in the context of location, urgency and the problem presenting itself. Furthermore, the high proportion of individuals who have received prior MH treatment among consumers emphasises the criticality of early intervention and community support. Specifically, understanding their connection to hospital admissions, presenting concerns such as suicidal thoughts, and factors influencing decisions to leave at their own risk or not wait. Given the substantial number of presentations involving individuals with previous contact or current service utilisation, alongside numerous social-related cases, additional research is necessary to comprehend the characteristics of these individuals and understand the gap in community healthcare delivery. Likewise, future research of service delivery innovations such as telehealth, virtual care and hospital-in-the-home models should prioritise the incorporation of lived experiences to provide a platform for those who utilise and engage with MH services.

14470349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/inm.13414 by Chloe Higgins - The University Of Newcastle , Wiley Online Library on [16/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

5.3 | Recommendation 3: Integration of Machine Learning, Decisional Tools and Innovation

The large volume of data and the complexity of every presentation require further analysis to understand the intricate interactions. Systems need to be developed to promote the identification of care requirements and forecast the appropriate care trajectory. With the significant increase in the demand for care, clinicians must turn towards tools to enhance and support decision-making (Higgins et al. 2023). Recent advances in Machine Learning (ML) may offer the opportunity to discover critical factors and complex interactions for people experiencing mental health problems. It is recommended that further research be conducted using ML to aid the discovery of service delivery complexity and promote clinical decision support when working with vulnerable populations (Gregor et al. 2023).

6 | Limitations

Several limitations to this study should be noted. First, a unique person identifier was not included in the dataset under analysis, which made it impossible to establish if a person had presented once or multiple times. The presenting problem data captured the primary problem and did not allow for complex interactions or diagnosis. The data did not include the legal status of ambulance and police presentations. The study did not include after-hours MH presentations that did not go directly to the ED, such as police arriving directly to the MH unit or direct community admissions. Due to the way that the data are coded and sourced, not all MH presentations may be coded correctly and may not be present in the data. Likewise, there may be instances where an individual may present and be coded as an MH problem initially but require non-mental healthcare, and the MH team may not have direct input while they are in ED, such as for a person presenting with intoxication. Finally, it is essential to note that data input and coding may have varied from site to site, which could have introduced inconsistencies in the data. These limitations should be considered when interpreting the findings of this study and may impact the generalisability of the results.

7 | Conclusion

In conclusion, investigating MH-related ED presentations on the Central Coast of NSW, Australia, has revealed several insights. Dominant help-seeking dynamics for MH conditions in ED are characterised by two main themes: Suicidal ideation and access and egress pathways. A high proportion of consumers had previously accessed MH services, demonstrating the high demand within the community for MH services, with the most common presenting problem being suicidal ideation. Finally, there are significant differences in the mode of arrival at the hospital between Site B and Site A, with Site B receiving more ambulance presentations, highlighting possible access issues for an already disadvantaged population. Future research should focus on three distinct themes: The high prevalence of suicidal ideation and the significant risk associated with departures from emergency departments underscores the importance of implementing safety plans for individuals in suicide-related distress across different clinical settings and delivery methods. Second, integrative and holistic models of care target primary health and social care providers together with public community MH services in response to the identified gaps in service delivery for people seeking MH care in ED need to be developed. Finally, the complexity and volume of data in mental healthcare necessitate further analysis and the development of ML systems to identify care requirements, forecast care trajectory, enhance decision-making, and discover critical factors and interactions, thereby promoting clinical decision support and service delivery understanding.

Author Contributions

O.H. – concept development, project design, data collection, data analysis, manuscript preparation. R.B.S.-R. – data analysis, contribution to manuscript. S.K.C. – data analysis, contribution to manuscript, supervision of project. R.L.W. – concept development, project design, data collection, data analysis, manuscript contribution, supervision of project.

Acknowledgements

Authors would like to acknowledge the support of Central Coast Local Health District. Open access publishing facilitated by RMIT University, as part of the Wiley - RMIT University agreement via the Council of Australian University Librarians.

Ethics Statement

Ethical approval was granted by the Hunter New England Human Research Ethics Committee 2022/ETH01597. Central Coast Local Health District granted site-specific approval 2022/STE03296.

Conflicts of Interest

Prof. R.L.W. is an Editorial Board Member of International Journal of Mental Health Nursing.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

Australian Bureau of Statistics. 2021a. "2021 Census All Persons QuickStats Central Coast (NSW)." https://abs.gov.au/census/find-census-data/quickstats/2021/LGA11650.

Australian Bureau of Statistics. 2021b. "Estimates of Aboriginal and Torres Strait Islander Australians." *ABS*. https://www.abs.gov.au/statistics/people/aboriginal-and-torres-strait-islander-peoples/estimates-aboriginal-and-torres-strait-islander-australians/latest-release.

Australian Institute of Health and Welfare. 2021. "Emergency Department Care." https://www.aihw.gov.au/reports-data/myhospitals/sectors/emergency-department-care.

Australian Institute of Health and Welfare. 2022. "Mental Health Services in Australia." https://www.aihw.gov.au/reports/mental-health-services/mental-health-services-in-australia/report-contents/emergency-department-mental-health-services.

Berendsen Russell, S., M. M. Dinh, and N. Bell. 2017. "Triage, Damned Triage... and Statistics: Sorting out Redundancy and Duplication Within an Emergency Department Presenting Problem Code Set to Enhance

14470349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/inm.13414 by Oliver Higgins - The University Of Newcastle , Wiley Online Library on [16/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Research Capacity." *Australasian Emergency Nursing Journal* 20, no. 1: 48–52. https://doi.org/10.1016/j.aenj.2016.09.004.

Bureau of Health Information. 2022. "Technical Supplement Emergency Department, Ambulance, Admitted Patients, Seclusion and Restraint, and Elective Surgery." Bureau of Health Information. https://www.bhi.nsw.gov.au/__data/assets/pdf_file/0009/715194/BHI_HQ47_OCT-DEC_2021_TECH_SUPP.pdf.

Central Coast Local Health District. 2017. "Clinical Services Plan 2017–2022." https://www.cclhd.health.nsw.gov.au/wp-content/uploads/CCLHDClinicalServicesPlan2017-2022.pdf.

Clarke, D. E., D. Dusome, and L. Hughes. 2007. "Emergency Department From the Mental Health Client's Perspective." *International Journal of Mental Health Nursing* 16, no. 2: 126–131. https://doi.org/10.1111/j.1447-0349.2007.00455.x.

Digel Vandyk, A., L. Young, C. MacPhee, and K. Gillis. 2018. "Exploring the Experiences of Persons Who Frequently Visit the Emergency Department for Mental Health-Related Reasons." *Qualitative Health Research* 28, no. 4: 587–599. https://doi.org/10.1177/1049732317746382.

Duggan, M., B. Harris, W. K. Chislett, and R. Calder. 2020. "Nowhere Else to Go: Why Australia's Health System Results in People With Mental Illness Getting 'Stuck' in Emergency Departments." Mitchell Institute Commissioned Report 2020, Victoria University. https://vuir.vu.edu.au/41956/1/nowhere-else-to-go-people-mental-illness-stuck-emergency-departments-report-mitchell-institute.pdf.

Ferguson, M., K. Rhodes, M. Loughhead, H. McIntyre, and N. Procter. 2022. "The Effectiveness of the Safety Planning Intervention for Adults Experiencing Suicide-Related Distress: A Systematic Review." *Archives of Suicide Research* 26, no. 3: 1022–1045. https://doi.org/10.1080/13811118.2021.1915217.

Fonseca Barbosa, J., and J. Gama Marques. 2023. "The Revolving Door Phenomenon in Severe Psychiatric Disorders: A Systematic Review." *International Journal of Social Psychiatry* 69, no. 5: 1075–1089. https://doi.org/10.1177/00207640221143282.

Góngora Alonso, S., I. Herrera Montano, J. L. M. Ayala, J. J. P. C. Rodrigues, M. Franco-Martín, and I. de la Torre Díez. 2023. "Machine Learning Models to Predict Readmission Risk of Patients With Schizophrenia in a Spanish Region." *International Journal of Mental Health and Addiction*. https://doi.org/10.1007/s11469-022-01001-x.

Gregor, S., M. Gunawardena, A. Imran, S. Okai-Ugbaje, C. Page Jeffery, and R. Wilson. 2023. "Indigenous Knowledge-Sharing Interventions in Australia and the Use of Information and Communication Technology: A Scoping Review." *Australasian Journal of Information Systems* 27. https://doi.org/10.3127/ajis.v27i0.4217.

Higgins, O., B. L. Short, S. K. Chalup, and R. L. Wilson. 2023. "Artificial Intelligence (AI) and Machine Learning (ML) Based Decision Support Systems in Mental Health: An Integrative Review." *International Journal of Mental Health Nursing* 32: 966–978. https://doi.org/10.1111/inm.13114.

Kang, H. 2013. "The Prevention and Handling of the Missing Data." *Korean Journal of Anesthesiology* 64, no. 5: 402–406. https://doi.org/10.4097/kjae.2013.64.5.402.

Liamputtong, P. 2021. *Research Methods and Evidence-Based Practice*. Docklands, Australia: Oxford University Press Australia & New Zealand.

Morphet, J., K. Innes, I. Munro, et al. 2012. "Managing People With Mental Health Presentations in Emergency Departments—A Service Exploration of the Issues Surrounding Responsiveness From a Mental Health Care Consumer and Carer Perspective." *Australasian Emergency Nursing Journal* 15, no. 3: 148–155. https://doi.org/10.1016/j.aenj.2012.05.003.

NSW Health. 2013. "Triage of Patients in NSW Emergency Departments." https://www1.health.nsw.gov.au/pds/ActivePDSDocuments/PD2013_047.pdf.

NSW Health. 2019. "Declared Mental Health Facilities." https://www.health.nsw.gov.au/mentalhealth/services/consumers/Pages/facilities.aspx.

NSW Health. 2021. "Emergency Department Care." https://www.health.nsw.gov.au/Performance/Pages/emergency.aspx.

NSW Health. 2023. "High or Very High Psychological Distress in Adults by LHD." https://www.healthstats.nsw.gov.au/#/indicator?name=-men-hidistress-phs&location=LHD&view=Trend&measure=prevalence&groups=&compare=LHD&filter=LHD,All%20LHDs,Central%20Coast%20LHD.

Nuij, C., W. van Ballegooijen, D. de Beurs, et al. 2021. "Safety Planning-Type Interventions for Suicide Prevention: Meta-Analysis." *British Journal of Psychiatry* 219, no. 2: 419–426. https://doi.org/10.1192/bjp.2021.50.

Perera, J., T. Wand, K. J. Bein, et al. 2018. "Presentations to NSW Emergency Departments With Self-Harm, Suicidal Ideation, or Intentional Poisoning, 2010–2014." *Medical Journal of Australia* 208, no. 8: 348–353. https://doi.org/10.5694/mja17.00589.

Sara, G., J. Wu, J. Uesi, et al. 2022. "Growth in Emergency Department Self-Harm or Suicidal Ideation Presentations in Young People: Comparing Trends Before and Since the COVID-19 First Wave in New South Wales, Australia." *Australian and New Zealand Journal of Psychiatry* 57, no. 1: 58–68. https://doi.org/10.1177/00048674221082518.

Sperandei, S., A. Page, P. Bandara, et al. 2022. "The Impact of the COVID-19 Pandemic on Hospital-Treated Self-Harm in Sydney (Australia)." *Australian and New Zealand Journal of Psychiatry* 56, no. 11: 1515–1522. https://doi.org/10.1177/00048674211068393.

The Independent Pricing and Regulatory Tribunal of NSW. 2015. "Efficiency of NSW Public Transport Services." C. f. I. Economics. https://www.ipart.nsw.gov.au/sites/default/files/documents/consultant_report_-_cie_-_efficiency_of_nsw_public_transport_services_-_december_2015.pdf.

Vigod, S. N., P. A. Kurdyak, C.-L. Dennis, et al. 2013. "Transitional Interventions to Reduce Early Psychiatric Readmissions in Adults: Systematic Review." *British Journal of Psychiatry* 202, no. 3: 187–194. https://doi.org/10.1192/bjp.bp.112.115030.

Wand, T., N. D'Abrew, C. Barnett, L. Acret, and K. White. 2015. "Evaluation of a Nurse Practitioner-Led Extended Hours Mental Health Liaison Nurse Service Based in the Emergency Department." *Australian Health Review* 39, no. 1: 1–8. https://doi.org/10.1071/AH14100.

Wardrop, R., J. Ranse, W. Chaboyer, J. T. Young, S. A. Kinner, and J. Crilly. 2023. "Profile and Outcomes of Emergency Department Mental Health Patient Presentations Based on Arrival Mode: A State-Wide Retrospective Cohort Study." *Journal of Emergency Nursing* 49, no. 6: 951–961. https://doi.org/10.1016/j.jen.2023.06.015.

14470349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/inm.13414 by Cibrel Higgins - The University Of Newcastle , Wiley Online Library on [16/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

7.2.2 Article Summary

The findings of this study have significant implications for improving MH care delivery and addressing the challenges faced by individuals seeking care in EDs. The high prevalence of suicidal ideation and the significant proportion of individuals departing without being seen or against medical advice highlights the urgency of implementing safety planning interventions and improving access to specialised MH services.

Furthermore, the study underscores the need for a more integrated and holistic approach to MH care, bridging the gap between primary care, community-based services, and EDs. By understanding the factors contributing to ED presentations and the associated outcomes, healthcare providers and policymakers can develop targeted interventions and service delivery models that better meet the needs of individuals with MH conditions.

Future research should focus on several key areas:

- Evaluating the effectiveness of safety planning interventions and their implementation across different clinical settings and delivery methods for individuals presenting with suicidal ideation or self-harm.
- Exploring integrative and collaborative models of care that involve primary health and social care providers, public community MH services, and EDs to improve continuity of care and address gaps in service delivery.
- Leveraging ML and predictive modelling techniques to identify critical risk factors, forecast care trajectories, and enhance clinical decision support for MH presentations in EDs.

- Examining the impact of socioeconomic factors, access to transportation, and geographical barriers on MH service utilisation and outcomes in regional and rural areas.

In summary, this study provides a detailed examination of the cohort seeking MH care in regional EDs, identifying the critical challenges these individuals face and the gaps in service delivery. The purpose of this analysis was to describe the population that was then used to build the ML model, which aimed to predict factors contributing to inpatient admissions. By addressing these challenges and leveraging advanced analytical techniques, healthcare systems can better meet the needs of vulnerable populations and improve MH outcomes. The following section builds on this foundation by using the same dataset to further explore the help-seeking behaviours of local Central Coast Aboriginal and Torres Strait Islander people in relation to their MH care needs.

7.3 Disproportionate Mental Health Presentations to Emergency Departments in a Coastal Regional Community in Australia of First Nation People

Higgins, O., Sheather-Reid, R. B., Chalup, S. K., & Wilson, R. L. (2024a). Disproportionate mental health presentations to emergency departments in a coastal regional community in Australia of First Nation people. *International Journal of Mental Health Nursing*. Advance online publication. https://doi.org/10.1111/inm.13362

	Higgins, O	Sheather-Reid, R. B	Chalup, S. K	Wilson, R. L
Conceptualization	X		X	X
Methodology	X		X	X
Software	X			
Validation	X	X		X
Formal analysis	X	X		X
Investigation	X	X		
Resources	X			
Data Curation	X			
Writing - Original Draft	X			

Writing - Review & Editing	X	X	X	X
Visualization	X			
Supervision			X	X
Project administration	X			
Funding acquisition				
Percentage	**70%**	**10%**	**5%**	**15%**

7.3.1 Article Introduction

The provision of MH care for Aboriginal and Torres Strait Islander communities in Australia remains an area of significant concern, marked by inadequacy and inequity. These First Nations populations confront a disproportionate burden of MH challenges, with higher rates of suicide, deliberate self-harm, and psychological distress compared to their non-Indigenous counterparts. Despite utilising some MH services at elevated rates, researchers and advocates assert that the delivery of culturally appropriate and accessible care continues to fall short. EDs play a crucial role in providing MH care, yet they present numerous challenges for individuals experiencing MH conditions. The fast-paced, clinical environment can exacerbate distress, leading to long wait times, overcrowding, and delays in treatment. These issues contribute to frustration, dissatisfaction, and adverse safety outcomes among consumers, highlighting the need for improved infrastructure, resources, and specialised knowledge in EDs.

Building upon the broader Study A, which examined sociodemographic factors of individuals seeking MH care in two regional EDs, this section narrows the focus to investigate MH service utilisation among the Aboriginal and Torres Strait Islander populations in the Central Coast region of New South Wales, Australia. This refined analysis aimed to understand the patterns of ED usage and the representation of First Nations individuals, a group that experiences significant disparities in MH outcomes (Australian Institute of Health and Welfare, 2022b). By examining the demographic profiles and service usage patterns, this research seeks to inform the development of culturally responsive care models and service delivery strategies that

address the unique needs of First Nations communities in the Central Coast region. The disproportionate burden of MH challenges faced by Aboriginal and Torres Strait Islander communities underscores the importance of this research. Existing disparities in access to care and culturally safe healthcare environments highlight the need for a deeper understanding of service usage patterns and the factors influencing ED utilisation among First Nations populations. This study drew upon theoretical perspectives that recognise the impact of cultural disruptions, intergenerational trauma, and the inherent biases within Western models of MH service delivery. By acknowledging these factors, the research aimed to contribute to the development of holistic, culturally enriched approaches that align with Indigenous concepts of social and emotional well-being.

Check for updates

Received: 29 February 2024 | Accepted: 12 May 2024

DOI: 10.1111/inm.13362

ORIGINAL ARTICLE

International Journal of Mental Health Nursing

Disproportionate mental health presentations to emergency departments in a coastal regional community in Australia of first nation people

Oliver Higgins[1,2] | Rachel B. Sheather-Reid[2] | Stephan K. Chalup[3] | Rhonda L. Wilson[1,2]

[1]RMIT University, Melbourne, Victoria, Australia

[2]Central Coast Local Health District, Gosford, New South Wales, Australia

[3]School of Information and Physical Sciences (Computer Science and Software Engineering), University of Newcastle, Newcastle, New South Wales, Australia

Correspondence
Oliver Higgins, C/O Rhonda Wilson, Central Coast Research Institute, 77a Holden St, Gosford, NSW 2250, Australia.
Email: oliver.higgins@newcastle.edu.au

Abstract

Emergency department (ED) presentations for mental health (MH) help-seeking have been rising rapidly in recent years. This research aims to identify the service usage demographic for people seeking MH care in the ED, specifically in this case, to understand the usage by First Nation people. This retrospective cohort study examined the sociodemographic and presentation characteristics of individuals seeking MH care in two EDs between 2016 and 2021. Data were collected using existing records and analysed using descriptive univariate analysis with statistical significance between the two sites determined using chi-squared test, $p<0.05$. The overall data presented in this analysis show an overall ED mental health presentation rate of 12.02% for those who identified as 'Aboriginal but not Torres Strait Islander origin', 0.36% as 'Both Aboriginal and Torres Strait Islander' and 0.27% as 'Torres Strait Islander' totalling 12.63%. This is an overrepresentation compared to the regional population of 4.9%. One site recorded 14.1% of ED presentations that identified as *Aboriginal and/or Torres Strait Islander*, over double the site's demographic of 6.3%. Given the disproportionately high representation of First Nation people in MH-related ED presentations, further research is required to prioritise a First Nation research perspective that draws on First Nation research methods, such as yarning and storytelling to understand the unique cultural needs and challenges experienced by First Nation people accessing MH care via ED. Understanding the demographic is but one step in supporting the Cultural Safety needs of First Nation people. Additionally, research should be designed, governed and led by First Nation researchers.

KEYWORDS

Aboriginal, emergency department, first nation, mental health, Torres Strait Islander

INTRODUCTION

Colonisation has disrupted traditional ways of life, eroded cultural practices and undermined social structures within First Nation communities (Paradies, 2016). This historical trauma has contributed to enduring disparities in health, education, employment and overall well-being experienced by First Nation populations (Paradies, 2016). Moreover, colonisation has left a legacy of intergenerational trauma, systemic discrimination and ongoing socio-economic disadvantage that continues to shape First Nation experiences and outcomes in contemporary society (Smallwood et al., 2021). Recognising and addressing these enduring impacts of colonisation is essential for advancing reconciliation, promoting social justice and fostering meaningful improvements in the lives of First Nation people (Taylor

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2024 The Author(s). *International Journal of Mental Health Nursing* published by John Wiley & Sons Australia, Ltd.

14470349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/inm.13362 by National Health And Medical Research Council, Wiley Online Library on [24/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

& Habibis, 2020). Additionally, First Nation communities confront a disproportionate burden of mental health (MH) challenges compared to their non-First Nation counterparts in Australia (Australian Institute of Health and Welfare, 2023). Suicide rates among First Nation populations are nearly twice as high, hospitalisation rates for deliberate self-harm are three times greater, and the prevalence of high/very high psychological distress is 2.4 times higher when compared to non-First Nation Australians (Australian Institute of Health and Welfare, 2023). From 2009–2010 to 2018–2019, the hospitalisation rate for MH-related conditions among First Nation Australians escalated by 52% (from 19 to 29 per 1000 hospitalisations) (Australian Institute of Health and Welfare, 2023). This increase was even more pronounced among First Nation females, with a notable 58% rise, and among First Nation males, it increased by 46% (Australian Institute of Health and Welfare, 2023). While First Nation Australians utilise some MH services at elevated rates compared to non-First Nation Australians, gauging whether this utilisation aligns with the underlying need remains a complex task. First Nation MH researchers and advocates assert that the provision of MH care for First Nation communities continues to fall short, marked by inadequacy and inequity (Dudgeon et al., 2020).

Emergency departments (ED) are complex and challenging environments for people with mental health conditions. They are excessively loud, busy, crowded and clinical, operating 24h per day with high levels of activity, exacerbating distress for many people (Clarke et al., 2007). The high volume of presentations to ED with MH problems can lead to long wait times, overcrowding and delays in treatment for those who present (Duggan et al., 2020), which can result in frustration and dissatisfaction among consumers (Digel Vandyk et al., 2018; Morphet et al., 2012). Compared to consumers presenting with other conditions, MH consumers experience significantly longer lengths of stay and wait times for admission to an inpatient bed in the ED (Duggan et al., 2020). General care emergency clinicians report a lack of strategies and knowledge to care for people requiring MH interventions (Derblom et al., 2021). The triage processes are crucial in ensuring that consumers with MH problems receive timely and appropriate care, but many EDs struggle to implement effective triage systems (Morphet et al., 2012).

The prolonged waiting periods for MH consumers have been linked with adverse safety and clinical consequences. These include heightened mortality rates, increased incidents of violence and aggression in the ED and instances of patients leaving before receiving care. Consequently, this contributes to a perception among consumers that their needs have not been acknowledged (Duggan et al., 2020). This issue has also been documented in research focusing on young First Nation people, who reported feeling unheard or not listened to (Cerreto, 2018). Despite the efforts to improve the consumer journey, many MH consumers still report unsatisfactory experiences in the ED, including long wait times, lack of specialised care and inappropriate treatment settings (Digel Vandyk et al., 2018; Morphet et al., 2012). Overall, the difficulties MH presentations pose in EDs highlight the need for improved infrastructure, resources and knowledge to offer specialised and timely care to persons with MH problems (Digel Vandyk et al., 2018).

The Central Coast of New South Wales is located close to Sydney, Australia. According to the Australian Bureau of Statistics (2021a), 4.9% of the population within the Central Coast identify as 'Aboriginal and/or Torres Strait Islander peoples', compared to 4.2% in NSW and 3.8% in Australia (Australian Bureau of Statistics, 2021e). The Central Coast experiences poorer health outcomes compared to both NSW and Australia, with a higher prevalence of MH conditions such as anxiety and depression, at 11.0%, compared to 8.0% in NSW and 8.8% in Australia (Australian Bureau of Statistics, 2021a) and higher than average increases in high psychological distress over 2015–2021 (NSW Health, 2023). Public health services require robust evidence outlining the response and processes associated with MH presentations in EDs. This evidence is crucial for informed planning and the effective and efficient delivery of services that meet the needs of the catchment population, all while staying within budgetary and resource constraints.

This research aims to identify the service usage demographic for people seeking MH care in the emergency departments, specifically in this case, to understand the usage by First Nation people.

The ED is often the only place available 24h a day for people experiencing mental ill health to seek help. However, many First Nation people report feeling unsafe when accessing health care due to actions undertaken by the Australian government since colonisation, such as the removal of First Nation children from their families (Ware, 2013). Significant barriers exist in access to healthcare for First Nation people in Australia, resulting in a frustrating experience for many First Nation people (Nolan-Isles et al., 2021), less positive health outcomes (Gwynne et al., 2018) and a likelihood of accessing services later in the disease process if at all (Campbell et al., 2018). Therefore, it is imperative that the service usage demographic for people seeking MH care in the ED, specifically in this case, First Nation people, is understood. This research will further enable the development of care and service delivery models that address cultural needs.

METHODS

Design and sample

The study elucidates the demographic profile of individuals seeking Mental Health (MH) care at Site A and Site B Emergency Departments (EDs) from 1 January

2016, to 30 December 2021. Employing a retrospective cohort design, existing data collected in routine care were utilised. The study adheres to a cross-sectional descriptive epidemiological method outlined by Liamputtong (2021, p. 212). This methodology was selected to provide a systematic framework for characterising the cohort in terms of temporal, spatial and demographic aspects (Liamputtong, 2021, p. 212). A descriptive univariate analysis was conducted to scrutinise individual elements.

Data collection and ethical considerations

The initial data collection process involved the review of pre-existing data spanning from 1 January 2016, to 30 December 2021, obviating the need for participant recruitment. Relevant ED data were identified by applying the following criteria to the presenting problem field: '%suici%' OR '%MH%' OR '%self harm%' OR '%psych%', OR '%Mental%'. A subsequent data collection phase was undertaken to extend the data set further and establish connections between ED presentations and local MH service data. This linkage involved extracting data from the electronic medical records and extending the time frame to encompass the period from 1 January 2015 to 30 December 2021. The research received approval from Hunter New England Human Research Ethics Committee, 2022/ETH01597 (approved on 18/11/2022), and site-specific approval from Central Coast Local Health District, granted on 14/12/2022. Subsequently, data extraction and linkage processes were undertaken by researchers on 16 December 2022. To ensure the research is safe, respectful, responsible, of high quality and of benefit to First Nation people and communities, this research was conducted in accordance with the National Health and Medical Research Council's (NHMRC) 'Ethical conduct in research with Aboriginal and Torres Strait Islander Peoples and communities: Guidelines for researchers and stakeholders' (National Health Medical Research Council, 2018).

Data analysis

The Central Coast Local Health District encompasses four hospitals, with two housing public Emergency Departments (EDs). These two locations exhibit distinct sociodemographic profiles and levels of public accessibility. To account for these disparities between the sites, the research team conducted a comparative analysis using chi-squared (χ^2) tests, with a statistical significance threshold set at $p<0.05$. To ensure data integrity, missing data were addressed through case deletion, as recommended by Kang (2013). This process resulted in a final data set comprising 26 616 records, with 515 records (1.93%) lost during the data cleansing procedure.

RESULTS

It is important to note that these data are a collection of presentations and not individual people. As Medical Record Numbers (MRN) were not included in the complete data set, the researchers could not interpret the data in a way that allowed exploration of the individuals' frequency and history of presenting to the EDs included in the study. As such, an individual may present once or several times, but the data cannot accurately reflect high service usage for individual people. Most presentations reported themselves as 'Neither Aboriginal nor Torres Strait Islander origin' at 86.44%, with the number of presentations identifying as 'Aboriginal and/or Torres Strait Islander origin' totalling 12.63%. At Site A, 11.2% of presentations identified as 'Aboriginal and/or Torres Strait Islander origin' compared to 14.1% at Site B, with a significant difference between the two sites ($p<0.0001$) (Table 1, Figure 1), and a higher percentage of presentations by First Nation people at Site B where the First Nation People population has higher residential density.

Overall, the age profile structure and presentation frequency provide useful information for service planning. While it is not possible to determine the number of individuals that may repeatedly present, the presentation numbers independently identify sufficient

TABLE 1 First Nation status.

First Nation	Site A	%	Site B	%	Total	%	p*
Aboriginal but not Torres Strait Islander origin	1356	10.55	1790	13.44	3146	12.02	0.0000*
Both Aboriginal & Torres Strait Islander origin	30	0.23	48	0.36	78	0.30	0.0596
Declined to Respond	10	0.08	13	0.10	23	0.09	0.5889
Neither Aboriginal nor Torres Strait Islander origin	11 275	87.74	11 343	85.18	22 618	86.44	0.0000*
Torres Strait Islander but not Aboriginal origin	44	0.34	36	0.27	80	0.31	0.2912
Unknown	135	1.05	86	0.65	221	0.84	0.0003*

Note: p is the comparison between Site A and Site B Emergency Department presentations.

14470349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/inm.13362 by National Health And Medical Research Council, Wiley Online Library on [26/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

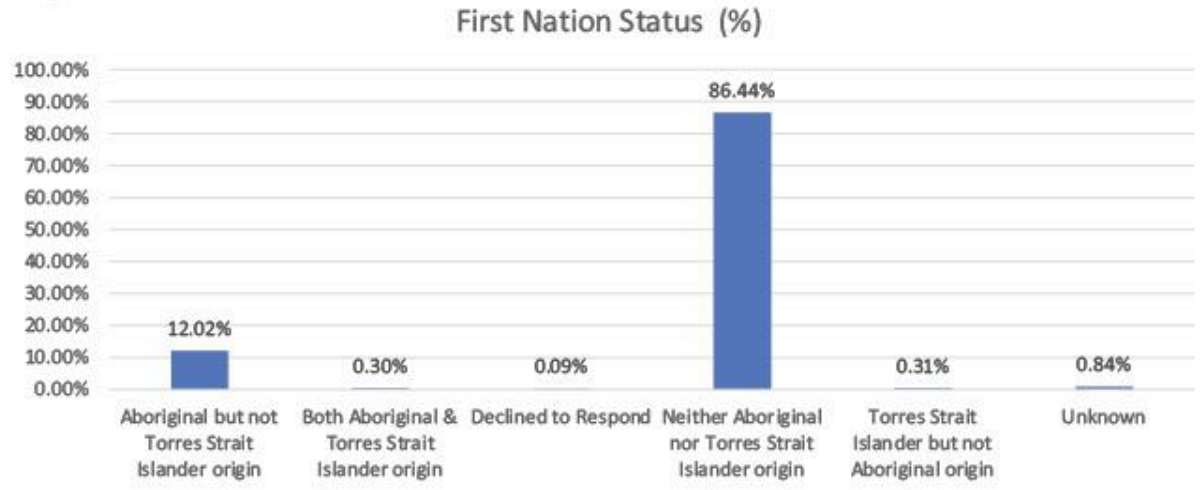


FIGURE 1 First Nation status %.

TABLE 2 Presentations to ED by age group and First Nations status.

Age	Non First Nation	%	First Nation	%	Total	%	p^*
0–11	350	1.53	82	2.48	432	1.65	0.0001*
12–17	3318	14.51	725	21.94	4043	15.45	0.0000*
18–24	4285	18.74	809	24.49	5094	19.47	0.0000*
25–34	4512	19.74	799	24.18	5311	20.30	0.0000*
35–44	4206	18.40	427	12.92	4633	17.71	0.0000*
45–54	3077	13.46	313	9.47	3390	12.96	0.0000*
55–64	1657	7.25	111	3.36	1768	6.76	0.0000*
65–74	801	3.50	29	0.88	830	3.17	0.0000*
75–84	446	1.95	8	0.24	454	1.74	0.0000*
Above 85	210	0.92	1	0.03	211	0.81	0.0000*

information to demonstrate a level of distress indicative of health response. The data presented in Table 2 highlight the percentage of individual presentations within each age bracket who attended the ED seeking MH care. The results reveal a significant discrepancy in healthcare access between non-First Nation and First Nation populations across all age groups. Notably, the percentage of First Nation occasions of service in the younger age groups consistently exceeds that of their non-First Nation counterparts. This pattern is particularly pronounced in the 12–17 years (Non- First Nation = 14.51% First Nation = 21.94%) and 18–24 years (Non-First Nation = 18.74%, First Nation = 24.49%) age groups, where First Nation people seeking help and access to care is markedly higher (Figure 2).

DISCUSSION

First Nation people represent approximately 3.8% of the Australian population (Australian Bureau of Statistics, 2021e) but account for 12.0% of MH-related ED presentations and 7.5% of all ED presentations across Australia (Australian Institute of Health and Welfare, 2022). On the Central Coast of NSW, 4.9% of the population identify as 'Aboriginal and/or Torres Strait Islander' (Australian Bureau of Statistics, 2021a). The overall data presented in this analysis show an overall ED mental health presentation rate of 12.0% for those who identified as 'Aboriginal but not Torres Strait Islander origin', 0.4% as 'Both Aboriginal and Torres Strait Islander' and 0.3% as 'Torres Strait Islander' totalling 12.7%. These results align with the Australian data (Australian Institute of Health and Welfare, 2022) and demonstrate a significant disproportionate representation on the Central Coast. However, further disproportion is evident when data are examined to reflect each local hospital catchment area separately. Site B has 6.3% (Australian Bureau of Statistics, 2021d) of its population identifying as 'Aboriginal and/or Torres Strait Islander', Site A 4.0% (Australian Bureau of Statistics, 2021b) and 3.4% for NSW (Australian Bureau of Statistics, 2021c).

14470349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/inm.13362 by National Health And Medical Research Council, Wiley Online Library on [26/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

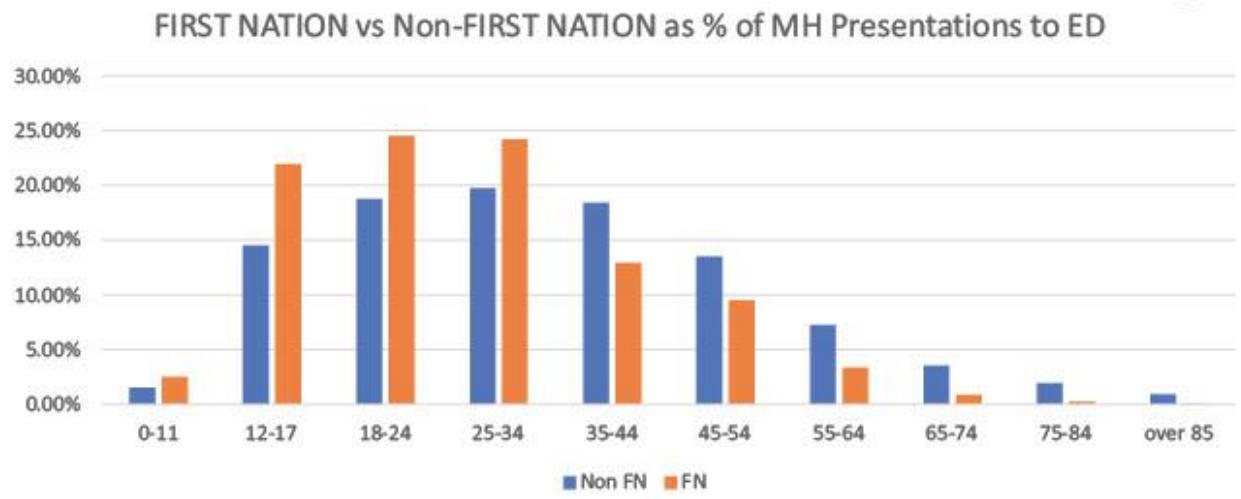


FIGURE 2 First Nation percentage comparison.

By contrast, Perera et al. (2018) conducted a retrospective, descriptive analysis of linked ED presentation data for NSW over five calendar years, 2010–2014 and observed that 6.7% of presentations for MH in 36 EDs identified as 'Aboriginal and/or Torres Strait Islander', proportionally similar to that of the overall presentations. However, the findings from this study show a much higher overall presentation for 'Aboriginal and/or Torres Strait Islander' consumers at Site B at 14.07% and Site A at 11.13%. With a *p*-value of <0.00001, this difference between the two sites is significant and raises concerns about the health outcomes and access to healthcare for First Nation people in the Central Coast of NSW, especially at Site B. Further research is needed to understand the root causes of this overrepresentation and develop targeted strategies to improve mental health outcomes for all First Nation people. This imbalance suggests a tailored response is needed to achieve equitable access to quality healthcare for all individuals, accommodating their specific background and/or ethnicity.

The age profiles presented in this research align with similar research conducted in Australia. It is known that First Nation Australians under 25 years were 1.4 times as likely as non-First Nation Australians to access Medicare Benefits Schedule (MBS) subsidised primary MH services (13.5% compared with 9.5%) (Australian Goverment Productivity Commision, 2020). In 2017–2018, there were 902900 service contacts recorded by community MH teams for people who identified as First Nation, proportionally 10% of all MH contacts across Australia (Australian Institute of Health and Welfare, 2023). The adjusted contact rate for First Nation people was 3.3 times the rate of non-Indigenous Australians and was higher across all age groups peaking for those aged 25–44 years (Australian Institute of Health and Welfare, 2023).

We propose several well-accepted theoretical perspectives that may assist in explaining our results. Firstly, we suggest that the results need to be considered in the context of poorer mental health outcomes related to cultural disruptions as the logical consequence of pervasive and intergenerational traumas related to colonisation for First Nation people (Wilson & Waqanaviti, 2021). An ontological consideration occurs within the inherent biases within default Western models of mental health service delivery and difficulty in aligning this with holistic Indigenous models that value culturally enriched ways of knowing, being, doing and belonging applied to concepts of social and emotional well-being (Gee et al., 2014; McGough et al., 2022).

Many First Nation people, for example, the Stolen Generations, experienced dislocation from their cultural lands and families and were required to assimilate, while others have endured generational housing, economic and social insecurity (Wilson & Waqanaviti, 2021). First, Nation people born before 1972 (>44–50 for the data period) experienced a period of extensive trauma and have significantly worse outcomes associated with mental health (Australian Institute of Health and Welfare, 2021). The access to care in the older age groups that is present in the results of this research is potentially related to feeling unsafe when accessing healthcare (Nolan-Isles et al., 2021; Ware, 2013) and echoes the findings of Campbell et al. (2018) in a reduced likelihood of access to services much later in the disease process, if at all. Thus, we can hypothesise that where populations have high-density representation, these populations are likely to experience higher levels of mental health disadvantage and will be likely to require higher levels of mental health care, and more often.

Limitations

Several limitations to this study should be noted. A unique person identifier was not included in the data set

14470349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/inm.13363 by National Health And Medical Research Council, Wiley Online Library on [26/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

under analysis, which made it impossible to establish if a person had presented once or multiple times. This study did not include after-hours MH presentations that did not go directly to the ED, such as police arriving at the MH unit or direct community admissions. Due to how the data are coded and sourced, not all MH presentations may be coded correctly and may not be present in the data. Likewise, there may be instances where an individual may present and be coded as an MH problem initially but require non-mental health care, and the MH team may not have direct input while they are in ED, such as for a person presenting with intoxication. These limitations should be considered when interpreting the findings of this study and may impact the generalisability of the results.

Recommendations

Improving First Nation cultural safety in ED

First Nation people are overrepresented in not just MH (12.0%), but ED presentations overall at 7.5% across Australia, yet account for only 3.3% of the population (Australian Institute of Health and Welfare, 2022). Priority 14 of Australia's National Strategy for 'Closing the Gap' highlights the importance of First Nation people social and emotional well-being (SEWB) (Commonwealth of Australia, 2020). The strategy reports on SEWB as measured by the number of suicides of First Nation people. Any recommendation must reflect care appropriate to and consistent with First Nation people cultural and spiritual beliefs and practices, including the involvement of traditional healers, Elders and other cultural healers (Commonwealth of Australia, 2020; Rooney et al., 2023).

In Australia, Cultural Safety has gained recognition as a vital framework for enhancing healthcare access and quality for First Nation communities and addressing health disparities (McGough et al., 2022). Due to the diverse context of First Nation communities, strategies led by communities are required for each approach, and caution should be used when applying previous approaches to different settings. Persistent misunderstandings surround this concept, highlighting the need for evaluation and research to generate evidence-based knowledge on effectively integrating Cultural Safety into clinical practice (McGough et al., 2022).

Health services and practitioners need to look beyond the traditional view of Cultural Safety and examine their behaviours, beliefs and practices that continue to support issues such as institutional racism (Ramsden, 2002). Moving beyond cultural sensitivity training to a model of Cultural Humility to understand the impact of colonisation and racism on healthcare for First Nation people. For example, in Nursing and Midwifery, programmes, such as *Murra Mullangari: Introduction to Cultural Safety and Cultural Humility*, aim to provide an understanding of the history, origins and concepts of Cultural Safety and Cultural Humility and offer insights into the culture of power and its implications on the Nursing workforce (Congress of Aboriginal and Torres Strait Islander Nurses and Midwives, 2022).

Given the disproportionate representation of First Nation people in MH-related ED presentations, it is imperative to prioritise a research perspective centred on First Nation communities. This necessitates not only employing First Nation research methods such as yarning and storytelling (Dudgeon et al., 2020; Walker et al., 2014), but also ensuring that research priorities, governance and benefits are led by First Nation people within their respective communities. This research should avoid 'Deficit Discourse', which refers to negative language and behaviour that represents people in terms of their deficiencies and failures (Lowitja Institute, 2018). This focus on failure and dysfunction can eclipse the complex reasons for inequalities and overlook diversity, capability and strength (Lowitja Institute, 2018). Instead, future research must seek to move towards a 'Strengths-based approach' to change the language of the narrative to one that aligns with existing First Nation approaches (Lowitja Institute, 2018). This approach is essential for comprehensively understanding the unique cultural needs and upstream factors influencing First Nation individuals seeking help and accessing MH care through ED.

CONCLUSION

There is a significant over-representation of First Nation people, particularly in the Site B hospital catchment area, and this highlights the need for targeted strategies to improve mental health outcomes and access to MH care for First Nation people. Future research should focus on the following:

- Service design, codesign and health practices must prioritise the Cultural Safety of First Nation people and draw on Indigenous research methods, such as yarning and storytelling, to understand presentation and access prevalence for MH care via ED.
- The intricate nature and sheer volume of data in mental health care underscore the need for extensive analysis and the advancement of service delivery systems. These systems are pivotal in identifying care needs, predicting care pathways, improving decision-making and uncovering crucial variables and their interrelationships. Ultimately, they contribute to bolstering clinical decision support and a deeper understanding of service delivery in the field of MH care for First Nation people.

14470349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/inm.13362 by National Health And Medical Research Council, Wiley Online Library on [24/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

RELEVANCE TO PRACTICE

Mental health nurses and other health professionals require an understanding of the demographic profile of people in the community where they practice. Recognising the population trends will assist in planning to improve supportive cultural safety needs for Aboriginal and Torres Strait Islander people and aligning service provision to match the requirements of the community of interest. Significant barriers persist in access to healthcare for First Nation people in Australia, resulting in a negative experience for many First Nation people. The service usage demographic for people seeking MH care in the ED must be understood to enable the development of care and service delivery models that address cultural safety needs.

AUTHOR CONTRIBUTIONS

OH was involved in concept development, project design, data collection, data analysis, manuscript preparation. RSR was involved in data analysis, contribution to manuscript. SC was involved in data analysis, contribution to manuscript, supervision of project. RW was involved in concept development, project design, data collection, data analysis, manuscript contribution, supervision of project.

ACKNOWLEDGEMENTS

Authors would like to acknowledge the support of Central Coast Local Health District. This research was conducted on traditional Darkinjung Country. We pay respect to elders past present and emerging. The authors wish to acknowledge that while many First Nations researchers and scholars are working towards improving mental health outcomes for First Nations people, it is not scientifically relevant to the reporting of this research to cite all of them in this report. However, we recognise their broadly related contributions to scholarship and research focussed on improving mental health outcomes for First Nations people. One of the authors (RW) identifies as an Aboriginal person, and all authors pay respect to First Nation mental health and social and emotional well-being researchers and scholars, joining with them in the urgent work of improving mental health outcomes for First Nations people. Open access publishing facilitated by RMIT University, as part of the Wiley - RMIT University agreement via the Council of Australian University Librarians.

FUNDING INFORMATION

Partial financial support was received from the NSW Ministry of Health as part of the Towards Zero Suicides initiative.

CONFLICT OF INTEREST STATEMENT

Prof. Rhonda Wilson is an Editorial Board Member of the International Journal of Mental Health Nursing.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ETHICS APPROVAL

Ethical approval was granted by the Hunter New England Human Research Ethics Committee 2022/ETH01597. Central Coast Local Health District granted site-specific approval 2022/STE03296.

ORCID

Oliver Higgins https://orcid.org/0000-0001-6545-145X
Rachel B. Sheather-Reid https://orcid.org/0009-0006-9282-2408
Stephan K. Chalup https://orcid.org/0000-0002-7886-3653
Rhonda L. Wilson https://orcid.org/0000-0001-9252-2321

REFERENCES

Australian Bureau of Statistics. (2021a) *2021 census all persons QuickStats central coast (NSW)*. Retrieved 05/02/2023 from https://abs.gov.au/census/find-census-data/quickstats/2021/LGA11650

Australian Bureau of Statistics. (2021b) *2021 Gosford, census all persons QuickStats*. Retrieved 14/02/2023 from https://abs.gov.au/census/find-census-data/quickstats/2021/SED10030

Australian Bureau of Statistics. (2021c) *2021 New South Wales, census all persons QuickStats*. Retrieved 14/02/2023 from https://www.abs.gov.au/census/find-census-data/quickstats/2021/1

Australian Bureau of Statistics. (2021d) *2021 Wyong, census all persons QuickStats*. Retrieved 14/02/2023 from https://abs.gov.au/census/find-census-data/quickstats/2021/SAL14464

Australian Bureau of Statistics. (2021e) *Estimates of Aboriginal and Torres Strait Islander Australians*. Abs. https://www.abs.gov.au/statistics/people/aboriginal-and-torres-strait-islander-peoples/estimates-aboriginal-and-torres-strait-islander-australians/latest-release

Australian Goverment Productivity Commision (2020) *Report on government services 2020*. Canberra. Retrieved from https://www.pc.gov.au/ongoing/report-on-government-services/2020/health/mental-health-management

Australian Institute of Health and Welfare. (2021) *Aboriginal and Torres Strait Islander Stolen Generations aged 50 and over: updated analyses for 2018–19*. Retrieved 25/11/2023 from https://www.aihw.gov.au/reports/indigenous-australians/indigenous-stolen-generations-50-and-over/contents/summary

Australian Institute of Health and Welfare. (2022) *Mental health services in Australia*. https://www.aihw.gov.au/reports/mental-health-services/mental-health-services-in-australia/report-contents/emergency-department-mental-health-services

Australian Institute of Health and Welfare. (2023) *3.10 Access to mental health services*. Australian Institute of Health and Welfare. Retrieved 25/11/2023 from https://www.indigenoushpf.gov.au/measures/3-10-access-mental-health-services

Campbell, M.A., Hunt, J., Scrimgeour, D.J., Davey, M. & Jones, V. (2018) Contribution of Aboriginal community-controlled health services to improving Aboriginal health: an evidence review. *Australian Health Review*, 42(2), 218–226.

Cerreto, A. (2018) *Ngaga-dji (hear me): Young voices creating change for justice*. Melbourne, Vic.: Koorie Youth Council.

14470349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/inm.13362 by National Health And Medical Research Council, Wiley Online Library on [26/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Clarke, D.E., Dusome, D. & Hughes, L. (2007) Emergency department from the mental health client's perspective. *International Journal of Mental Health Nursing*, 16(2), 126–131. Available from: https://doi.org/10.1111/j.1447-0349.2007.00455.x

Commonwealth of Australia. (2020) *National agreement on closing the gap*. https://www.closingthegap.gov.au/sites/default/files/2021-05/ctg-national-agreement_apr-21.pdf

Congress of Aboriginal and Torres Strait Islander Nurses and Midwives. (2022) *Murra Mullangari: introduction to cultural safety and cultural humility*. Retrieved 06/04/2024 from https://hnc.org.au/news-story/murra-mullangari-introduction-to-cultural-safety-and-cultural-humility-xhp-x1/

Derblom, K., Molin, J., Gabrielsson, S. & Lindgren, B.-M. (2021) 'Acknowledge me as a capable person': how people with mental ill health describe their experiences with general emergency care staff – a qualitative interview study. *International Journal of Mental Health Nursing*, 30(6), 1539–1549. Available from: https://doi.org/10.1111/inm.12903

Digel Vandyk, A., Young, L., MacPhee, C. & Gillis, K. (2018) Exploring the experiences of persons who frequently visit the emergency Department for Mental Health-Related Reasons. *Qualitative Health Research*, 28(4), 587–599. Available from: https://doi.org/10.1177/1049732317746382

Dudgeon, P., Bray, A., Darlaston-Jones, D. & Walker, R. (2020) *Aboriginal participatory action research: an indigenous research methodology strengthening decolonisation and social and emotional wellbeing* (Lowitja Institute Discussion Paper Series, Issue 1). https://www.lowitja.org.au/content/Document/Lowitja-Publishing/LI_Discussion_Paper_P-Dudgeon_FINAL3.pdf

Duggan, M., Harris, B., Chislett, W.K. & Calder, R. (2020) *Nowhere else to go: why Australia's health system results in people with mental illness getting 'stuck' in emergency departments*. Mitchell Institute Commissioned report 2020, Victoria University. https://vuir.vu.edu.au/41956/1/nowhere-else-to-go-people-mental-illness-stuck-emergency-departments-report-mitchell-institute.pdf

Gee, G., Dudgeon, P., Schultz, C., Hart, A. & Kelly, K. (2014) Aboriginal and Torres Strait islander social and emotional wellbeing. In: Dudgeon, P., Milroy, H. & Walker, R. (Eds.) *Working together: Aboriginal and Torres Strait Islander mental health and wellbeing principles and practice*, Vol. 2. Barton, ACT: Australian Government Department of the Prime Minister and Cabinet, pp. 55–68.

Gwynne, K., Jeffries, T. & Lincoln, M. (2018) Improving the efficacy of healthcare services for Aboriginal Australians. *Australian Health Review*, 43(3), 314–322.

Kang, H. (2013) The prevention and handling of the missing data. *Korean Journal of Anesthesiology*, 64(5), 402–406. Available from: https://doi.org/10.4097/kjae.2013.64.5.402

Liamputtong, P. (2021) *Research methods and evidence-based practice*. Australia: Oxford University Press Australia & New Zealand.

Lowitja Institute. (2018) *Deficit discourse and Aboriginal and Torres Strait Islander health policy*. Australia: Lowitja Institute.

McGough, S., Wynaden, D., Gower, S., Duggan, R. & Wilson, R. (2022) There is no health without cultural safety: why cultural safety matters. *Contemporary Nurse*, 58(1), 33–42. Available from: https://doi.org/10.1080/10376178.2022.2027254

Morphet, J., Innes, K., Munro, I., O'Brien, A., Gaskin, C.J., Reed, F. et al. (2012) Managing people with mental health presentations in emergency departments—a service exploration of the issues surrounding responsiveness from a mental health care consumer and carer perspective. *Australasian Emergency Nursing Journal*, 15(3), 148–155. Available from: https://doi.org/10.1016/j.aenj.2012.05.003

National Health Medical Research Council. (2018) *Ethical conduct in research with Aboriginal and Torres Strait Islander peoples and communities: guidelines for researchers and stakeholders*. Australia: National Health and Medical Research Council.

Nolan-Isles, D., Macniven, R., Hunter, K., Gwynn, J., Lincoln, M., Moir, R. et al. (2021) Enablers and barriers to accessing healthcare services for Aboriginal People in New South Wales, Australia. *International Journal of Environmental Research and Public Health*, 18(6), 3014. Available from: https://doi.org/10.3390/ijerph18063014

NSW Health. (2023) *High or very high psychological distress in adults by LHD*. Retrieved 14/03/2023 from https://www.healthstats.nsw.gov.au/#/indicator?name=-men-hidistress-phs&location=LHD&view=Trend&measure=prevalence&groups=&compare=LHD&filter=LHD,All%20LHDs,Central%20Coast%20LHD

Paradies, Y. (2016) Colonisation, racism and indigenous health. *Journal of Population Research*, 33(1), 83–96.

Perera, J., Wand, T., Bein, K.J., Chalkley, D., Ivers, R., Steinbeck, K.S. et al. (2018) Presentations to NSW emergency departments with self-harm, suicidal ideation, or intentional poisoning, 2010–2014. *Medical Journal of Australia*, 208(8), 348–353. Available from: https://doi.org/10.5694/mja17.00589

Ramsden, I. (2002) *Cultural safety and nursing education in Aotearoa and Te Waipounamu*. Wellington, New Zealand: Victoria University of Wellington.

Rooney, E.J., Wilson, R.L. & Johnson, A. (2023) Integration of traditional therapies for First Nations people within western healthcare: an integrative review. *Contemporary Nurse*, 1–17, 294–310. Available from: https://doi.org/10.1080/10376178.2023.2276718

Smallwood, R., Woods, C., Power, T. & Usher, K. (2021) Understanding the impact of historical trauma due to colonization on the health and well-being of indigenous young peoples: a systematic scoping review. *Journal of Transcultural Nursing*, 32(1), 59–68. Available from: https://doi.org/10.1177/1043659620935955

Taylor, P.S. & Habibis, D. (2020) Widening the gap: white ignorance, race relations and the consequences for Aboriginal people in Australia. *Australian Journal of Social Issues*, 55(3), 354–371. Available from: https://doi.org/10.1002/ajs4.106

Walker, M., Fredericks, B., Mills, K. & Anderson, D. (2014) "Yarning" as a method for community-based health research with indigenous women: the indigenous Women's wellness research program. *Health Care for Women International*, 35(10), 1216–1226. Available from: https://doi.org/10.1080/07399332.2013.815754

Ware, V.A. (2013) *Improving the accessibility of health Services in Urban and Regional Settings for indigenous people*. Australia: AIHW.

Wilson, R.L. & Waqanaviti, K. (2021) Chapter 13 emotional and social well-being for First Nations people in the mental health context. In: Fredericks, B. (Ed.) *Yatdjulgin: Aboriginal and Torres Strait Islander nursing and midwifery care*. Cambridge: Cambridge University Press, pp. 281–306.

How to cite this article: Higgins, O., Sheather-Reid, R.B., Chalup, S.K. & Wilson, R.L. (2024) Disproportionate mental health presentations to emergency departments in a coastal regional community in Australia of first nation people. *International Journal of Mental Health Nursing*, 00, 1–8. Available from: https://doi.org/10.1111/inm.13362

14470349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/inm.13362 by National Health And Medical Research Council, Wiley Online Library on [24/05/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

7.3.2 Article Summary

The over-representation of First Nations peoples, particularly youth and young adults, in mental health–related ED presentations is a stark testament to the profound transgenerational impacts of colonisation, dispossession, and marginalisation (Australian Institute of Health and Welfare, 2021a; Paradies, 2016). These findings underscore the urgent need for a multifaceted, culturally grounded approach to addressing the MH crisis within Aboriginal and Torres Strait Islander communities.

Future research endeavours must prioritise collaboration with First Nations communities, embracing Indigenous research methodologies and co-design principles to develop culturally safe and responsive MH services. By centring the voices, experiences, and traditional knowledge of First Nations peoples, researchers and healthcare providers can gain a deeper understanding of the unique socio-cultural determinants of MH and develop holistic, trauma-informed interventions that resonate with First Nations knowledge (Dudgeon et al., 2020; Gee et al., 2014; Milroy et al., 2024; Wilson & Waqanaviti, 2021).

Moreover, fostering cultural safety within EDs and broader healthcare settings is paramount. This entails ongoing cultural competency training for healthcare professionals, the recruitment and retention of First Nations healthcare workers, and the integration of traditional healing practices alongside Western medical models. By creating inclusive and culturally affirming spaces, healthcare systems can begin to rebuild trust and dismantle the systemic barriers that have historically impeded First Nations peoples' access to quality MH care (Dudgeon et al., 2020; Gee et al., 2014; Milroy et al., 2024; Wilson & Waqanaviti, 2021).

Ultimately, addressing the MH crisis within First Nations communities requires a multi-agency approach that addresses the social determinants of health, promotes cultural revitalisation and self-determination, and dismantles the systemic racism and marginalisation that have perpetuated intergenerational trauma. Only through a sustained, collaborative effort that centres the voices and experiences of First Nations peoples can we hope to achieve equitable MH outcomes and foster healing for communities long afflicted by the enduring legacies of colonisation.

7.4 Machine Learning Model Reveals Determinators for Admission to Acute Mental Health Wards from Emergency Department Presentations

Higgins, O., Chalup, S. K., & Wilson, R. L. (2024c). Machine learning model reveals determinators for admission to acute mental health wards from emergency department presentations. *International Journal of Mental Health Nursing*. Advance online publication. https://doi.org/10.1111/inm.13402

	Higgins, O	Chalup, S. K	Wilson, R. L
Conceptualization	X	X	X
Methodology	X	X	X
Software	X		
Validation	X		
Formal analysis	X		X
Investigation	X		X
Resources	X		
Data Curation	X		
Writing - Original Draft	X		
Writing - Review & Editing	X	X	X
Visualization	X		
Supervision		X	X
Project administration	X		
Funding acquisition			
Percentage	**70%**	**10%**	**20%**

7.4.1 Article Introduction

This study focused on identifying key factors associated with admissions to acute MH inpatient facilities following ED presentations for MH concerns, particularly suicidality. Efficient management of MH presentations in EDs is a significant challenge both in Australia and globally, often due to extended assessment times and the perception among ED staff of being inadequately equipped to manage complex MH cases (Clarke et al., 2014). This can lead to tensions between MH and ED nursing teams, complicating care delivery (Broadbent et al., 2020). This study utilised Python and the InterpretML package to analyse ED presentations and uncover correlations between key presentation features—such as the presenting problem, triage category, and prior healthcare utilisation—and the likelihood of admission to an acute MH ward. The focus of the study was on transparency and interpretability rather than model performance, aiming to provide clinicians with clear, interpretable insights. These insights are intended to enhance clinical decision making, improve triage processes, and address missed care issues within EDs, ultimately facilitating more effective resource allocation and alleviating the burden on healthcare staff.

This study built on the previous study findings by deepening the understanding of the factors influencing MH admissions from ED presentations. In Study A, the sociodemographic characteristics and presentation features of individuals seeking MH care were identified, providing a broad overview of the population and the key challenges faced in regional EDs. Study B then focused specifically on the disparities in MH service use among Aboriginal and Torres Strait Islander people, highlighting the over-representation of these populations in MH-related ED presentations and the need for culturally responsive care.

This analysis served to provide a platform for Study D, where the focus shifted to developing and testing machine learning models to predict MH admissions. The correlations and insights

revealed in Study C served as crucial inputs for Study D, guiding the selection of relevant features and ensuring the models developed in Study D were informed by the real-world patterns and challenges identified earlier. Thus, Study C not only consolidated the findings from Studies A and B but also lay the groundwork for predictive modelling in Study D, aiming to improve decision making and triage processes in MH care. By identifying and explaining these key correlations, the research contributes to a deeper understanding of the factors driving admissions for MH concerns, particularly suicidality, and supports the development of AI-driven decision support systems that build clinician trust through interpretability and transparency.

Check for updates

International Journal of Mental Health Nursing

WILEY

International Journal of Mental Health Nursing

ORIGINAL ARTICLE OPEN ACCESS

Machine Learning Model Reveals Determinators for Admission to Acute Mental Health Wards From Emergency Department Presentations

Oliver Higgins[1,2,3] | Stephan K. Chalup[4] | Rhonda L. Wilson[1,2,3]

[1]RMIT University, Melbourne, Victoria, Australia | [2]Central Coast Research Institute, Gosford, Australia | [3]Central Coast Local Health District, Gosford, New South Wales, Australia | [4]Data Science, School of Information and Physical Sciences (Computer Science and Software Engineering), University of Newcastle, Callaghan, New South Wales, Australia

Correspondence: Oliver Higgins (oliver.higgins@health.nsw.gov)

Received: 13 April 2024 | **Revised:** 31 May 2024 | **Accepted:** 27 July 2024

Funding: Partial financial support was received from the NSW Ministry of Health as part of the Towards Zero Suicides initiative.

Keywords: Aboriginal and Torres Strait Islander | emergency department | first nation | mental health

ABSTRACT

This research addresses the critical issue of identifying factors contributing to admissions to acute mental health (MH) wards for individuals presenting to the emergency department (ED) with MH concerns as their primary issue, notably suicidality. This study aims to leverage machine learning (ML) models to assess the likelihood of admission to acute MH wards for this vulnerable population. Data collection for this study used existing ED data from 1 January 2016 to 31 December 2021. Data selection was based on specific criteria related to the presenting problem. Analysis was conducted using Python and the Interpretable Machine Learning (InterpretML) machine learning library. InterpretML calculates overall importance based on the mean absolute score, which was used to measure the impact of each feature on admission. A person's 'Age' and 'Triage category' are ranked significantly higher than 'Facility identifier', 'Presenting problem' and 'Active Client'. The contribution of other presentation features on admission shows a minimal effect. Aligning the models closely with service delivery will help services understand their service users and provide insight into financial and clinical variations. Suicidal ideation negatively correlates to admission yet represents the largest number of presentations. The nurse's role at triage is a critical factor in assessing the needs of the presenting individual. The gap that emerges in this context is significant; MH triage requires a complex understanding of MH and presents a significant challenge in the ED. Further research is required to explore the role that ML can provide in assisting clinicians in assessment.

1 | Introduction

Globally, there is a need for improved mental health (MH) care with increasing rates of depression, anxiety and other MH conditions (World Health Organization 2022). Approximately 5% of the adult world population have depression, and approximately 700 000 people die by suicide each year (World Health Organization 2022). In 2021, there were 3144 deaths by suicide in Australia, an average of nine deaths per day (Australian Institute of Health and Welfare 2023b). Many people seek initial, acute or emergency MH assistance at their local hospital emergency department (ED), placing a growing burden on public health facilities. The increasing demands suggest health service providers require innovative solutions to manage. One potential approach for enhancing the analysis and interpretation of regularly collected data pertaining to consumer presentations in EDs involves the application of techniques such as machine learning (ML). ML is a type of

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2024 The Author(s). *International Journal of Mental Health Nursing* published by John Wiley & Sons Australia, Ltd.

14470349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/inm.13402 by Cochrane Hungary, Wiley Online Library on [29/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

artificial intelligence that enables computers to learn from data and improve their performance on tasks without being explicitly programmed (Molnar 2020).

Managing consumers who seek care for MH problems in the ED can be complex and challenging, with ED triage nursing a required and discrete skill. MH triage inconsistencies and delayed clinician access have been reported as barriers (Morphet et al. 2012). Consumers are often unsure about how care is prioritised, with long waits often following arrival to ED (Fatovich and Hirsch 2003) and this can lead to dissatisfaction among many ED consumers (Morphet et al. 2012). Previous research has identified three main themes that characterise help-seeking dynamics for MH conditions in ED: First Nations MH, suicidal ideation (SI) and access and egress pathways (Higgins et al. 2024b). While the research focused on a regional local health district in Australia, it found that First Nations peoples were overrepresented, accounting for 12.63% of MH presentations, despite a population representation of 4.9% in the catchment area (Higgins et al. 2024b). The findings suggest that the reasons underlying First Nations peoples' access and egress disparity dynamics remain elusive and require further investigation. Significantly, a substantial number of presentations involve potentially life-threatening situations such as SI (38.19%) with many presenting by ambulance (45.91%) (Higgins et al. 2024b). Additionally, the number of people who 'Did not wait' or 'Left at own risk' accounted for 10.20% of departures from ED (Higgins et al. 2024b).

This paper is part of a wider study into the use of ML in MH and aims to identify factors that contribute to admission to an MH inpatient facility for people who present to ED for MH care. Hospital EDs fulfil a crucial role in the treatment of individuals experiencing MH problems (Morphet et al. 2012), from first-time service users to those already undergoing MH treatment in the community, all of whom may require urgent or immediate attention (Morphet et al. 2012; Sunderji, de Bibiana, and Stergiopoulos 2015). The efficient handling of individuals seeking help at EDs for MH concerns is a matter of significant interest both in Australia and worldwide (Morphet et al. 2012). The timely and accurate triage assessment, along with appropriate referral and intervention, relies on the knowledge and use of the MH triage scale and experience of MH acuity (Broadbent, Moxham, and Dwyer 2020). MH presentations often pose challenges to consumer flow in EDs as they typically require lengthier assessment times. ED staff frequently express feeling ill-prepared to handle these consumers (Clarke, Dusome, and Hughes 2007) and can result in conflicting perspectives between MH and ED nursing groups (Broadbent, Moxham, and Dwyer 2020). The focus for MH for ED nurses, however well-intentioned and prepared, is quickly eroded by the time constraints and challenges of the day-to-day ED environment and turnover of presentations (Hall et al. 2016).

1.1 | Research Problem

This research aims to prepare, test and evaluate the use of ML models to identify which factors may contribute to admission to an acute MH ward for those presenting to ED with MH as the primary presenting problem, particularly suicidality.

1.2 | Background

1.2.1 | Presentations, Admissions and Readmissions

This research aims to better understand the complexities, factors and relationships of those who present to ED with MH issues and which factors relate to admission. To date, only a limited amount of research describes this issue. Geng et al. (2020) undertook a national survey of Chinese psychiatric hospitals from 19 to 31 March 2019. The data from 41 MH hospitals, 196 children and adolescents, were analysed, finding that psychotic symptoms, depressive symptoms and self-injury/suicide were the most common reasons for admission (Geng et al. 2020). Walker et al. (2019) found that involuntary hospitalisation was associated with male, single marital status, unemployment, receiving welfare, psychotic disorder or bipolar disorder and previous involuntary admission. Lay, Kawohl, and Rössler (2019) conducted a randomised control trial that followed 168 consumers with severe MH issues and researched the factors of readmission in four psychiatric hospitals in Zurich, Switzerland, over a 2-year period after discharge. The regression models identified a strong association between a previous compulsory (involuntary) admission and subsequent readmission, indicating that previous involuntary admission may predict future (involuntary) admission (Lay, Kawohl, and Rössler 2019). This research aims to build on this knowledge providing a retrospective cross-sectional descriptive cohort study to describe the factors contributing to ED presentations for MH care.

1.3 | Research Setting and Context

This paper presents research outcomes within a coastal regional local health service in New South Wales (NSW), Australia, located just north of Sydney. The health service encompasses two public hospitals with EDs, catering to a population of 346 596 residents across a land area spanning 1681 km^2 (Australian Bureau of Statistics 2021). The public MH service delivery programme includes acute inpatient and community-based care. These services are organised into three categories: child and youth MH (spanning ages 0–12 and 12–17), adult MH (ages 18–65) and older persons MH (ages 65+). Notably, the health service distinguishes itself from many other NSW EDs by providing specialised MH nurses and medical staff for assessments at the EDs around the clock, 7 days a week. The health service operates across two sites: Site A comprises 24 acute beds and six high-dependency MH beds, while Site B offers 25 acute beds, 10 high-dependency beds and 15 beds specifically designed for older persons (65+). Additionally, Site B houses a Psychiatric Emergency Care Centre (PECC) equipped with four beds and adjacent to the ED. Neighbouring health districts provide access to subacute, non-acute and rehabilitation beds for individuals requiring extended or specialised MH care. It is important to note that the service lacks dedicated MH beds for adolescents aged 12–17. Instead, referrals for adolescent inpatient services are directed to specialised

inpatient units in neighbouring districts (Central Coast Local Health District 2017).

2 | Methods

2.1 | Design and Data Collection

The primary data collection involved reviewing existing data from 1 January 2016 to 31 December 2021. The ED data were selected for inclusion based on specific terms related to the presenting problem. A second data collection was performed to determine whether the individual presentation was associated with current or past care and is defined in the research as 'MH history'. Ethical approval for the study was granted by the Hunter New England Human Research Ethics Committee, with reference number 2022/ETH01597, and the approval was given on 18 November 2022. Site-specific approval was also obtained from the Central Coast Local Health District on 14 December 2022. Before any analysis, the medical record numbers were encrypted, and subsequently, they were completely removed from the data set. The study examines the following fields within the data set pertinent to MH-related ED presentation which are presented in Table 1.

For the initial data investigation, we utilised several tools and programming languages. Structured query language was used for managing and querying databases, while Python, a versatile programming language popular for data analysis and ML tasks, played a central role in our analysis. Excel, a spreadsheet application commonly used for data organisation, analysis and visualisation, helped in structuring and reviewing the data. Additionally, Seaborn, a Python library designed for creating informative and attractive statistical graphics, was employed to visualise the data effectively. Further information on the data collection and extraction methods can be found in Higgins et al. (2024b). Using Python, the data were brought into a pandas data frame (McKinney 2010) and modelled with Interpretable Machine Learning (InterpretML) (Nori et al. 2019), discussed later in this document. A complete list of variables used is located in Table 1. The initial review of the model noted that the 'site' feature contributed significantly to admission (Figures 1 and 2). Therefore, the data were split, and separate models were developed for each hospital site.

2.2 | Data Definitions

The MH context of the ED presentation is established through the use of the presenting problem. In the electronic medical

TABLE 1 | Variables.

Item	Description
Age	Numerical value indicating the persons age at the time of presentation
Gender	Gender reported at the time of presentation
First Nations peoples' status	A person disclosed if they identify as an 'Aboriginal and/or Torres Strait Islander' person status at the time of presentation.
Time of the day	The time of the day, 0–23, that a person presented to ED
Day of the week	The day in which a person presented, Monday through Sunday.
Month	The month of the presentation, January to December.
Mode of arrival	Police, ambulance, etc.
Triage category	NSW Health Triage Categories indicating urgency of care. Resuscitation (1) Emergency (2) Urgent (3) Semi-urgent (4) Non-urgent (5)
Presenting problem	Altered Mental Status Anxious Behavioural Disturbance Depression Eating Disorder Hallucinations Memory Impairment Mental Health Problem Self Harm Suicidal Ideation
Hospital facility	(Site A/Site B)
Active client	At the time of the presentation to the ED
Previous client	Previous service user preceding the ED presentation back to 1 January 2015

14470349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/inm.13403 by Oliver Higgins - The University Of Newcastle , Wiley Online Library on [20/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

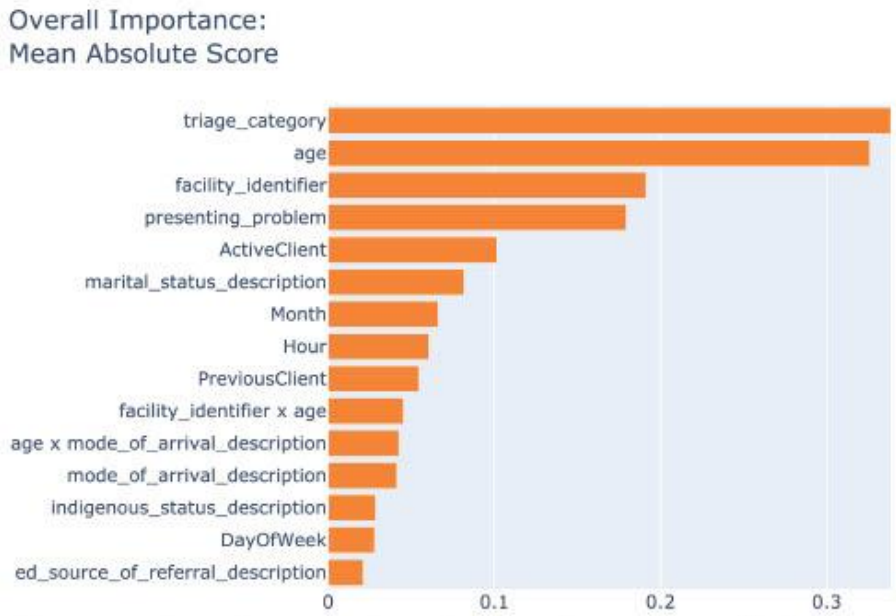


FIGURE 1 | Overall importance. The overall importance of various features in a data set, as measured by the mean absolute score from an InterpretML analysis. The features are ranked from highest to lowest importance based on their mean absolute scores.

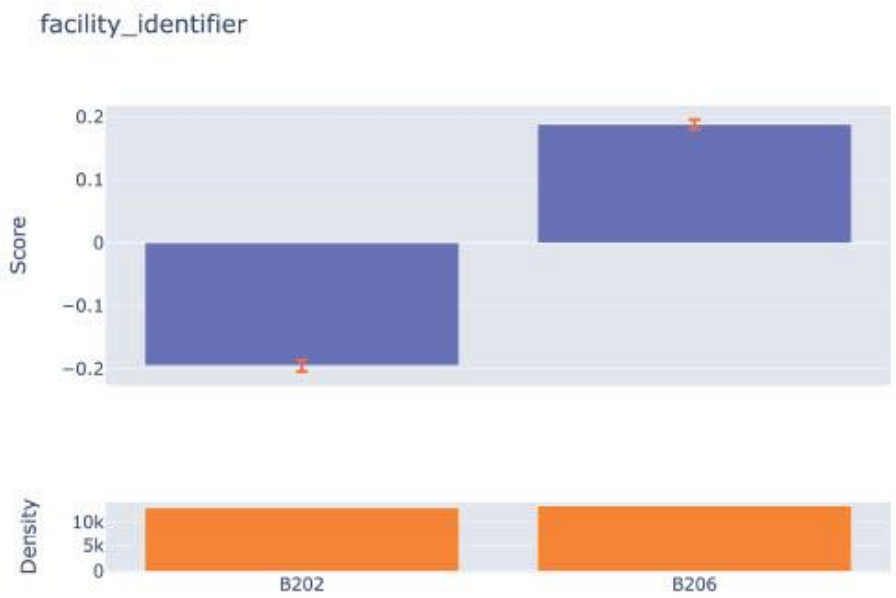


FIGURE 2 | Facility Identifier. The top plot displays the 'facility_identifier' feature. The *y*-axis shows the score values, and the *x*-axis shows two categories, 'Site A' and 'Site B'. The plot indicates that the 'facility_identifier' feature has a positive score value for one of the sites, while the score is negative for the other site. The bottom plot displays the density of data for the 'facility_identifier' feature. The *y*-axis shows the density values, and the *x*-axis shows the same two categories as in the top plot, 'Site A' and 'Site B'.

record, the 'presenting problem' refers to a field that records the consumer's description of their symptoms or medical condition during triage at the ED (Bureau of Health Information 2022). Despite calls for standardisation of this input (Berendsen Russell, Dinh, and Bell 2017), the literature lacks clear definitions or explanations. However, previous works have set a precedent in employing the presenting problem field for specific purposes (Sara et al. 2022; Sperandei et al. 2022; Wand et al. 2015). For the purpose of this discussion, an admission is derived by the value of 'Admitted to the ward/inpatient unit, not a Critical Care Ward' in the 'Mode of separation' field of the dataset (NSW Health 2017). 'MH history' is defined as any involvement with

14470349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/inm.13402 by Chloe Higgins - The University Of Newcastle , Wiley Online Library on [30/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

a Central Coast Local Health District Mental Health Service before the date of presentation, within the time period from 1 January 2015 to 31 December 2021.

2.3 | InterpretML

The prepared data set used in Higgins et al. (2024b) was further analysed using the InterpretML python package to determine the most significant factors contributing to admission. InterpretML is a framework for building interpretable models to help clinicians make informed decisions about consumer care. Traditional ML models, such as deep neural networks, can be very accurate but are often considered 'black boxes' because it can be challenging to understand how they arrive at their predictions (Rai 2019). This lack of transparency can be a significant obstacle in healthcare, where decisions can have adverse consequences, and consumers have the right to understand how and why decisions are made (Molnar 2020). One of the key features of InterpretML is the ability to visualise the model and its predictions using various techniques, such as feature importance plots, partial dependence plots and individual conditional expectation plots (Nori et al. 2019). These plots provide an understanding of the impact each feature contributes to the outcome and can identify areas where improvements can be made in consumer care (Molnar 2020). Using InterpretML with tools such as Log Odds Ratio, we can build transparent, accurate and effective models to improve consumer outcomes.

2.4 | InterpretML Mean Absolute Score

InterpretML helps identify complex interactions, and the overall importance graph is one of the methods it provides for interpreting ML models (Nori et al. 2019). InterpretML calculates overall importance based on the mean absolute score (MAS), which measures the impact of each feature on the model's predictions (InterpretML Documentation 2023). To calculate the MAS, InterpretML performs a series of experiments where the value of each feature is randomly permuted across the data set. The MAS is then calculated for each feature by comparing the model's predictions before and after the feature has been permuted (InterpretML Documentation 2023). The larger the MAS for a given feature, the more important that feature is in the model's predictions. The overall importance graph in InterpretML visualises the MASs for all the features in the model, ranking them in order of importance (InterpretML Documentation 2023). The graph provides a way to quickly identify which features have the most significant impact on the model's predictions and can be a valuable tool for interpreting the model's behaviour. To better understand each element, InterpretML provides a visual interpretation of each element and its relationship to admission using log odds. Log odds is a way of representing the probability of an event as a logarithmic function of the odds of that event occurring (Tabachnick, Fidell, and Ullman 2013). Log odds ratio (OR) is commonly used to compare the odds of an event in two different groups, such as a treatment group and a control group. The Log OR is simply the natural logarithm of the odds ratio, this means that if the odds of an event are 4:1, the log odds would be (4/1) = 1.39 (Tabachnick, Fidell, and Ullman 2013). By incorporating Log OR into the model, we can compare the odds of the outcome between different groups, assess the impact of each predictor on the outcome and compare the probability of different outcomes on a single scale.

3 | Results

The results are presented using Log OR for each component comparison. The metric of MAS is used to determine a feature's contribution and influence relating to admission. Figure 1 is the overall ranking of the importance of the variables used in the ML in the relationship to admission. Both 'Age' (0.3272) and 'Triage category' (0.3396) are ranked significantly higher than 'Facility identifier' (0.1915) and 'Presenting problem' (0.1804). 'Active Client' is ranked fifth as the contribution factor drops off quickly. Given that 'facility identifier' is such a decisive factor (Figures 1 and 2) and Boolean, the data sets were split into the two sites, and individual models were subsequently built. Figure 3A (Site A) and Figure 3B (Site B) present the overall importance of each site's model. Interestingly, the first two items, 'Triage category' and 'Age' are inversed in their overall weightings for the two sites. Site A has 'Triage category' first (0.3032) and 'Age' at 0.2429. For Site B, the 'Age' factor is much more robust in its relationship to admission (0.4101) and 'Triage category' at 0.3544. While the 'Triage category' is second on Site B's list, it still has a higher MAS than Site A. Presenting problem is ranked third on both lists, with the MAS contributing factor slightly higher at Site A (0.1705) than Site B (0.1829). After this, the features become more disparate between the two locations but similar to the overall model, their MAS contribution to admission falls off rapidly.

3.1 | Age Category

The MAS presents that age is a significant contributing factor to an admission from an ED presentation. Figure 4 shows the overall Log OR score in the top half; a positive score implies an increase in the model's weight towards admission, and a negative value implies that they would not be admitted. The bar graph at the bottom of Figure 4 shows the density of the population samples in the data set, with a high negative correlation to admission for presentations under 18 and an increase towards a positive correlation at the age of 65. Note that the outliers are for the under 10's and over 85's groups. When comparing the two sites, there is a considerable difference in the age-related correlations. Site A (Figure 5A) shows a positive correlation in the 5–9.75 age that drops down to a negative before starting to increase in the late teens, flattening through the 20–30s, then slowly climbing as the age increases. Site B (Figure 5B) shows a very different profile. The 0–16.4 range shows a significant negative correlation towards admission, rising steadily through 18 years, where it starts to level off. At approximately 30, the correlation starts to move towards an increased likelihood of admission, with a clear and definite step occurring at 65 years. The ML model identified

14470349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ajsw.13402 by Oliver Higgins - The University Of Newcastle , Wiley Online Library on [30/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

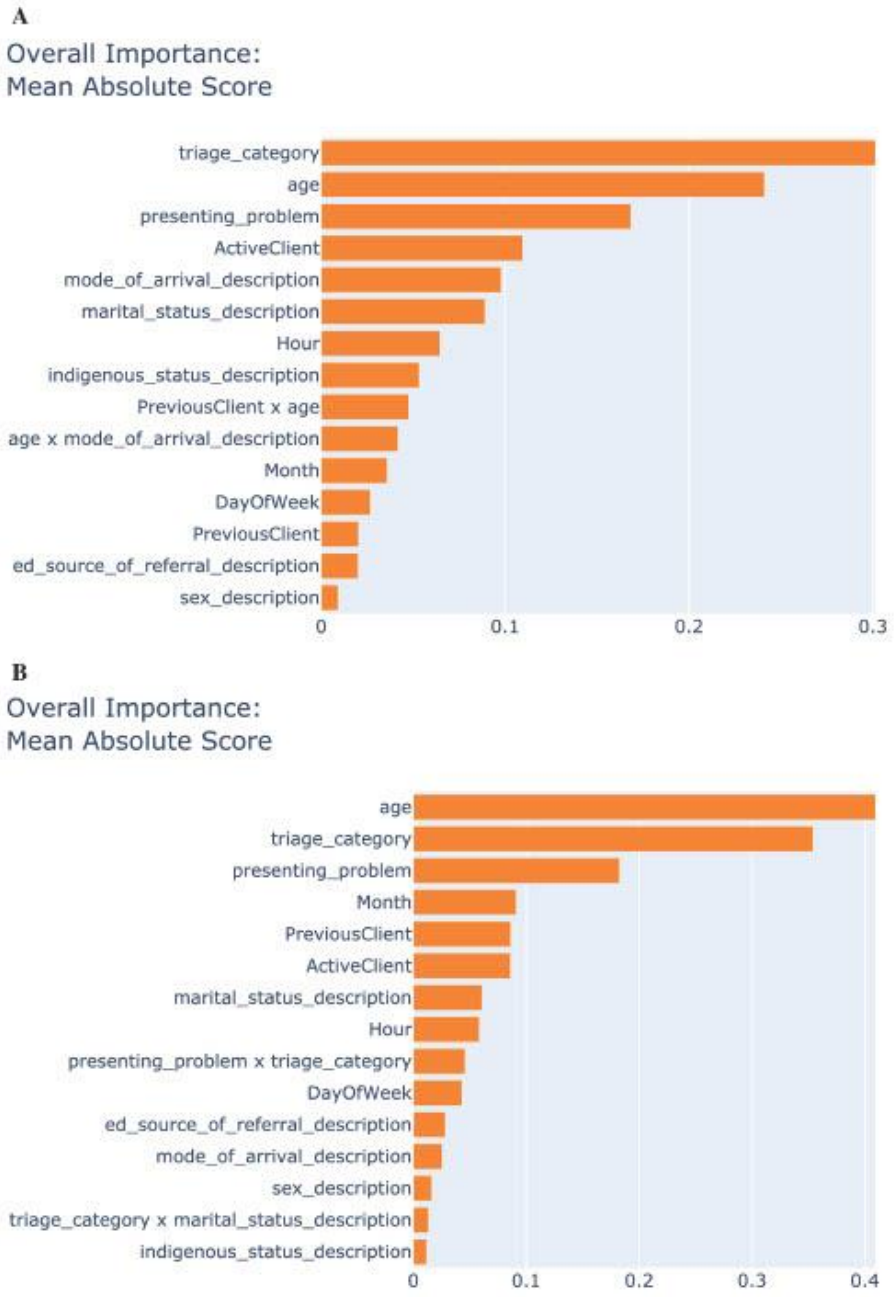


FIGURE 3 | (A) Overall. The overall importance of various features in a data set, as measured by the mean absolute score from an InterpretML analysis. (B) The features are ranked from highest to lowest importance based on their mean absolute scores. The overall importance of various features in a data set, as measured by the mean absolute score from an InterpretML analysis. The features are ranked from highest to lowest importance based on their mean absolute scores.

that increasing age significantly contributed to the likelihood of admission (Figure 1). However, given the lack of data relating to socioeconomic and other factors, age becomes broad and may represent a more generic characterisation of the consumer and their health (Caruana et al. 2015); an older person is more likely to have higher health needs and comorbidities that necessitate admission. However, age is a critical factor that may explain the model (Caruana et al. 2015). There are points in the age data that show a clear 'step' in the correlation relationship (Figure 4).

3.2 | Triage Category

Triage is similarly weighted to age and is the only element deemed a clinical assessment. The overall figure shows a high

14470349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/inm.13402 by Clive Higgins - The University Of Newcastle , Wiley Online Library on [29/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

FIGURE 4 | Age overall. Two charts—the top chart displays a line graph depicting the relationship between 'age' and the likelihood of admission to a mental health inpatient ward, while the bottom chart displays a bar graph showing the frequency or count of various categories along the *x*-axis.

and stepped correlation with admission, with a triage code of 1 showing a positive correlation down to 5 having a negative correlation, as would be expected, with both sites exhibiting very similar characteristics (Figures 6 and 7A,B).

3.3 | Presenting Problem Category

Overall, 'Altered Mental Status' (0.3737), 'Behavioural Disturbance' (0.3065), 'Eating Disorder' (0.4245), 'Hallucinations' (0.3826) and 'Memory Impairment' (0.2667) all have a positive correlation to admission. 'Anxiety' (−0.5135), 'Depression' (−0.0613), 'Self Harm' (SH) (−0.0997) and 'Suicidal Ideation' (SI) (−0.0945) have a negative correlation (Figure 8). It should be noted that while SI and SH have a negative correlation, their overall influence on the model is minimal. Both sites present very similar correlations in the presenting problem data. Site B (Figure 9B) has a slightly higher positive correlation for 'Hallucinations' than Site A (Figure 9A). Interestingly, while both sites have a negative correlation for SI and SH for admission, the weight they contribute is inverse. SI has a stronger negative correlation at Site A than Site B and SH at Site A compared to Site B (Figure 9A,B).

3.4 | First Nations Category

According to the model, a person's 'Indigenous status' plays a minor role in determining a person's likelihood of admission. However, it should be noted that the subcategory of 'Neither Aboriginal nor Torres Strait Islander origin', people that do not identify as First Nations, is by far the largest group and has a flat correlation profile for admission. The next significant subcategory is 'Aboriginal but not Torres Strait Islander origin', which has a minor correlation with admission. Both sites show very similar results; however, it is noted that there is a stronger negative correlation at Site A (Figure 10A) for those that identify as 'Aboriginal but not Torres Strait Islander origin' compared to Site B (Figure 10B).

3.5 | Service User

There is a positive correlation between a presentation being an active service consumer and resulting in admission. In comparison, those that were not active at the time of presentation have a minimal negative correlation.

3.6 | Marital Status

The ML model found a positive correlation with admission if the person stated they were 'Divorced', 'Married (including de facto)', 'Separated', 'Unknown' or 'Widowed'. 'Never married' and 'Declined to respond' had a negative impact on admission. However, it should be noted that these data included people under the age of 16 and may not be a true reflection of the effect of 'Never married' or 'Declined to respond'.

3.7 | Month, Day of the Week and Time of the Day

The ML model found a minimal correlation towards the winter months; however, the volume of presentations is lower during this time. When comparing the admitted/not admitted groups, it can be observed that there is a consistent volume of data in presentations over 5 years that resulted in an admission. By contrast, there are significantly higher presentations to ED occurring in the summer months, with the number of people that present to ED changing over the year. However, the number of people admitted remains consistent despite these fluctuations (Figure 11). Even though the number of presentations fluctuates over the period in question, the admission rate remains flat. Overall,

14470349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/inm.13402 by Clive Higgins - The University Of Newcastle , Wiley Online Library on [20/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

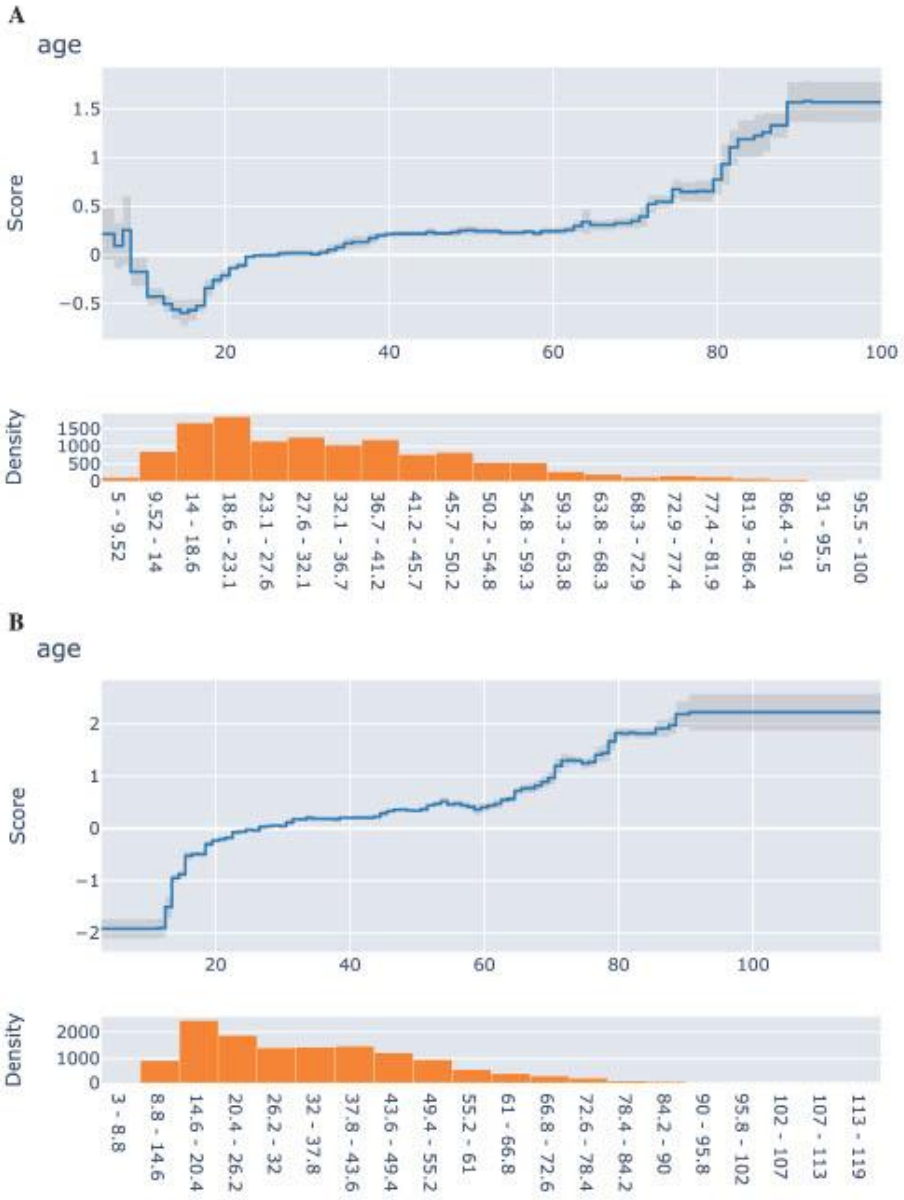


FIGURE 5 | (A) Age. Two charts—the top chart displays a line graph depicting the relationship between 'age' and the likelihood of admission to a mental health inpatient ward, while the bottom chart displays a bar graph showing the frequency or count of various categories along the *x*-axis. (B) depicts two charts—the top chart displays a line graph depicting the relationship between 'age' and the likelihood of admission to a mental health inpatient ward, while the bottom chart displays a bar graph showing the frequency or count of various categories along the *x*-axis.

according to the ML model, the time of day that a person presents to the hospital has only a small impact. A slightly high positive correlation occurs between 5:00 AM and 8:00 AM, and a small negative correlation occurs between 10:00 PM and 3:00 AM.

3.8 | Mode of Arrival

The ML model found a high correlation for 'Air ambulance', 'Internal Ambulance', 'Internal Bed/Wheelchair' and 'Other' for *Mode of arrival*. However, the incidence of these occurring in the data was very small. 'State Ambulance' had a small positive correlation, and 'Private vehicle' had a small negative correlation.

4 | Discussion

The findings of this research indicate that the ML techniques used are suitable to highlight site-specific dynamics and offer service providers an opportunity to tailor and target service provision aligned to community requirements. The ML model provided has given a deeper understanding of the complexities of ED presentation for MH, with two different models emerging when

14470349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/inm.13402 by Oliver Higgins - The University Of Newcastle , Wiley Online Library on [30/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

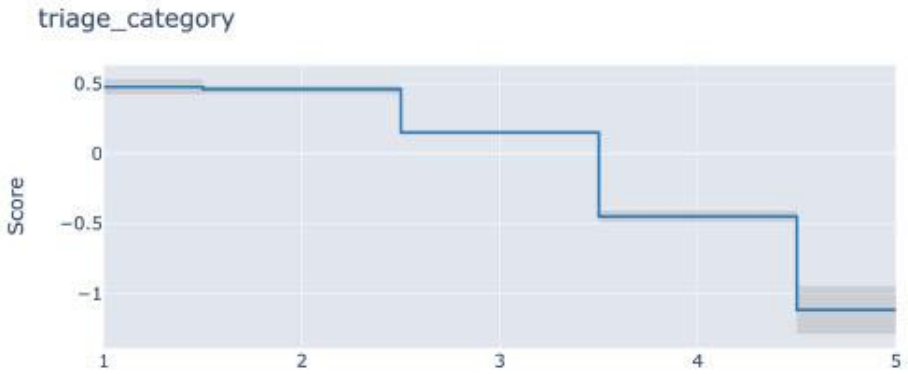


FIGURE 6 | Triage category (overall). Resuscitation (1), emergency (2), urgent (3), semi-urgent (4), and non-urgent (5). A positive value has a higher likelihood of admission.

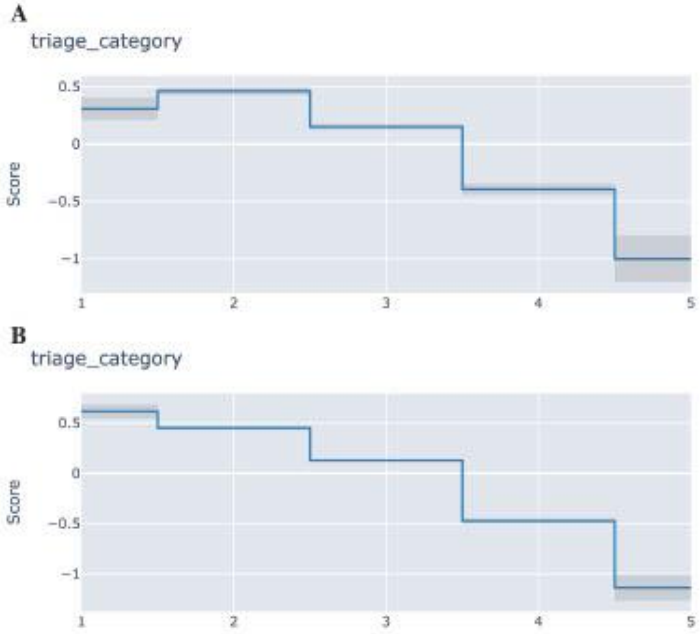


FIGURE 7 | (A) Triage. Resuscitation (1), emergency (2), urgent (3), semi-urgent (4) and non-urgent (5). A positive value has a higher likelihood of admission. (B) Resuscitation (1), emergency (2), urgent (3), semi-urgent (4) and non-urgent (5). A positive value has a higher likelihood of admission.

examining the different sites and the relationship to service delivery. The results have highlighted that age, triage, location and the presenting problem significantly influence admission.

Separating the data into the two site models reveals a similar step pattern that mirrors their service models. It should be noted that the service delivery models used in the locations in this research are based on age groups. Overall, there are no under 18 inpatient MH beds, 69 18–65 MH beds and 15 over 65 MH beds. The 15 65+ beds are located at Site B as well as four PECC beds. Site A only has 30 18–65 beds but does have access to a general paediatric inpatient ward. The resulting age models (Figure 5A,B) show an admission profile structured similar to the available beds profile, indicative of the service delivery model. One possible explanation for this is that the service delivery models change at these points, similar to variations found by Caruana et al. (2015). It may be clinicians, consciously or subconsciously, treating consumers in these brackets differently, ultimately affecting the likelihood of admission (Caruana et al. 2015). This leads to two critical points: models need to be built in place relevant to the service availability, and any service delivery model changes will result in the model being void or require the model to be edited accordingly (Caruana et al. 2015).

4.1 | Triage

In NSW, ED presentations are rated using the NSW Health triage scale. People who are critically ill and require immediate treatment are classified as Code 1 (NSW Health 2013) and someone who requires less urgent care, such as minor illness or symptoms that have been present for more than 1 week, is classified as Code 5 (NSW Health 2013). The triage code data in this data

14470349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ajm.13402 by Olivia Higgins - The University Of Newcastle , Wiley Online Library on [30/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

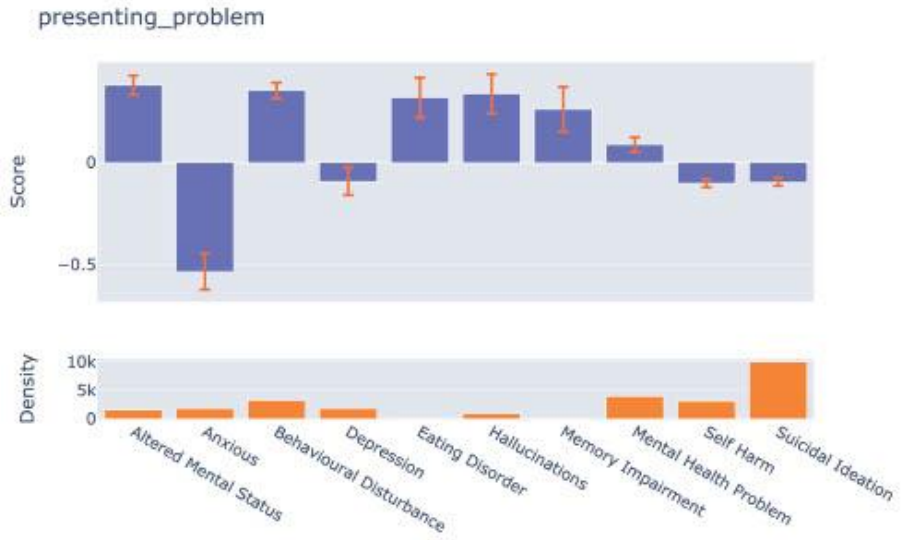


FIGURE 8 | Presenting problem. Plot of the 'presenting_problem' feature. The *y*-axis represents the score values, and the *x*-axis displays the different categories of the 'presenting_problem' feature. The plot consists of a series of points, each representing a score value for a particular 'presenting_problem' category. Positive scores indicate an increased likelihood of admission, while negative sources indicate a decreased likelihood.

set aligns with the NSW Health guidelines (NSW Health 2013). There is a positive correlation for admission with Code 1, progressing to a negative correlation for Code 5. Additionally, triage and ED access are local, and it is difficult to compare results in context, even in NSW, as models of triage, preparation and education of emergency nurses are highly variable, and auditing of triage practice occurs infrequently (Varndell, Hodge, and Fry 2019). Furthermore, the importance and responsibility for timely and accurate data entry by the triage nurse should be noted. Data integrity is a fundamental nursing skill that nurses are accustomed to collecting at the bedside (e.g., vital signs), but an enhanced appreciation of the ramifications of accurate data collection is paramount to understanding the importance and value of this data entry. The collection needs to be done with precision integrity due to clinical outcomes associated with the data and will allow for more accurate ML models to be built in the future.

The triage of people with MH problems requires high-level competencies in terms of knowledge, skills and attitudes in unpredictable and complex patient situations (Stigter-Outshoven et al. 2024). Within Australia, triage is an autonomous role predominantly conducted by specialist emergency nurses (Varndell, Hodge, and Fry 2019) and initiates the assessment process for a person who presents to the ED with an MH problem. Timely and accurate assessment, followed by appropriate referral and response, depends on the clinician's understanding of the MH triage and acuity (Broadbent, Moxham, and Dwyer 2020). The triage process is further impacted by conflicting orientations between the two nursing groups with different professional cultures (Broadbent, Moxham, and Dwyer 2020). ED triage nurses have expressed fear, frustration, lack of knowledge and confidence when dealing with people who present with MH problems (Perrone McIntosh 2021). Hall et al. (2016) demonstrated that extended MH professional development training is beneficial for ED triage nurses; however, the nurses in the study acknowledge that their role is primarily driven by time constraints and the volume of people that present to the ED. The resulting gap that emerges in this context is significant, MH triage is complex and requires an extensive understanding of MH (Broadbent, Moxham, and Dwyer 2020). It represents a significant challenge to ED models of care (MoC) that are challenged to support the volume of people seeking MH care, many of whom leave without treatment (Higgins et al. 2024a, 2024b). Given the complexity of the MH presentations, introducing the MH nurse earlier at the point of triage may benefit the ED triage nurse by reducing their workload and, more importantly, provide the person seeking MH care with a clinician who possesses the required MH knowledge and training to understand their needs. Furthermore, Australia's current comprehensive pre-registration nursing education curriculum is perceived as inadequate for preparing nurses for work in MH settings (Lakeman et al. 2024). Therefore, specialist MH nurses with post-graduate qualifications are best suited to supporting MH nursing triage. Addressing triage issues for MH in the ED MoC will require the development of an integrated MH triage MoC that would incorporate the required MH knowledge and alternative pathways to ED, and provide a clear path for the MH needs of the people who need care.

4.2 | Access

The findings for age reporting are important in the context of service delivery and the subsequent differences between the two sites. Admission is ultimately a function of bed availability, and it stands to reason that if a particular bed type is available, it is likely to result in proximal admission. The density of the young people presenting is significantly higher than the older age groups. The lack of Child and Adolescent Mental Health Service (CAMHS) beds and the high density

14470349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/inm.13402 by Curtin University , Wiley Online Library on [20/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

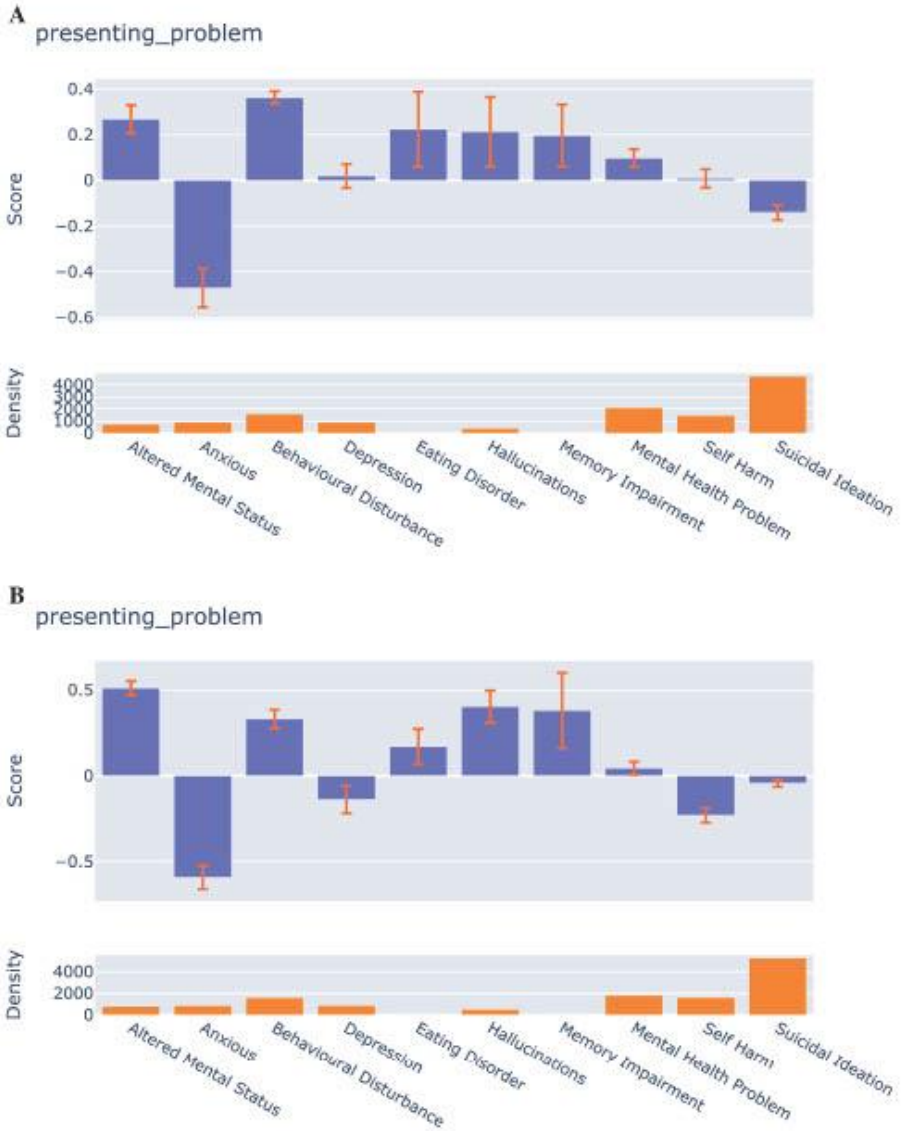


FIGURE 9 | (A) Presenting Problem. Plot of the 'presenting_problem' feature. The *y*-axis represents the score values, and the *x*-axis displays the different categories of the 'presenting_problem' feature. The plot consists of a series of points, each representing a score value for a particular 'presenting_problem' category. Positive scores indicate an increased likelihood of admission, while negative sources indicate a decreased likelihood. Plot of the 'presenting_problem' feature. The *y*-axis represents the score values, and the *x*(B)-axis displays the different categories of the 'presenting_problem' feature. The plot consists of a series of points, each representing a score value for a particular 'presenting_problem' category. Positive scores indicate an increased likelihood of admission, while negative sources indicate a decreased likelihood.

of under 18 presentations are likely combined to facilitate the negative correlation in admission (Figure 4). Therefore, there is a need to examine access to acute MH care for young people in the context of ED presentations. Recent research recommendations by Suicide Prevention Australia call for co-designed youth-specific environments to provide young people with medical assistance and compassionate care (Suicide Prevention Australia 2022). This care should be delivered in a way that gives clear processes, improves accessibility and is supported by specific suicide response staff, including youth peer workers (Suicide Prevention Australia 2022). While significant steps have been made with non-clinical spaces, there is still a significant need to integrate these new models into the current MoC and extend the services beyond the current limitations to provide 24/7 support (Shams 2023).

4.3 | Young Peoples Access to Mental Health Care

In Australia in 2022, 77 deaths by suicide occurred for young people (aged 17 and below), with 83.1% occurring in the 15–17 age bracket (Australian Institute of Health and Welfare 2023a). This is an increasing concern, as rates of death by suicide among 15- to 17-year-olds ranged between a low of 3.2 per

14670343, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ajsw.13402 by Oliver Higgins - The University Of Newcastle , Wiley Online Library on [29/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

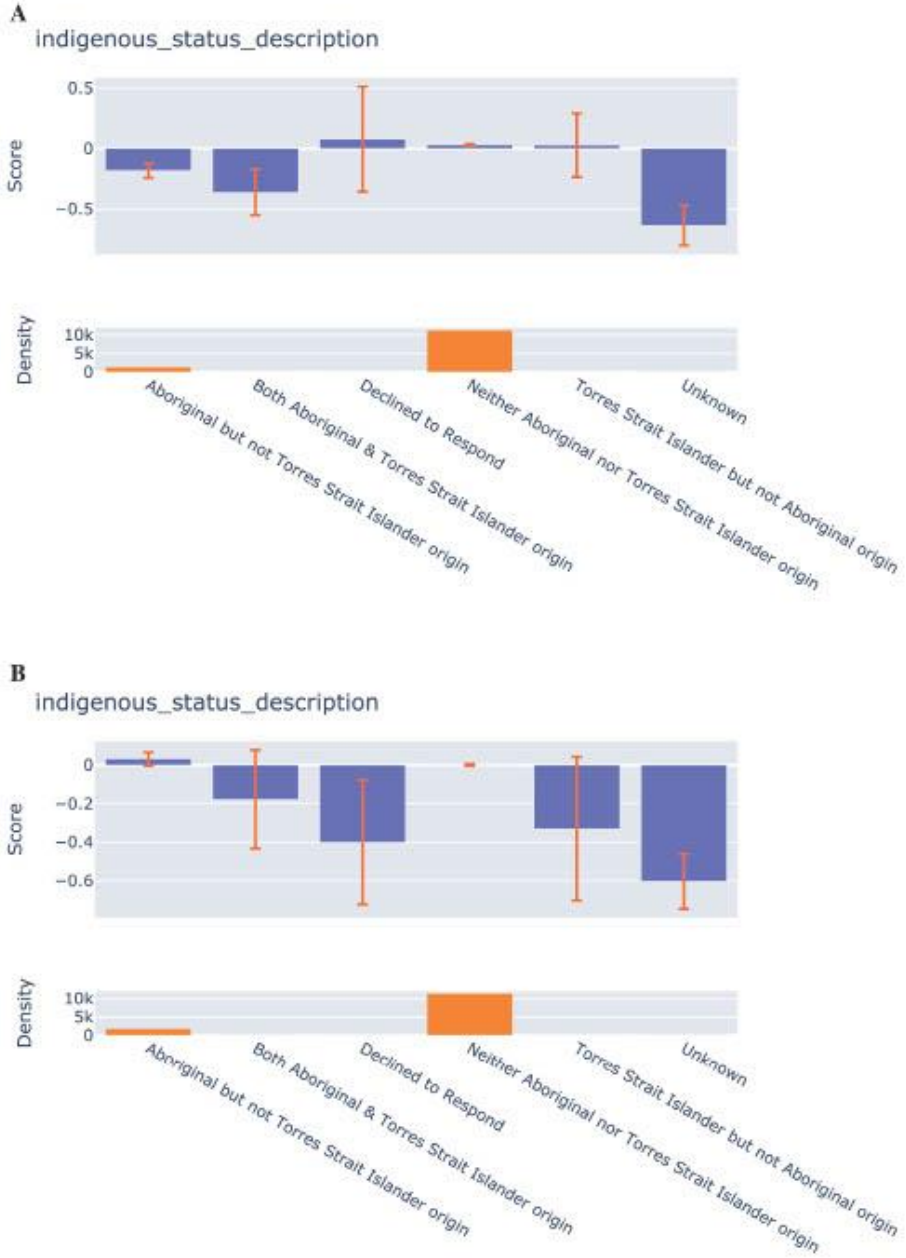


FIGURE 10 | (A) Plot of the 'indigenous_status_description' feature. The *y*-axis represents the score values, and the *x*-axis displays the different categories of the 'indigenous_status_description' feature. The plot consists of a series of points, each representing a score value for a particular 'indigenous_status_description' category. Positive scores indicate an increased likelihood of admission, while negative sources indicate a decreased likelihood. Plot of the 'indigenous_status_description' feature. The *y*-axis represents the score values, and the *x*-axis displays the different categories of the 'indigenous_status_description' feature. The plot consists of a series of points, each representing a score value for a particular 'indigenous_status_description' category. Positive scores indicate an increased likelihood of admission, while negative sources indicate a decreased likelihood.

100 000 population in 2004 and a high of 8.9 in 2021 (Australian Institute of Health and Welfare 2023a). Figure 4 shows a large volume of young people that present to the ED and a significant negative correlation to admission. The healthcare research setting in question does not have any dedicated CAMHS beds; however, State-based CAMHS beds exist in neighbouring local health districts.

However, even if a small number of beds were added, the volume of presentations would still exceed capacity and not address the depth of the need. For services without beds, there is a need to better understand the presentation cohort and provide services that align with the need. Borschmann et al. (2018) found that the bulk of admissions (78.2%) to a Melbourne CAMHS IPU were for suicidality, indicating the high need for young

14470349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/inm.13402 by Oliver Higgins - The University Of Newcastle , Wiley Online Library on [29/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

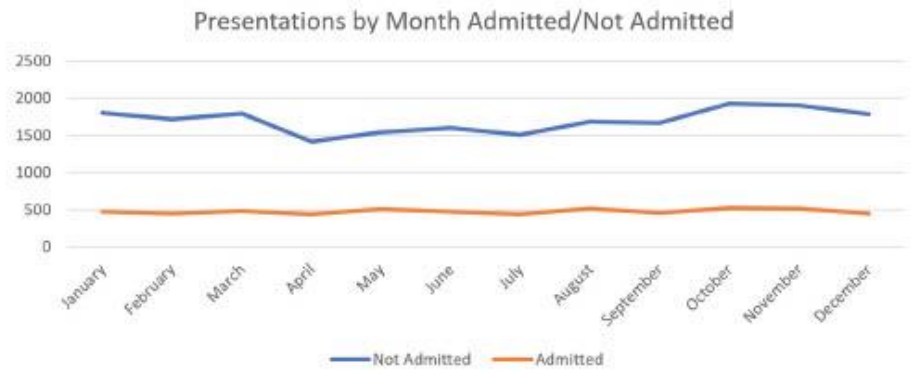


FIGURE 11 | Month. This chart depicts the number of presentations by month, divided into those that were admitted and those that were not admitted across all years. The blue line represents the number of presentations that were not admitted, while the orange line represents the number of presentations that were admitted. The data show a clear seasonal pattern, with the number of presentations peaking during the winter months (December, January, February) and declining during the summer months (June, July, August).

people's safety in times of being unwell. If there are no beds or even a low number of them, services should focus on suicidality and keeping young people safe as a whole of health response. Second, if units are built, the design needs to focus on the needs of young people and not simply be a satellite of general paediatric or adult MH beds. Environments need to be built considering young people's requirements and designed to promote physical security while supporting interactions between users and systems that promote relational security (Wilson, Hutton, and Foureur 2023). The users of the space should be able to rely on the built environment to ensure their physical security and psychological safety and be designed to accommodate a wide range of diversity and acuity (Wilson, Hutton, and Foureur 2023). The aesthetics of the space should align with the promotion of recovery in the context of person-centred and trauma-informed MoC (Wilson, Hutton, and Foureur 2023). The design principles should be conscious of the cultural safety of young First Nations people with cultural safety frameworks, such as those used in Australia (McGough et al. 2022), being incorporated to improve the access and quality of health care for First Nations people and to improve disparities in health care outcomes.

4.4 | Presenting Problem

'Presenting problem' is the most comprehensive MH data available in this data set and is collected at the point of triage before an MH assessment has been completed. The presenting problem description is comprised of the following options for MH; *Altered Mental Status*, *Anxious*, *Behavioural Disturbance*, *Depression*, *Eating Disorder*, *Hallucinations*, *Memory Impairment*, *Mental Health Problem*, *Self-Harm and Suicidal Ideation*. The ML model rated this item third overall for importance overall and across both sites. 'Altered MH status', 'Behavioural disturbance' and 'Hallucinations' are reported to correlate positively with admission. 'Eating disorder' and 'Memory impairment' also have a positive correlation but a very low density in the data. These problems are symptoms associated with psychotic-related mental illness (World Health Organization (WHO) 2021). SI negatively correlates to admission, yet it represents a high volume of presentations, accounting for 38.19% (Higgins et al. 2024b).

Both sites have a higher than average (per 100 000) death-by-suicide rate (Site A 12.9 and Site B 16.2) compared to the NSW state 11.2 and 12.6 for Australia for the period of 2016–2020 (Australian Institute of Health and Welfare 2023b). SI at Site B has an almost negligible influence on admission according to the model, which may be indicative of the difficulty in assessing people with this problem. However, this is present with low admission tendencies. There are likely to be contributing factors and relationships in the SI data beyond the scope of this paper.

5 | Limitations

Several limitations must be considered. First, the data rely on the presenting problem to identify MH presentations. This may exclude presentations that relate to intentional overdose, drug and alcohol issues or complex comorbidities. Second, in this context, SI is a Boolean result and does not reflect the spectrum of distress associated with this presentation. The extracted data period includes the COVID-19 pandemic and may show variations during the data collection phase. The data could be re-examined with a longitudinal lens and establish variation over time. While the data set determines whether a person is a current or previous consumer of the MH services, it does not differentiate if a person had a previous presentation to ED. The nature of the data used in the study only provides a simple Boolean response of the person's primary presenting problem, thereby limiting the conclusions that can be drawn. The presenting problem itself only has a limited number of options available, and clear definitions are not readily available. It is important to note that if the service delivery model changes, such as the addition of beds, the resulting model that has been built will no longer be accurate. Finally, there is some ambiguity in the data for the presenting problem, especially when SI is listed as the presenting problem. A data set that incorporates secondary presenting problems would provide a richer insight into this data.

14470349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/inm.13402 by Oliver Higgins - The University Of Newcastle , Wiley Online Library on [29/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

6 | Recommendations

R1/For Mental Health Nurses (MHN), ML has the capacity to help them make informed decisions earlier, improve the patient journey, assist in care navigation and provide efficiency in the triage process. MHNs must engage early in the development process to ensure that clean data are entered and clinically valid sources are used when building any model that relates to consumer care.

R2/Health services should follow a similar methodological approach using available data sets to validate and replicate the findings. Aligning the models closely with service delivery will help services understand their service users and provide insight into financial and clinical variations.

R3/The research highlights an incongruity in the high number of under 18 presentations and no beds, whereas, in other age groups, there are relatively lower presentations and higher bed utilisation and availability. There is a need to research similar health service profiles that have access to CAMHS beds to compare their service delivery model. Services should consider revisions of their CAMHS service models based on service profiles to establish bed: presentation ratios and consider alternative models, such as Youth Hospital in the Home (Towicz et al. 2021), where CAMHS beds are unavailable.

R4/SI negatively correlates to admission yet represents a high volume of presentations. While the ML model negatively correlates with admission and SI, this is across a substantial amount of the data. The service profile does not provide support for those presenting with SI. Furthermore, qualitative and peer-led research is required to understand this phenomenon's complexity and the needs of this population, especially among at-risk populations such as First Nations and young people.

R5/Further research is required to explore the role that ML can provide in reducing wait times by assisting in the prediction of admission or other suitable pathways. This work would require the development of different ML models, such as Random Forest or Support Vector Machines, in addition to InterpretML's EBM's, to establish the feasibility of using these models in practice.

R6/Addressing triage issues for MH will require the redesign of the current ED MoC to integrate MH nurses at the point of triage. The MoC would acknowledge the complexity of MH triage and incorporate the required MH knowledge requirements, returning valuable time to the ED triage nurses and, most importantly, providing a patient-centred focus for those requesting help for their MH.

7 | Conclusion

This research provides specific insight into the complexities for people seeking MH care in the ED across two sites in NSW, Australia. Many factors contribute to a person's experience and need for care; however, there is evidence that there are features that significantly correlate with a person's likelihood of being admitted in conjunction with the care delivery model of the health service. Significantly, this research has highlighted the gap that exists at the ED point of triage and the role that MH nurses play in addressing this need. Disparities in youth presentations and bed availability highlight the need for services to explore CAMHS bed profiles, revise service models and consider alternatives like Youth Hospital in the Home. Additionally, the complex correlation between SI and admission warrants further qualitative research, particularly focusing on at-risk populations and should be peer-led. Clinicians should actively contribute to the development process of MoC and ML models, emphasising clean data and clinically valid sources. Noting that any models built to assess or predict must be done using local data and rebuilt when service delivery models change. Health services are advised to align methodologies with available data sets, integrating ML findings closely with service delivery to understand service users and identify clinical variations. Finally, further investigation is required into the applicability and accuracy of ML models' ability to predict if a person requires admission from the ED for MH problems. In summary, this research provides insights into the complexities of MH care requirements in ED, emphasising the need for local data, adaptable models and deeper qualitative exploration of the lived experience of those who present.

8 | Relevance to Clinical Practice

The research sets forth a deeper understanding of the factors that contribute to admission via the ED. Understanding the needs of these people will not only help design service provision but also provide further evidence in understanding the large number of people who present with SI. The nurse's role at triage is a critical factor in assessing the needs of the presenting individual. The gap that emerges in this context is significant; MH triage requires a complex understanding of MH and presents a significant challenge in the ED MoC that does not support the volume of people seeking care. Recording their initial clinical impression/finding with clinical accuracy establishes quality and clean data collection at the outset and is the first link in the chain of logic that informs the decision-making capability within the ML model. The implications of quality data input are a crucial step in forming the chain of logic and ideally result in accurate and timely decisional support processes.

Author Contributions

Oliver Higgins was involved in concept development, project design, data collection, data analysis, manuscript preparation. Stephan K. Chalupwas involved in data analysis, contribution to manuscript, supervision of project. Rhonda L. Wilson was involved in concept development, project design, data collection, data analysis, manuscript contribution, supervision of project.

Acknowledgements

Authors would like to acknowledge the support of Central Coast Local Health District.

Ethics Statement

Ethical approval was granted by the Hunter New England Human Research Ethics Committee 2022/ETH01597. Central Coast Local Health District granted site-specific approval 2022/STE03296.

14470349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/inm.13400 by Oliver Higgins - The University Of Newcastle , Wiley Online Library on [29/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14470349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/inm.13402 by Clive Higgins - The University Of Newcastle , Wiley Online Library on [20/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Conflicts of Interest

Prof. Rhonda Wilson is an Editorial Board Member of the International Journal of Mental Health Nursing.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

Australian Bureau of Statistics. 2021. "2021 Census All Persons QuickStats Central Coast (NSW)." https://abs.gov.au/census/find-census-data/quickstats/2021/LGA11650.

Australian Institute of Health and Welfare. 2023a. "Deaths by Suicide Among Young People." https://www.aihw.gov.au/suicide-self-harm-monitoring/data/populations-age-groups/suicide-among-young-people.

Australian Institute of Health and Welfare. 2023b. "Suicide & Self-Harm Monitoring." https://www.aihw.gov.au/suicide-self-harm-monitoring/data/data-downloads.

Berendsen Russell, S., M. M. Dinh, and N. Bell. 2017. "Triage, Damned Triage ... and Statistics: Sorting Out Redundancy and Duplication Within an Emergency Department Presenting Problem Code Set to Enhance Research Capacity." *Australasian Emergency Nursing Journal* 20, no. 1: 48–52. https://doi.org/10.1016/j.aenj.2016.09.004.

Borschmann, R., P. Stark, C. Prakash, and S. M. Sawyer. 2018. "Risk Profile of Young People Admitted to Hospital for Suicidal Behaviour in Melbourne, Australia." *Journal of Paediatrics and Child Health* 54, no. 11: 1213–1220. https://doi.org/10.1111/jpc.13938.

Broadbent, M., L. Moxham, and T. Dwyer. 2020. "Understanding Nurses Perspectives of Acuity in the Process of Emergency Mental Health Triage: A Qualitative Study." *Contemporary Nurse* 56, no. 3: 280–295. https://doi.org/10.1080/10376178.2020.1841014.

Bureau of Health Information. 2022. "Technical Supplement Emergency Department, Ambulance, Admitted Patients, Seclusion and Restraint, and Elective Surgery." *Bureau of Health Information* 1: 5. https://www.bhi.nsw.gov.au/__data/assets/pdf_file/0009/715194/BHI_HQ47_OCT-DEC_2021_TECH_SUPP.pdf.

Caruana, R., Y. Lou, J. Gehrke, P. Koch, M. Sturm, and N. Elhadad. 2015. "Intelligible Models for Healthcare: Predicting Pneumonia Risk and Hospital 30-Day Readmission." In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1721–1730. https://doi.org/10.1145/2783258.2788613.

Central Coast Local Health District. 2017. Clinical Services Plan 2017-2022. https://www.cclhd.health.nsw.gov.au/wp-content/uploads/CCLHD ClinicalServicesPlan2017-2022.pdf.

Clarke, D. E., D. Dusome, and L. Hughes. 2007. "Emergency Department From the Mental Health Client's Perspective." *International Journal of Mental Health Nursing* 16, no. 2: 126–131. https://doi.org/10.1111/j.1447-0349.2007.00455.x.

Fatovich, D., and R. Hirsch. 2003. "Entry Overload, Emergency Department Overcrowding, and Ambulance Bypass." *Emergency Medicine Journal* 20, no. 5: 406–409.

Geng, F., F. Jiang, R. Conrad, et al. 2020. "Factors Associated With Involuntary Psychiatric Hospitalization of Youths in China Based on a Nationally Representative Sample." *Frontiers in Psychiatry* 11: 1–7. https://doi.org/10.3389/fpsyt.2020.607464.

Hall, A., B. McKenna, V. Dearie, T. Maguire, R. Charleston, and T. Furness. 2016. "Educating Emergency Department Nurses About Trauma Informed Care for People Presenting With Mental Health Crisis: A Pilot Study." *BMC Nursing* 15, no. 1: 21. https://doi.org/10.1186/s12912-016-0141-y.

Higgins, O., R. B. Sheather-Reid, S. K. Chalup, and R. L. Wilson. 2024a. "Disproportionate Mental Health Presentations to Emergency Departments in a Coastal Regional Community in Australia of First Nation People." *International Journal of Mental Health Nursing*. Published ahead of print, May 26, 2024. https://doi.org/10.1111/inm.13362.

Higgins, O., R. B. Sheather-Reid, S. K. Chalup, and R. L. Wilson. 2024b. "Sociodemographic Factors and Presentation Features of Individuals Seeking MH Care in EDs: A Retrospective Cohort Study." *International Journal of Mental Health Nursing (Under Review)*.

InterpretML Documentation. 2023. "Overall Importance." https://interpret.ml/docs/overall_importance/.

Lakeman, R., K. Foster, B. Happell, M. Hazelton, L. Moxham, and J. Hurley. 2024. "Informing the Development of a Fit-for-Purpose Mental Health Nursing Curriculum: A Survey of Mental Health Nurse Academics in Australia." *International Journal of Mental Health Nursing* 33, no. 1: 93–103. https://doi.org/10.1111/inm.13226.

Lay, B., W. Kawohl, and W. Rössler. 2019. "Predictors of Compulsory Re-Admission to Psychiatric Inpatient Care." *Frontiers in Psychiatry* 10: 120. https://doi.org/10.3389/fpsyt.2019.00120.

McGough, S., D. Wynaden, S. Gower, R. Duggan, and R. Wilson. 2022. "There Is No Health Without Cultural Safety: Why Cultural Safety Matters." *Contemporary Nurse* 58, no. 1: 33–42. https://doi.org/10.1080/10376178.2022.2027254.

McKinney, W. 2010. Data structures for Statistical Computing in Python. Proceedings of the 9th Python in Science Conference.

Molnar, C. 2020. "Interpretable Machine Learning." *Leanpub* 1: 15–29. https://christophm.github.io/interpretable-ml-book/.

Morphet, J., K. Innes, I. Munro, et al. 2012. "Managing People With Mental Health Presentations in Emergency Departments—A Service Exploration of the Issues Surrounding Responsiveness From a Mental Health Care Consumer and Carer Perspective." *Australasian Emergency Nursing Journal* 15, no. 3: 148–155. https://doi.org/10.1016/j.aenj.2012.05.003.

Nori, H., S. Jenkins, P. Koch, and R. Caruana. 2019. "Interpretml: A Unified Framework for Machine Learning Interpretability." *arXiv preprint arXiv*: 1909. 09223.

NSW Health. 2013. "Triage of Patients in NSW Emergency Departments." https://www1.health.nsw.gov.au/pds/ActivePDSDocuments/PD2013_047.pdf.

NSW Health. 2017. "ED Mode of Separation Code Set Update 2017/18." *NSW Health*, 1–5. https://www1.health.nsw.gov.au/pds/ActivePDSDocuments/IB2017_028.pdf.

Perrone McIntosh, J. T. 2021. "Emergency Department Nurses' Care of Psychiatric Patients: A Scoping Review." *International Emergency Nursing* 54: 100929. https://doi.org/10.1016/j.ienj.2020.100929.

Rai, A. 2019. "Explainable AI: From Black Box to Glass Box." *Journal of the Academy of Marketing Science* 48, no. 1: 137–141. https://doi.org/10.1007/s11747-019-00710-5.

Sara, G., J. Wu, J. Uesi, et al. 2022. "Growth in Emergency Department Self-Harm or Suicidal Ideation Presentations in Young People: Comparing Trends Before and Since the COVID-19 First Wave in New South Wales, Australia." *Australian and New Zealand Journal of Psychiatry* 57, no. 1: 58–68. https://doi.org/10.1177/00048674221082518.

Shams, H. 2023. *Young People in Mental Health Distress Need Better Alternatives to Emergency Departments, Advocates Say*. Mad in America. https://www.abc.net.au/news/2023-02-18/emergency-departments-not-set-up-for-youth-mental-health-report/101855202.

Sperandei, S., A. Page, P. Bandara, et al. 2022. "The Impact of the COVID-19 Pandemic on Hospital-Treated Self-Harm in Sydney (Australia)." *Australian and New Zealand Journal of Psychiatry* 56, no. 11: 1515–1522. https://doi.org/10.1177/00048674211068393.

Stigter-Outshoven, C., G. Van de Glind, L. J. Wieberdink, R. van Zelm, and A. Braam. 2024. "Competencies Emergency and Mental Health Nurses Need in Triage in Acute Mental Health Care: A Narrative Review." *Journal of Emergency Nursing* 50, no. 1: 55–71. https://doi.org/10.1016/j.jen.2023.08.005.

Suicide Prevention Australia. 2022. "In Their Words: How to Support Young People in Suicidal Distress." https://www.suicidepreventionaust.org/wp-content/uploads/2022/08/SPA_Youth-Report_In-your-words_12-August-2022-1.pdf.

Sunderji, N., J. T. de Bibiana, and V. Stergiopoulos. 2015. "Urgent Psychiatric Services: A Scoping Review." *Canadian Journal of Psychiatry* 60, no. 9: 393–402. https://doi.org/10.1177/070674371506000904.

Tabachnick, B. G., L. S. Fidell, and J. B. Ullman. 2013. *Using Multivariate Statistics*. Vol. 6, 497–516. Boston, MA: Pearson.

Towicz, M., W. Yang, S. Moylan, R. Tindall, and M. Berk. 2021. "Hospital-in-the-Home as a Model for Mental Health Care Delivery: A Narrative Review." *Psychiatric Services* 72, no. 12: 1415–1427. https://doi.org/10.1176/appi.ps.202000763.

Varndell, W., A. Hodge, and M. Fry. 2019. "Triage in Australian Emergency Departments: Results of a New South Wales Survey." *Australasian Emergency Care* 22, no. 2: 81–86. https://doi.org/10.1016/j.auec.2019.01.003.

Walker, S., E. Mackay, P. Barnett, et al. 2019. "Clinical and Social Factors Associated With Increased Risk for Involuntary Psychiatric Hospitalisation: A Systematic Review, Meta-Analysis, and Narrative Synthesis." *Lancet Psychiatry* 6, no. 12: 1039–1053. https://doi.org/10.1016/S2215-0366(19)30406-7.

Wand, T., N. D'Abrew, C. Barnett, L. Acret, and K. White. 2015. "Evaluation of a Nurse Practitioner-Led Extended Hours Mental Health Liaison Nurse Service Based in the Emergency Department." *Australian Health Review* 39, no. 1: 1–8. https://doi.org/10.1071/AH14100.

Wilson, R. L., A. Hutton, and M. Foureur. 2023. "Promoting Mental Health Recovery by Design: Physical, Procedural, and Relational Security in the Context of the Mental Health Built Environment." *International Journal of Mental Health Nursing* 32, no. 1: 147–161. https://doi.org/10.1111/inm.13070.

World Health Organization. 2022. "World Mental Health Report: Transforming Mental Health for all." *World Health Organization* 1: 46. https://apps.who.int/iris/handle/10665/356119.

World Health Organization (WHO). 2021. *International Classification of Diseases Eleventh Revision (ICD-11)*. Geneva, Switzerland: World Health Organization.

14470349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/inm.13402 by Clive Higgins - The University Of Newcastle , Wiley Online Library on [20/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

7.4.2 Article Summary

The research findings highlight the potential of ML techniques to provide insights into the complexities of ED presentations for MH care, with site-specific dynamics emerging from the analysis. Age, triage category, location, and the presenting problem were identified as significant factors influencing admission. However, the results also revealed disparities, such as the high volume of youth presentations and the lack of dedicated Child and Adolescent Mental Health Service (CAMHS) beds, highlighting the need for revised service models and alternative approaches like Youth Hospital in the Home.

Furthermore, the complex correlation between suicidal ideation and admission warrants further qualitative research, particularly focusing on at-risk populations like First Nations and young people. The study emphasises the importance of clinicians contributing to the development process of models of care and ML models, ensuring clean data and clinically valid sources. To address the identified challenges, the research recommends revising the current ED model of care to integrate MH nurses at the point of triage, acknowledging the complexity of MH triage and incorporating the required MH knowledge. Additionally, health services should align methodologies with available datasets, integrating ML findings closely with service delivery to understand service users and identify clinical variations.

This study laid a strong foundation for Study D by providing critical insights into the factors influencing admissions from EDs for MH care. By identifying key variables such as age, triage category, location, and presenting problem, Study C revealed the complex dynamics shaping admission decisions, particularly the high volume of youth presentations. These findings highlight the significance of integrating ML techniques to enhance understanding of ED presentations, as well as the necessity of refining care pathways, especially for vulnerable populations such as First Nations and young people presenting with suicidal ideation.

Further investigation is needed into the applicability and accuracy of ML models' ability to predict whether a person requires admission from the ED for MH problems. The research provides insights into the complexities of MH care requirements in EDs, emphasising the need for local data, adaptable models, and deeper qualitative exploration of the lived experience of those who present. The correlations uncovered in Study C formed the basis for the predictive modelling work in Study D. While Study C focused on uncovering relationships between presentation features and admissions, Study D took this further by developing and testing ML models to predict whether an individual will require admission to a MH inpatient facility. Study D built upon these findings to create actionable tools that improve clinical decision making, address service gaps, and enhance the triage process, ultimately aiming to streamline admissions and optimise resource allocation in EDs. The study's implications for clinical practice lie in understanding the factors that contribute to admission via the ED, which can inform service provision and provide evidence for understanding the large number of people who present with suicidal ideation. The nurse's role at triage is critical in assessing the needs of the presenting individual, and accurate data collection at this stage is crucial for informing decision-making capabilities within ML models.

7.5 Using Machine Learning to Assist Decision Making in the Assessment of Mental Health Consumers Presenting to Emergency Departments

Higgins, O., Wilson, R. L., & Chalup, S. K. (2024d). Using machine learning to assist decision making in the assessment of mental health consumers presenting to emergency departments. *Digital Health*. https://doi.org/10.1177/20552076241287364

	Higgins, O	Chalup, S. K	Wilson, R. L
Conceptualization	X	X	X
Methodology	X	X	
Software	X	X	
Validation	X	X	

Formal analysis	X		
Investigation	X		
Resources	X		
Data Curation	X		
Writing - Original Draft	X		
Writing - Review & Editing	X	X	X
Visualization	X		
Supervision		X	X
Project administration	X		
Funding acquisition			
Percentage	**70%**	**20%**	**10%**

7.5.1 Article Introduction

This research aims to assess the predictability of admissions to a MH inpatient ward using ML modelling based on routine data collected at the point of triage in the ED. Suicide and self-harm represent major global public health concerns with devastating impacts (World Health Organization, 2022b). In Australia, there is an average of nine deaths by suicide each day (Australian Institute of Health and Welfare, 2023c). Many individuals seek initial or acute MH assistance at hospital EDs, which can be loud, bright environments operating 24/7 (Clarke et al., 2007). However, ED staff often feel ill-prepared to handle MH consumers, with long wait times and consumer dissatisfaction common issues (Clarke et al., 2014). Applying ML techniques to analyse ED data on MH presentations offers a promising approach for early detection and prevention of adverse outcomes like suicide and self-harm (Iorfino et al., 2020; Raita et al., 2019). Previous research has demonstrated the potential of ML in predicting suicidal ideation, self-harm behaviours, and need for admission, though challenges remain around bias, transparency, and gaining clinician trust in these models (Iorfino et al., 2020; Raita et al., 2019).

This study built directly upon the findings of Study C by advancing the application of ML techniques to predict admissions to MH inpatient wards from ED presentations. While Study C focused on uncovering correlations between key presentation features—such as age, triage

category, and presenting problem—and admission outcomes, Study D took this further by assessing the predictability of admissions using ML models. Study C highlighted significant factors influencing admissions and the need for accurate data collection and clinicians' involvement, which served as the foundation for developing predictive models in Study D.

The study aimed to develop and test both interpretable ML models using techniques like explainable boosting machines (EBMs) as well as more traditional non-linear models like support vector machines (SVMs). A key focus was ensuring transparency and interpretability to facilitate clinicians' acceptance of the models for practical use in MH care decision making.

Check for updates

Original Research Article

DIGITAL HEALTH

DIGITAL HEALTH
Volume 10: 1–17
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20552076241287364
journals.sagepub.com/home/dhj
Sage

Using machine learning to assist decision making in the assessment of mental health patients presenting to emergency departments

Oliver Higgins[1,2,3], Rhonda L. Wilson[1,2,3] and Stephan K. Chalup[4]

Abstract

Objective: The objective of this study was to assess the predictability of admissions to a MH inpatient ward using ML models, based on routine data collected during triage in EDs. This research sought to identify the most effective ML model for this purpose while considering the practical implications of model interpretability for clinical use.

Methods: The study utilised existing data from January 2016 to December 2021. After data pre-processing, an exploratory analysis revealed the non-linear nature of the dataset. Six different ML models were tested: Random Forest, XGBoost, CatBoost, k-Nearest Neighbours (kNN), Explainable Boosting Machine (EBM) using InterpretML, and Support Vector Machine using Support Vector Classification (SVC). The performance of these models was evaluated using various metrics including the Matthews Correlation Coefficient (MCC).

Results: Among the models evaluated, the CatBoost model achieved the highest MCC score of 0.1952, demonstrating superior balanced accuracy and predictive power, particularly in correctly identifying positive cases. The InterpretML model also performed well, with an MCC score of 0.1914. While CatBoost showed strong predictive capabilities, its complexity poses challenges for clinical interpretation. Conversely, the InterpretML model, though slightly less powerful, offers better transparency and is more practical for clinical use.

Conclusion: The findings suggest that the CatBoost model is a compelling choice for scenarios prioritising the detection of positive cases. However, the InterpretML model's ease of interpretation makes it more suitable for clinical application. Integrating explanation methods like SHAP with non-linear models could enhance model transparency and foster clinician trust. Further research is recommended to refine non-linear models within decision support systems, explore multi-source data integration, understand clinician attitudes towards ML, and develop real-time data collection systems. This study highlights the potential of ML in predicting MH admissions from ED data while stressing the importance of interpretability, ethical considerations, and ongoing validation for successful clinical implementation.

Keywords

Mental health, emergency department, first nation, aboriginal and torres strait islander

Submission date: 2 June 2024; Acceptance date: 10 September 2024

Introduction

Suicide and self-harm are global public health concerns with devastating consequences for individuals, families, and communities.[1] Approximately 5% of the adult world population have depression,[1] and approximately 700,000 people die by suicide each year.[1] In Australia, there was

[1]School of Health and Biomedical Sciences, RMIT University, Melbourne, Australia
[2]Department of Mental Health, Central Coast Local Health District, Gosford, NSW, Australia
[3]Central Coast Research Institute, Gosford, NSW, Australia
[4]School of Information and Physical Sciences (Computer Science and Software Engineering), University of Newcastle, Australia

Corresponding author:
Oliver Higgins, C/O Rhonda Wilson, 77a Holden St, Gosford, NSW 2250 Australia.
Email: Oliver.Higgins@rmit.edu.au

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (https://us.sagepub.com/en-us/nam/open-access-at-sage).

an average of nine deaths by suicide each day in 2021.[2] Many people seek initial, acute or emergency mental health (MH) assistance at their local hospital emergency department (ED) with recent research identifying three main themes that characterise help-seeking dynamics for MH conditions in ED: First Nations MH, suicidal ideation (SI), and access and egress pathways.[3,4] EDs are busy environments populated by clinical, administrative, security, visiting staff (ambulance, police and other agencies), and with patients typically arriving by their own means, via ambulance, police, carers, or family. They are characterised with loud noise, bright lights, busy activity and usually open every day across the 24 h period.[5] ED staff frequently express feeling unequipped to provide care for MH patients,[5] with longer waiting periods common following arrival to ED for MH presentations[6] leading to dissatisfaction among many MH patients in ED.[7] For those unwell enough to require admission or transfer to an inpatient facility, the stay in the ED is generally more extended than those discharged from ED.[8] One potential approach to detect MH presentations earlier in the ED experience is to enhance the analysis and interpretation of regularly collected data pertaining to patient presentations in EDs. It involves the application of Machine Learning (ML), which has emerged as a powerful tool in healthcare, offering promising avenues for predicting significant risks such as suicide and self-harm.[9–13] Therefore, the aim of this research is to assess the predictability of admissions to a MH inpatient ward using ML modelling from routine data collected at the point of triage in the ED.

However, the usage of ML is not without its challenges and ethical considerations. One of the crucial issues is the potential for bias in ML models when dealing with vulnerable populations and the possibility of producing further discrimination.[14] Ensuring transparency and interpretability of ML models is equally important, as it is essential for gaining support among clinicians and patients.[15] Previous research into the use ML in MH for the prediction of SI and self-harm behaviours has yielded positive results, for example, Reale, Novak[13] researched the development of an intervention that used predictive analytics to inform care teams about their patient's risk of suicide attempts. Iorfino, Ho[11] developed self-harm monitoring ML models, which helped identify a large subpopulation that would benefit from targeted interventions and helped identify factors that contribute to self-harming behaviours. Similarly, research has been undertaken into the use of ML for the prediction of admission by Hong, Haimovich[10] testing three different ML models in the ED setting that used a combination of triage and patient history data. Raita, Goto[12] tested four machine learning (ML) models that all outperformed the benchmark set in the study. The models were built to predict outcomes related to critical care and hospitalisation, and improved clinicians' decision-making process.[12] In a recent systematic review of clinical decision support systems (CDSS) in ED, Fernandes, Vieira[9] reported that the use CDSS improved decision-making by health professionals and led to better clinical management and patient outcomes. However, Fernandes, Vieira[9] go on to state that more than half of the studies lacked clear implementation and performance measures. Contrasting the previous studies, this research will use a minimal dataset collected at the point of entry to the ED where only triage diagnostics have occurred. This study seeks to determine if basic presentation information can be used to predict the need for admission to an acute inpatient unit. However, recent research has demonstrated the reluctance of clinicians to trust ML due to the inability to understand the underlying mechanisms of the prediction.[15,16] This research will test six different ML models, including interpretable ML (InterpretML using Explainable Boosting Machine (EBM) is a tree-based, cyclic gradient boosting Generalized Additive Model) and traditional models such as Support Vector Machine using Support Vector Classification (SVC) with Radial Bias Function Kernel.

Importance of trust and transparency

Trustworthiness in ML innovations is a priority for governments, researchers and clinicians; however, trust and confidence have been highlighted as significant barriers to the acceptance of ML within a clinical setting.[15] The exclusive emphasis on demonstrating trust without ensuring reliability and validity during practice may lead to negative experiences for clinical users and hinder acceptance.[16] It is essential that ML interventions be described in terms of competence, reliability and validity as would be expected of other clinical tools where the quality of care and the safety of a person are the highest priority.[16] Clinicians should be presented with treatment recommendations that describe the validity and confidence of prediction, with the final decision for care always made by a clinician.[16]

Method

This study utilises a retrospective cohort analysis[17] combined with predictive modelling using ML techniques[18] and aims to assess the predictability of MH inpatient admissions using various ML models. The study was conducted in two public EDs within the Central Coast Local Health District of New South Wales, Australia. The population and sample included all ED presentations where a MH issue was recorded as the primary presenting problem. The primary data collection for this study involved reviewing existing records from January 1, 2016, to December 30, 2021 resulting in an initial dataset of 26,681 records. Missing records in this dataset were determined to be *missing completely at random* (MCAR), and the standard listwise deletion approach was applied.[19] This method led

to the removal of 515 rows and reduced the final dataset to ($n = 26{,}166$) records. A second round of data collection was conducted to determine if the patient's presentation was linked to current or past treatment with the service, which is referred to as "MH history" in the research. The study received ethical approval from the Hunter New England Human Research Ethics Committee, under the reference number 2022/ETH01597, on November 18, 2022. Additionally, site-specific approval was obtained from the Central Coast Local Health District on December 14, 2022. The requirement for informed consent to participate has been waived by the Hunter New England Human Research Ethics Committee, where it has been deemed that consent would be impossible or impracticable to obtain. The initial data investigation used SQL, Python, Excel, and Seaborne.[4] Data were cleaned, and missing data was handled as per.[4] Further information on the data collection and extraction methods can be found in[4] and exploratory analysis of the data.[4] Pearson's correlation coefficient[20] was used to determine the lineal/nonlinear nature of the data, the correlation coefficient and linear regression r-squared with a lineal/non-linear tolerance set at 0.01. The tolerance value represents the minimum threshold for the R-squared value of the linear regression model to be considered as indicating a linear relationship.[20] The R-squared value measures the proportion of the variance in the dependent variable explained by the independent variable(s). If the R-squared value is above the specified tolerance threshold, the code concludes that the relationship is linear; otherwise, it is considered non-linear. The resulting test concludes that the dataset is non-linear in nature due to the small number (11/120) of correlations.

This research will test six different types of ML models, Random Forest, XGBoost, CatBoost, k-Nearest Neighbours (kNN), Explainable Boosting Machine (EBM) as InterpretML, and Support Vector Machine using Support Vector Classification (SVC) with Radial Bias Function Kernel to provide a comprehensive evaluation due to the inherent differences in their underlying methodologies, strengths, and limitations. This approach ensures that various aspects of the data and the problem are addressed, leading to a more robust and well-rounded understanding of overall model performance in the context of the research.

Random forest

Random Forest is an ensemble learning method primarily used for classification and regression tasks. Developed by Breiman,[21] it constructs a multitude of decision trees during training and outputs the mode of the classes for classification or the mean prediction for regression. Each tree in the forest is built using a subset of the data and a subset of features, which helps to ensure that the trees are not highly correlated. The randomness introduced in the selection of data samples and features leads to model diversity, thereby improving the overall prediction performance and robustness against overfitting compared to a single decision tree.[22] The algorithm works by splitting nodes in each decision tree using the best among a subset of predictors randomly chosen at that node. This random selection of features leads to a decrease in variance without increasing bias, resulting in a model that performs well on unseen data. The method's efficiency and accuracy in handling large datasets and its ability to model complex interactions make it a popular choice in various applications, including healthcare and finance.[23]

XGBoost

XGBoost, or eXtreme Gradient Boosting, is an efficient and scalable implementation of gradient boosting frameworks.[24] It is designed to be highly efficient, flexible, and portable. The model builds trees sequentially, where each new tree corrects the errors made by the previous ones. This iterative process aims to minimise a loss function by using a gradient descent algorithm. XGBoost incorporates a regularisation term in its objective function, which helps prevent overfitting, making it a robust choice for both classification and regression problems. It also supports parallel and distributed computing, which significantly speeds up the learning process, making it suitable for large datasets. Additionally, XGBoost provides functionalities for cross-validation and automatic handling of missing data, further enhancing its versatility and application in a wide range of ML tasks.[25]

Catboost

CatBoost is another gradient boosting framework that stands out for its ability to handle categorical features natively.[26] Traditional ML algorithms often require categorical variables to be converted into numerical values, which can lead to information loss. CatBoost, developed by Yandex, addresses this by using an ordered boosting process and incorporating category-specific transformations to capture the information contained in categorical features effectively.[26] The model uses an innovative method of calculating leaf values during the boosting process, which helps in reducing the prediction shift caused by biased leaf value estimation. CatBoost's ability to handle categorical data without preprocessing steps, along with its robust regularisation techniques, make it a powerful tool for a variety of applications, including those with structured and unstructured data.[27]

k-Nearest neighbours (kNN)

The k-nearest neighbours (kNN) algorithm is a simple, yet effective, instance-based learning method used for

classification and regression. Proposed by Cover and Hart in 1967,[28] it is a non-parametric method that predicts the output for a query instance based on the majority class (for classification) or average outcome (for regression) of its k nearest neighbours in the feature space. One of the primary advantages of kNN is its simplicity and ease of implementation. However, it can be computationally expensive for large datasets because it requires calculating distances between the query instance and all other points in the dataset. Moreover, kNN is sensitive to the local structure of the data and the choice of k, where a small k can lead to overfitting, while a large k may result in underfitting. Despite these challenges, kNN is widely used in various fields, including pattern recognition, data mining, and image processing, due to its effectiveness in scenarios where the data distribution is unknown.[29]

Interpretml

InterpretML[30] is a framework for creating interpretable models that can assist clinicians with making informed decisions about patient care using algorithms such as Explainable Boosting Machine (EBM), a tree-based, cyclic gradient boosting Generalized Additive Model (GAM). Traditional ML models, such as deep neural networks, can be very accurate but are often considered "black boxes" because it can be challenging to understand how they arrive at their predictions.[31] This lack of transparency can be an obstacle in healthcare, where patients have the right to understand decision-making.[32] InterpretML allows visualising models and their predictions via feature importance, partial dependence, and individual conditional expectation plots.[33] These plots can help identify the impact of each feature on the outcome and areas for improvement.[32] For additional information pertaining to the use of InterpretML on this dataset, please refer to.[34]

Support vector machines for classification

Support Vector Machines (SVM) are ML algorithms that can be used for classification, regression and clustering.[35–37] In the case of two-class classification, SVMs' fundamental goal is to establish a hyperplane that maximises the margin of separation between the two classes, taking noise and outliers into account. By using kernel functions SVMs can produce non-linear decision boundaries and, with a suitable kernel, can also handle non-vectorial data. Due to their training based on constrained optimisation they can be applied also in cases where little data is available. This makes them a versatile tool for analysing the often relatively small but complex datasets that can be encountered in healthcare research. This research will use an SVM for classification using the SVC with a Radial Bias Function Kernel.

Data description

The ED presentations related to MH and suicidal behaviour between January 1, 2016, and December 30, 2021. The ED data were selected based on specific terms in the presenting problem, such as *'%suici%', '%MH%', '%self harm%', '%psych%', or '%Mental%'*. Additionally, data from electronic medical records were extracted to link the ED presentations with MH service data, indicating whether the individual received current or previous care. The dataset contained the following features: *age, active/previous service user, presenting problem, facility identifier, first nations status, source of referral, marital status, model of arrival, gender, triage category, time of the day, day of the week, month, referred to on departure (including admission)* and *mode of separation.*

The extracted dataset contained 26,681 records, with a relatively small proportion of missing data for various variables, such as *ED source of referral, mode of separation, marital status, mode of arrival,* and *referred to on departure.* Although small amounts of missing data can impact study results, the missing data in this case were considered missing completely at random (MCAR). The most straightforward approach, listwise or case deletion, was employed, resulting in the removal of entire missing records from the dataset. This approach reduced the final record count to 26,166, with a loss of 515 records (1.93%).[34] The final data set contained 26,166 records of people who presented to ED with MH as their primary presenting problem; however, there was a significant class imbalance in the target of *admission* (Yes, $n=5774$, No, $n=20392$). The baseline characteristics for these presentations are described in Table 1.

Data transformations for modelling

The following steps were taken to prepare the data set for ML modelling. Logarithmic transformation was applied to the data to calculate the natural logarithm of each element, allowing for a compressed range of values.[38] The logarithmic transformation process is used to address issues related to skewed data distributions and variations in magnitude and is required when handling variables with varying dynamic ranges.[38] The pandas function *pd.get_dummies()* is employed with the pandas DataFrame[39] to perform *one-hot encoding* on categorical variables. This transformation results in the creation of binary columns indicating the presence or absence of specific categories within the original dataset. Notably, the parameter *drop_first = True* is specified, leading to the exclusion of one dummy column for each categorical variable to mitigate multicollinearity concerns. Scaling was performed using *Standard Scaler* from scikit-learn to standardise specific columns using *to_scale* within the pandas DataFrame. The *fit_transform* method is applied,

Table 1. Description of data set based on admission.

Admitted	NO	%	YES	%	Total	%
n	20392	77.93%	5774	22.07%	26166	
Age						
0–11	391	1.92%	41	0.71%	432	1.65%
12–17	3646	17.88%	397	6.88%	4043	15.45%
18–24	4243	20.81%	851	14.74%	5094	19.47%
25–34	4158	20.39%	1153	19.97%	5311	20.30%
35–44	3456	16.95%	1177	20.38%	4633	17.71%
45–54	2401	11.77%	989	17.13%	3390	12.96%
55–64	1278	6.27%	490	8.49%	1768	6.76%
65–74	495	2.43%	335	5.80%	830	3.17%
75–84	251	1.23%	203	3.52%	454	1.74%
over 85	73	0.36%	138	2.39%	211	0.81%
Presenting Problem						
Altered Mental Status	1013	4.97%	497	8.61%	1510	5.77%
Anxious	1607	7.88%	157	2.72%	1764	6.74%
Behavioural Disturbance	2125	10.42%	1041	18.03%	3166	12.10%
Depression	1509	7.40%	262	4.54%	1771	6.77%
Eating Disorder	18	0.09%	10	0.17%	28	0.11%
Hallucinations	579	2.84%	333	5.77%	912	3.49%
Memory Impairment	18	0.09%	10	0.17%	28	0.11%
Mental Health Problem	3076	15.08%	845	14.63%	3921	14.99%
Self Harm	2563	12.57%	509	8.82%	3072	11.74%
Suicidal Ideation	7884	38.66%	2110	36.54%	9994	38.19%
Mode of Arrival						
Air Ambulance Service	2	0.01%	2	0.03%	4	0.02%
Community/Public Transport	31	0.15%	6	0.10%	37	0.14%
Internal Ambulance / Transport	4	0.02%	6	0.10%	10	0.04%

(continued)

Table 1. Continued.

Admitted	NO	%	YES	%	Total	%
Internal Bed / Wheelchair	4	0.02%	4	0.07%	8	0.03%
No transport (walked in)	168	0.82%	22	0.38%	190	0.73%
Other eg Undertakers/Contractors	39	0.19%	21	0.36%	60	0.23%
Police/Correctional Services Vehicle	1425	6.99%	482	8.35%	1907	7.29%
Private vehicle	9689	47.51%	2221	38.47%	11910	45.52%
Retrieval (including NETS)	22	0.11%	6	0.10%	28	0.11%
State Ambulance Vehicle	9008	44.17%	3004	52.03%	12012	45.91%
Gender						
Female	10391	50.96%	2857	49.48%	13248	50.63%
Indeterminate	1	0.00%	1	0.02%	2	0.01%
Male	9996	49.02%	2916	50.50%	12912	49.35%
Unknown	4	0.02%	0	0.00%	4	0.02%
Triage Category						
Resuscitation (1)	214	1.05%	118	2.04%	332	1.27%
Emergency (2)	3952	19.38%	1767	30.60%	5719	21.86%
Urgent (3)	8385	41.12%	2543	44.04%	10928	41.76%
Semi Urgent (4)	7396	36.27%	1311	22.71%	8707	33.28%
Non Urgent (5)	445	2.18%	35	0.61%	480	1.83%
Mode of Separation						
Admitted Left at own risk	25	0.12%	0	0.00%	25	0.10%
Admitted Transferred to another hospital	16	0.08%	0	0.00%	16	0.06%
Admitted and discharged as inpatient within ED	0	0.00%	5774	100.00%	5774	22.07%
Admitted to ward / inpatient unit	198	0.97%	0	0.00%	198	0.76%
Admitted: To Critical Care Ward (including ICU1/ICU2/CCU/COU/NICU)	98	0.48%	0	0.00%	98	0.37%
Admitted: Via Operating Suite	8	0.04%	0	0.00%	8	0.03%
Departed: Did not wait	1754	8.60%	0	0.00%	1754	6.70%
Departed: Left at own risk	139	0.68%	0	0.00%	139	0.53%

(continued)

Table 1. Continued.

Admitted	NO	%	YES	%	Total	%
Departed: Transferred to another hospital w/out 1st being admitted to hospital transferred from	915	4.49%	0	0.00%	915	3.50%
Departed: Treatment completed	417	2.04%	0	0.00%	417	1.59%
Departed: for other Clinical Service Location	16822	82.49%	0	0.00%	16822	64.29%

wherein the scaler computes the mean and standard deviation necessary for standardisation and then transforms the selected columns. This process ensures that the numerical features in the specified columns have a standardised distribution with zero mean and unit variance.[39] A test/train split was executed using the *train_test_split* function from scikit-learn, with a specified test size of 20%. This division allocates 80% of the data for training the model and reserves the remaining 20% for evaluating the model's performance on unseen data. The choice of a 20% test size is a common practice in model evaluation.[39] This strategic split allows for robust training and assessment of the model's generalisation capabilities, providing valuable insights into its performance on new, previously unseen observations.[40] To address the class imbalance in the dataset, the *RandomOverSampler* from the imbalanced-learn library was employed with the parameter *sampling_strategy='minority'* after the test, train split to ensure that the oversampled data is not present in the test dataset, which is not appropriate for testing the model and leaves the final testing data unchanged. This oversampling technique focuses on the minority class, generating synthetic samples to balance the class distribution.[41] The *fit_resample* method is then applied to the feature matrix and target variable, efficiently handling the oversampling process. Implementing random oversampling on the minority class mitigates the impact of class imbalance, preventing model bias towards the majority class and enhancing overall predictive performance.[41]

Machine learning models

Given the previous indication of the nature of the lineal data, it was decided to train both linear and non-linear models on the data and compare their performance. Models were built using InterpretML and SVC, *GridsearchCV* was used to establish parameters, with the resulting models being evaluated using F1, recall (sensitivity), precision, a confusion matrix, receiver operating characteristic (ROC), negative predictive value (NPV), false positive rate, false discovery rate, and false negative rate. A confusion matrix is used in classification problems to evaluate the performance of a machine-learning model by summarising the outcomes of predictions[39] and represents the number of correct and incorrect predictions made by the ML model being tested. The matrix has four main components: True Positive (TP), which is an instance where the model correctly predicts the positive class; True Negative (TN), instances where the model correctly predicts the negative class. False Positive (FP), instances where the model incorrectly predicts the positive class (Type I error), and False Negative (FN), instances where the model incorrectly predicts the negative class (Type II error).[42] The confusion matrix is useful for understanding the model's performance across different classes by providing insights into the model's strengths and weaknesses. Performance metrics such as accuracy, precision, recall (sensitivity), specificity, and the F1 score can be derived from these values.[39] The ROC curve is a graph that displays the relationship between the recall (sensitivity) and the false positive rate of a model for different probability thresholds.[42] Recall also referred to as sensitivity or true positive rate (TPR), measures the proportion of actual positive instances correctly identified by the ML model. In contrast, the false positive rate (FPR) represents the ratio of negative instances incorrectly classified as positive. The ROC curve is plotted by comparing the TPR against the FPR at different threshold values.[42] A perfect model would have a ROC curve that reaches the top-left corner of the graph (TPR = 1, FPR = 0), while an ineffective model would produce a curve close to the diagonal line (45-degree angle).[39] AUC-ROC (Area Under the ROC Curve) is a numerical metric that summarises the model's overall performance in distinguishing between positive and negative classes. A higher AUC-ROC indicates better model performance.[39] ROC curves and AUC-ROC are commonly used to assess model performance, especially in situations such as health care where sensitivity and specificity are both important.

Working with healthcare data presents several challenges, such as siloing, segmentation, privacy ethics, and health policy.[43] Compared to other sectors, such as finance, there are limitations on the size of the data available for use in healthcare. The data used in this study presents similar issues and only represents a small portion of presentations to the ED and lacks the depth of clinical narrative that is present in the wider system. This dataset's limited

number of records poses a significant challenge for machine ML techniques, which typically perform better with larger datasets. However, this limitation is outweighed by the critical importance of accurate prediction in the MH ED context, where misclassifications can have severe consequences for patient well-being and safety. False negatives, where patients in need of admission are missed, can result in delayed treatment, worsening of conditions, or even life-threatening situations. False positives, where patients are incorrectly identified as requiring inpatient admission, can lead to unnecessary hospitalisations, increased healthcare costs, and potential emotional distress for patients and their families. Clinicians and developers must carefully balance these risks and consider the potential consequences of each type of misclassification. In practice, clinicians are trained to prioritise caution over taking risks, as the consequences of missing a patient in need of intensive treatment can be far more severe than the inconvenience and costs associated with an unnecessary admission. In this context, the use of specificity as an evaluation metric becomes more relevant than precision. While precision measures the proportion of correct positive predictions, specificity measures the proportion of negative instances that are correctly identified.[42] In the MH ED setting, where false negatives are particularly concerning, specificity provides a more direct assessment of the model's ability to identify patients who do not require inpatient admission accurately. Specificity measures the proportion of true negatives that are correctly identified by the model, demonstrating the ability of the model to accurately identify patients who do not require inpatient admission. A high specificity value is indicative that the model is effective at correctly identifying and filtering out negative cases (those who can be safely discharged or treated in the community). The ROC curve offers a powerful tool for clinicians to evaluate and select appropriate classification models for the MH ED context. By plotting the true positive rate (sensitivity) against the false positive rate (1 - specificity) for classification thresholds, the ROC curve allows clinicians and developers to visualise the model's ability to correctly identify positive cases (patients requiring admission) and its tendency to generate false positives. By carefully considering the ROC curve and the associated trade-offs, clinicians and developers can select a classifier that balances the risks of false positives and false negatives in a manner that aligns with their clinical objectives and the specific context of the MH ED.

Results

The performance of ML models in healthcare is critical, as these models assist in clinical decision-making by predicting patient outcomes. In evaluating the six ML models (EBM InterpretML, Random Forest, XGBoost, CatBoost, k-Nearest Neighbors (kNN), and Support Vector Machine using Support Vector Classification (SVC) with Radial Bias Function Kernel) a variety of metrics are utilised, including accuracy, AUC-ROC, precision, recall, F1 score, specificity, NPV, false positive rate, false discovery rate, false negative rate and Matthews Correlation Coefficient (MCC). The overall results can be found in Table 2, with the overall classifier accuracy demonstrated in Figure 1 and the overall classifier ROC comparison is in Figure 2. The following section breaks down the performance of each classifier model performance.

EBM InterpretML achieves an accuracy of 0.7883, MCC of 0.1914 and a ROC/AUC of 0.7167 (Table 2, Figures 3 and 4), indicating moderate discriminative power. The model shows a recall of 0.1082 and a precision of 0.6158, reflecting a relatively low sensitivity but moderate precision. The F1 score of 0.1841 indicates a need for improvement in balancing recall and precision. The model's specificity of 0.9809 suggests excellent performance in identifying true negatives, while the false positive rate is low at 0.0191. However, the false discovery rate is 0.3842, and the false negative rate is high at 0.8918, pointing to challenges in correctly identifying positive cases.

Random Forest presents a slightly lower accuracy of 0.7776, MCC of 0.1453 and a ROC/AUC of 0.6820 (Table 2, Figures 5 and 6). It achieves a recall of 0.1152 and a precision of 0.4836, resulting in an F1 score of 0.1860. The specificity is 0.9652, demonstrating good capability in correctly identifying true negatives, although the false positive rate is slightly higher at 0.0348. The false discovery rate of 0.5164 and the false negative rate of 0.8848 indicate a significant number of misclassified positive cases.

XGBoost shows an accuracy of 0.7814, MCC of 0.1928 and a ROC/AUC of 0.6986 (Table 2, Figures 7 and 8). The model's recall is 0.1602, and its precision is 0.5153, leading to an F1 score of 0.2444. Its specificity is 0.9573, with a false positive rate of 0.0427. The false discovery rate is 0.4847, while the false negative rate is 0.8398, reflecting moderate challenges in detecting true positive cases.

CatBoost stands out with the highest accuracy of 0.7900, and MCC of 0.1952 and a ROC/AUC of 0.7129 (Table 2, Figures 9 and 10), indicating superior overall performance. Its recall is 0.1307, and precision is 0.5720, resulting in an F1 score of 0.2128. The specificity of 0.9723 and the false positive rate of 0.0277 are commendable, showing a low rate of false positives. However, the false discovery rate is 0.4280, and the false negative rate is 0.8693, indicating room for improvement in positive case detection.

k-Nearest Neighbours (kNN) has an accuracy of 0.7553, MCC of 0.1154 and an ROC/AUC of 0.5950 (Table 2, Figures 11 and 12), the lowest among the models evaluated. The recall is 0.1558, and precision is 0.3704, resulting in an F1 score of 0.2194. The specificity is 0.9250, with a higher false positive rate of 0.0750. The model also exhibits a high false discovery rate of 0.6296 and a false negative rate of

Table 2. Model performance.

	EBM InterpretML	Random Forest	XGBoost	Catboost	kNN	SVC
Accuracy	0.7883	0.7776	0.7814	0.7900	0.7553	0.7780
ROC/AUC	0.7167	0.6820	0.6986	0.7129	0.5950	0.5960
TN	4001	3937	3905	3966	3773	4068
FP	78	142	174	113	306	11
FN	1030	1022	970	1004	975	1151
TP	125	133	185	151	180	4
Recall Sensitivity	0.1082	0.1152	0.1602	0.1307	0.1558	0.0035
Precision	0.6158	0.4836	0.5153	0.5720	0.3704	0.2667
F1	0.1841	0.1860	0.2444	0.2128	0.2194	0.0068
Specificity	0.9809	0.9652	0.9573	0.9723	0.9250	0.9973
NPV	0.7953	0.7939	0.8010	0.7980	0.7947	0.7795
False Positive Rate	0.0191	0.0348	0.0427	0.0277	0.0750	0.0027
False Discovery Rate	0.3842	0.5164	0.4847	0.4280	0.6296	0.7333
False Negative Rate	0.8918	0.8848	0.8398	0.8693	0.8442	0.9965
MCC	0.1914	0.1453	0.1928	0.1952	0.1154	0.0059

0.8442, indicating substantial issues in identifying positive cases.

Finally, the SVC model, with an accuracy of 0.778, MCC of 0.0059 and an ROC/AUC of 0.5960 (Table 2, Figures 13 and 14), presents unique characteristics. It achieves a recall of only 0.0035, indicating a very low sensitivity, and a precision of 0.2667, resulting in a minimal F1 score of 0.0068. However, the specificity is extremely high at 0.9973, and the false positive rate is very low at 0.0027, suggesting the model is highly conservative in identifying positives. The false discovery rate is 0.7333, and the false negative rate is 0.9965, reflecting severe challenges in positive case detection.

Discussion

CatBoost emerges as the model with the highest accuracy and balance between recall and precision, although improvements are needed in positive case detection. Models like EBM InterpretML and Random Forest demonstrate good specificity but struggle with recall and false negatives, which are critical in healthcare applications, however unlike Random Forest, EBM InterpretML does offer the ability to provide further insight into its recommendation, making it suitable for general healthcare applications where transparency is as important as the prediction provided. The comparison between the EBM InterpretML, CatBoost, and Random Forest models illustrates the trade-offs between interpretability and predictive performance in healthcare applications. EBM InterpretML exhibits a moderate level of accuracy at 0.7883, with a balanced distribution between true positives and false positives, indicating moderate effectiveness in both positive and negative predictions. Its primary advantage lies in its interpretability, making it a suitable choice for applications where understanding the decision-making process is crucial. However, EBM InterpretML's lower recall (sensitivity) of 0.1082 compared to the CatBoost model suggests a tendency to miss more positive cases. This is an important consideration in the context of care delivery, where failing to identify patients in need of intervention could have significant consequences. The model's precision is 0.6158, and it achieves a specificity of 0.9809, indicating excellent performance in identifying true negatives.

The CatBoost model demonstrates a higher ability to correctly identify positive cases, such as patients requiring

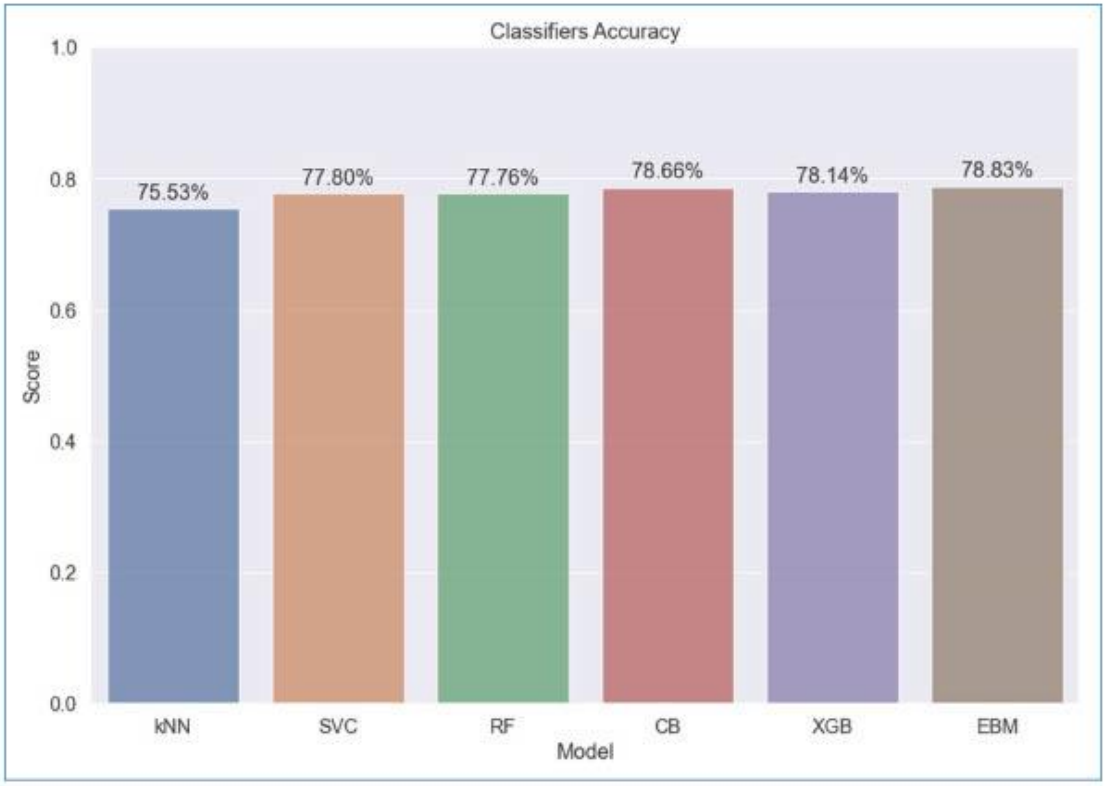


Figure 1. Classifier accuracy.

admission to a mental health inpatient ward. It achieves the highest accuracy among the evaluated models at 0.7900, indicating strong overall performance. CatBoost's recall of 0.1307 compared to EBM InterpretML suggests it is more effective at capturing true positive cases, reducing the likelihood of missing patients who require critical care. This makes CatBoost a compelling choice for scenarios where the priority is to maximise the detection of positive cases, even if it means sacrificing some interpretability. CatBoost's precision is 0.5720, with a specificity of 0.9723, reflecting its capability to avoid false positives while effectively identifying true positives.

In contrast, the Random Forest model presents a balanced approach with its accuracy of 0.7776 and recall of 0.1152. While it does not achieve the highest sensitivity, it performs reasonably well across various metrics, offering a compromise between interpretability and predictive power. Random Forest's recall is slightly higher than EBM InterpretML, suggesting a better ability to capture positive cases, although it still lags behind CatBoost in this regard. The model's precision is 0.4836, and its specificity is 0.9652, indicating good performance in identifying true negatives, which is crucial for minimising unnecessary interventions.

The choice between EBM InterpretML, CatBoost, and Random Forest should be informed by specific clinical priorities. If interpretability is crucial, EBM InterpretML is advantageous; however, CatBoost offers superior performance in capturing true positive cases if the goal is to maximise the identification of patients requiring admission. With its balanced approach, the Random Forest model provides a viable middle ground for situations where both interpretability and predictive accuracy are important considerations. The choice of model should be guided by the specific clinical context and the importance of minimising false negatives or false positives, depending on the consequences of each error type. The results of this study demonstrate the effectiveness of ML models in the prediction of admissions to a MH inpatient ward based on routine data collected at the triage stage in the ED. This finding aligns with previous research[10,12] that demonstrates the potential of ML models to enhance decision-making processes and improve patient outcomes in the ED setting.

While some of the models exhibit positive predictive performance, ensuring transparency and interpretability is crucial for gaining trust and acceptance among clinicians and patients.[15] This is where the EBM InterpretML model offers a strong overall balance, making it suitable for

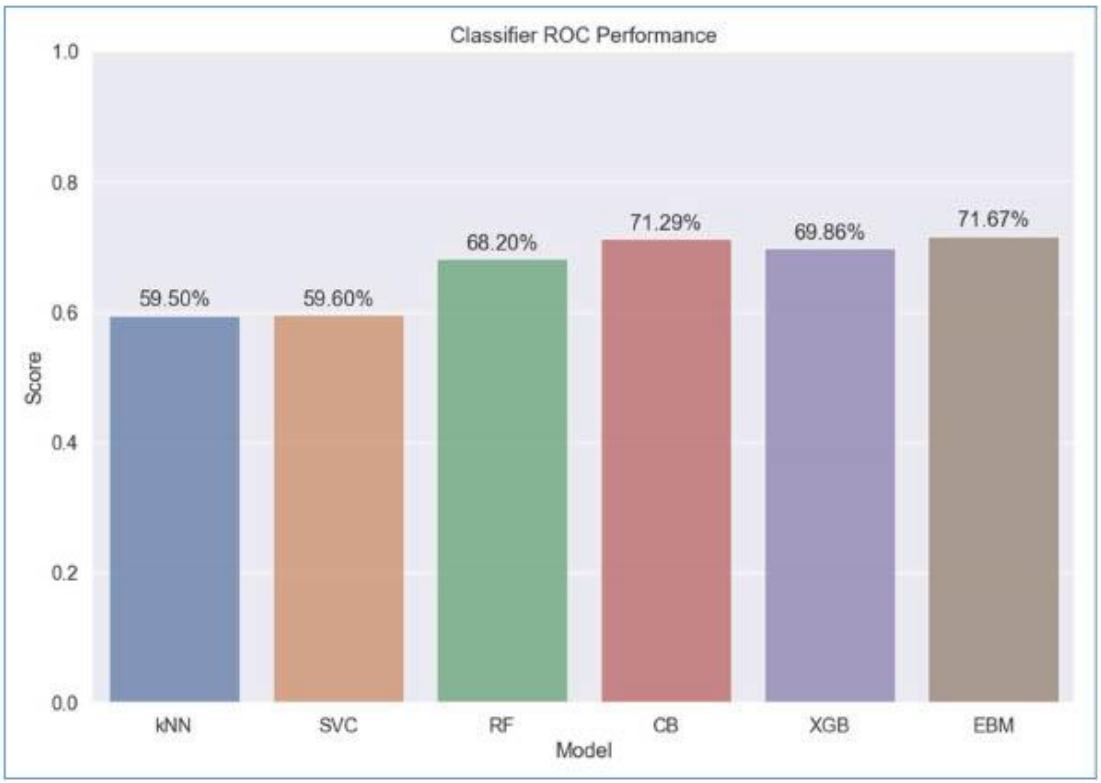


Figure 2. Classifier ROC.

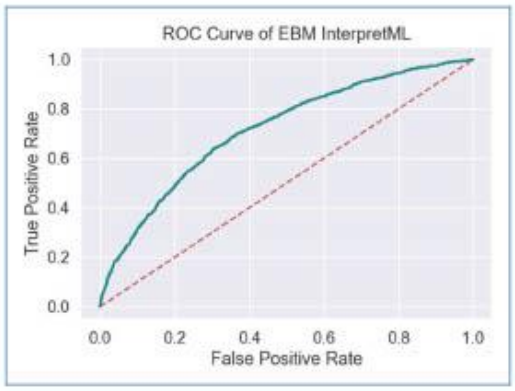


Figure 3. ROC of EBM interpretML.

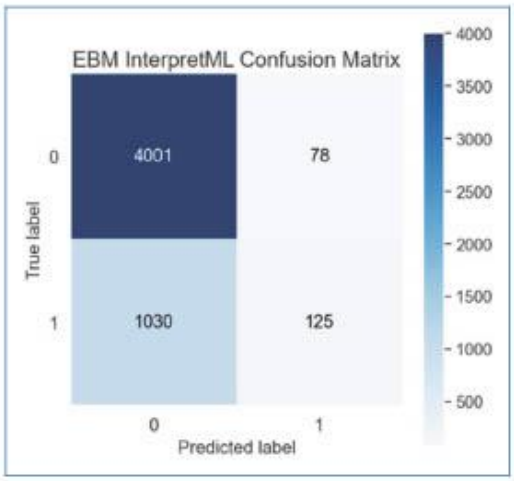


Figure 4. Confusion matrix of EBM interpretML.

applications such as healthcare, offering clinicians deeper insight into the data and how recommendations are made. InterpretML facilitates the interpretation of ML models by identifying complex interactions through methods such as the overall importance graph.[33] This graph uses the mean absolute score (MAS) to assess each feature's impact on model predictions.[44] MAS is calculated by permuting the values of each feature in the dataset and comparing the model's predictions before and after permutation. Features with higher MAS are deemed more important to the model's predictions. The overall importance graph

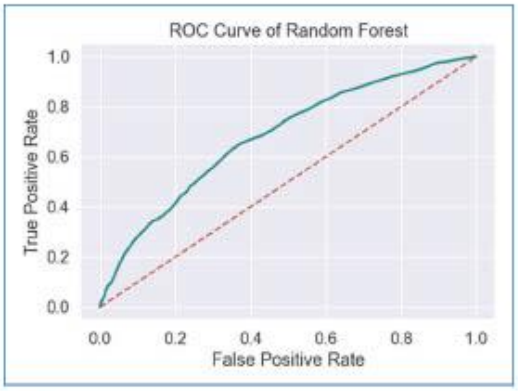


Figure 5. ROC random forest.

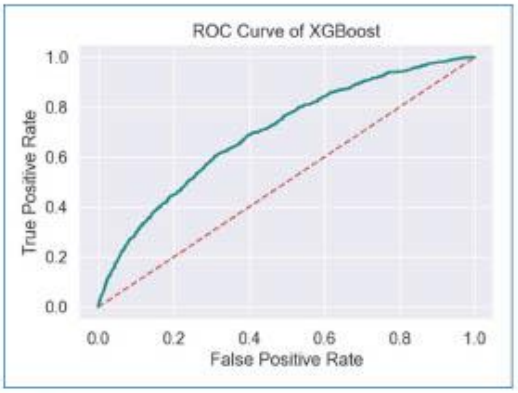


Figure 7. ROC XGBoost.

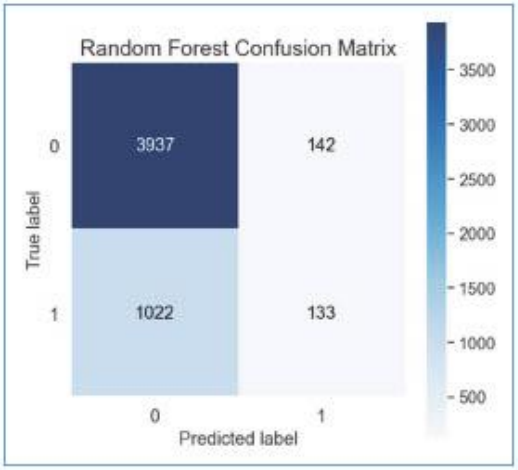


Figure 6. Confusion matrix of RF.

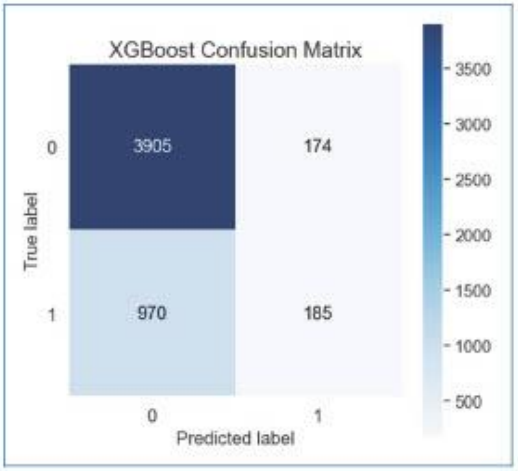


Figure 8. Confusion Matrix of XGBoost.

ranks features by their MAS, allowing for quick identification of those with the most significant impact. Additionally, InterpretML uses Log odds to interpret the relationship of each element to outcomes visually. Log odds represent the probability of an event as a logarithmic function of its odds.[45] The log odds ratio (OR) compares the odds of an event occurring between two groups, such as treatment and control. By integrating Log OR, the model can assess the impact of each predictor, compare outcomes across different groups, and provide a unified probability scale.

Figures 15 and 16 provide a demonstration of the capabilities of the InterprtML package. Figure 15 provides the overall global importance of the features of the model and their relative predictive power and influence on the model. Figure 16 is a demonstration of one prediction from the unseen test dataset. In this case, the most heavily weighted features that the model used to make a recommendation were the individual's need for admission based on their triage category, the fact that they are an active client of the service, their age, the facility they presented to, their marital status (divorced), and their presenting problem of suicidal ideation. This output (Figure 16) enables the clinician to examine the recommendation (in this case, admit) to see if there are confounding factors present that would require further assessment and time with the individual to assess their needs. Based on this information, it may be possible enhance clinical practice by offering quicker assessments or by providing less experienced clinicians with the necessary tools to guide their practice in identifying important presentation factors that require further exploration. Visualisations such as Figure 16 can assist clinicians in understanding the underlying factors contributing to the model's predictions, fostering trust and enabling more informed decision-making processes. This transparency is crucial in any ML innovation gaining acceptance among healthcare professionals.[15,16] Further detail regarding

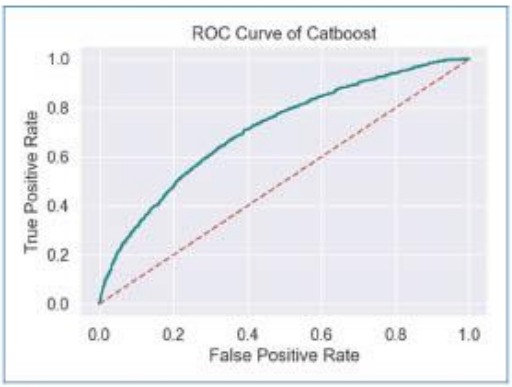


Figure 9. ROC CatBoost.

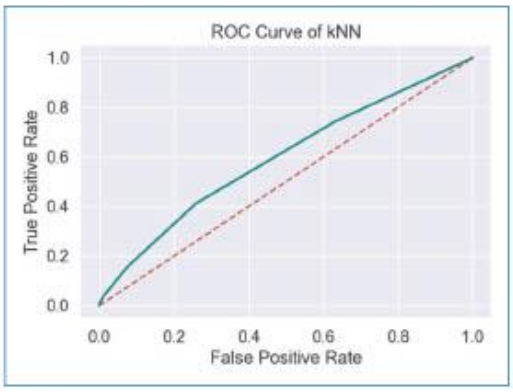


Figure 11. ROC kNN.

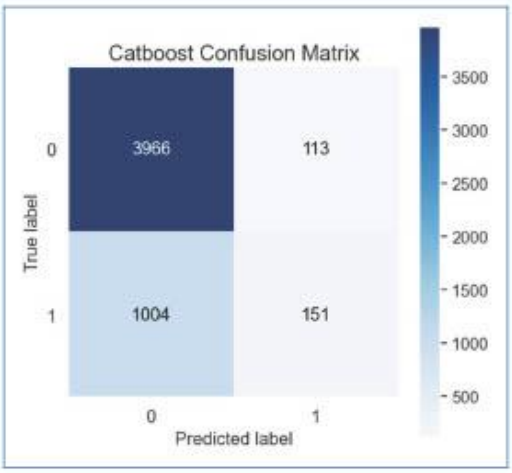


Figure 10. Confusion matrix of CatBoost.

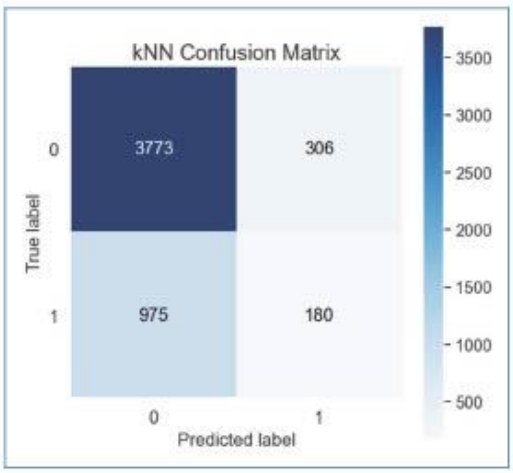


Figure 12. Confusion matrix of kNN.

InterpretML and its use in this data asset can be found at Higgins, Chalup[34]

Limitations

One limitation of this study is the reliance solely on the initial data collected during the patient's presentation to triage in the ED and may not be representative of all MH presentations due to issues such as triage coding or alternative admission pathways. Furthermore, while this data provides valuable insights into the patient's condition at the time of triage, it may not capture the complete picture of the patient's overall health status, medical history, and various psychosocial factors that could influence the need for admission to a MH inpatient ward. By limiting the data sources to the initial triage information, this research may overlook crucial details that could enhance the predictive capabilities of the ML models. Integrating additional data sources, such as electronic medical records (EMRs), patient-reported outcomes, and social determinants of health, could potentially provide a more comprehensive and holistic assessment of the patient's needs. EMRs contain a wealth of information, including past medical histories, diagnoses, treatments, and laboratory results, which could offer valuable context for the ML models to consider.[10] Furthermore, incorporating patient-reported outcomes, such as questionnaires or self-assessments, could provide insights into the patient's subjective experiences, symptoms, and overall well-being, which may not be fully captured in the initial triage data. It is important to note that the data used in the development of these models was from January 1, 2016, to December 30, 2021. This period included the COVID-19 pandemic, which included two major lockdowns[46] for the region in question and is noted as a period of increased MH stress

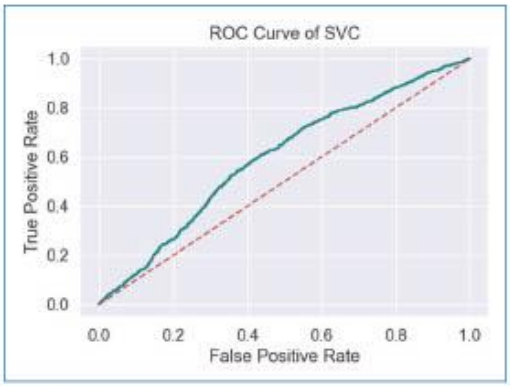


Figure 13. ROC of SVC.

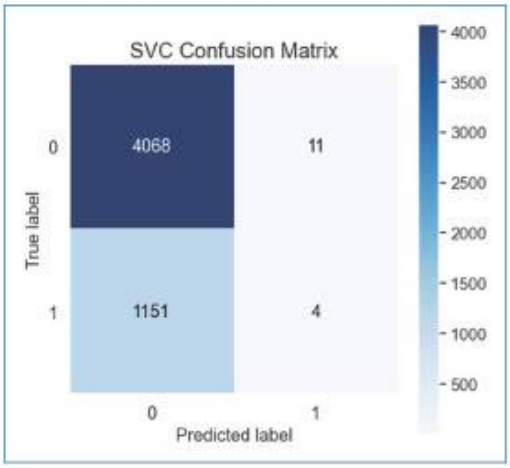


Figure 14. Confusion matrix of SVC.

for Australian people.[47] This must be taken into consideration and may limit the generalisability and applicability of the study's findings to diverse patient populations and healthcare settings. Finally, it must be noted that models should ideally be tested on an external dataset to assess generalisability. This is a broader issue in the development of ML models, especially in healthcare, as the absence of open data and code in many studies makes it difficult to reproduce and validate findings.[48,49] This lack of transparency also limits the ability to compare results across studies using varying data labels and evaluation metrics.[48]

Recommendations

R1 Further development and testing of the InterpretML model in a CDSS to aid triage decision-making processes in the ED is required before any system is deployed using the methods discussed. However, the model's predictive performance has demonstrated an ability that can assist clinicians in identifying patients who require admission to a MH inpatient ward, potentially reducing wait times and improving the overall quality of care. This approach can enhance trust and acceptance among healthcare professionals, addressing the concerns[15,16] raised regarding the barriers to ML adoption in clinical applications.

R2 Further research is needed to validate the CatBoost, Random Forest and the InterpretML model's performance across multiple healthcare settings and diverse patient populations. This should include continuous monitoring and refinement of the model, which are essential to ensure its reliability and validity in real-world clinical contexts.[16] This should include the development of a comprehensive implementation plan that addresses ethical considerations, data privacy, and potential biases. To ensure a transparent and inclusive process, stakeholders, such as clinicians, patients, and policymakers, must be present throughout the development cycle.[15]

R3 Future research could explore the integration of multiple data sources, including EMRs, patient-reported outcomes, and social determinants of health, into the ML modelling process. By leveraging a more comprehensive dataset, the ML models may better capture the complexities of MH conditions and provide more informed and tailored predictions for admission to MH inpatient wards. Furthermore, real-time data collection and monitoring would enable the continuous evaluation and refinement of the ML models' performance. By tracking the accuracy of predictions against actual outcomes, researchers and clinicians can identify areas for improvement and make necessary adjustments to the models, ensuring their long-term reliability and validity.

R4 Further research is required to understand the barriers to AI integration with clinicians and explore their concerns about the interpretability and transparency of ML models. Tools such as the General Attitudes towards Artificial Intelligence Scale (GAAIS) survey[50] may help to better understand the apprehensions or negative attitudes towards AI among MH clinicians and help guide the development of targeted educational and training initiatives to address these concerns. This would address the clinician's need to understand the model's decision-making process, thereby increasing their trust and acceptance of the ML technology.

R5 Real-time data collection and monitoring are crucial for implementing and continuously improving ML models. The models can be regularly updated and refined by capturing patient data as it becomes available, ensuring they reflect the most current trends and patterns in MH presentations and outcomes. It is recommended that a robust data collection and monitoring system that enables real-time capture of patient data in the ED and MH inpatient settings be developed. This system should integrate seamlessly with the existing EMR infrastructure, facilitating efficient and accurate data extraction for the ML models.

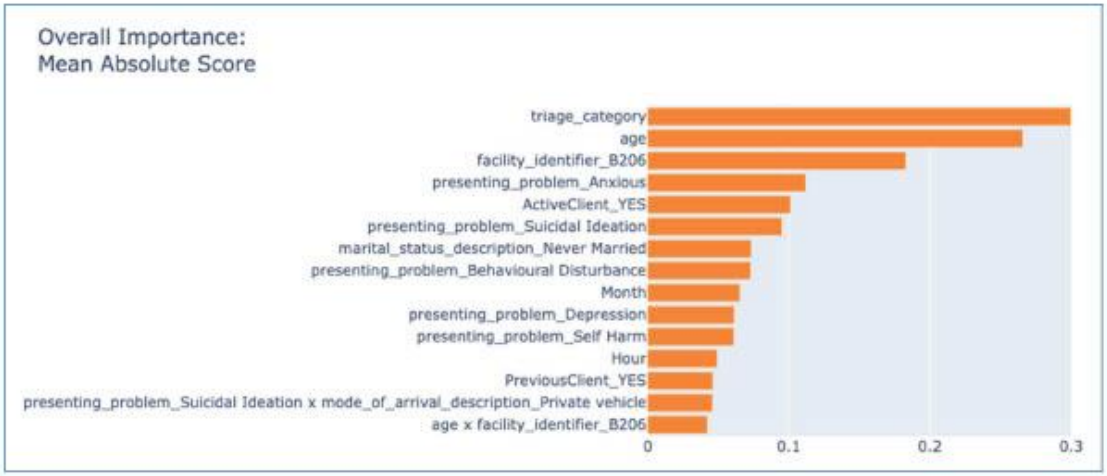


Figure 15. EBM overall importance.

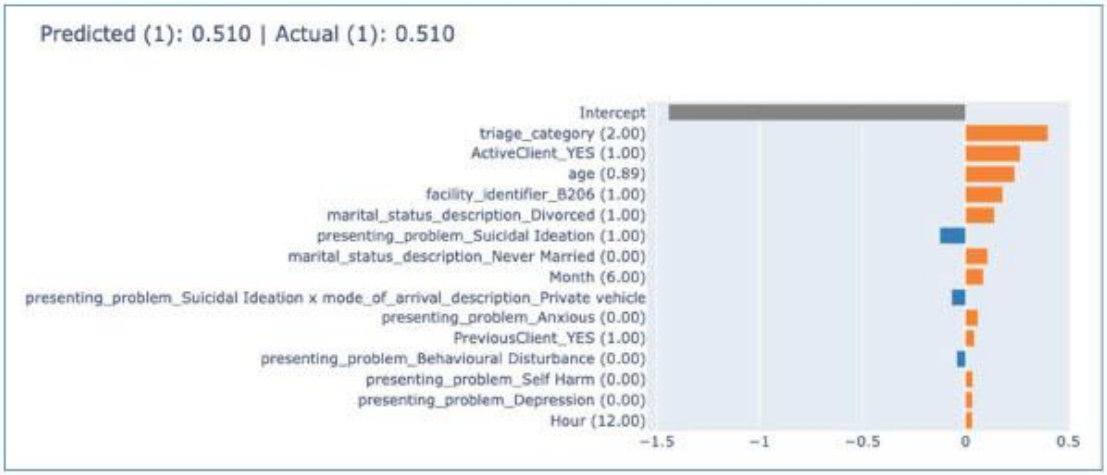


Figure 16. EBM individual prediction.

Conclusion

This research has demonstrated the potential of Machine Learning in predicting admissions to a MH inpatient ward based on routine data collected at the triage stage in the ED. The positive predictive performance of the models, combined with the interpretability provided by the InterpretML framework, offers a promising approach to enhancing decision-making processes and optimising resource allocation in MH care. However, the successful implementation of these ML models in clinical practice requires a multifaceted approach that addresses concerns about trust and transparency, ethical considerations, and continuous validation and refinement. By collaborating with subject matter experts and incorporating feedback from clinicians and patients, these ML innovations can be effectively integrated into existing healthcare systems, ultimately improving patient outcomes and enhancing the quality of MH care delivery.

Acknowledgements: Authors would like to acknowledge the support of Central Coast Local Health District

Authorship statement: OH – concept development, project design, data collection, data analysis, manuscript preparation.
RW – concept development, contribution to manuscript, supervision of project.
SC – data analysis, project design, data collection, data analysis, manuscript contribution, supervision of project

Consent to participate: The requirement for informed consent to participate has been waived by the Hunter New England Human Research Ethics Committee, where it has been deemed that consent would be impossible or impracticable to obtain.

Declaration of conflicting interests: The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethics approval: Ethical approval was granted by the Hunter New England Human Research Ethics Committee 2022/ETH01597. Central Coast Local Health District granted site-specific approval 2022/STE03296

Funding: Partial financial support was received from NSW Ministry of Health as part of the Towards Zero Suicides initiative.

Guarantor statement: OH is the guarantor who has taken full responsibility for this article, including the accuracy and appropriateness of the reference list.

ORCID iDs: Oliver Higgins https://orcid.org/0000-0001-6545-145X
Rhonda L. Wilson https://orcid.org/0000-0001-9252-2321

References

1. World Health Organization. *World mental health report: transforming mental health for all*. Geneva: World Health Organization, 2022.
2. Australian Institute of Health and Welfare. Suicide & self-harm monitoring, https://www.aihw.gov.au/suicide-self-harm-monitoring/data/data-downloads (2023, accessed 11/08/2023 2023).
3. Higgins O, Sheather-Reid RB, Chalup SK, et al. Disproportionate mental health presentations to Emergency Departments in a coastal regional community in Australia of First Nation people. *Int J Ment Health Nurs* 2024: 1–8. DOI: 10.1111/inm.13362
4. Higgins O, Sheather-Reid RB, Chalup SK, et al. Sociodemographic factors and presentation features of individuals seeking MH care in EDs: a retrospective cohort study. *Int J Ment Health Nurs* 2024: 1–12. DOI: 10.1111/inm.13414
5. Clarke DE, Dusome D and Hughes L. Emergency department from the mental health client's perspective. *Int J Ment Health Nurs* 2007; 16: 126–131.
6. Fatovich D and Hirsch R. Entry overload, emergency department overcrowding, and ambulance bypass. *Emerg Med J* 2003; 20: 406–409.
7. Morphet J, Innes K, Munro I, et al. Managing people with mental health presentations in emergency departments—A service exploration of the issues surrounding responsiveness from a mental health care consumer and carer perspective. *Australas Emerg Nurs J* 2012; 15: 148–155.
8. Pearlmutter MD, Dwyer KH, Burke LG, et al. Analysis of emergency department length of stay for mental health patients at ten Massachusetts emergency departments. *Ann Emerg Med* 2017; 70: 193–202.e116.
9. Fernandes M, Vieira SM, Leite F, et al. Clinical decision support systems for triage in the emergency department using intelligent systems: a review. *Artif Intell Med* 2020; 102: 101762.
10. Hong WS, Haimovich AD and Taylor RA. Predicting hospital admission at emergency department triage using machine learning. *PLOS ONE* 2018; 13: e0201016.
11. Iorfino F, Ho N, Carpenter JS, et al. Predicting self-harm within six months after initial presentation to youth mental health services: a machine learning study. *PLOS ONE* 2020; 15: e0243467.
12. Raita Y, Goto T, Faridi MK, et al. Emergency department triage prediction of clinical outcomes using machine learning models. *Crit Care* 2019; 23: 64.
13. Reale C, Novak LL, Robinson K, et al. User-Centered design of a machine learning intervention for suicide risk prediction in a military setting. *AMIA Annu Symp Proc* 2020; 2020: 1050–1058.
14. Gianfrancesco MA, Tamang S, Yazdany J, et al. Potential biases in machine learning algorithms using electronic health record data. *JAMA Intern Med* 2018; 178: 1544–1547.
15. Higgins O, Short BL, Chalup SK, et al. Artificial intelligence (AI) and machine learning (ML) based decision support systems in mental health: An integrative review. *Int J Ment Health Nurs* 2023; 32: 966–978.
16. Higgins O, Chalup SK and Wilson RL. Artificial intelligence in nursing: trustworthy or reliable? *J Res Nurs* 2024; 29: 143–153.
17. Liamputtong P. *Research Methods and Evidence-Based Practice*. Docklands: Oxford University Press Australia & New Zealand, 2021.
18. Kuhn M. *Applied Predictive Modeling*. New York, NY: Springer, 2013.
19. Kang H. The prevention and handling of the missing data. *Korean J Anesthesiol* 2013; 64: 402–406.
20. Francis G and Garing A. *Foundations of Statistics, Pearson Original Edition*. Melbourne, Australia: Pearson Education Australia, 2015.
21. Breiman L. Random forests. *Mach Learn* 2001; 45: 5–32.
22. Liaw A and Wiener M. Classification and regression by randomForest. *R News* 2002; 2: 18–22.
23. Biau G and Scornet E. A random forest guided tour. *Test* 2016; 25: 197–227.
24. Chen T and Guestrin C. Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining 2016, pp.785–794.
25. Chen T, He T, Benesty M, et al. Xgboost: extreme gradient boosting. *R Package Version* 04-2 2024; 1: 1–4.
26. Prokhorenkova L, Gusev G, Vorobev A, et al. CatBoost: unbiased boosting with categorical features. *Adv Neural Inf Process Syst* 2018; 31.
27. Dorogush AV, Ershov V and Gulin A. CatBoost: gradient boosting with categorical features support. *arXiv preprint arXiv:181011363* 2018.
28. Cover T and Hart P. Nearest neighbor pattern classification. *IEEE Trans Inf Theory* 1967; 13: 21–27.
29. Hastie T, Tibshirani R and Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer, 2013.
30. InterpretML Documentation. InterpretML Documentation, https://interpret.ml/docs/ (2023, accessed 06/04/2023 2023).

31. Rai A. Explainable AI: from black box to glass box. *J Acad Mark Sci* 2019; 48: 137–141.
32. Molnar C. *Interpretable Machine Learning*. Leanpub, 2020.
33. Nori H, Jenkins S, Koch P, et al. Interpretml: A unified framework for machine learning interpretability. *arXiv preprint arXiv:190909223* 2019.
34. Higgins O, Chalup SK and Wilson RL. Machine learning model reveals determinators for admission to acute mental health wards from emergency department presentations. *Int J Ment Health Nurs* 29 August 2024; 2024: 1–16.
35. Ben-Hur A, Horn D, Siegelmann HT, et al. Support vector clustering. *J Mach Learn Res* 2001; 2: 125–137.
36. Schölkopf B and Smola AJ. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. Cambridge, MA: MIT Press, 2002.
37. Vapnik V. *The nature of statistical learning theory*. New York, NY: Springer Science & Business Media, 2013.
38. Dong G and Liu H. *Feature engineering for machine learning and data analytics*. First edition. ed. Boca Raton, FL: CRC Press/Taylor & Francis Group, 2018.
39. Paper D. *Hands-on Scikit-Learn for Machine Learning Applications Data Science Fundamentals with Python*. 1st 2020. ed. Berkeley, CA: Apress, 2020.
40. Chen Q, Zhang-James Y, Barnett EJ, et al. Predicting suicide attempt or suicide death following a visit to psychiatric specialty care: a machine learning study using Swedish national registry data. *PLoS Med* 2020; 17: e1003416.
41. Zolbanin HM, Delen D, Crosby D, et al. A predictive analytics-based decision support system for drug courts. *Inf Syst Front* 2019; 22: 1323–1342.
42. Chicco D, Tötsch N and Jurman G. The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData Min* 2021; 14: 13.
43. Consoli S, Recupero DR and Petkovic M. *Data science for healthcare*. Switzerland: Springer, 2019.
44. InterpretML Documentation. Overall Importance, https://interpret.ml/docs/overall_importance/ (2023, accessed 06/04/2023 2023).
45. Tabachnick BG, Fidell LS and Ullman JB. *Using Multivariate Statistics*. London: Pearson, 2018.
46. NSW Health. Public Health Orders and restrictions, https://www.health.nsw.gov.au/Infectious/covid-19/Pages/public-health-orders.aspx (2022, accessed 09/03/2024 2024).
47. Melbourne Institute. *Coping with COVID-19: rethinking Australia, Chapter 4: Heightened Mental Distress: Can Addressing Financial Stress Help*. Melbourne: Melbourne Institute, 2020.
48. Malgaroli M, Hull TD, Zech JM, et al. Natural language processing for mental health interventions: a systematic review and research framework. *Transl Psychiatry* 2023; 13: 309.
49. Chandler C, Foltz PW and Elvevåg B. Improving the applicability of AI for psychiatric applications through human-in-the-loop methodologies. *Schizophr Bull* 2022; 48: 949–957.. 2022/06/01
50. Schepman A and Rodway P. Initial validation of the general attitudes towards artificial intelligence scale. *Computers in Human Behavior Reports* 2020; 1: 100014.

7.5.2 Article Summary

The results demonstrate the superior predictive performance of the non-linear models compared to the interpretable EBM in correctly identifying patients requiring inpatient admission. However, the EBM provided a more balanced performance across true/false positives and greater interpretability. While some models exhibited positive prediction capabilities, integrating interpretation techniques such as SHapley Additive exPlanations (SHAP) is crucial for ensuring transparency and fostering clinician trust in the model. The limited dataset poses challenges but is outweighed by the critical importance of accurate prediction in the MH ED context where misclassifications can severely impact patient well-being.

Key recommendations include further development and prospective testing of the non-linear model with SHAP interpretation in a clinical decision support system, validating performance across diverse settings, integrating richer data sources like electronic medical records, understanding clinicians' attitudes towards ML, and enabling real-time data collection for continuous model refinement. Ultimately, this research highlights the potential of ML models, when developed and implemented responsibly with a focus on transparency, to enhance decision making around MH inpatient admissions and improve quality of care. However, successful real-world deployment requires a multifaceted approach including collaborating with clinicians, patients, and other stakeholders to address trust barriers and ethical considerations.

7.6 Chapter Summary

This thesis has provided a comprehensive examination of the complexities surrounding MH care delivery in EDs in Australia, focusing on sociodemographic factors, presentation features, and the application of ML to enhance service provision. The findings underscore the critical role of EDs as frontline access points for MH care, while also highlighting significant challenges such as overcrowding, long wait times, and the need for specialised resources.

7.6.1 Sociodemographic Factors and Presentation Features

The analysis of sociodemographic factors and presentation characteristics of individuals seeking MH care in two regional EDs in New South Wales from 2016 to 2021 revealed several key insights. Suicidal ideation emerged as the most common presenting problem, indicating a pressing need for targeted interventions and safety planning. Additionally, the significant proportion of individuals departing without being seen or against medical advice underscores the urgency of improving access to specialised MH services. These findings call for a more integrated and holistic approach to MH care, bridging gaps between primary care, community-based services, and EDs.

7.6.2 Disproportionate Mental Health Presentations Among First Nations People

The study highlighted the disproportionate burden of MH challenges faced by Aboriginal and Torres Strait Islander communities, who experience higher rates of suicide, self-harm, and psychological distress compared to non-Indigenous populations. The over-representation of First Nations peoples in MH-related ED presentations underscores the need for culturally responsive care models. Recommendations include fostering cultural safety within EDs, integrating traditional healing practices, and developing trauma-informed interventions that resonate with Indigenous worldviews. By centring the voices and experiences of First Nations

peoples, healthcare systems can begin to rebuild trust and dismantle systemic barriers to quality MH care.

7.6.3 Machine Learning and Determinants of Admission to Acute Mental Health Wards

This study applied ML techniques to analyse ED presentations and identify the key factors influencing admissions to acute MH inpatient wards. The findings revealed significant correlations between features such as age, triage category, presenting problem, and prior healthcare utilisation, highlighting their impact on admission decisions. The study also brought attention to site-specific dynamics, such as the high volume of youth presentations and the lack of dedicated CAMHS beds, which pointed to the need for alternative care models like Youth Hospital in the Home. Additionally, the complex relationship between suicidal ideation and admission outcomes was identified, suggesting the need for further qualitative research on at-risk populations, particularly First Nations peoples and young people. The results underscore the potential of ML to support clinicians in making more informed and transparent decisions regarding MH admissions, ultimately contributing to improved service delivery and addressing key issues such as missed care and resource allocation in EDs.

7.6.4 Machine Learning Models to Assist Decision Making in the Assessment of Mental Health Consumers

The research demonstrated the potential of ML models to predict admissions to MH inpatient facilities from ED presentations. Non-linear models were found to outperform linear models in predictive accuracy. However, the importance of transparency, interpretability, and clinician trust in these models was emphasised. The integration of ML models into clinical decision support systems can enhance triage processes, optimise resource allocation, and improve overall care delivery. Further research is needed to ensure these models are ethically sound, clinically valid, and effectively integrated into existing healthcare workflows.

7.6.5 Implications and Future Directions

Addressing these areas will require a sustained, collaborative effort that centres on the voices and experiences of those most affected by MH challenges. By adopting a comprehensive, integrated approach to MH care, healthcare systems can better meet the needs of vulnerable populations, promote positive MH outcomes, and ensure that individuals experiencing MH crises receive the timely and appropriate care they need.

Overall, this research underscores the importance of innovative approaches to MH care delivery, including the application of advanced analytical techniques and the development of culturally responsive care models. By addressing the identified gaps and leveraging the potential of ML, healthcare providers and policymakers can improve service delivery, enhance patient outcomes, and promote a more equitable and effective MH care system.

Chapter 8 Discussion

8.1 Introduction

This thesis has presented a substantial body of literature that demonstrates the use of machine learning to assist decision making in the assessment of mental health consumers presenting to emergency departments. This research has improved understand of the complex characteristics of people who present with mental health (MH) issues in the emergency department (ED) environment and has helped to understand how these features contribute to the admission of people to MH units. The aim of this research was to develop, test, and evaluate machine learning (ML) models for predicting the likelihood of clients presenting to an ED with MH concerns, particularly suicidality, being admitted to an acute MH inpatient ward. The research objectives included enhancing clinicians' decision making by providing complementary data insights, reducing patient wait times in the ED, and improving the overall experience of MH services for clients by ensuring high-quality and safe care. Additionally, this study sought to deepen the understanding of artificial intelligence (AI) and ML adoption in MH contexts, examining how these technologies can be effectively integrated into clinical practice. This research has demonstrated the positive results that the use of ML for the prediction of admission can yield. It has also demonstrated that the use of ML has the potential to cause more harm if not used in an ethical manner and that its adoption will be limited if clinicians do not engage with it and trust its predictions. ML can potentially make significant changes in people's lives; however, without transparent models and clear ways of using them in practice, they will require significant planning and implementation to achieve adoption. This discussion chapter will provide further discourse on the key findings of this thesis.

8.2 AI in Complex Healthcare Environments

The initial letter that the research team received and the study "Interpretations of innovation: The intersection of technological advancement and psychosis", presented in Chapter 2 of this thesis, demonstrates the importance of keeping the person at the centre of the work that clinicians and healthcare services do. However, in order to progress the use of AI in MH, we must first discuss the potential benefits, challenges, and ethical considerations surrounding the integration of AI in health care, particularly in MH nursing. While AI has the potential to revolutionise health care by improving efficiency, consistency, and patient care (Australian College of Nursing, 2024; Pailaha, 2023; Saban & Dubovi; Van der Gaag et al., 2023; Wilson, Higgins et al., 2023), there is considerable research that cautions against the naive adoption of AI, emphasising the need for careful consideration of potential risks and limitations (Australian College of Nursing, 2024; Coiera & Magrabi, 2023; Saban & Dubovi, 2024; Van der Gaag et al., 2023; Wilson, Higgins et al., 2023). Any intervention that uses AI in MH must be approached with a degree of caution as AI may not always accurately interpret complex human behaviour, particularly in MH, and relying solely on AI for diagnosis or treatment without human oversight could have negative consequences (Saban & Dubovi, 2024; Wilson, Higgins et al., 2023). It is imperative that we emphasise the importance of human oversight in AI-driven MH care and be open about the potential for misinterpretation of complex human behaviour (Saban & Dubovi, 2024; Wilson, Higgins et al., 2023). AI systems may struggle to accurately interpret the nuances of human emotions, behaviours, and communication, which are crucial aspects of MH assessment and treatment (Saban & Dubovi, 2024; Wilson, Higgins et al., 2023). An over-reliance on AI for diagnosis or treatment recommendations without human oversight could result in several critical issues, for example:

- *Misdiagnosis:* Due to the complexities of human behaviour and the potential for AI to misinterpret subtle cues, there is a risk of inaccurate diagnoses, which could lead to

inappropriate or even harmful treatment plans (Saban & Dubovi, 2024; Wilson, Higgins et al., 2023).

- *Inappropriate interventions:* AI systems, without the capacity for nuanced understanding, may recommend interventions that are not suitable for a patient's unique circumstances or cultural background (Saban & Dubovi, 2024; Wilson, Higgins et al., 2023).

- *Failure to recognise underlying issues:* AI's focus on data analysis might overlook underlying psychological or social factors contributing to a patient's MH issues (Saban & Dubovi, 2024; Wilson, Higgins et al., 2023). Human clinicians are better equipped to recognise these complexities and provide holistic care (Saban & Dubovi, 2024; Wilson, Higgins et al., 2023).

To demonstrate AI's limitations in understanding complex situations, Saban and Dubovi (2024) compared the performance of ChatGPT, a generative AI tool, with human clinicians (expert nurses and novice nursing students). They found that ChatGPT exhibited indecisiveness when presented with clinical scenarios requiring nuanced judgement (Saban & Dubovi, 2024). Specifically, it struggled to accurately assess the urgency of patient situations, often suggesting unnecessary diagnostic tests, which indicates a tendency towards over-triage (Saban & Dubovi, 2024). This example highlights the importance of a cautious approach to integrating AI into clinical practice, particularly in MH, where human interaction and interpretation are crucial. It also highlights the need for continuous monitoring, evaluation, and human oversight to ensure patient safety, prevent misinterpretations of complex human behaviour, and guide ethical AI development in MH care (Saban & Dubovi, 2024; Wilson, Higgins et al., 2023).

AI can potentially transform health care, but it also presents significant challenges that must be addressed to ensure ethical and effective implementation. Another major concern is the risk of

codifying existing biases in AI systems, which could lead to discriminatory outcomes, especially when algorithms are trained on biased data (Van der Gaag et al., 2023). Additionally, the negative perceptions associated with AI, such as concerns about surveillance and impersonal decision making, highlight the importance of using terms like “decision support” to emphasise human oversight (Van der Gaag et al., 2023). Accountability and transparency in AI-driven decisions are also critical, as unclear responsibility can erode trust in regulatory processes and the healthcare system, as discussed in Chapter 4 (Van der Gaag et al., 2023). Privacy and data security concerns arise from the handling of sensitive patient data, necessitating robust governance models and security measures (Australian College of Nursing, 2024; Coiera & Magrabi, 2023). Furthermore, ensuring informed consent for AI-driven interventions is essential to maintaining ethical practice, alongside efforts to prevent AI from exacerbating existing healthcare inequities (Coiera & Magrabi, 2023; Hantel et al., 2024). The “human touch” in nursing care must be preserved, with AI augmenting rather than replacing empathetic and compassionate interaction (Coiera & Magrabi, 2023; Pailaha, 2023). Increased education and training for nurses on AI technologies, ethical considerations, and data governance are recommended to mitigate these risks. Ongoing collaboration between nurses, AI developers, regulators, and the public is essential to implement AI safely and ethically, benefiting both patients and the nursing profession (Australian College of Nursing, 2024; Coiera & Magrabi, 2023; Hantel et al., 2024; Saban & Dubovi, 2024)

8.3 Overview of Sociodemographic Factors and Presentation Features

The findings of Study A (Higgins, Sheather-Reid et al., 2024c) provide a further understanding of the sociodemographic factors and presentation characteristics of individuals seeking MH care in EDs from one local health district in New South Wales, Australia, from 2016 to 2021. The study identified several key sociodemographic factors and presentation features associated

with MH care in EDs, including age, gender, mode of arrival, presenting problem, and outcomes such as admission to inpatient MH facilities.

Suicidal ideation emerged as the most common presenting problem, accounting for a substantial percentage of ED visits. Suicide continues to remain a significant global public health issue; approximately 700,000 people die by suicide each year (World Health Organization, 2022b). In the United States from 1999 to 2018, suicides increased by 35% (Hedegaard et al., 2020) before reversing direction for two consecutive years, but rising again in 2021 (Martínez-Alés et al., 2022). Study A found that a significant proportion of ED presentations for MH reasons involved young adults and adolescents, with a notable prevalence of suicidal ideation among these groups. This aligns with existing research indicating that younger populations are at a higher risk for MH crises, often leading to ED visits (Benton et al., 2021; Sara et al., 2019; Slankamenac et al., 2020).

The challenges of ED overcrowding and prolonged wait times identified in this study are well-documented in the literature (Clarke et al., 2007; Fatovich & Hirsch, 2003; Wand et al., 2021). In Study A, a notable percentage of individuals left the ED without being seen or against medical advice. This phenomenon may reflect systemic issues such as long wait times, inadequate staffing, and insufficient MH resources within EDs (Clarke et al., 2007; Fatovich & Hirsch, 2003; Wand et al., 2021). Understanding and addressing these factors will be crucial for improving patient care and outcomes, as these issues not only exacerbate patient distress but also strain healthcare resources and compromise the quality of care (Harrou et al., 2020; Ortíz-Barrios & Alfaro-Saíz, 2020; Rowe & Knox, 2023). Ortíz-Barrios and Alfaro-Saíz (2020) suggest the use of integrated care models that include operations research methods, quality-based techniques, and data-driven approaches, which may help improve triage systems by mitigating these problems and enhancing the efficiency of ED operations. The adoption of integrated care models that bridge the gap between EDs, primary care, and community-based

MH services should strive to facilitate continuity of care, reduce repeat ED visits, and improve overall patient outcomes (Beaudin et al., 2022). This underscores the need for systemic changes to implement and support collaborative care frameworks effectively (Rugkåsa et al., 2020).

8.4 Implications for ED Practices and Policies

EDs often serve as the initial point of contact for individuals experiencing MH crises, including those with suicidal ideation. This research has demonstrated the high demand for MH services in ED, especially for people who are suicidal., highlighting the critical need for effective triage and intervention strategies.

8.4.1 Triage and Intervention

Effective triage is crucial in identifying patients at immediate risk and ensuring timely intervention. Current triage systems assign priority based on clinical urgency (NSW Health, 2013), but they may not adequately address the nuanced needs of MH patients (Morphet et al., 2012; Wand et al., 2021), with many EDs struggling to implement successful triage methods in regard to MH (Berendsen Russell et al., 2017; Morphet et al., 2012) . The triage of people who present to ED with MH problems requires high-level competencies in terms of knowledge, skills, and attitudes in what can often be unpredictable and complex patient situations (Stigter-Outshoven et al., 2024) and relies on the knowledge and use of the MH triage scale and experience of MH acuity (Broadbent et al., 2020). The challenges posed by MH presentations in EDs underscore the need for greater infrastructure, resources, knowledge and training to provide specific and appropriate care to people with MH problems (Boudreaux et al., 2016; Digel Vandyk et al., 2018).

8.4.1.1 Current Practices and Areas for Improvement

People who present with MH problems often pose challenges to consumer flow in EDs as they usually require extended assessment times (Morphet et al., 2012). ED staff repeatedly convey feeling ill-prepared to handle these consumers (Clarke et al., 2007) and may be hindered by conflicting perspectives between MH and ED clinical groups (Broadbent et al., 2020). The Australasian College for Emergency Medicine (2018) found that ED staff did not feel adequately prepared to handle MH emergencies, pointing to a gap in training and resources. However, it is difficult to compare triage results in context, as models of triage, organisation and the culture of emergency clinicians are highly variable (Varndell et al., 2019). The focus on MH of ED clinicians, however well-meant and trained, is quickly eroded by the time limitations and workflow of the day-to-day ED setting and the turnover of people presenting, not just those with MH issues (Hall et al., 2016). Those who attend the ED for MH problems report that they are often unsure about how care is prioritised, with long waits often following arrival at ED (Fatovich & Hirsch, 2003), and this can lead to discontent and frustration among many ED consumers (Morphet et al., 2012).

8.4.1.2 Integrating Safety Planning and Specialised Mental Health Resources

Safety planning is an evidence-based intervention that helps individuals recognise warning signs, employ coping strategies, and access emergency contacts and resources. A meta-analysis by Nuij et al. (2021) examined the effectiveness of safety planning-type interventions in the reduction of suicidal behaviour. Their analysis included six studies with a total of 3,536 participants and included various settings, such as EDs, military hospitals, and case management services. The interventions involved safety plans that included coping strategies, sources of support, and personal warning signs of a suicidal crisis (Nuij et al., 2021). The results of this research showed that the use of safety planning is associated with a significant reduction

in suicidal behaviour, with the risk of suicidal behaviour for participants who received a safety planning intervention being 0.57 times the risk of other patients and an overall decrease of 43% in the risk of suicidal behaviour (Nuij et al., 2021), showing that this warrants further exploration.

Recent research conducted by Lohani et al. (2024) has provided insight into the potential benefits of incorporating higher levels of patient–clinician collaboration during suicide risk assessment processes. Lohani et al. (2024) randomly assigned participants (n = 82) to one of four interventions that varied in patient–clinician collaboration. Their findings suggest that clinician-led collaborative risk assessment approaches can effectively augment the effects of safety planning-type interventions among patients exhibiting elevated risk for suicide when compared to traditional checklist-based assessment approaches. Lohani et al. (2024) recommend employing a narrative assessment technique, which fosters greater collaboration between the clinician and the patient and can significantly reduce the risk of suicidal ideation. Specifically, their findings indicate that pairing narrative assessment with a safety plan intervention or a crisis response plan can lead to a notable reduction in suicidal thoughts among at-risk individuals. The study underscores the importance of collaborative risk assessments in enhancing the efficacy of safety planning-type interventions. By actively involving patients in the assessment process and encouraging open dialogue, clinicians can better understand the individual's unique circumstances, challenges, and perspectives. This collaborative approach promotes a more holistic and personalised assessment and empowers patients to actively participate in their own care, fostering a sense of agency and self-determination (Lohani et al., 2024). Overall, the research by Lohani et al. (2024) highlights the potential benefits of adopting a patient-centred, collaborative approach to suicide risk assessment, emphasising the synergistic effects of narrative assessment techniques and safety planning interventions in mitigating suicidal ideation and promoting better MH outcomes.

The incorporation of specialised MH resources, such as in-house MH services and crisis intervention teams, can improve care and result in a marked improvement in the experiences of patients and staff. Improvements may also involve standardising screening tools and requiring MH training for all ED staff (Wand et al., 2021). Evans et al. (2019) conducted a comprehensive systematic review to examine various consultation liaison (CL) models within ED setting. They identified and categorised CL services into four distinct models, each demonstrating unique impacts on patient care and operational efficiency, and noted that CL models where MH personnel were integrated directly into the ED team or triage processes showed significant benefits (Evans et al., 2019). These integrated models were associated with reduced patient waiting times, potentially lowering the incidence of patients leaving without being seen. Additionally, these models yielded high levels of staff satisfaction, suggesting that the integration of MH personnel into the ED team enhances both operational efficiency and workplace morale (Evans et al., 2019). Secondly, the establishment of co-located MH spaces or the presence of MH personnel within the ED also contributed to decreased patient waiting times. This structural approach enabled faster access to MH services, thereby improving patient flow within the ED (Evans et al., 2019). While integrating MH clinicians in the ED triage setting may appear promising in reducing waiting times and improving staff satisfaction, further research is required to evaluate the efficacy, cost-effectiveness, and patient satisfaction associated with the various CL models.

8.5 Disproportionate Mental Health Presentations Among First Nations People

Aboriginal and Torres Strait Islander people in Australia exhibit disproportionately high rates of ED presentations for MH issues (Australian Institute of Health and Welfare, 2022a). This over-representation is reflective of broader health disparities, cultural disruptions, intergenerational trauma, and systemic biases within the healthcare system (Australian Institute

of Health and Welfare, 2022a; Paradies, 2016; Wilson & Waqanaviti, 2021). First Nations Australians are more likely to suffer from chronic diseases such as diabetes, cardiovascular disease, and respiratory conditions, which often intersect with MH problems, leading to complex presentations in EDs (Australian Institute of Health and Welfare, 2022a). Furthermore, there is a significant gap in life expectancy between First Nations and non-First Nations Australians, largely due to preventable health conditions and inadequate access to healthcare services (Nolan-Isles et al., 2021). Paradies (2016) reports that systemic biases, including racism and discrimination within healthcare settings, deter First Nations Australians from seeking help until crises occur. The healthcare system's failure to provide culturally safe and competent care has led to mistrust and an underutilisation of services (Durey & Thompson, 2012; Paradies, 2016). These issues are further compounded by a history of cultural disruptions that include the loss of land and cultural practices and have had a profound impact on the MH of First Nations populations (Gee et al., 2014). Colonisation has left a legacy of forced removals and assimilation policies that have resulted in intergenerational trauma and the manifestation of higher rates of MH issues (Atkinson et al., 2014; Wilson & Waqanaviti, 2021). The Stolen Generations, in particular, have had lasting effects on family structures and community cohesion, further contributing to MH crises (Haebich, 2011; Wilson & Waqanaviti, 2021). The rate of ED presentations for MH issues among Aboriginal and Torres Strait Islander people is significantly higher than that of non-First Nations populations, as seen in Chapter 7 (Australian Institute of Health and Welfare, 2022a). First Nations people's use of EDs as a primary point of care for MH issues is indicative of broader systemic health inequities (Kelaher et al., 2014) such as limited access to primary healthcare services due to geographical, financial, and cultural barriers (Browne et al., 2015).

8.5.1 Recommendations for Culturally Responsive Care Models

Study B (Higgins, Sheather-Reid et al., 2024a) examined the Aboriginal and Torres Strait Islander people of the Central Coast NSW who presented to the ED for MH issues between 2016 and 2021. It is important that the findings be contextualised in relation to the poorer MH outcomes associated with cultural disruptions, which can be viewed as a logical consequence of the pervasive intergenerational trauma stemming from the colonisation experienced by First Nations peoples (Wilson & Waqanaviti, 2021). Furthermore, we need to consider the biases inherent in the default Western models of mental health service delivery and the challenge of aligning these models with holistic Indigenous frameworks that prioritise culturally enriched ways of understanding, living, acting, and fitting in, especially in relation to concepts of social and emotional well-being (Gee et al., 2014; McGough et al., 2022).

Many First Nations people, such as the Stolen Generations, have experienced dislocation from their cultural lands and families, being required to assimilate (Wilson & Waqanaviti, 2021). Others have endured generational housing, economic, and social insecurity (Wilson & Waqanaviti, 2021). First Nations individuals born before 1972 (aged 44–50 or older during the data collection period) experienced an extensive period of trauma and have significantly worse outcomes associated with MH (Australian Institute of Health and Welfare, 2021a). The access to care among older age groups, as evident in the results of this research, is potentially related to feelings of insecurity when accessing healthcare services (Nolan-Isles et al., 2021; Ware, 2013), which echoes the findings of Campbell et al. (2018).

The over-representation of Aboriginal and Torres Strait Islander populations in ED presentations for MH issues necessitates a multifaceted approach. Addressing broader health disparities, the impacts of cultural disruptions, intergenerational trauma, and systemic biases, along with implementing culturally responsive care models, is essential for achieving equitable

MH care for First Nations Australians. It requires culturally responsive care models, for example the integration of traditional healing practices alongside Western medical models to provide a holistic approach to MH care (Gee et al., 2014). We must collaborate with First Nations communities to co-design MH services to ensure they are culturally appropriate and meet the specific needs of the community (Dudgeon et al., 2014). This should include engaging with local First Nations Elders and healers to incorporate culturally significant practices into MH care provision (Bainbridge et al., 2015). By adopting these recommendations, the healthcare system can better meet the needs of Aboriginal and Torres Strait Islander people, improving their MH outcomes and overall well-being.

8.6 Application of Machine Learning in Mental Health Care

Study D (Higgins, et al., 2024d) provides valuable insights into the application of ML models for predicting admissions to MH inpatient wards based on routine data collected at triage in EDs. The study compared the performance of six different ML models including the explainable boosting machine (EBM) from the InterpretML framework, kNN, XGBoost, CatBoost, random forest, and the support vector machine (SVM) using support vector classification (SVC) with a radial basis function kernel.

8.6.1 Analysis of the Performance of ML Models in Predicting Admissions to Inpatient Mental Health Facilities

In this study CatBoost was the most accurate model, offering a good balance between recall and precision, though improvements in positive case detection are needed. Models such as EBM InterpretML and random forest exhibited good specificity but faced challenges with recall and false negatives, crucial in healthcare applications. Unlike random forest, EBM InterpretML provided insights into its recommendations, making it suitable for general healthcare applications where transparency is crucial alongside prediction.

The comparison between EBM InterpretML, CatBoost, and random forest highlights the trade-offs between interpretability and predictive performance in health care. EBM InterpretML had a moderate accuracy of 0.7883 and a balanced distribution of true and false positives, indicating moderate effectiveness in both positive and negative predictions. Its primary advantage was interpretability, making it suitable for applications where understanding the decision-making process is essential. However, EBM InterpretML's lower recall of 0.1082, compared to CatBoost, suggests it missed more positive cases, a significant consideration in health care where missing patients needing intervention could have severe consequences. The model's precision was 0.6158, and its specificity was 0.9809, indicating excellent true negative identification.

CatBoost excelled at correctly identifying positive cases, such as patients requiring MH inpatient admission. It achieved the highest accuracy at 0.7900, demonstrating strong overall performance. With a recall of 0.1307, CatBoost was more effective at capturing true positives, reducing the likelihood of missing critical care patients, making it suitable for scenarios prioritising maximising positive case detection over interpretability. CatBoost's precision was 0.5720, and its specificity was 0.9723, reflecting its ability to avoid false positives while effectively identifying true positives.

In contrast, the random forest model offers a balanced approach with an accuracy of 0.7776 and recall of 0.1152. Although it did not have the highest sensitivity, it performed well across various metrics, balancing interpretability and predictive power. Random forest's recall was slightly higher than EBM InterpretML, suggesting a better ability to capture positive cases, though it still lagged behind CatBoost. The model's precision was 0.4836, and its specificity was 0.9652, indicating good true negative identification, crucial for minimising unnecessary interventions.

Linear models, such as the EBM InterpretML, are generally easier to interpret and can provide insights into the relationships between features and outcomes (Molnar, 2020). However, their performance can be limited when dealing with complex, non-linear data patterns (Molnar, 2020). In contrast, non-linear models like the CatBoost and random forest can effectively capture intricate relationships, leading to improved predictive accuracy (Schölkopf & Smola, 2002). However, this increased complexity often comes at the cost of reduced interpretability, which can hinder clinicians' trust in and acceptance of the model's predictions (L. A. Brown et al., 2020; Eltorai et al., 2024).

8.6.2 Importance of Transparency, Interpretability, and Clinician Trust

Ensuring transparency and interpretability is crucial for gaining clinicians' trust and acceptance of ML models in healthcare settings. Without a clear understanding of the model's decision-making process, clinicians may be hesitant to rely on its predictions, especially in sensitive areas like MH care (Evans et al., 2019). The literature review highlighted the importance of addressing this barrier to the adoption of ML innovations in clinical practice. To address the interpretability challenge, researchers can employ techniques such as SHapley Additive exPlanations (SHAP), or InterpretML, which provide insights into the contributions of individual features to the model's predictions (Lundberg & Lee, 2017) in addition to interpretable models such as EBM InterpretML. By integrating explanations with models that are difficult to understand, clinicians can gain valuable insights into the factors influencing the predictions of admission to MH inpatient wards, fostering trust and enabling more informed decision-making processes. Additionally, engaging clinicians throughout ML models' development and implementation phases is crucial as this collaboration can address concerns, incorporate subject matter expertise, and ensure that the models align with clinical workflows and priorities. Integrating ML models into clinical decision support systems (CDSSs) holds promise for enhancing decision-making processes in the ED setting, as CDSSs in the ED have

been shown to improve clinical management and patient outcomes (Fernandes et al., 2020). While further research is required, incorporating the various models and the SHAP explainer into a CDSS has the potential to provide clinicians the opportunity to leverage the predictive power of ML while benefiting from interpretable explanations, potentially leading to more efficient triage and resource allocation for MH patients.

8.6.3 Ethical Considerations and Potential Biases in ML Applications

While ML models offer promising applications in health care, it is essential to address ethical considerations and potential biases. As Gianfrancesco et al. (2018) emphasise, ML algorithms trained on electronic health record data may perpetuate or amplify existing biases present in the data. This is particularly concerning when dealing with vulnerable populations, such as individuals with MH conditions or groups such as First Nations people, as inherent system biases could lead to disparities in care and further discrimination (Gianfrancesco et al., 2018). To mitigate potential biases, researchers must employ rigorous data pre-processing techniques, such as identifying and addressing any imbalances or under-representation of certain groups in the training data. Once models have been built, they require continuous monitoring and evaluation of the model's performance across diverse patient populations to ensure fairness and accuracy (Gianfrancesco et al., 2018). It is also essential to consider the potential implications of model predictions on different demographic groups and, if disparities in model outcomes are identified, appropriate measures need to be taken to address them, such as adjusting the model's prediction thresholds or implementing additional safeguards to ensure equitable treatment (Gianfrancesco et al., 2018).

Study D (Higgins et al., 2024d) highlighted the potential of ML models, particularly the non-linear models, in predicting admissions to MH inpatient wards based on routine ED data. However, it also underscored the importance of transparency, interpretability, and clinician

trust if health services are to be successful in the implementation of these predictive models. Strategies such as integrating SHAP explanations and actively engaging clinicians throughout the development process can enhance model transparency and foster acceptance. Additionally, addressing ethical considerations and potential biases is crucial to ensure fairness and accuracy in predictive modelling, particularly when dealing with vulnerable populations. By focusing on these challenges, ML models can be effectively integrated into clinical settings, ultimately improving the quality of care and resource allocation for MH patients in the ED setting.

8.6.4 Addressing Missed Care Through Machine Learning and Future Research

This research provides valuable insights into how the integration of ML and advanced analytical techniques can address the issue of missed care in clinical settings. Missed care—defined as essential elements of patient care that are delayed, omitted, or incomplete—remains a significant challenge in healthcare systems globally, driven by factors such as high patient volumes, understaffing, and the growing complexity of patient needs (Ball et al., 2018; Jones et al., 2015). This issue is particularly acute in areas where time pressures, overcrowding, and the need for quick decision making often result in gaps in care, contributing to workforce burnout and adversely affecting patient outcomes (Mudallal et al., 2017).

Study C, which applied ML models to identify key factors influencing admissions to MH inpatient facilities, offers a crucial foundation for addressing missed care. By uncovering correlations between variables such as triage category, presenting problem, and prior healthcare utilisation, Study C provides actionable insights that support clinicians in making informed, data-driven decisions. This addresses one of the root causes of missed care: the cognitive overload experienced by healthcare staff in busy ED environments. With interpretable ML models, clinicians can access clear, real-time insights that prioritise urgent cases, enabling more effective allocation of time and resources. This would contribute to reducing omissions and

delays, particularly in complex mental health cases, where prolonged assessments are often necessary (Clarke et al., 2014).

Building on these findings, Study D focused on developing predictive models for MH admissions, further advancing the potential of AI-driven CDSS to mitigate missed care. ML's predictive capabilities allow for the early identification of high-risk patients, enabling better triage and resource allocation. By automating parts of the decision-making process, ML models can relieve clinicians of administrative burdens, freeing up time for direct patient care and reducing the risk of missed care (Topaz & Pruinelli, 2017). Additionally, these models help ensure that critical interventions are delivered in a timely manner, enhancing both care delivery and patient safety (Seibert et al., 2021).

The implications of these studies are significant. Integrating ML models into care workflows can help address the increasing demand for services while alleviating workforce challenges, such as clinician burnout, which are known contributors to missed care (Mudallal et al., 2017). These technologies offer scalable solutions for improving care delivery, optimising resource allocation, and enhancing patient safety, particularly in overstretched environments where human resources are limited (Bates et al., 2014).

Future research should focus on refining and validating these ML models in real-world clinical settings, with a particular emphasis on addressing ethical considerations such as bias and ensuring equity in care (Floridi et al., 2018). Additionally, exploring the lived experiences of patients and healthcare providers will provide a deeper understanding of how AI-driven solutions can be integrated into clinical practice to foster trust and enhance care quality (Iorfino et al., 2020; Raita et al., 2019).

This research has highlighted the significant potential of AI and ML to challenge the issue of missed care on a global scale, offering a data-driven solution to enhance decision making,

alleviate clinician burnout, and improve patient outcomes (Seibert et al., 2021). By incorporating these technologies, healthcare systems can begin to address the widespread challenges associated with missed care and deliver more timely and effective care to vulnerable populations. However, addressing the issue of missed care on a global level requires a holistic approach that recognises the interconnectedness of nurse burnout, rising healthcare demand, and patient outcomes. Simply using AI will not address the underlying root causes, such as staffing shortages, limited resources, and systemic inefficiencies; it is crucial for healthcare systems to reduce the adverse effects on both nurses and patients.

8.7 Recommendations for Future Research, Clinical Practice and Policy

8.7.1 Health Service Systems

8.7.1.1 Mental Health Clinicians at the Point of Triage

Integrating specialised MH clinicians at the triage stage may significantly enhance the quality of care provided to patients presenting with MH issues. Specialised MH clinicians are trained to assess, triage, and manage MH crises effectively, providing timely and appropriate interventions. Their presence in EDs can reduce wait times, improve patient outcomes, and alleviate the pressure on general ED staff, who may not have specialised training in MH care (Evans et al., 2019). Further research that extends the work of Evans et al. (2019) could develop more comprehensive ED MH models of care in collaboration with advanced practice nurses, such as nurse practitioners (Nicholls et al., 2011).

8.7.1.2 Peer-Led Qualitative Research in ED and Safety Planning to Address Suicidal Ideation

Based on the findings of this research, further qualitative and peer-led research is recommended to better understand the complexity of this phenomenon and address the specific needs of at-

risk populations, particularly First Nations communities, young people, and individuals presenting with suicidal ideations. This research should encompass the use of suicide prevention therapeutic interventions, such as safety planning (Ferguson et al., 2022; Nuij et al., 2021), and utilise qualitative peer-led research methodologies to build understanding of the multifaceted nature of suicidal ideations, which cannot be fully captured through quantitative methods alone. This research is crucial for uncovering the experiences and perspectives of those affected, providing unique insights, fostering trust within these communities, and delivering innovations such as telehealth, virtual care, virtual reality, and hospital in the home models (Bell et al., 2024).

8.7.2 Improving First Nations Cultural Safety in ED

Health services and clinicians must go beyond the traditional perspective of cultural safety and critically examine their behaviours, beliefs, and practices that perpetuate issues such as institutional racism (Ramsden, 2002). It is imperative to transition from cultural sensitivity training to a model of cultural humility, which fosters an understanding of the impact of colonisation and racism on health care for First Nations peoples (Ramsden, 2002). For example, in the field of nursing and midwifery, programs such as Murra Mullangari: Introduction to Cultural Safety and Cultural Humility aim to provide an understanding of the history, origins, and concepts of cultural safety and cultural humility while offering insights into the culture of power and its implications on the nursing workforce (Congress of Aboriginal and Torres Strait Islander Nurses and Midwives, 2022).

Given the disproportionate representation of First Nations peoples in MH-related ED presentations, it is imperative to prioritise a research perspective centred on First Nations communities. This necessitates employing First Nations research methods such as yarning and storytelling (Dudgeon et al., 2020; Walker et al., 2014) and ensuring that First Nations people

lead research priorities, governance, and benefits within their respective communities (Harfield et al., 2020). This research should avoid "deficit discourse", which refers to negative language and behaviour that represents people in terms of their deficiencies and failures (Lowitja Institute, 2018), as the focus on failure and dysfunction can obscure the complex reasons for inequalities and overlook diversity, capability, and strength (Lowitja Institute, 2018). Instead, future research must strive to adopt a "strengths-based approach" to shift the language of the narrative to one that aligns with existing First Nations approaches (Lowitja Institute, 2018) and reflect care appropriate to and consistent with First Nations peoples' spiritual beliefs and cultural practices, including the involvement of traditional healers and local Elders (Rooney et al., 2023). This approach is essential for comprehensively understanding the unique cultural needs and upstream factors influencing First Nations individuals seeking help and accessing MH care through EDs.

8.7.3 Machine Learning Recommendations

8.7.3.1 Transparency and Explainability

Further development and evaluation of various models of ML-based CDSSs to enhance triage decision-making processes in the ED are necessary before deploying any such model. Despite this, the high predictive accuracy of the models tested in this research demonstrates significant potential to assist clinicians in identifying patients who require admission to a MH inpatient ward, potentially reducing wait times and improving overall care quality. Implementing the models in clinical practice would necessitate the integration of the EBM InterpretML or SHAP explainer alongside the non-linear models to provide clinicians with interpretable visualisations and explanations of the models' predictions. This strategy can improve trust and acceptance among healthcare professionals, addressing the concerns regarding barriers to ML adoption in clinical settings. Models should employ an explainable design, such as InterpretML, and be

integrated into digital patient record interfaces. Additionally, further research is required to understand the impact of ML in care delivery on individuals with lived experiences of mental illness. ML interventions should be characterised by competence, reliability, and validity, aligning with the standards expected of other clinical tools where quality and safety are paramount. A comprehensive description of decision logic functions is essential to support the implementation of AI-based interventions and any implementation should include clear details explaining how clinical features contributed to the outcomes while communicating multiple recommendations that demonstrate confidence in these recommendations. Treatment recommendations should be presented to clinicians with validated predictions, ensuring that clinicians retain the final decision-making authority regarding the appropriate course of treatment.

8.7.3.2 Clinician Barriers and Involvement

For MH clinicians, ML possesses the potential to facilitate early decision making, enhance patient experiences, assist in care coordination, and optimise the triage process. It is imperative that MH nurses engage proactively in the developmental stages to ensure the incorporation of accurate data and the utilisation of clinically valid sources in any model related to patient care. Nonetheless, significant challenges hinder the adoption of ML in clinical practice, notably issues related to clinicians' trust and confidence, and the transparency of these systems (Eltorai et al., 2024). Effective clinician participation necessitates involvement in every phase of design and development, from initial specification to final evaluation. Integration efforts should clearly define actions, incorporate the clinical measures informing outcomes, convey multiple recommendations, and bolster confidence in the proposed recommendations. Further research is essential to comprehend the obstacles to AI integration among clinicians and to explore their concerns regarding the interpretability and transparency of ML models. Instruments such as the General Attitudes towards Artificial Intelligence Scale (GAAIS) (Schepman & Rodway,

2020) could be instrumental in understanding the apprehensions or negative perceptions of ML among MH clinicians. This insight could be used in the development of targeted educational and training programs aimed at addressing these concerns. Such initiatives would fulfil clinicians' need to comprehend the decision-making processes of ML models, thereby fostering increased trust and acceptance of ML technologies. By addressing these concerns, such programs would promote increased transparency, enabling clinicians to comprehend how AI-driven decisions are made, which in turn could foster greater trust and acceptance of ML technologies. Study E's exploration of clinicians' attitudes towards AI and its findings on the necessary conditions for trust in AI systems will be critical in ensuring that the integration of these technologies is both effective and ethically sound. This understanding is essential for developing AI tools that not only enhance clinical decision making but also are embraced by the very professionals they are designed to support.

8.7.3.3 Integration of Machine Learning and Decision Tools

Aligning ML models with service delivery processes will enable services to better understand their users and provide insights into financial and clinical variations. Future research should investigate the integration of multiple data sources, including electronic medical records (EMRs), patient-reported outcomes, and social determinants of health, into the ML modelling process. By leveraging a more comprehensive dataset, ML models can more accurately capture the complexities of MH conditions and offer more informed and tailored predictions for admission to MH inpatient wards. Real-time data collection and monitoring are essential for the effective implementation and continuous enhancement of ML models. By capturing patient data as it becomes available, models can be regularly updated and refined to reflect the latest trends and patterns in MH presentations. It is recommended that a comprehensive data collection and monitoring system be developed, enabling real-time capture of patient data in both the ED and MH settings in the context of AI/ML. This system should integrate seamlessly

with the existing EMR infrastructure, facilitating efficient and accurate data extraction for the ML models. Finally, further research is required to validate the performance of the linear and non-linear models across various healthcare settings and diverse patient populations. This research should involve continuous monitoring and refinement of the model to ensure its reliability and validity in real-world clinical contexts. Developing a comprehensive implementation plan is also crucial, addressing ethical considerations, data privacy, and potential biases. It is essential to involve stakeholders, including clinicians, patients, and policymakers, throughout the development cycle to maintain a transparent and inclusive process.

8.8 Reflection on Research Limitations

It is imperative to acknowledge the inherent limitations that may impact the study's findings and their applicability. While this research provides valuable contributions to the field, various constraints, such as data generalisability, potential biases, and the evolving nature of technology, present challenges that must be carefully considered. This section critically reflects on these limitations, discussing their implications for the research outcomes and suggesting ways to address or mitigate them in future studies. Understanding these limitations is crucial for situating the research within the broader context of clinical practice and AI in MH services.

8.8.1 Data Constraints and Generalisability

One of the primary limitations of this study lies in the data constraints, particularly concerning the generalisability of the findings. The data used in this research was region-specific, which inherently limits the extent to which the results can be applied to other regions or populations. While the methodological approach employed in this study could be replicated in different regions using similar datasets, the findings themselves may not be universally applicable. This

limitation underscores the need for caution when attempting to generalise the results beyond the specific context of the study (Creswell & Creswell, 2018).

8.8.2 Potential Biases in Data Collection and Analysis

Another significant limitation relates to potential biases in data collection and analysis. The data utilised in this research was sourced from triage nurses rather than MH-trained nurses. This distinction is crucial as it may impact the overall quality and relevance of the data, potentially introducing biases that could affect the study's outcomes. While this issue partly stems from systemic limitations, it is essential to acknowledge that the data's origin may influence the results, thereby affecting the study's validity (Polit & Beck, 2021).

8.8.3 Clinical Validation and Data Quality

The clinical validation of the findings is another area that warrants discussion. The data used in this research, derived from triage nurses, may not represent the highest possible quality of clinical data. This limitation, while recognised, is a reflection of the available data rather than a flaw in the research design. Future research should strive to incorporate data from MH-trained nurses or other relevant clinical sources to enhance the validity and reliability of the findings (Greenhalgh, 2019).

8.8.4 Data Type and Complexity

The research was constrained by the type of data available, which consisted primarily of simple presentation data rather than complex clinical data. The inclusion of more nuanced data, such as clinical notes, could have provided deeper insights and enriched the analysis. However, incorporating such data would necessitate the use of more sophisticated AI algorithms, which may introduce additional complexities and challenges (Esteva et al., 2021). Future studies

should consider leveraging advanced AI techniques to analyse richer clinical datasets, potentially leading to more comprehensive and actionable insights.

8.8.5 Legal Status and Mental Health Act Implications

Another limitation is the lack of information within the dataset on legal issues, particularly concerning individuals who are brought to the ED by the police or under the *Mental Health Act* (NSW Health, 2015). Although such cases do not always result in admission, they legally require further assessment, which could provide valuable data for the analysis. The inclusion of legal status information could enhance the understanding of the broader context in which individuals are assessed and treated, thereby enriching the study's findings (Szmukler et al., 2014).

8.8.6 Technological and Data Availability Changes

Finally, it is important to acknowledge the rapid rate of technological change, particularly in the field of AI, as well as the increasing availability of data. Since the commencement of this project, advancements in AI and data availability may have outpaced the methods and datasets used in this study. This limitation highlights the need for ongoing research that keeps pace with technological advancements, ensuring that the findings remain relevant and applicable in a rapidly evolving field (Topol, 2019a).

Chapter 9 Conclusion

9.1 Introduction

This thesis explored the integration of artificial intelligence (AI) and machine learning (ML) within emergency departments (ED) to address critical challenges such as overcrowding, long wait times, and limited specialised resources. The research demonstrated how AI and ML can support clinical decision making and potentially improve patient outcomes in these high-pressure environments.

A major contribution of this thesis lies in its comprehensive examination of the ethical, cultural, and practical implications of deploying AI in MH care. By highlighting the importance of trust, transparency, and cultural sensitivity, the research sets a foundation for the responsible development and implementation of AI-based tools in clinical settings. The findings underscore the need for AI systems that complement, rather than replace, human clinical judgement, and emphasise the role of diverse stakeholders in shaping these technologies to ensure they meet the highest standards of care.

The research also presents a vision for the future of ED MH care delivery, where AI and ML are integrated into a comprehensive, patient-centred approach. This vision includes the development of predictive modelling and decision support systems that can assist clinicians in providing timely and effective care, particularly in resource-constrained settings. By leveraging the strengths of AI while maintaining a focus on ethical considerations and patient needs, healthcare systems can evolve to better serve vulnerable populations and improve overall MH outcomes.

In reflecting on the overall contribution of this thesis to the field, it is clear that the integration of AI into MH care is both promising and challenging. The research provides valuable insights

into how these technologies can be harnessed to address longstanding issues in MH care delivery, while also cautioning against the potential risks and ethical dilemmas that must be carefully managed.

9.2 Addressing the Research Question: Developing and Evaluating ML Models for Mental Health Admissions

The thesis has successfully addressed the research question, *How can machine learning models be developed, tested, and evaluated to predict whether clients presenting to an emergency department with a mental health concern, particularly suicidality, will require admission to an acute mental health inpatient ward?*, through a systematic exploration of multiple studies, each contributing to different aspects of the model development, testing, and evaluation.

In addressing the research question, the literature review (Higgins, Short et al., 2023) and the exploration of the importance of trust in AI (Higgins, Chalup et al., 2024a) provided essential foundations by highlighting the critical role of transparency, interpretability, and clinician trust in the adoption of AI and ML models in health care. These works emphasised that for AI-based clinical decision support systems (CDSS) to be effective, particularly in settings like EDs, they must be designed in ways that are not only accurate but also transparent and understandable to clinicians, ensuring that these technologies are trusted and effectively integrated into clinical practice.

Study A laid the foundation by examining the sociodemographic factors and presentation characteristics of individuals seeking MH care in EDs (Higgins, Sheather-Reid et al., 2024c). This provided essential context regarding the patient population and their presenting concerns, identifying key factors such as age, triage category, and presenting problem that were critical for building predictive models.

Study B expanded the research by focusing on the unique challenges faced by Aboriginal and Torres Strait Islander people, a group disproportionately affected by MH issues (Higgins, Sheather-Reid et al., 2024a). This provided further insight into how specific population characteristics might influence admission decisions, enriching the dataset used in ML model development.

Study C took the next step by applying ML techniques to ED data to uncover correlations between presentation features and the likelihood of admission to MH inpatient facilities (Higgins, Chalup et al., 2024b). The study used interpretable ML models, such as those generated by the InterpretML package, to reveal significant patterns in the data. These models identified that key factors like suicidal ideation, triage level, and prior healthcare utilisation were influential in predicting admissions, thereby informing the criteria for future predictive models.

Study D built upon these findings by testing and evaluating more complex ML models, such as explainable boosting machines (EBMs) and support vector machines (SVMs), to predict admissions with a focus on transparency and clinical trust (Higgins et al., 2024d). Study D emphasised the importance of developing models that are both accurate and interpretable, allowing clinicians to understand and trust the decision-making process behind the predictions.

Study E plans to further develop the final aspect of the research question by examining clinicians' attitudes towards AI and ML integration. It will highlight the importance of trust, transparency, and interpretability in the adoption of ML models in clinical practice, particularly in high-stakes environments like EDs, where rapid and reliable decision making is essential.

9.3 Achieving the Aims: Integrating AI and ML into Mental Health Care

In addition to the research question, this thesis has rigorously addressed its central aims of complementing and augmenting clinicians' decision making, enhancing the patient experience within MH services while ensuring the delivery of high-quality and safe care, and advancing the understanding of AI and ML adoption in mental health contexts. These objectives have been achieved through a comprehensive integration of findings from both the literature reviews and empirical studies, offering a nuanced exploration of how AI and ML can be ethically and effectively implemented in MH care.

9.3.1 Complement and Augment Clinicians' Evaluation and Decision Making

The literature review and subsequent research have underscored the transformative potential of AI- and ML-based CDSS in MH care. These systems present the potential to alleviate clinicians' intellectual burden by assisting in identifying missed care and improving decision making, particularly in complex cases like suicidality. By using interpretable ML models, such as those implemented with the InterpretML tool (Nori et al., 2019), clinicians can better understand and trust the recommendations provided, as these systems present a transparent chain of logic and confidence levels. However, the literature also emphasises the importance of addressing transparency, interpretability, and bias to foster clinician trust. The findings of this thesis contribute to this aim by developing ML models that are designed with these factors in mind, ensuring that AI complements rather than replaces clinical decision making.

9.3.2 Enhance the Client's Experience of Mental Health Services While Providing High-Quality and Safe Care

The integrative review highlighted that AI- and ML-based CDSS have the potential to mitigate some of the resource constraints and workforce challenges facing the MH sector (Higgins,

Short et al., 2023). Study A (Higgins, Sheather-Reid et al., 2024c) identified key factors affecting MH admissions, such as the prevalence of suicidal ideation and inefficiencies in the ED, which negatively impact client satisfaction. Study B (Higgins, Sheather-Reid et al., 2024a) highlighted the need for culturally responsive care, particularly for Aboriginal and Torres Strait Islander populations, whose experiences in the current system are often inadequate. Study C (Higgins, Chalup et al., 2024b) demonstrated how interpretable ML models can assist in triaging patients more effectively, ensuring those in need receive timely care, thus improving both the client's experience and care quality. Study D (Higgins et al., 2024) built on this by refining ML models to predict if a person requires an admission, further optimising resource allocation and reducing wait times. The literature supports these findings, indicating that AI can help address challenges like missed care and clinician burnout while enhancing client outcomes (Jones et al., 2015; Mudallal et al., 2017; Seibert et al., 2021). By reducing clinicians' workload and streamlining decision-making processes, these technologies can assist in the provision of improved patient outcomes, particularly in ED settings where timely interventions are crucial. Furthermore, the literature emphasises that AI system predictions, particularly for patients with mental health conditions like psychosis, must be interpreted by clinicians who act with empathy and align these insights with patients' needs. The use of design thinking (DT) principles (Brown, 2008) has been advocated as a way to ensure AI interventions are both effective and support empathetic care. Through this lens, the research contributes to enhancing the patient experience by promoting AI solutions that are designed to provide safe, reliable, and patient-centred care.

9.3.3 Understand What is Known About Artificial Intelligence and Machine Learning Adoption in a Mental Health Context

Study E, along with studies included in this thesis (Higgins, Short et al., 2023; Higgins, Chalup et al., 2024a), aims to provide insight into the attitudes of clinicians towards AI adoption, especially in the MH context. The literature indicates that the successful implementation of AI in health care hinges on building trust, transparency, and interpretability into the models used (Molnar, 2020; Ryan, 2020; Schepman & Rodway, 2020). Additionally, the theoretical discussion of trust in AI stresses that these technologies must meet the same standards of reliability, competence, and validity as any other clinical tool to gain widespread acceptance. The research also explores the importance of ongoing dialogue between clinicians, patients, and policymakers to ensure ethical AI deployment, as outlined by global and national ethical frameworks, such as those from the World Health Organization (2021a) and the Australian Government's Department of Industry, Science and Resources (2021). By addressing these critical aspects of AI adoption, the research advances the understanding of how AI/ML can be ethically and effectively integrated into MH care.

In summary, the thesis has developed, tested, and evaluated ML models that predict whether clients presenting to EDs with mental health concerns, especially suicidality, will require admission to an acute inpatient ward. Through a combination of empirical data analysis, ML model development, and an exploration of clinicians' perspectives, the research provides a comprehensive framework for implementing AI-driven decision support systems that are both effective and trusted by healthcare professionals.

9.4 Addressing Missed Care Through AI-Driven Solutions in Mental Health Care

This research has made significant contributions to addressing the issue of missed care. The testing of a ML model (Higgins et al., 2024d) and the study of the trustworthiness of AI

(Higgins, Chalup et al., 2024a) provided a critical underpinning argument by highlighting the importance of transparency, interpretability, and clinician trust in AI systems. This focus on transparency, interpretability, and clinician trust is crucial for addressing missed care because it ensures that AI-driven tools are not only accurate but also reliable and understandable by healthcare professionals. When clinicians trust the decision-making processes of AI systems, they are more likely to adopt these technologies in their workflows, allowing them to make timely, informed decisions that reduce care omissions and improve patient outcomes. Without this trust, even the most advanced AI systems may fail to be utilised effectively, perpetuating the issue of missed care.

The studies of who presents to EDs with MH issues (Higgins, Sheather-Reid et al., 2024c) and ED presentations of First Nations people (Higgins, Sheather-Reid et al., 2024a) provided foundational insights by identifying key sociodemographic factors and highlighting disparities, particularly among Aboriginal and Torres Strait Islander populations, where care gaps are more pronounced due to systemic inequities and a lack of culturally responsive services. Study C (Higgins, Chalup et al., 2024b) further demonstrated how ML models can be used to identify factors influencing MH admissions, enabling clinicians to make more informed decisions that reduce delays in care delivery. By improving triage and resource allocation, these ML models can help address missed care by prioritising urgent cases, ensuring patients with critical needs, such as those experiencing suicidality, are not overlooked (Seibert et al., 2021). The final study for this thesis (Higgins et al., 2024d) extended this work by refining ML models to predict which patients would require admission, further supporting clinical decision making and reducing the administrative burden on staff, which often contributes to care omissions. Together, this body of research shows how AI and ML can be effectively integrated into healthcare workflows to mitigate missed care, reduce clinician burnout, and ultimately enhance patient outcomes by providing more timely and comprehensive care.

The most important outcome of this research is highlighting the critical role of trust and transparency in the integration of AI and ML within MH care, particularly in ED. The research underscores that, for AI and ML systems to be successfully adopted in these settings, they must be designed with transparency and interpretability at their core, allowing clinicians to understand, scrutinise, and trust the decision-making processes behind AI recommendations. By fostering trust, these systems can complement clinical decision making, ultimately enhancing patient outcomes while ensuring that the ethical, equitable, and safe care of vulnerable populations remains central to their deployment.

As a closing statement, this thesis highlights the importance of continued research and innovation in MH services. The rapid advancements in AI and ML present both opportunities and challenges, and it is essential that future research continues to explore how these technologies can be used to enhance care while upholding the principles of ethical and humane treatment. By fostering collaboration between technological and clinical fields, and by involving patients and communities in the development of AI tools, the field of MH care can move towards a future where AI-driven innovations play a crucial role in improving patient outcomes and ensuring equitable access to care. In conclusion, this thesis contributes to a deeper understanding of the complexities and potential of AI in MH care, offering a foundation for future research and innovation. The findings advocate a balanced approach that integrates AI into clinical practice in a way that enhances human expertise, addresses ethical challenges, and ultimately leads to a more effective and patient-centred MH care system.

9.5 Summary of Thesis

This thesis has provided a comprehensive examination of the intersection between AI and MH care, with a particular focus on the application of ML in clinical settings such as EDs. Throughout the research, a recurring emphasis has been placed on the necessity of adopting a

comprehensive, integrated approach to MH care—one that combines technological innovation with empathetic, user-centred care models.

Chapter 1: Introduction. The introductory chapter established the critical issues facing MH care delivery in EDs and the potential role of AI and ML in addressing these challenges. The chapter outlined the study's aims and significance, emphasising the urgent need for innovative solutions to improve the efficiency and effectiveness of MH services. It also highlighted the importance of ethical considerations and the inclusion of lived experiences and minority groups, framing AI and ML as tools that could enhance clinical decision making and improve patient outcomes.

Chapter 2: Background. This chapter provided a foundational context for the research, focusing on the transformative role of AI and ML in MH care. It explored both the opportunities and challenges these technologies present, particularly for individuals experiencing psychosis. The chapter also examined the influence of commercial determinants in shaping AI's development and the implications of this for patient-centred care. Emphasising the need for ethical, safe, and transparent AI deployment, the chapter directly addressed the research aim of enhancing the client's experience in MH services.

Chapter 3: Critical Review of the Literature. This chapter offered a critical review of existing literature on AI- and ML-based CDSS in MH care. It explored the role of trust, transparency, and clinicians' engagement in the successful implementation of these technologies. The review identified key challenges, such as algorithmic bias, accountability, and the potential for over-reliance on technology. The chapter provided recommendations for developing AI/ML systems that align with clinical workflows and maintain human oversight, contributing to the project's overall aims.

Chapter 4: Theoretical Foundations of Trust. This chapter investigated the theoretical dimensions of trust in the context of AI adoption in MH care. It argued that AI should be viewed not as a tool to be inherently trusted but as a clinical instrument whose reliability and validity must be rigorously evaluated. The chapter emphasised the importance of transparency and interpretability in AI systems to foster clinician trust and ensure AI supports, rather than replaces, human decision making. This analysis was crucial for advancing the study's goals of enhancing clinicians' decision making and improving patient care.

Chapter 5: Update to Literature Review. Revisiting the original literature review, this chapter incorporated recent advancements in AI, particularly the rise of large language models (LLMs). It explored how these developments have influenced health care, with a focus on improving clinical decision making and reducing missed care. The chapter also discussed the ethical implications of LLMs, such as transparency and trust, reaffirming the importance of balancing AI innovation with workforce improvements to ensure high-quality care.

Chapter 6: Methods. This chapter outlined the methodological approach used to develop, test, and evaluate ML models for predicting MH admissions from ED presentations. Using design thinking (DT) and retrospective cohort methods, the research integrated sociodemographic data and ML techniques to provide a robust framework for understanding ED presentations and admissions. Ethical considerations, including cultural sensitivity, were central to the methodology, ensuring the research adhered to best practices in healthcare innovation.

Chapter 7: Results. The results chapter presented findings on MH care delivery in EDs, focusing on sociodemographic factors, presentation characteristics, and the application of ML in improving service provision. The analysis revealed key factors influencing admissions, such as suicidal ideation and triage level, and highlighted the potential of ML models to enhance clinical decision making, reduce overcrowding, and improve patient outcomes.

Chapter 8: Discussion. This chapter synthesised the research findings, addressing how ML models can predict admissions to MH inpatient wards from ED presentations. It discussed the importance of transparency and clinician trust in the effective integration of AI in clinical practice. The chapter also outlined future research directions, emphasising the need for continued exploration of AI's role in improving MH services while maintaining ethical standards and clinician engagement.

Chapter 9: Conclusion. The concluding chapter summarised the thesis' contributions to MH care, particularly in integrating AI and ML into EDs. It highlighted the importance of trust, transparency, and clinicians' involvement in developing AI tools. The chapter called for continued research to ensure AI enhances, rather than replaces, human clinical judgement, and promotes the responsible and ethical deployment of AI technologies in MH care.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., & Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden. P. . . . Zheng, X. (2016). Tensorflow: A system for large-scale machine learning. *arXiv*. arXiv.1605.08695

Abd-alrazaq, A., Alajlani, M., Ali, N., Denecke, K., Bewick, B., & Househ, M. (2021). Perceptions and opinions of patients about mental health chatbots: Scoping review. *Journal of Medical Internet Research*, *23*(1), Article 17828.

Abrams, N. (2017). What was HAL? IBM, Jewishness and Stanley Kubrick's 2001: A Space Odyssey (1968). *Historical Journal of Film, Radio and Television*, *37*(3), 416-435. https://doi.org/10.1080/01439685.2017.1342328

Aiken, L. H., Sloane, D., Griffiths, P., Rafferty, A. M., Bruyneel, L., McHugh, M., Maier, C. B., Moreno-Casbas, T., Ball, J. E., Ausserhofer, D., & Sermeus, W. (2017). Nursing skill mix in European hospitals: Cross-sectional study of the association with mortality, patient ratings, and quality of care. *BMJ Quality & Safety*, *26*(7), 559-568. https://doi.org/10.1136/bmjqs-2016-005567

Airenti, G. (2015). The cognitive bases of anthropomorphism: From relatedness to empathy. *International Journal of Social Robotics*, *7*(1), 117-127. https://doi.org/10.1007/s12369-014-0263-x

Aiva. (2016, October 28). *"Genesis" symphonic fantasy in A minor, Op. 21* [video]. YouTube. https://www.youtube.com/watch?v=Ebnd03x137A

Aljanabi, M., Ghazi, M., Ali, A. H., & Abed, S. A. (2023). ChatGPT: Open possibilities. *Iraqi Journal for Computer Science and Mathematics*, *4*(1), 62-64.

Alowais, S. A., Alghamdi, S. S., Alsuhebany, N., Alqahtani, T., Alshaya, A. I., Almohareb, S. N., Aldairem, A., Alrashed, M., Bin Saleh, K., & Badreldin, H. A. (2023). Revolutionizing healthcare: The role of artificial intelligence in clinical practice. *BMC Medical Education*, *23*(1), Article 689.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Association.

American Psychiatric Association. (2015). *Schizophrenia spectrum and other psychotic disorders : DSM-5® selections*. American Psychiatric Publishing.

American Psychological Association. (2023). *Guidelines for the practice of telepsychology.* https://www.apa.org/practice/guidelines/telepsychology

Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. (2016). Concrete problems in AI safety. *arXiv.* arXiv:1606.06565

Andrews, J. T., & Siddiqi, A. A. (2011). *Into the cosmos: Space exploration and Soviet culture*. University of Pittsburgh Press.

Aramaki, E., Wakamiya, S., Yada, S., & Nakamura, Y. (2022). Natural language processing: From bedside to everywhere. *Yearbook of Medical Informatics*, *31*(1), 243-253. https://doi.org/10.1055/s-0042-1742510

Asimov, I. (1952). Runaround. In I. Asimov, *I, robot* (Vol. 1, pp. 70-91). Gnome Press.

Atkinson, J., Nelson, J., Brooks, R., Atkinson, C., & Ryan, K. (2014). Addressing individual and community transgenerational trauma. In P. Dudgeon, H. Milroy, & R. Walker (Eds.), *Working together: Aboriginal and Torres Strait Islander mental health and wellbeing principles and practice* (2nd ed., pp. 289-307). Commonwealth of Australia.

Austin EE, Cheek C, Richardson L, Testa L, Dominello A, Long JC, Carrigan A, Ellis LA, Norman A, Murphy M, Smith K, Gillies D, Clay-Williams R. *Improving emergency department care for adults presenting with mental illness: a systematic review of strategies and their impact on outcomes, experience, and performance*. Front Psychiatry. 2024 Feb 29;15 doi: 10.3389/fpsyt.2024.1368129.

Australasian College for Emergency Medicine. (2018). *The long wait: An analysis of mental health presentations to Australian emergency departments.*

Australian Bureau of Statistics. (2021a). *2021 Census all persons QuickStats Central Coast (NSW)*. https://abs.gov.au/census/find-census-data/quickstats/2021/LGA11650

Australian Bureau of Statistics. (2021b). *2021 Gosford, Census all persons QuickStats*. https://abs.gov.au/census/find-census-data/quickstats/2021/SED10030

Australian Bureau of Statistics. (2021c). *2021 New South Wales, Census all persons QuickStats*. https://www.abs.gov.au/census/find-census-data/quickstats/2021/1

Australian Bureau of Statistics. (2021d). *2021 Wyong, Census all persons QuickStats*. https://abs.gov.au/census/find-census-data/quickstats/2021/SAL14464

Australian Bureau of Statistics. (2021e). *Estimates of Aboriginal and Torres Strait Islander Australians*. https://www.abs.gov.au/statistics/people/aboriginal-and-torres-strait-islander-peoples/estimates-aboriginal-and-torres-strait-islander-australians/latest-release

Australian Bureau of Statistics. (2022). *Standards for statistics on cultural and language diversity*. https://www.abs.gov.au/statistics/standards/standards-statistics-cultural-and-language-diversity/latest-release

Australian College of Nursing. (2024). *ACN position statement: Artificial intelligence (AI).*

Australian Commission on Safety and Quality in Health Care. (2020). *National safety and quality digital mental health standards.* https://www.safetyandquality.gov.au/publications-and-resources/resource-library/national-safety-and-quality-digital-mental-health-standards

Australian Commission on Safety and Quality in Health Care. (2024). *National Safety and Quality Health Service Standards*. https://www.safetyandquality.gov.au/standards

Australian Institute of Health and Welfare. (2021a). *Aboriginal and Torres Strait Islander Stolen Generations aged 50 and over: Updated analyses for 2018–19.* https://www.aihw.gov.au/reports/indigenous-australians/indigenous-stolen-generations-50-and-over/contents/summary

Australian Institute of Health and Welfare. (2021b). *Emergency department care*. https://www.aihw.gov.au/reports-data/myhospitals/sectors/emergency-department-care

Australian Institute of Health and Welfare. (2021c). *Mental health services in Australia*. https://www.aihw.gov.au/reports/mental-health-services/mental-health-services-in-australia/report-contents/emergency-department-mental-health-services

Australian Institute of Health and Welfare. (2022a). *Indigenous health and wellbeing*. https://www.aihw.gov.au/reports/australias-health/indigenous-health-and-wellbeing

Australian Institute of Health and Welfare. (2022b). *Mental health services in Australia*. https://www.aihw.gov.au/reports/mental-health-services/mental-health-services-in-australia/report-contents/emergency-department-mental-health-services

Australian Institute of Health and Welfare. (2022c). *Mental health workforce*. https://www.aihw.gov.au/mental-health/topic-areas/workforce

Australian Institute of Health and Welfare. (2023a). *3.10 Access to mental health services*. https://www.indigenoushpf.gov.au/measures/3-10-access-mental-health-services

Australian Institute of Health and Welfare. (2023b). *Deaths by suicide among young people*. https://www.aihw.gov.au/suicide-self-harm-monitoring/data/populations-age-groups/suicide-among-young-people

Australian Institute of Health and Welfare. (2023c). *Suicide & self-harm monitoring*. https://www.aihw.gov.au/suicide-self-harm-monitoring/data/data-downloads

Badcock, P. B. (2012). Evolutionary systems theory: A unifying meta-theory of psychological science. *Review of General Psychology*, *16*(1), 10-23.

Baddal, B., Taner, F., & Uzun Ozsahin, D. (2024). Harnessing of artificial intelligence for the diagnosis and prevention of hospital-acquired infections: a systematic review. *Diagnostics (Basel)*, *14*(5). https://doi.org/10.3390/diagnostics14050484

Bainbridge, R., McCalman, J., Clifford, A., & Tsey, K. (2015). *Cultural competency in the delivery of health services for Indigenous people*. Closing the Gap Clearinghouse.

Ball, J. E., Bruyneel, L., Aiken, L. H., Sermeus, W., Sloane, D. M., Rafferty, A. M., Lindqvist, R., Tishelman, C., & Griffiths, P. (2018). Post-operative mortality, missed care and nurse staffing in nine countries: A cross-sectional study. *International Journal of Nursing Studies*, *78*, 10-15. https://doi.org/10.1016/j.ijnurstu.2017.08.004

Ball, J. E., Murrells, T., Rafferty, A. M., Morrow, E., & Griffiths, P. (2014). "Care left undone" during nursing shifts: Associations with workload and perceived quality of care. *BMJ Quality & Safety*, *23*(2), 116-125. https://doi.org/10.1136/bmjqs-2012-001767

Ban, T. A. (2007). Fifty years chlorpromazine: A historical perspective. *Neuropsychiatric Disease and Treatment*, *3*(4), 495-500. https://www.ncbi.nlm.nih.gov/pubmed/19300578

Barthelmess, U., & Furbach, U. (2014). Do we need Asimov's laws? *arXiv.* arXiv:1405.0961

Bates, D. W., Landman, A., & Levine, D. M. (2018). Health apps and health policy: What is needed? *JAMA*, *320*(19), 1975-1976.

Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A., & Escobar, G. (2014). Big data in health care: Using analytics to identify and manage high-risk and high-cost patients. *Health Affairs (Millwood)*, *33*(7), 1123-1131. https://doi.org/10.1377/hlthaff.2014.0041

Bazzano, A. N., Martin, J., Hicks, E., Faughnan, M., & Murphy, L. (2017). Human-centred design in global health: A scoping review of applications and contexts. *PLOS ONE*, *12*(11), Article e0186744.

Beaudin, J., Chouinard, M.-C., Girard, A., Houle, J., Ellefsen, É., & Hudon, C. (2022). Integrated self-management support provided by primary care nurses to persons with chronic diseases and common mental disorders: A scoping review. *BMC Nursing*, *21*(1), Article 212.

Bell, I. H., Pot-Kolder, R., Rizzo, A., Rus-Calafell, M., Cardi, V., Cella, M., Ward, T., Riches, S., Reinoso, M., Thompson, A., Alvarez-Jimenez, M., & Valmaggia, L.

(2024). Advances in the use of virtual reality to treat mental health conditions. *Nature Reviews Psychology*, *3*(8), 552-567. https://doi.org/10.1038/s44159-024-00334-9

Ben-Hur, A., Horn, D., Siegelmann, H. T., & Vapnik, V. (2001). Support vector clustering. *Journal of Machine Learning Research*, *2*, 125-137.

Ben-Zeev, D., Buck, B., Kopelovich, S., & Meller, S. (2019). A technology-assisted life of recovery from psychosis. *NPJ Schizophrenia*, *5*(1), Article 15. https://doi.org/10.1038/s41537-019-0083-y

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 610-623). Association for Computing Machinery.

Benrimoh, D., Tanguay-Sela, M., Perlman, K., Israel, S., Mehltretter, J., Armstrong, C., Fratila, R., Parikh, S. V., Karp, J. F., Heller, K., Vahia, I. V., Blumberger, D. M., Karama, S., Vigod, S. N., Myhr, G., Martins, R., Rollins, C., Popescu, C., Lundrigan, E., . . . Margolese, H. C. (2020). Using a simulation centre to evaluate preliminary acceptability and impact of an artificial intelligence-powered clinical decision support system for depression treatment on the physician-patient interaction. *BJPsych Open*, *7*(1), Article e22. https://doi.org/10.1192/bjo.2020.127

Bentley, K., Zuromski, K., Fortgang, R., Madsen, E., Kessler, D., Lee, H., Nock, M., Reis, B., Castro, V., & Smoller, J. (2021). Implementing machine learning algorithms for suicide risk prediction in clinical practice: A focus group study. *PsyArXiv*. https://doi.org/10.31234/osf.io/6m5qd

Benton, T. D., Boyd, R. C., & Njoroge, W. F. (2021). Addressing the global crisis of child and adolescent mental health. *JAMA Pediatrics*, *175*(11), 1108-1110.

Berendsen Russell, S., Dinh, M. M., & Bell, N. (2017). Triage, damned triage … and statistics: Sorting out redundancy and duplication within an Emergency Department Presenting Problem Code Set to enhance research capacity. *Australasian Emergency Nursing Journal*, *20*(1), 48-52. https://doi.org/10.1016/j.aenj.2016.09.004

Berlinger, J. (2016, August 4). Gavin Long: Who is Baton Rouge cop killer? *CNN*. https://edition.cnn.com/2016/07/18/us/who-is-gavin-long/

Bertalanffy, L. v. (1968). *General system theory: Foundations, development, applications*. G. Braziller.

Bertsekas, D. (2019). *Reinforcement and optimal control*. Athena Scientific.

Bhatt, U., & Shams, Z. (2021). Trust in artificial intelligence: Clinicians are essential. In A. B. Bhatt (Ed.), *Healthcare information technology for cardiovascular medicine: Telemedicine & digital health* (pp. 127-141). Springer International Publishing. https://doi.org/10.1007/978-3-030-81030-6_10

Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, *25*, 197-227.

Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4). Springer.

Borschmann, R., Stark, P., Prakash, C., & Sawyer, S. M. (2018). Risk profile of young people admitted to hospital for suicidal behaviour in Melbourne, Australia. *Journal of Paediatrics and Child Health*, *54*(11), 1213-1220. https://doi.org/10.1111/jpc.13938

Botelho, G., & Sterling, J. (2013, September 26). FBI: Navy Yard shooter "delusional," said "low frequency attacks" drove him to kill. *CNN*. https://edition.cnn.com/2013/09/25/us/washington-navy-yard-investigation/

Botha, F., Butterworth, P., & Wilkins, R. (2020). Heightened mental distress: Can addressing financial stress help? In B. Broadway, A. A. Payne, & N. Salamanca (Eds.), *Coping with COVID-19: Rethinking* Australia (pp. 13-16). Melbourne Institute. https://melbourneinstitute.unimelb.edu.au/__data/assets/pdf_file/0011/3565145/Chapter-4-Heightened-Mental-Distress.pdf

Botha, F., Butterworth, P., & Wilkins, R. (2022). Evaluating how mental health changed in Australia through the COVID-19 pandemic: Findings from the "Taking the Pulse of the Nation" (TTPN) survey. *International Journal of Environmental Research and Public Health*, *19*(1). https://doi.org/10.3390/ijerph19010558

Botsman, R. (2018). *Who can you trust? How technology brought us together and why it might drive us apart*. Penguin.

Boudreaux, E. D., Camargo, C. A., Jr., Arias, S. A., Sullivan, A. F., Allen, M. H., Goldstein, A. B., Manton, A. P., Espinola, J. A., & Miller, I. W. (2016). Improving suicide risk screening and detection in the emergency department. *American Journal of Preventive Medicine, 50*(4), 445-453. https://doi.org/10.1016/j.amepre.2015.09.029

Boulanger, M., Dethier, M., Gendre, F., & Blairy, S. (2013). Identity in schizophrenia: A study of trait self-knowledge. *Psychiatry Research*, *209*(3), 367-374. https://doi.org/10.1016/j.psychres.2013.04.002

Braitenberg, V. (1986). *Vehicles: Experiments in synthetic psychology*. MIT Press.

Braithwaite, V. (2020). Beyond the bubble that is Robodebt: How governments that lose integrity threaten democracy. *Australian Journal of Social Issues*, *55*(3), 242-259.

Brandt, S. (2009). *The harvest of a century: Discoveries of modern physics in 100 episodes*. Oxford University Press.

Braun, V., & Clarke, V. (2021). *Thematic analysis: A practical guide*. Sage.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5-32.

Broadbent, M., Moxham, L., & Dwyer, T. (2020). Understanding nurses perspectives of acuity in the process of emergency mental health triage: A qualitative study. *Contemporary Nurse*, *56*(3), 280-295. https://doi.org/10.1080/10376178.2020.1841014

Brown, L. A., Benhamou, K., May, A. M., Mu, W., & Berk, R. (2020). Machine learning algorithms in suicide prevention: Clinician interpretations as barriers to implementation. *Journal Clinical Psychiatry, 81*(3). https://doi.org/10.4088/JCP.19m12970

Brown, T. (2008). Design thinking. *Harvard Business Review*, *86*(6), 84-92.

Brown, T. (2009). *Change by design: How design thinking transforms organizations and inspires innovation*. HarperCollins.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., & Askell, A. (2020). Language models are few-shot learners. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, & H. Lin (Eds.), *Advances in neural information processing systems* (Vol. 33, pp. 1877-1901). Curran Associates.

Browne, A. J., Varcoe, C., Ford-Gilboe, M., & Wathen, C. N. (2015). EQUIP Healthcare: An overview of a multi-component intervention to enhance equity-oriented care in primary health care settings. *International Journal for Equity in Health*, *14*, Article 152. https://doi.org/10.1186/s12939-015-0271-y

Brüne, M. (2005). "Theory of mind" in schizophrenia: A review of the literature. *Schizophrenia Bulletin*, *31*(1), 21-42. https://doi.org/10.1093/schbul/sbi002

Bruns, A., Harrington, S., & Hurcombe, E. (2020). "Corona? 5G? or both?": The dynamics of COVID-19/5G conspiracy theories on Facebook. *Media International Australia*, *177*(1), 12-29. https://doi.org/10.1177/1329878x20946113

Buchanan, C., Howitt, M. L., Wilson, R., Booth, R. G., Risling, T., & Bamford, M. (2020). Predicted influences of artificial intelligence on the domains of nursing: Scoping review. *JMIR Nursing*, *3*(1), Article e23939. https://doi.org/10.2196/23939

Bureau of Health Information. (2022). *Technical supplement: Emergency department, ambulance, admitted patients, seclusion and restraint, and elective surgery*.

https://www.bhi.nsw.gov.au/__data/assets/pdf_file/0009/715194/BHI_HQ47_OCT-DEC_2021_TECH_SUPP.pdf

Burke, T. A., Ammerman, B. A., & Jacobucci, R. (2019). The use of machine learning in the study of suicidal and non-suicidal self-injurious thoughts and behaviors: A systematic review. *Journal of Affective Disorders, 245*, 869-884. https://doi.org/10.1016/j.jad.2018.11.073

Butera, S., Brasseur, N., Filion, N., Bruyneel, A., & Smith, P. (2021). Prevalence and associated factors of burnout risk among intensive care and emergency nurses before and during the coronavirus disease 2019 pandemic: A cross-sectional study in Belgium. *Journal of Emergency Nursing*, *47*(6), 879-891. https://doi.org/10.1016/j.jen.2021.08.007

Byrne, L., Happell, B., & Reid-Searl, K. (2017). Risky business: Lived experience mental health practice, nurses as potential allies. *International Journal of Mental Health Nursing*, *26*(3), 285-292.

Bzdok, D., & Meyer-Lindenberg, A. (2018). Machine learning for precision psychiatry: Opportunities and challenges. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *3*(3), 223-230. https://doi.org/10.1016/j.bpsc.2017.11.007

Campbell, M. A., Hunt, J., Scrimgeour, D. J., Davey, M., & Jones, V. (2018). Contribution of Aboriginal community-controlled health services to improving Aboriginal health: An evidence review. *Australian Health Review*, *42*(2), 218-226.

Campesato, O. (2020). *Artificial intelligence, machine learning, and deep learning*. Mercury Learning & Information.

Carr, S. (2020). "AI gone mental": Engagement and ethics in data-driven technology for mental health. *Journal of Mental Health*, *29*(2), 125-130. https://doi.org/10.1080/09638237.2020.1714011

Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., & Elhadad, N. (2015). Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1721-1730). Association for Computing Machinery. http://doi.acm.org/10.1145/2783258.2788613

Celi, L. A., Cellini, J., Charpignon, M.-L., Dee, E. C., Dernoncourt, F., Eber, R., Mitchell, W. G., Moukheiber, L., Schirmer, J., & Situ, J. (2022). Sources of bias in artificial intelligence that perpetuate healthcare disparities—A global review. *PLOS Digital Health*, *1*(3), Article e0000022.

Central Coast Local Health District. *About Us*. Retrieved 01/01/2022 from https://www.cclhd.health.nsw.gov.au/about-us/

Central Coast Local Health District. (2018). *Research—Management of clinical and non-clinical data PR2017_099*.

Central Coast Local Health District. (2022). *Adult mental health*. https://www.cclhd.health.nsw.gov.au/services/mental-health/adult-mental-health/

Centre for International Economics. (2015). *Efficiency of NSW public transport services*. https://www.ipart.nsw.gov.au/sites/default/files/documents/consultant_report_-_cie_-_efficiency_of_nsw_public_transport_services_-_december_2015.pdf

Cermolacce, M., Sass, L., & Parnas, J. (2010). What is bizarre in bizarre delusions? A critical review. *Schizophrenia Bulletin*, *36*(4), 667-679. https://doi.org/10.1093/schbul/sbq001

Cerreto, A. (2018). *Ngaga-dji (hear me): Young voices creating change for justice*. Koorie Youth Council.

Chandler, C., Foltz, P. W., & Elvevåg, B. (2022). Improving the applicability of AI for psychiatric applications through human-in-the-loop methodologies. *Schizophrenia Bulletin*, *48*(5), 949-957. https://doi.org/10.1093/schbul/sbac038

Chang, C.-H., Tan, S., Lengerich, B. J., Goldenberg, A., & Caruana, R. (2020). How interpretable and trustworthy are GAMs? *arXiv.* arXiv:2006.06466

Char, D. S., Abramoff, M. D., & Feudtner, C. (2020). Identifying ethical considerations for machine learning healthcare applications. *American Journal of Bioethics*, *20*(11), 7-17. https://doi.org/10.1080/15265161.2020.1819469

Charalambous, C., Koulori, A., Vasilopoulos, A., & Roupa, Z. (2018). Evaluation of the validity and reliability of the Waterlow Pressure Ulcer Risk Assessment Scale. *Medical Archives, 72*(2), 141-144. https://doi.org/10.5455/medarh.2018.72.141-144

Charles, J. (2019). *Jack Charles*. Penguin Random House.

Chen, Q., Zhang-James, Y., Barnett, E. J., Lichtenstein, P., Jokinen, J., D;Onofrio, B. M., Faraone, S. V., Larsson, H., & Fazel, S. (2020). Predicting suicide attempt or suicide death following a visit to psychiatric specialty care: A machine learning study using Swedish national registry data. *PLoS Medicine, 17*(11), Article e1003416. https://dx.doi.org/10.1371/journal.pmed.1003416

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 785-794). Association for Computing Machinery.

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., & Zhou, T. (2024). *Xgboost: Extreme gradient boosting*. R Package Version 0.4-2, 1-4.

Chicco, D., Tötsch, N., & Jurman, G. (2021). The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData Mining, 14*(1), Article 13. https://doi.org/10.1186/s13040-021-00244-z

Chor, W. P. D., Ng, W. M., Cheng, L., Situ, W., Chong, J. W., Ng, L. Y. A., Mok, P. L., Yau, Y. W., & Lin, Z. (2021). Burnout amongst emergency healthcare workers during the COVID-19 pandemic: A multi-center study. *American Journal of Emergency Medicine, 46,* 700-702. https://doi.org/10.1016/j.ajem.2020.10.040

Chui, M., Manyika, J., & Miremadi, M. (2016, July 8). Where machines could replace humans—And where they can't (yet). *McKinsey Quarterly,* 1-12.

Clarke, D., Usick, R., Sanderson, A., Giles-Smith, L., & Baker, J. (2014). Emergency department staff attitudes towards mental health consumers: A literature review and thematic content analysis. *International Journal of Mental Health Nursing, 23*(3), 273-284.

Clarke, D. E., Dusome, D., & Hughes, L. (2007). Emergency department from the mental health client's perspective. *International Journal of Mental Health Nursing, 16*(2), 126-131. https://doi.org/10.1111/j.1447-0349.2007.00455.x

Cleary, M., West, S., & Hungerford, C. (2023). Mental health nurses overcoming adversity: A discursive review. *Issues in Mental Health Nursing, 44*(10), 944-950.

Coiera, E., & Magrabi, F. (2023). *A national policy roadmap for artificial intelligence in healthcare*. Australian Alliance for Artificial Intelligence in Healthcare.

Collins, H., Calvo, S., Greenberg, K., Forman Neall, L., & Morrison, S. (2016). Information needs in the precision medicine era: How genetics home reference can help. *Interactive Journal of Medical Research, 5*(2), Article e13. https://doi.org/10.2196/ijmr.5199

Commonwealth of Australia. (2020). *National Agreement on Closing the Gap*. https://www.closingthegap.gov.au/sites/default/files/2021-05/ctg-national-agreement_apr-21.pdf

Congress of Aboriginal and Torres Strait Islander Nurses and Midwives. (2022). *Murra Mullangari: Introduction to cultural safety and cultural humility*.

https://hnc.org.au/news-story/murra-mullangari-introduction-to-cultural-safety-and-cultural-humility-xhp-x1/

Consoli, S., Recupero, D. R., & Petkovic, M. (2019). *Data science for healthcare*. Springer.

Cooban, A. (2021, June 30). *Elon Musk suggests Starlink will be available worldwide from August – a month earlier than planned*. Business Insider. https://www.businessinsider.com.au/starlink-global-coverage-elon-musk-spacex-august-2021-6

Corrigan, P. W., Druss, B. G., & Perlick, D. A. (2014). The impact of mental illness stigma on seeking and participating in mental health care. *Psychological Science in the Public Interest*, *15*(2), 37-70. https://doi.org/10.1177/1529100614531398

Costa DSJ, Mercieca-Bebber R, Tesson S, Seidler Z, Lopez AL. *Patient, client, consumer, survivor or other alternatives? A scoping review of preferred terms for labelling individuals who access healthcare across settings*. BMJ Open. 2019 Mar 7;9(3): doi: 10.1136/bmjopen-2018-025166.

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, *13*(1), 21-27. https://doi.org/10.1109/TIT.1967.1053964

Crawford, K. (2021). *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Sage.

Dans, E. (2021, June 11). *The future of autonomous vehicles: Product or service?* Forbes. https://www.forbes.com/sites/enriquedans/2021/06/11/the-future-of-autonomous-vehicles-product-orservice/?sh=6c56fe0d5892

Darcy, A. (2023). Why generative AI is not yet ready for mental healthcare. *Woebot Health*.

Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future Healthcare Journal*, *6*(2), 94-98.

De Mol, L. (2019). Turing machines. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2019 ed.). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2019/entries/turing-machine/

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 248-255). IEEE.

Department of Health. (2020). *National suicide prevention strategy for Australia's health system: 2020–2023.* Australian Government.

https://www.health.gov.au/resources/publications/national-suicide-prevention-strategy-for-australias-health-system-2020-2023

Department of Health. (2021). *National digital mental health framework.* Australian Government. https://www.health.gov.au/resources/publications/national-digital-mental-health-framework

Department of Health and Aged Care. (2024). *Nursing Supply and Demand Study*. https://hwd.health.gov.au/resources/primary/nursing-supply-and-demand-study-2023-2035.pdf

Department of Industry, Science and Resources. (2019, updated 2024). *Australia's AI Ethics Principles*. https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles

Department of Industry, Science and Resources (2019, updated 2024). *Australia's Artificial Intelligence Ethics Principles* : https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-principles

Derblom, K., Molin, J., Gabrielsson, S., & Lindgren, B.-M. (2021). "Acknowledge me as a capable person": How people with mental ill health describe their experiences with general emergency care staff—A qualitative interview study. *International Journal of Mental Health Nursing*, *30*(6), 1539-1549. https://doi.org/10.1111/inm.12903

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv.* arXiv:1810.04805

Digel Vandyk, A., Young, L., MacPhee, C., & Gillis, K. (2018). Exploring the experiences of persons who frequently visit the emergency department for mental health-related reasons. *Qualitative Health Research*, *28*(4), 587-599. https://doi.org/10.1177/1049732317746382

Dixon, L. B., Holoshitz, Y., & Nossel, I. (2016). Treatment engagement of individuals experiencing mental illness: Review and update. *World Psychiatry*, *15*(1), 13-20. https://doi.org/10.1002/wps.20306

Dobson, H., Malpas, C. B., Burrell, A. J., Gurvich, C., Chen, L., Kulkarni, J., & Winton-Brown, T. (2021). Burnout and psychological distress amongst Australian healthcare workers during the COVID-19 pandemic. *Australasian Psychiatry*, *29*(1), 26-30.

Dodgson, J. E. (2022). The pandemic has brought too much change: Too many preprints; Too many retractions. *Journal of Human Lactation*, *38*(2), 207-208. https://doi.org/10.1177/08903344221082571

Dong, G., & Liu, H. (Eds.). (2018). *Feature engineering for machine learning and data analytics*. CRC Press. https://doi.org/10.1201/9781315181080

Dorogush, A. V., Ershov, V., & Gulin, A. (2018). CatBoost: Gradient boosting with categorical features support. *arXiv.* arXiv:1810.11363

Dudgeon, P., Bray, A., Darlaston-Jones, D., & Walker, R. (2020). *Aboriginal participatory action research: An Indigenous research methodology strengthening decolonisation and social and emotional wellbeing* (Discussion Paper Series, Issue 1). Lowitja Institute. https://www.lowitja.org.au/content/Document/Lowitja-Publishing/LI_Discussion_Paper_P-Dudgeon_FINAL3.pdf

Dudgeon, P., Walker, R., Scrine, C., Cox, K., D'Anna, D., Dunkley, C., Kelly, K., & Hams, K. (2014). Enhancing wellbeing, empowerment, healing and leadership. In P. Dudgeon, H. Milroy, & R. Walker (Eds.), *Working together: Aboriginal and Torres Strait Islander mental health and wellbeing principles and practice* (2nd ed., pp. 437-448). Commonwealth of Australia.

Duggan, M., Harris, B., Chislett, W. K., & Calder, R. (2020). *Nowhere else to go: Why Australia's health system results in people with mental illness getting 'stuck' in emergency departments*. Victoria University. https://vuir.vu.edu.au/41956/1/nowhere-else-to-go-people-mental-illness-stuck-emergency-departments-report-mitchell-institute.pdf

Durey, A., & Thompson, S. C. (2012). Reducing the health disparities of Indigenous Australians: Time to change focus. *BMC Health Services Research*, *12*, 1-11.

Edwards, B. (2018, June 8). *Birth of a standard: The Intel 8086 microprocessor turns 40 today*. PCWorld. https://www.pcworld.com/article/146957/article.html

Eltorai, A. E. M., McKinney, S. E., Rockenbach, M. A. B. C., Karuppiah, S., Bizzo, B. C., & Andriole, K. P. (2024). Primary care provider perspectives on the value of opportunistic CT screening. *Clinical Imaging*, *112*, Article 110210. https://doi.org/10.1016/j.clinimag.2024.110210

eMarketer. (2020). *Purchases via smart speakers are not taking off*. https://www.emarketer.com/content/purchases-via-smart-speakers-are-not-taking-off

Engel, G. L. (1977). The need for a new medical model: A challenge for biomedicine. *Science*, *196*(4286), 129-136.

Engel, G. L. (1980). The clinical application of the biopsychosocial model. *American Journal of Psychiatry*, *137*(5), 535-544.

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, *114*(4), 864-886. https://doi.org/10.1037/0033-295X.114.4.864

Esteva, A., Chou, K., Yeung, S., Naik, N., Madani, A., Mottaghi, A., Liu, Y., Topol, E., Dean, J., & Socher, R. (2021). Deep learning-enabled medical computer vision. *NPJ Digital Medicine*, *4*(1), Article 5.

Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., & Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine*, *25*(1), 24-29.

European Commission. (2023). *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: EU action plan for digital health* 2023–2030. https://digital-strategy.ec.europa.eu/en/policies/ehealth

Evans, R., Connell, J., Ablard, S., Rimmer, M., O'Keeffe, C., & Mason, S. (2019). The impact of different liaison psychiatry models on the emergency department: A systematic review of the international evidence. *Journal of Psychosomatic Research*, *119*, 53-64.

Everymind. (2022). *The Way Back Support Service NSW*. https://everymind.org.au/programs/twb-support-service-nsw

Fairburn, C. G., & Patel, V. (2017). The impact of digital technology on psychological treatments and their dissemination. *Behaviour Research and Therapy*, *88*, 19-25.

Fatovich, D., & Hirsch, R. (2003). Entry overload, emergency department overcrowding, and ambulance bypass. *Emergency Medicine Journal*, *20*(5), 406-409.

Feldman, R. C., Aldana, E., & Stein, K. (2019). Artificial intelligence in the health care space: How we can trust what we cannot know. *Stanford Law & Policy Review*, *30*, 399-419.

Fenwick, J. (2018). *Understanding Kubrick's 2001: A Space Odyssey: Representation and interpretation*. Intellect Books.

Ferguson, M., Rhodes, K., Loughhead, M., McIntyre, H., & Procter, N. (2022). The effectiveness of the safety planning intervention for adults experiencing suicide-related distress: A systematic review. *Archives of Suicide Research*, *26*(3), 1022-1045. https://doi.org/10.1080/13811118.2021.1915217

Fernandes, M., Vieira, S. M., Leite, F., Palos, C., Finkelstein, S., & Sousa, J. M. C. (2020). Clinical Decision Support Systems for Triage in the Emergency Department using

Intelligent Systems: a Review. *Artificial Intelligence in Medicine, 102*, 101762. https://doi.org/10.1016/j.artmed.2019.101762

Fischer, M., Coogan, A. N., Faltraco, F., & Thome, J. (2020). COVID-19 paranoia in a patient suffering from schizophrenic psychosis—A case report. *Psychiatry Research, 288*, Article 113001. https://doi.org/10.1016/j.psychres.2020.113001

Fisk, M., Livingstone, A., & Pit, S. W. (2020). Telehealth in the context of COVID-19: Changing perspectives in Australia, the United Kingdom, and the United States. *Journal of Medical Internet Research, 22*(6), Article e19264.

Fiske, A., Henningsen, P., & Buyx, A. (2019). Your robot therapist will see you now: Ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. *Journal of Medical Internet Research, 21*(5), Article e13216. https://doi.org/10.2196/13216

Fleet, R. P., Lavoie, K. L., Martel, J. P., Dupuis, G., Marchand, A., & Beitman, B. D. (2003). Two-year follow-up status of emergency department patients with chest pain: Was it panic disorder? *Canadian Journal of Emergency Medicine, 5*(4), 247-254. https://doi.org/10.1017/s1481803500008447

Floridi, L., & Cowls, J. (2022). A unified framework of five principles for AI in society. In S. Carta (Ed.), *Machine learning and the city: Applications in architecture and urban design* (pp. 535-545). Wiley Blackwell.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People–An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds Mach (Dordr), 28*(4), 689-707. https://doi.org/10.1007/s11023-018-9482-5

Fonseca Barbosa, J., & Gama Marques, J. (2023). The revolving door phenomenon in severe psychiatric disorders: A systematic review. *International Journal of Social Psychiatry, 69*(5), 1075-1089. https://doi.org/10.1177/00207640221143282

Fonseka, T. M., Bhat, V., & Kennedy, S. H. (2019). The utility of artificial intelligence in suicide risk prediction and the management of suicidal behaviors. *Australian & New Zealand Journal of Psychiatry, 53*(10), 954-964. https://doi.org/10.1177/0004867419864428

Francis, C., Wilson, R. L., & Higgins, O. (2024).Suicidal Ideation in ED. *ACMHN's 48th International Mental Health Nursing Conference, Mental Health Nursing, 32*(S1), 1-2. https://onlinelibrary.wiley.com/doi/abs/10.1111/inm.13209

Francis, G., & Garing, A. (2015). *Foundations of statistics*. Pearson Education Australia.

French, C. C., Santomauro, J., Hamilton, V., Fox, R., & Thalbourne, M. A. (2008). Psychological aspects of the alien contact experience. *Cortex, 44*(10), 1387-1395. https://doi.org/10.1016/j.cortex.2007.11.011

Frey, W. R., Patton, D. U., Gaskell, M. B., & McGregor, K. A. (2020). Artificial intelligence and inclusion: Formerly gang-involved youth as domain experts for analyzing unstructured Twitter data. *Social Science Computer Review*, *38*(1), 42-56.

Frith, C. D., & Corcoran, R. (1996). Exploring "theory of mind" in people with schizophrenia. *Psychological Medicine, 26*(3), 521-530. https://doi.org/10.1017/s0033291700035601

Fukushima, K. (1980). Neocognitron: A self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, *36*(4), 193-202. https://doi.org/10.1007/BF00344251

Gabbatt, A., & Laughland, O. (2022, January 14). Hospitals in half of US states close to capacity as Omicron continues surge. *The Guardian*. https://www.theguardian.com/us-news/2022/jan/14/us-hospitals-capacity-staffing-shortages-omicron-covid

Gaedtke, A. (2017). *Modernism and the machinery of madness: Psychosis, technology, and narrative worlds*. Cambridge University Press.

Gee, G., Dudgeon, P., Schultz, C., Hart, A., & Kelly, K. (2014). Aboriginal and Torres Strait Islander social and emotional wellbeing. In P. Dudgeon, H. Milroy, & R. Walker (Eds.), *Working together: Aboriginal and Torres Strait Islander mental health and wellbeing principles and practice* (2nd ed., pp. 55-68). Commonwealth of Australia.

Geng, F., Jiang, F., Conrad, R., Liu, T., Liu, Y., Liu, H., & Tang, Y.-l. (2020). Factors associated with involuntary psychiatric hospitalization of youths in China based on a nationally representative sample. *Frontiers in Psychiatry, 11*. https://doi.org/10.3389/fpsyt.2020.607464

Gerke, S., Minssen, T., & Cohen, G. (2020). Ethical and legal challenges of artificial intelligence-driven healthcare. In A. Bohr & K. Memarzadeh (Eds.), *Artificial intelligence in healthcare* (pp. 295-336). Academic Press. https://doi.org/10.1016/B978-0-12-818438-7.00012-5

Ghaemi, S. N. (2009). The rise and fall of the biopsychosocial model. *British Journal of Psychiatry, 195*(1), 3-4.

Gianfrancesco, M. A., Tamang, S., Yazdany, J., & Schmajuk, G. (2018). Potential biases in machine learning algorithms using electronic health record data. *JAMA Internal Medicine*, *178*(11), 1544-1547. https://doi.org/10.1001/jamainternmed.2018.3763

Gildberg, F., & Wilson, R. (2023). Scientific models for qualitative research: A textual thematic analysis coding system—Part 1. *Nurse Researcher*, *31*(3), 36-42. https://doi.org/10.7748/nr.2023.e1893

Goffman, E. (1968). *Asylums: Essays on the social situation of mental patients and other inmates*. AldineTransaction.

Góngora Alonso, S., Herrera Montano, I., Ayala, J. L. M., Rodrigues, J. J. P. C., Franco-Martín, M., & de la Torre Díez, I. (2024). Machine Learning Models to Predict Readmission Risk of Patients with Schizophrenia in a Spanish Region. *International Journal of Mental Health and Addiction, 22*, 2508–2527. https://doi.org/10.1007/s11469-022-01001-x

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Gooding, P., Brown, L. X. Z., Myrick, K., Ubozoh, K., Horton, J., Bossewitch, J., Vásquez Encalada, A., & Katterl, S. (2022). *Digital futures in mind: Reflecting on technological experiments in mental health and crisis support*. University of Melbourne. https://melbourne.figshare.com/articles/report/Digital_Futures_in_Mind_Reflecting_on_Technological_Experiments_in_Mental_Health_and_Crisis_Support/21113899

Gooding, P., & Kariotis, T. (2021). Ethics and law in research on algorithmic and data-driven technology in mental health care: Scoping review. *JMIR Mental Health*, *8*(6), Article e24668. https://doi.org/10.2196/24668

Greenhalgh, T. (2019). *How to read a paper: The basics of evidence-based medicine and healthcare*. John Wiley & Sons.

Gregor, S., Imran, A., Okai-Ugbaje, I., Gunawardena, M., Jeffery, C. P., & Wilson, R. L. (2023). Indigenous knowledge-sharing interventions in Australia and the use of information and communication technology: A scoping review. *Australian Journal of Information Systems, 27.* https://doi.org/10.3127/ajis.v27i0.4217

Griffiths, P., Maruotti, A., Recio Saucedo, A., Redfern, O. C., Ball, J. E., Briggs, J., Dall'Ora, C., Schmidt, P. E., & Smith, G. B. (2019). Nurse staffing, nursing assistants and hospital mortality: Retrospective longitudinal cohort study. *BMJ Quality & Safety*, *28*(8), 609-617. https://doi.org/10.1136/bmjqs-2018-008043

Gualano, M. R., Sinigaglia, T., Lo Moro, G., Rousset, S., Cremona, A., Bert, F., & Siliquini, R. (2021). The Burden of Burnout among Healthcare Professionals of Intensive Care Units and Emergency Departments during the COVID-19 Pandemic: A Systematic Review. *International Journal of Environmental Research and Public Health, 18*(15), Article 8172. https://doi.org/10.3390/ijerph18158172

Gwynne, K., Jeffries, T., & Lincoln, M. (2018). Improving the efficacy of healthcare services for Aboriginal Australians. *Australian Health Review, 43*(3), 314-322.

Habli, I., Lawton, T., & Porter, Z. (2020). Artificial intelligence in health care: Accountability and safety. *Bulletin of the World Health Organization*, *98*(4), 251-256. https://doi.org/10.2471/BLT.19.237487

Haebich, A. (2011). Forgetting Indigenous histories: Cases from the history of Australia's Stolen Generations. *Journal of Social History, 44*(4), 1033-1046.

Hah, H., & Goldin, D. S. (2021). How clinicians perceive artificial intelligence-assisted technologies in diagnostic decision making: Mixed methods approach. *Journal of Medical Internet Research, 23*(12), Article e33540. https://doi.org/10.2196/33540

Hall, A., McKenna, B., Dearie, V., Maguire, T., Charleston, R., & Furness, T. (2016). Educating emergency department nurses about trauma informed care for people presenting with mental health crisis: A pilot study. *BMC Nursing, 15*(1), Article 21. https://doi.org/10.1186/s12912-016-0141-y

Hanson, J. (2016). *The social media revolution: An economic encyclopedia of friending, following, texting, and connecting*. ABC-CLIO.

Hantel, A., Walsh, T. P., Marron, J. M., Kehl, K. L., Sharp, R., Van Allen, E., & Abel, G. A. (2024). Perspectives of oncologists on the ethical implications of using artificial intelligence for cancer care. *JAMA Network Open, 7*(3), Article e244077. https://doi.org/10.1001/jamanetworkopen.2024.4077

Haque, M. D. R., & Rubya, S. (2023). An overview of chatbot-based mobile mental health apps: Insights from app description and user reviews. *JMIR mHealth and uHealth, 11*, Article e44838. https://doi.org/10.2196/44838

Harfield, S., Pearson, O., Morey, K., Kite, E., Canuto, K., Glover, K., Gomersall, J. S., Carter, D., Davy, C., Aromataris, E., & Braunack-Mayer, A. (2020). Assessing the quality of health research from an Indigenous perspective: The Aboriginal and Torres Strait Islander quality appraisal tool. *BMC Medical Research Methodology, 20*(1), Article 79. https://doi.org/10.1186/s12874-020-00959-3

Harrington, L., Siegert, R. J., & McClure, J. (2005). Theory of mind in schizophrenia: A critical review. *Cognitive Neuropsychiatry, 10*(4), 249-286. https://doi.org/10.1080/13546800444000056

Harris, P., Vardaxis, N., & Purtell, L. (2024). *Mosby's Dictionary of Medicine, Nursing and Health Professions-4th ANZ Edition-E-Book*. Elsevier Health Sciences.

Harrou, F., Dairi, A., Kadri, F., & Sun, Y. (2020). Forecasting emergency department overcrowding: A deep learning framework. *Chaos, Solitons & Fractals*, *139*, Article 110247.

Hastie, T., Tibshirani, R., & Friedman, J. (2013). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.

Health Canada. (n.d.). *Digital health.* Government of Canada. https://digitalhealthcanada.com/about/strategy/ Accessed 15/02/2025

Hedegaard, H., Curtin, S. C., & Warner, M. (2020). Increase in suicide mortality in the United States, 1999-2018. *NCHS Data Brief, 362*, 1-8.

Heiss, A. (2018). *Growing up Aboriginal in Australia*. Black Inc.

Henderson, C., Evans-Lacko, S., & Thornicroft, G. (2013). Mental illness stigma, help seeking, and public health programs. *American Journal of Public Health, 103*(5), 777-780. https://doi.org/10.2105/ajph.2012.301056

Hendrickson, K. E., III, Corfield, J. J., & Danver, S. L. (2014). *The encyclopedia of the Industrial Revolution in world history*. Rowman & Littlefield.

Herwix, A., Haj-Bolouri, A., Rossi, M., Tremblay, M. C., Purao, S., & Gregor, S. (2022). Ethics in information systems and design science research: Five perspectives. *Communications of the Association for Information Systems, 50*(1), 589-616.

Higgins, O., Chalup, S. K., Short, B. L., & Wilson, R. L. (2021a, October 13). *Disruptions of the familiar: Psychosis echoes technological advancement* [Panel presentation]. TheMHS 2021 Hope into Action, Melbourne.

Higgins, O., Chalup, S. K., Short, B. L., & Wilson, R. L. (2021, December 7-8). *Interpretations of innovation: The intersection of technological advancement and psychosis* [Paper presentation]. University of Newcastle School of Nursing and Midwifery Research Symposium, Newcastle.

Higgins, O., Chalup, S. K., Short, B. L., & Wilson, R. L. (2022a). Interpretations of innovation: The intersection of technological advancement and psychosis. *JMIR Preprints*, 37033. https://doi.org/10.2196/preprints.37033

Higgins, O., Chalup, S. K., Short, B. L., & Wilson, R. L. (2022b, September 7–9). *Interpretations of innovation: The intersection of technological advancement and psychosis* [Paper presentation]. ACMHN's 46th International Mental Health Nursing Conference, Mental Health Nursing in a Climate of Change. https://onlinelibrary.wiley.com/doi/abs/10.1111/inm.13050

Higgins, O., Chalup, S. K., Short, B. L., & Wilson, R. L. (2022c, December 15) *Interpretations of innovation: The intersection of technological advancement and psychosis* [Paper presentation]. Research & Innovation Symposium, Central Coast Local Health District, Gosford.

Higgins, O., Chalup, S. K., Short, B. L., & Wilson, R. L. (2022d, September 29 to October 1). *Interpretations of innovation: The intersection of technological advancement and psychosis* [Paper presentation]. International Association for Youth Mental Health, Copenhagen, Denmark.

Higgins, O., Chalup, S. K., Short, B. L., & Wilson, R. L. (2023a, September 13-15). Artificial intelligence in nursing: Trustworthy or reliable? *ACMHN's 47th International Mental Health Nursing Conference*, https://doi.org/10.1111/inm.13209

Higgins, O., Chalup, S. K., Short, B. L., & Wilson, R. L. (2023b). Interpretations of innovation: The role of technology in explanation seeking related to psychosis. *Perspectives in Psychiatric Care, 2023*, Article 464934. https://doi.org/10.1155/2023/4464934

Higgins, O., Chalup, S. K., & Wilson, R. L. (2022a, June 17). *Artificial intelligence in nursing: Trustworthy or reliable?* [Paper presentation]. 1st Workshop on Artificial Intelligence in Nursing: Advances, Methods and Path Forward, Halifax, Canada.

Higgins, O., Chalup, S. K., & Wilson, R. L. (2022b, June 17). *Artificial intelligence in nursing: trustworthy or reliable?* [Paper presentation]. Forensic Mental Health Research Symposium, Middelfart, Denmark.

Higgins, O., Chalup, S. K., & Wilson, R. L. (2022c, June 17). *Artificial intelligence in nursing: Trustworthy or reliable?* University of Southern Denmark Research Symposium, Middelfart, Denmark.

Higgins, O., Chalup, S. K., & Wilson, R. L. (2024a). Artificial intelligence in nursing: Trustworthy or reliable? *Journal of Research in Nursing*, *29*(2), 143-153. https://doi.org/10.1177/17449871231215696

Higgins, O., Chalup, S. K., & Wilson, R. L. (2024b). Machine learning model reveals determinators for admission to acute mental health wards from emergency department

presentations. *International Journal of Mental Health Nursing*. Advance online publication. https://doi.org/10.1111/inm.13402

Higgins, O., Chalup, S. K., & Wilson, R. L. (2024c, October 29 - November 1). Machine learning reveals determinators for admission to Acute Mental Health Wards from Emergency Department presentations, *ACMHN's 48th International Mental Health Nursing Conference,* https://onlinelibrary.wiley.com/doi/abs/10.1111/inm.13209

Higgins, O., Sheather-Reid, R. B., Chalup, S. K., & Wilson, R. L. (2024a). Disproportionate mental health presentations to emergency departments in a coastal regional community in Australia of First Nation people. *International Journal of Mental Health Nursing*. Advance online publication. https://doi.org/10.1111/inm.13362

Higgins, O., Sheather-Reid, R. B., Chalup, S. K., & Wilson, R. L. (2024b, October 29 - November 1).Disproportionate mental health presentations to Emergency Departments in coastal regional Australia of First Nations people. *ACMHN's 48th International Mental Health Nursing Conference, Mental Health Nursing* https://onlinelibrary.wiley.com/doi/abs/10.1111/inm.13209

Higgins, O., Sheather-Reid, R. B., Chalup, S. K., & Wilson, R. L. (2024c). Sociodemographic factors and presentation features of individuals seeking mental health care in emergancy departments: A retrospective cohort study. *International Journal of Mental Health Nursing*. Advance online publication. https://doi.org/10.1111/inm.13414

Higgins, O., Short, B. L., Chalup, S. K., & Wilson, R. L. (2022). Artificial intelligence and machine learning based decision support systems in mental health. *Prospero,* CRD42022352360. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42022352360

Higgins, O., Short, B. L., Chalup, S. K., & Wilson, R. L. (2023). Artificial intelligence (AI) and machine learning (ML) based decision support systems in mental health: An integrative review. *International Journal of Mental Health Nursing*, *32*(4), 966-978. https://doi.org/10.1111/inm.13114

Higgins, O., & Wilson, R. L. (2023). Commercial determinants and therapeutic chatbots: A mental health nursing perspective. *International Journal of Mental Health Nursing*, *32*(6), 1509-1511.

Higgins, O., Wilson, R. L., & Chalup, S. K. (2024d). Using machine learning to assist decision making in the assessment of mental health consumers presenting to emergency departments. *Digital Health*.

Higgins, O., & Wilson, R. L., (2025). Integrating Artificial Intelligence (AI) with Workforce Solutions for Sustainable Care: A follow up to Artificial intelligence and machine learning (ML) based decision support systems in mental health. *International Journal of Mental Health Nursing*. Accepted 14/02/2025

Hirjak, D., & Fuchs, T. (2010). Delusions of technical alien control: A phenomenological description of three cases. *Psychopathology*, *43*(2), 96-103. https://doi.org/10.1159/000274178

HLEG AI. (2019). *Ethics guidelines for trustworthy AI*.

Holley, P., & Larimer, S. (2014, November 20). FSU gunman was in "state of crisis" during shooting, investigators say. *The Washington Post.* https://www.washingtonpost.com/news/post-nation/wp/2014/11/20/fsu-gunman-remembered-as-hard-worker-kindest-sweetest-person-by-baffled-friends-acquaintances/

Holmes, E. A., O'Connor, R. C., Perry, V. H., Tracey, I., Wessely, S., Arseneault, L., Ballard, C., Christensen, H., Cohen Silver, R., Everall, I., Ford, T., John, A., Kabir, T., King, K., Madan, I., Michie, S., Przybylski, A. K., Shafran, R., Sweeney, A., . . . Bullmore, E. (2020). Multidisciplinary research priorities for the COVID-19 pandemic: A call for action for mental health science. *The Lancet Psychiatry*, *7*(6), 547-560. https://doi.org/10.1016/S2215-0366(20)30168-1

Hong, W. S., Haimovich, A. D., & Taylor, R. A. (2018). Predicting hospital admission at emergency department triage using machine learning. *PLOS ONE*, *13*(7), Article e0201016. https://doi.org/10.1371/journal.pone.0201016

Hoy, M. B. (2018). Alexa, Siri, Cortana, and more: An introduction to voice assistants. *Medical Reference Services Quarterly*, *37*(1), 81-88. https://doi.org/10.1080/02763869.2018.1404391

Huang, Y., Wan, L. J., Ye, H., Jha, M., Wang, J., Li, Y., Zhang, X., & Chen, D. (2024). New solutions on LLM acceleration, optimization, and application. *arXiv*. arXiv:2406.10903.

InfoComm Media Development Authority. (2019). *Singapore wins international award for its artificial intelligence governance and ethics initiatives* [Press release]. https://www.imda.gov.sg/newsand-events/Media-Room/Media-Releases/2019/singapore-wins-international-award-for-its-artificialintelligence-governance-and-ethics-initiatives

International Council of Nurses. (2021). *The ICN Code of Ethics for Nurses*. https://www.icn.ch/sites/default/files/inline-files/ICN_Code-of-Ethics_EN_Web.pdf

InterpretML. (2023a). *InterpretML documentation*. https://interpret.ml/docs/

InterpretML. (2023b). *Overall importance*. https://interpret.ml/docs/overall_importance/

Iorfino, F., Ho, N., Carpenter, J. S., Cross, S. P., Davenport, T. A., Hermens, D. F., Yee, H., Nichles, A., Zmicerevska, N., & Guastella, A. (2020). Predicting self-harm within six months after initial presentation to youth mental health services: A machine learning study. *PLOS ONE, 15*(12), Article e0243467.

Isaac, R. J., & Brakel, S. J. (1992). Subverting good intentions: A brief history of mental health law reform. *Cornell Journal of Law & Public Policy, 2*, 89-119.

Isaak, J., & Hanna, M. J. (2018). User data privacy: Facebook, Cambridge Analytica, and privacy protection. *Computer, 51*(8), 56-59. https://doi.org/10.1109/mc.2018.3191268

Jacobs, M., Pradier, M. F., McCoy, T. H., Jr., Perlis, R. H., Doshi-Velez, F., & Gajos, K. Z. (2021). How machine-learning recommendations influence clinician treatment selections: The example of the antidepressant selection. *Translational Psychiatry, 11*(1), Article 108. https://doi.org/10.1038/s41398-021-01224-x

James, M., Painter, J., Buckingham, B., & Stewart, M. W. (2018). A review and update of the Health of the Nation Outcome Scales (HoNOS). *BJPsych Bulletin, 42*(2), 63-68. https://doi.org/10.1192/bjb.2017.17

Jen, N. (2017). Design thinking is bullshit. Filmed 2017 at 99U Conference, New York, New York. Video.

Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., & Wang, Y. (2017). Artificial intelligence in healthcare: Past, present and future. *Stroke and Vascular Neurology, 2*(4).

Johansson-Sköldberg, U., Woodilla, J., & Çetinkaya, M. (2013). Design thinking: Past, present and possible futures. *Creativity and Innovation Management, 22*(2), 121-146.

Jones, A., Read, J., & Wood, L. (2021). A retrospective case study of the thematic content of psychotic experiences in a first episode psychosis population. *Journal of Mental Health, 30*(4), 509-517. https://doi.org/10.1080/09638237.2020.1755024

Jones, T. L., Hamilton, P., & Murry, N. (2015). Unfinished nursing care, missed care, and implicitly rationed care: State of the science review. *International Journal of Nursing Studies, 52*(6), 1121-1137. https://doi.org/10.1016/j.ijnurstu.2015.02.012

Jones v. Ault, 67 F.R.D. 124 (United States District Court for the Southern District of Georgia 1974). https://cite.case.law/frd/67/124/

Kalisch, B. J., Tschannen, D., & Lee, K. H. (2012). Missed nursing care, staffing, and patient falls. *Journal of Nursing Care Quality*, *27*(1), 6-12. https://doi.org/10.1097/NCQ.0b013e318225aa23

Kalisch, B. J., Xie, B., & Ronis, D. L. (2013). Train-the-trainer intervention to increase nursing teamwork and decrease missed nursing care in acute care patient units. *Nursing Research*, *62*(6), 405-413. https://doi.org/10.1097/NNR.0b013e3182a7a15d

Kamel Rahimi, A., Pienaar, O., Ghadimi, M., Canfell, O. J., Pole, J. D., Shrapnel, S., van der Vegt, A. H., & Sullivan, C. (2024). Implementing AI in hospitals to achieve a learning health system: Systematic review of current enablers and barriers. *Journal of Medical Internet Research*, *26*, Article e49655. https://doi.org/10.2196/49655

Kang, H. (2013). The prevention and handling of the missing data. *Korean Journal of Anesthesiology*, *64*(5), 402-406. https://doi.org/10.4097/kjae.2013.64.5.402

Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., & Amodei, D. (2020). Scaling laws for neural language models. *arXiv.* arXiv:2001.08361

Karahanoğlu, A., & Bakırlıoğlu, Y. (2022). Evaluation of the usefulness of path of long-term user experience model in design process. *Behaviour & Information Technology*, *41*(4), 777-795. https://doi.org/10.1080/0144929X.2020.1836256

Kelaher, M. A., Ferdinand, A. S., & Paradies, Y. (2014). Experiencing racism in health care: The mental health impacts for Victorian Aboriginal communities. *Medical Journal of Australia*, *201*(1), 44-47. https://doi.org/10.5694/mja13.10503

Kelleher, J. D., Mac Namee, B., & D'arcy, A. (2020). *Fundamentals of machine learning for predictive data analytics: Algorithms, worked examples, and case studies*. MIT Press.

Kemp, S. (2021). *Digital 2021 April global statshot report*. DataReportal. https://datareportal.com/reports/digital-2021-april-global-statshot

Khamisa, N., Oldenburg, B., Peltzer, K., & Ilic, D. (2015). Work related stress, burnout, job satisfaction and general health of nurses. *International Journal of Environmental Research and Public Health, 12*(1), 652-666. https://doi.org/10.3390/ijerph120100652

Khan, M. A., & Salah, K. (2018). IoT security: Review, blockchain solutions, and open challenges. *Future Generation Computer Systems*, *82*, 395-411. https://doi.org/10.1016/j.future.2017.11.022

Knaak, S., Mantler, E., & Szeto, A. (2017). Mental illness-related stigma in healthcare: Barriers to access and care and evidence-based solutions. *Healthcare Management Forum*, *30*(2), 111-116. https://doi.org/10.1177_0840470416679413

Knight, W. (2017, April 11). *The dark secret at the heart of AI*. MIT Technology Review. https://www.technologyreview.com/2017/04/11/5113/the-dark-secret-at-the-heart-of-ai/

Kolko, J. (2015, September). Design thinking comes of age. *Harvard Business Review*.

Köppen, E., & Meinel, C. (2015). Empathy via design thinking: Creation of sense and knowledge. In H. Plattner, C. Meinel, & L. Leifer (Eds.), *Design thinking research: Building innovators* (pp. 15-28). Springer.

Kuhn, M. (2013). *Applied predictive modeling*. Springer.

Lakeman, R., Foster, K., Happell, B., Hazelton, M., Moxham, L., & Hurley, J. (2024). Informing the development of a fit-for-purpose mental health nursing curriculum: A survey of mental health nurse academics in Australia. *International Journal of Mental Health Nursing*, *33*(1), 93-103. https://doi.org/10.1111/inm.13226

Lay, B., Kawohl, W., & Rössler, W. (2019). Predictors of compulsory re-admission to psychiatric inpatient care. *Frontiers in Psychiatry, 10*, Article 120. https://doi.org/10.3389/fpsyt.2019.00120

Leigh, J., & Roberts, D. (2018). Critical exploration of the new NMC standards of proficiency for registered nurses. *British Journal of Nursing*, *27*(18), 1068-1072.

Lewrick, M., Link, P., & Leifer, L. (2018). *The design thinking playbook: Mindful digital transformation of teams, products, services, businesses and ecosystems*. Wiley.

Lheureux, B., Schulte, W. R., & Velosa, A. (2017). *Hype cycle for the internet of things, 2017*. Gartner. https://www.gartner.com/en/documents/3770369/hype-cycle-for-the-internet-of-things-2017

Liamputtong, P. (2021). *Research methods and evidence-based practice*. Oxford University Press Australia & New Zealand.

Liang, Y., & Lee, S. A. (2017). Fear of autonomous robots and artificial intelligence: Evidence from national representative data with probability sampling. *International Journal of Social Robotics*, *9*(3), 379-384. https://doi.org/10.1007/s12369-017-0401-3

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*(3), 18-22.

Liedtka, J. (2015). Perspective: Linking design thinking with innovation outcomes through cognitive bias reduction. *Journal of Product Innovation Management*, *32*(6), 925-938.

Lin, S. (2022). A clinician's guide to artificial intelligence (AI): Why and how primary care should lead the health care AI revolution. *Journal of the American Board of Family Medicine*, *35*(1), 175-184. https://doi.org/10.3122/jabfm.2022.01.210226

Lingel, J., & Crawford, K. (2020). “Alexa, tell me about your mother”: The history of the secretary and the end of secrecy. *Catalyst: Feminism, Theory, Technoscience*, *6*(1).

Liu, H.-Y. (2018). The power structure of artificial intelligence. *Law, Innovation and Technology*, *10*(2), 197-229. https://doi.org/10.1080/17579961.2018.1527480

Liu, K., Zhang, W., Li, W., Wang, T., & Zheng, Y. (2023). Effectiveness of virtual reality in nursing education: A systematic review and meta-analysis. *BMC Medical Education*, *23*(1), Article 710. https://doi.org/10.1186/s12909-023-04662-x

Liu, Y. F., Joly, R., Turchioe, M. R., Benda, N., Hermann, A., Beecy, A., Pathak, J., & Zhang, Y. Y. (2024). Preparing for the bedside—Optimizing a postpartum depression risk prediction model for clinical implementation in a health system. *Journal of the American Medical Informatics Association*, *31*(6), 1258-1267. https://doi.org/10.1093/jamia/ocae056

Lohani, M., Bryan, C. J., Elsey, J. S., Dutton, S., Findley, S. P., Langenecker, S. A., West, K., & Baker, J. C. (2024). Collaboration matters: A randomized controlled trial of patient-clinician collaboration in suicide risk assessment and intervention. *Journal of Affective Disorders*, *360*, 387-393. https://doi.org/10.1016/j.jad.2024.06.004

Lowitja Institute. (2018). *Deficit discourse and Aboriginal and Torres Strait Islander health policy*.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in neural information processing systems* (Vol. 30). Curran Associates.

Lutz, W., Deisenhofer, A. K., Rubel, J., Bennemann, B., Giesemann, J., Poster, K., & Schwartz, B. (2022). Prospective evaluation of a clinical decision support system in psychological therapy. *Journal of Consulting and Clinical Psychology*, *90*(1), 90–106. https://doi.org/10.1037/ccp0000642

Lysen, F., & Wyatt, S. (2024). Refusing participation: Hesitations about designing responsible patient engagement with artificial intelligence in healthcare. *Journal of Responsible Innovation*, *11*(1), Article 2300161.

Malgaroli, M., Hull, T. D., Zech, J. M., & Althoff, T. (2023). Natural language processing for mental health interventions: A systematic review and research framework. *Translational Psychiatry*, *13*(1), Article 309.

Marques, I. C. P., & Ferreira, J. J. M. (2020). Digital transformation in the area of health: Systematic review of 45 years of evolution. *Health and Technology*, *10*(3), 575-586. https://doi.org/10.1007/s12553-019-00402-8

Marr, B. (2019). *How is AI used In healthcare—5 powerful real-world examples that show the latest advances*. Forbes. https://www.forbes.com/sites/bernardmarr/2018/07/27/how-is-ai-used-in-healthcare-5-powerful-real-world-examples-that-show-the-latest-advances/

Martínez-Alés, G., Jiang, T., Keyes, K. M., & Gradus, J. L. (2022). The recent rise of suicide mortality in the United States. *Annual Review of Public Health*, *43*, 99-116. https://doi.org/10.1146/annurev-publhealth-051920-123206

Maunder, R. G., Heeney, N. D., Greenberg, R. A., Jeffs, L. P., Wiesenfeld, L. A., Johnstone, J., & Hunter, J. J. (2023). The relationship between moral distress, burnout, and considering leaving a hospital job during the COVID-19 pandemic: A longitudinal survey. *BMC Nursing*, *22*(1), Article 243.

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955. *AI Magazine*, *27*(4), 12-12.

McCradden, M. D., Anderson, J. A., A. Stephenson, E., Drysdale, E., Erdman, L., Goldenberg, A., & Zlotnik Shaul, R. (2022). *A research ethics framework for the clinical translation of healthcare machine learning*. The American Journal of Bioethics, 22(5), 8–22.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, *5*(4), 115-133.

McDowell, J. C. (2020). The low earth orbit satellite population and impacts of the SpaceX Starlink constellation. *Astrophysical Journal Letters*, *892*(2), L36.

McGough, S., Wynaden, D., Gower, S., Duggan, R., & Wilson, R. (2022). There is no health without cultural safety: Why cultural safety matters. *Contemporary Nurse*, *58*(1), 33-42. https://doi.org/10.1080/10376178.2022.2027254

McLeod, C. (2021). Trust. In E. N. Zalta (Ed.), *Stanford encyclopedia of philosophy archive* (Fall 2021 ed.). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2021/entries/trust/

Mereness, D. (1951). Preparation of the nurse for the psychiatric team. *American Journal of Nursing*, *51*(5), 320-322. https://doi.org/10.2307/3459409

Meskó, B. (2023). The impact of multimodal large language models on health care's future. *Journal of Medical Internet Research*, *25*, Article e52865.

Mignon, S. I. (2019). *Social work and mental health : Evidence-based policy and practice.* Springer.

Milroy, H., Kashyap, S., Collova, J., Mitchell, M., Ryder, A., Cox, Z., Coleman, M., Taran, M., Cuesta Briand, B., & Gee, G. (2024). Walking together in friendship: Learning about cultural safety in mainstream mental health services through Aboriginal participatory action research. *Australia & New Zealand Journal of Psychiatry, 58*(6), 498-505. https://doi.org/10.1177/00048674241246444

Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep learning for healthcare: Review, opportunities and challenges. *Briefings in Bioinformatics, 19*(6), 1236-1246. https://doi.org/10.1093/bib/bbx044

Miyazono, K. (2015). Delusions as harmful malfunctioning beliefs. *Consciousness and Cognition, 33,* 561-573. https://doi.org/10.1016/j.concog.2014.10.008

Moggia, D., Schaffrath, J., Bommer, J., Weinmann-Lutz, B., & Lutz, W. (2022). Data-informed case formulation with the Trier Treatment Navigator. *Revista De Psicoterapia, 33*(123), 151-171. https://doi.org/10.33898/rdp.v33i123.35971

Molnar, C. (2020). *Interpretable machine learning*. Leanpub. https://christophm.github.io/interpretable-ml-book/

Moreno, C., Wykes, T., Galderisi, S., Nordentoft, M., Crossley, N., Jones, N., Cannon, M., Correll, C. U., Byrne, L., Carr, S., Chen, E. Y. H., Gorwood, P., Johnson, S., Kärkkäinen, H., Krystal, J. H., Lee, J., Lieberman, J., López-Jaramillo, C., Männikkö, M., . . . Arango, C. (2020). How mental health care should change as a consequence of the COVID-19 pandemic. *The Lancet Psychiatry*, *7*(9), 813-824. https://doi.org/10.1016/S2215-0366(20)30307-2

Morphet, J., Innes, K., Munro, I., O'Brien, A., Gaskin, C. J., Reed, F., & Kudinoff, T. (2012). Managing people with mental health presentations in emergency departments—A service exploration of the issues surrounding responsiveness from a mental health care consumer and carer perspective. *Australasian Emergency Nursing Journal*, *15*(3), 148-155. https://doi.org/10.1016/j.aenj.2012.05.003

Morris, N. (2018, January 1). This secret experiment tricked psychiatrists into diagnosing sane people as having schizophrenia. *Washington Post*. https://www.washingtonpost.com/national/health-science/an-experiment-fooled-

psychiatrists-into-treating-sane-people-as-if-they-were-insane/2017/12/29/c6c9c3ea-d5f7-11e7-b62d-d9345ced896d_story.html

Mudallal, R. H., Othman, W. M., & Al Hassan, N. F. (2017). Nurses' burnout: The influence of leader empowering behaviors, work conditions, and demographic traits. *Inquiry*, *54*. https://doi.org/10.1177/0046958017724944

Mueller, S. T. (2020). Cognitive anthropomorphism of AI: How humans and computers classify images. *Ergonomics in Design: The Quarterly of Human Factors Applications*, *28*(3), 12-19. https://doi.org/10.1177/1064804620920870

Murphy, K. P. (2023). *Probabilistic machine learning: Advanced topics.* MIT press.

Mustian, J. (2014, March 12). Man who killed hostages in north Louisiana bank had mental illness. *Baton Rouge Morning Advocate*. https://web.archive.org/web/20140313181529/http://theadvocate.com/home/8612131-125/la-bank-hostages-killer-had-no

National Center for Biotechnology Information. (2016a). *Machine learning*. https://www.ncbi.nlm.nih.gov/mesh/2010029

National Center for Biotechnology Information. (2016b). *Machine learning—MeSH*. https://www.ncbi.nlm.nih.gov/mesh/2010029

National Health and Medical Research Council. (2018a). *Ethical conduct in research with Aboriginal and Torres Strait Islander Peoples and communities: Guidelines for researchers and stakeholders*.

National Health and Medical Research Council. (2018b). *National Statement on Ethical Conduct in Human Research 2007 (updated 2018)*. https://www.nhmrc.gov.au/about-us/publications/national-statement-ethical-conduct-human-research-2007-updated-2018

National Indigenous Australians Agency. (2023). *Culture and empowering communities*. https://www.niaa.gov.au/our-work/culture-and-empowering-communities

National Mental Health Commission. (2024). *National Report Card 2023*. https://www.mentalhealthcommission.gov.au/publications/national-report-card-2023

Navlani, A. (2019). *KNN classification using Scikit-learn*. Datacamp. https://www.datacamp.com/community/tutorials/k-nearest-neighbor-classification-scikit-learn

Newton, D. (2014). *Contemporary world issues: Science and political controversy: A reference handbook*. ABC-CLIO.

Ngan, C. K., Paat, Y. F., & Green, R. (2022). HDSS: A healthcare decision support system on combining domain knowledge and data analytics for predicting potential risk of mental health. *International Journal of Applied Decision Sciences*, *15*(4), 465-491. https://doi.org/10.1504/IJADS.2022.123844

Nicholls, D., Gaynor, N., Shafiei, T., Bosanac, P., & Farrell, G. (2011). Mental health nursing in emergency departments: The case for a nurse practitioner role. *Journal of Clinical Nursing*, *20*(3-4), 530-536. https://doi.org/10.1111/j.1365-2702.2010.03504.x

Noever, D., & McKee, F. (2023). Chatbots as problem solvers: Playing twenty questions with role reversals. *arXiv.* arXiv:2301.01743

Nolan-Isles, D., Macniven, R., Hunter, K., Gwynn, J., Lincoln, M., Moir, R., Dimitropoulos, Y., Taylor, D., Agius, T., Finlayson, H., Martin, R., Ward, K., Tobin, S., & Gwynne, K. (2021). Enablers and barriers to accessing healthcare services for Aboriginal people in New South Wales, Australia. *International Journal of Environmental Research and Public Health, 18*(6), Article 3014. https://doi.org/10.3390/ijerph18063014

Nori, H., Jenkins, S., Koch, P., & Caruana, R. (2019). InterpretML: A unified framework for machine learning interpretability. *arXiv.* arXiv:1909.09223

New South Wales Government. (2002). *Health Records and Information Privacy Act 2002 No 71*. https://legislation.nsw.gov.au/view/html/inforce/current/act-2002-071

NSW Health. (2013). *Triage of patients in NSW emergency departments*. https://www1.health.nsw.gov.au/pds/ActivePDSDocuments/PD2013_047.pdf

NSW Health. (2015). *Mental health for emergency departments*.

NSW Health. (2017). *ED Mode of Separation Code Set update 2017/18*. https://www1.health.nsw.gov.au/pds/ActivePDSDocuments/IB2017_028.pdf

NSW Health. (2019). *Declared mental health facilities*. https://www.health.nsw.gov.au/mentalhealth/services/consumers/Pages/facilities.aspx

NSW Health. (2020). *Strategic framework for suicide prevention in NSW 2018–2023*. https://www.health.nsw.gov.au/mentalhealth/resources/Pages/strategic-framework-implementation-plan.aspx

NSW Health. (2021a). *Emergency department care*. https://www.health.nsw.gov.au/Performance/Pages/emergency.aspx

NSW Health. (2021b). *Safe Haven*. https://www.health.nsw.gov.au/towardszerosuicides/Pages/safe-haven.aspx

NSW Health. (2022). *Public health orders and restrictions*. https://www.health.nsw.gov.au/Infectious/covid-19/Pages/public-health-orders.aspx

NSW Health. (2023). *High or very high psychological distress in adults by LHD*. https://www.healthstats.nsw.gov.au/#/indicator?name=-men-hidistress-phs&location=LHD&view=Trend&measure=prevalence&groups=&compare=LHD&filter=LHD,All%20LHDs,Central%20Coast%20LHD

Nuij, C., van Ballegooijen, W., de Beurs, D., Juniar, D., Erlangsen, A., Portzky, G., O'Connor, R. C., Smit, J. H., Kerkhof, A., & Riper, H. (2021). Safety planning-type interventions for suicide prevention: Meta-analysis. *British Journal of Psychiatry, 219*(2), 419-426. https://doi.org/10.1192/bjp.2021.50

Nursing and Midwifery Board of Australia. (2016). *Registered nurse standards for practice*. https://www.nursingmidwiferyboard.gov.au/codes-guidelines-statements/professional-standards/registered-nurse-standards-for-practice.aspx

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science, 366*(6464), 447-453. https://doi.org/10.1126/science.aax2342

Office of the Australian Information Commissioner. (2019). *Australian Privacy Principles guidelines* (Chapter 2, Australian Privacy Principles). https://www.oaic.gov.au/__data/assets/pdf_file/0019/8108/australian-privacy-principles-guidelines.pdf

Office of the Australian Information Commissioner. (2014). *Approval of guidelines issued under section 95A of the Privacy Act 1988*. https://www.legislation.gov.au/Details/F2014L00243

Olson, C., & Kemery, K. (2019). Voice report: Consumer adoption of voice technology and digital assistants. Seattle: Microsoft..

OpenAI. (2022). *ChatGPT: Optimizing language models for dialogue*. https://openai.com/blog/chatgpt/

OpenAI. (2023). *ChatGPT*. https://openai.com/

Ortíz-Barrios, M. A., & Alfaro-Saíz, J.-J. (2020). Methodological approaches to support process improvement in emergency departments: A systematic review. *International Journal of Environmental Research and Public Health, 17*(8), Article 2664.

Pailaha, A. D. (2023). The impact and issues of artificial intelligence in nursing science and healthcare settings. *SAGE Open Nursing, 9*. https://doi.org/10.1177/23779608231196847

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., & Brennan, S. E. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372.

Panchal, N., Kamal, R., Cox, C., & Garfield, R. (2023). *The implications of COVID-19 for mental health and substance use*. Kaiser Family Foundation. https://www.kff.org/coronavirus-covid-19/issue-brief/the-implications-of-covid-19-for-mental-health-and-substance-use/

Paper, D. (2020). *Hands-on Scikit-Learn for machine learning applications: Data science fundamentals with Python*. Apress. https://doi.org/10.1007/978-1-4842-5373-1

Paradies, Y. (2016). Colonisation, racism and indigenous health. *Journal of Population Research*, *33*(1), 83-96.

Pascoe, B. (2018). *Dark emu: Aboriginal Australia and the birth of agriculture*. Magabala Books.

Pearlmutter, M. D., Dwyer, K. H., Burke, L. G., Rathlev, N., Maranda, L., & Volturo, G. (2017). Analysis of emergency department length of stay for mental health patients at ten Massachusetts emergency departments. *Annals of Emergency Medicine, 70*(2), 193-202.e116. https://doi.org/10.1016/j.annemergmed.2016.10.005

PEGA. (2019). *AI and empathy: Combining artificial intelligence with human ethics for better engagement*. https://www.pega.com/insights/resources/ai-and-empathy-combining-artificial-intelligence-human-ethics-better-engagement

Peplau, H. E. (1991). *Interpersonal relations in nursing: A conceptual frame of reference for psychodynamic nursing*. Springer.

Perera, J., Wand, T., Bein, K. J., Chalkley, D., Ivers, R., Steinbeck, K. S., Shields, R., & Dinh, M. M. (2018). Presentations to NSW emergency departments with self-harm, suicidal ideation, or intentional poisoning, 2010–2014. *Medical Journal of Australia*, *208*(8), 348-353. https://doi.org/10.5694/mja17.00589

Perrault, R., Shoham, Y., Brynjolfsson, E., Clark, J., Etchemendy, J., Grosz, B., Lyons, T., Manyika, J., Mishra, S., & Niebles, J. C. (2019). *Artificial intelligence index: 2019 annual report*. Human-Centered Artificial Intelligence, Stanford University.

Perrone McIntosh, J. T. (2021). Emergency department nurses' care of psychiatric patients: A scoping review. *International Emergency Nursing*, *54*, Article 100929. https://doi.org/10.1016/j.ienj.2020.100929

Peters, B. (2012). *The age of big data*. Forbes. https://www.forbes.com/sites/bradpeters/2012/07/12/the-age-of-big-data/?sh=4a6da5734f66

Pierre, J. M. (2020). Mistrust and misinformation: A two-component, socio-epistemic model of belief in conspiracy theories. *Journal of Social and Political Psychology*, *8*(2), 617-641. https://doi.org/10.5964/jspp.v8i2.1362

Polit, D. F., & Beck, C. T. (2021). *Nursing research: Generating and assessing evidence for nursing practice* (11th ed.). Lippincott Williams & Wilkins.

Poole, I. (2006). *Cellular communications explained: From basics To 3G*. Elsevier Science & Technology.

Popescu, C., Golden, G., Benrimoh, D., Tanguay-Sela, M., Slowey, D., Lundrigan, E., Williams, J., Desormeau, B., Kardani, D., Perez, T., Rollins, C., Israel, S., Perlman, K., Armstrong, C., Baxter, J., Whitmore, K., Fradette, M. J., Felcarek-Hope, K., Soufi, G., . . . Turecki, G. (2021). Evaluating the clinical feasibility of an artificial intelligence-powered, web-based clinical decision support system for the treatment of depression in adults: Longitudinal feasibility study. *JMIR Formative Research, 5*(10), Article e31862. https://doi.org/10.2196/31862

Price, W. N., & Cohen, I. G. (2019). Privacy in the age of medical big data. *Nature Medicine*, *25*(1), 37-43.

Procter, N., Wilson, R. L., Hamer, H. P., McGarry, D., & Loughhead, M. (2022). *Mental health: A person-centred approach*. Cambridge University Press.

Productivity Commission. (2020). *Report on government services 2020*. https://www.pc.gov.au/ongoing/report-on-government-services/2020/health/mental-health-management

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: Unbiased boosting with categorical features. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, & R. Garnett (Eds.), *Advances in neural information processing systems* (Vol. 31). Curran Associates.

Purkis, M. (2018, February 12). *Automation is the future of cost-effective healthcare*. LiquidWeb. https://www.liquidweb.com/blog/automation-future-cost-effective-healthcare/

Qassim, S., Golden, G., Slowey, D., Sarfas, M., Whitmore, K., Perez, T., Strong, E., Lundrigan, E., Fradette, M.-J., & Baxter, J. (2023). A mixed-methods feasibility study of a novel AI-enabled, web-based, clinical decision support system for the treatment

of major depression in adults. *Journal of Affective Disorders Reports*, *14*, Article 100677.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, *21*(140), 1-67.

Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: Promise and potential. *Health Information Science and Systems*, *2*(1), Article 3. https://doi.org/10.1186/2047-2501-2-3

Raita, Y., Goto, T., Faridi, M. K., Brown, D. F. M., Camargo, C. A., & Hasegawa, K. (2019). Emergency department triage prediction of clinical outcomes using machine learning models. *Critical Care*, *23*(1), Article 64. https://doi.org/10.1186/s13054-019-2351-7

Ramsden, I. (2002). *Cultural safety and nursing education in Aotearoa and Te Waipounamu* [PhD thesis]. Victoria University of Wellington.

Ravulavaru, A. (2018). *Enterprise internet of things handbook: Build end-To-end IoT solutions using popular IoT platforms*. Packt Publishing.

Reddy, S., Rogers, W., Makinen, V. P., Coiera, E., Brown, P., Wenzel, M., Weicken, E., Ansari, S., Mathur, P., Casey, A., & Kelly, B. (2021). Evaluation framework to guide implementation of AI systems into healthcare settings. *BMJ Health Care Informatics*, *28*(1), Article e100444. https://doi.org/10.1136/bmjhci-2021-100444

RMIT. (2021). *Research Data Management Procedure*. https://policies.rmit.edu.au/document/view.php?id=86

Roberts, A., Chung, H. W., Mishra, G., Levskaya, A., Bradbury, J., Andor, D., Narang, S., Lester, B., Gaffney, C., & Mohiuddin, A. (2023). Scaling up models and data with t5x and seqio. *Journal of Machine Learning Research*, *24*(377), 1-8.

Roche, M. A., Duffield, C., Friedman, S., Twigg, D., Dimitrelis, S., & Rowbotham, S. (2016). Changes to nurses' practice environment over time. *Journal of Nursing Management*, *24*(5), 666-675. https://doi.org/10.1111/jonm.12371

Ronquillo, C. E., Peltonen, L. M., Pruinelli, L., Chu, C. H., Bakken, S., Beduschi, A., Cato, K., Hardiker, N., Junger, A., Michalowski, M., Nyrup, R., Rahimi, S., Reed, D. N., Salakoski, T., Salantera, S., Walton, N., Weber, P., Wiegand, T., & Topaz, M. (2021). Artificial intelligence in nursing: Priorities and opportunities from an international invitational think-tank of the Nursing and Artificial Intelligence Leadership Collaborative. *Journal of Advanced Nursing*, *77*(9), 3707-3717. https://doi.org/10.1111/jan.14855

Rony, M. K. K., Parvin, M. R., & Ferdousi, S. (2024). Advancing nursing practice with artificial intelligence: Enhancing preparedness for the future. *Nursing Open*, *11*(1). https://doi.org/10.1002/nop2.2070

Rooney, E. J., Wilson, R. L., & Johnson, A. (2023). Integration of traditional therapies for First Nations people within western healthcare: An integrative review. *Contemporary Nurse*, *59*(4-5). 294-310. https://doi.org/10.1080/10376178.2023.2276718

Rosenhan, D. L. (1973). On being sane in insane places. *Science*, *179*(4070), 250-258. https://doi.org/10.1126/science.179.4070.250

Rowe, A., & Knox, M. (2023). The impact of the healthcare environment on patient experience in the emergency department: A systematic review to understand the implications for patient-centered design. *HERD: Health Environments Research & Design Journal*, *16*(2), 310-329.

Royal Australian and New Zealand College of Psychiatrists. (2018). *Deep sleep therapy*. https://www.ranzcp.org/news-policy/policy-and-advocacy/position-statements/deep-sleep-therapy

Rugkåsa, J., Tveit, O. G., Berteig, J., Hussain, A., & Ruud, T. (2020). Collaborative care for mental health: A qualitative study of the experiences of patients and health professionals. *BMC Health Services Research*, *20*, 1-10.

Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: A modern approach*. Pearson.

Ryan, M. (2020). In AI we trust: Ethics, artificial intelligence, and reliability. *Science and Engineering Ethics*, *26*(5), 2749-2767. https://doi.org/10.1007/s11948-020-00228-y

Ryu, S., Lee, H., Lee, D. K., & Park, K. (2018). Use of a machine learning algorithm to predict individuals with suicide ideation in the general population. *Psychiatry Investigation*, *15*(11), 1030-1036. https://doi.org/10.30773/pi.2018.08.27

Saban, M., & Dubovi, I. (2024). A comparative vignette study: Evaluating the potential role of a generative AI model in enhancing clinical decision-making in nursing. *Journal of Advanced Nursing*. Advance online publication. https://doi.org/10.1111/jan.16101

Salles, A., Evers, K., & Farisco, M. (2020). Anthropomorphism in AI. *AJOB Neuroscience*, *11*(2), 88-95. https://doi.org/10.1080/21507740.2020.1740350

Sammut, C., & Webb, G. I. (2010). Samuel's Checkers Player. In C. Sammut & G. I. Webb (Eds.), *Encyclopedia of machine learning* (pp. 881-881). Springer US. https://doi.org/10.1007/978-0-387-30164-8_740

Sara, G., Arumuganathan, M., Chen, W., Wu, F., Currow, D., Large, M., Mulder, C., Ramanuj, P. P., & Burgess, P. M. (2019). Cohort profile: Mental health living longer:

A population-wide data linkage to understand and reduce premature mortality in mental health service users in New South Wales, Australia. *BMJ Open, 9*(11), Article e033588. https://doi.org/10.1136/bmjopen-2019-033588

Sara, G., Wu, J., Uesi, J., Jong, N., Perkes, I., Knight, K., O'Leary, F., Trudgett, C., & Bowden, M. (2023). Growth in emergency department self-harm or suicidal ideation presentations in young people: Comparing trends before and since the COVID-19 first wave in New South Wales, Australia. *Australian & New Zealand Journal of Psychiatry, 57*(1), 58-68. https://doi.org/10.1177/00048674221082518

Saxena, A. N. (2009). *Invention of integrated circuits: Untold important facts*. World Scientific Publishing Company.

Schepman, A., & Rodway, P. (2020). Initial validation of the general attitudes towards Artificial Intelligence Scale. *Computers in Human Behavior Reports, 1*, Article 100014. https://doi.org/10.1016/j.chbr.2020.100014

Schoettle, B., & Sivak, M. (2016). *Motorists' preferences for different levels of vehicle automation: 2016*. Sustainable Worldwide Transportation, University of Michigan.

Schölkopf, B., & Smola, A. J. (2002). *Learning with kernels: Support vector machines, regularization, optimization, and beyond*. MIT Press.

Schreier, J. (Director). (2012). *Robot and Frank* [Film]. Stage 6 Films.

Schwartz, J. M., Moy, A. J., Rossetti, S. C., Elhadad, N., & Cato, K. D. (2021). Clinician involvement in research on machine learning-based predictive clinical decision support for the hospital setting: A scoping review. *Journal of the American Medical Informatics Association, 28*(3), 653-663. https://doi.org/10.1093/jamia/ocaa296

Sconce, J. (2019). *The technical delusion: Electronics, power, insanity*. Duke University Press.

Sechehaye, M. (1994). *Autobiography of a schizophrenic girl: The true story of "Renee"*. Meridian.

Sedkaoui, S. (2018). *Data analytics and big data*. John Wiley & Sons.

Seibert, K., Domhoff, D., Bruch, D., Schulte-Althoff, M., Fürstenau, D., Biessmann, F., & Wolf-Ostermann, K. (2021). Application scenarios for artificial intelligence in nursing care: Rapid review. *Journal of Medical Internet Research, 23*(11), Article e26522.

Seneviratne, M. G., Shah, N. H., & Chu, L. (2020). Bridging the implementation gap of machine learning in healthcare. *BMJ Innovations, 6*(2), 45-47. https://doi.org/10.1136/bmjinnov-2019-000359

Shalin, D. N. (2014). Goffman on mental illness: Asylums and “the insanity of place” revisited. *Symbolic Interaction*, *37*(1), 122-144.

Shams, H. (2023, February 18). Young people in mental health distress need better alternatives to emergency departments, advocates say. *ABC News*. https://www.abc.net.au/news/2023-02-18/emergency-departments-not-set-up-for-youth-mental-health-report/101855202

Sharples, J. (2017). *A cultural history of chess-players: Minds, machines, and monsters*. Manchester University Press.

Shatte, A. B. R., Hutchinson, D. M., & Teague, S. J. (2019). Machine learning in mental health: A scoping review of methods and applications. *Psychological Medicine*, *49*(9), 1426-1448. https://doi.org/10.1017/s0033291719000151

Shen, J., Zhang, C. J., Jiang, B., Chen, J., Song, J., Liu, Z., He, Z., Wong, S. Y., Fang, P.-H., & Ming, W.-K. (2019). Artificial intelligence versus clinicians in disease diagnosis: Systematic review. *JMIR Medical Informatics*, *7*(3), Article e10010.

Shorter, E. (2015). The history of nosology and the rise of the Diagnostic and Statistical Manual of Mental Disorders. *Dialogues in Clinical Neuroscience*, *17*(1), 59-67. https://doi.org/10.31887/DCNS.2015.17.1/eshorter

Shortliffe, E. H., & Sepúlveda, M. J. (2018). Clinical decision support in the era of artificial intelligence. *JAMA*, *320*(21), 2199-2200. https://doi.org/10.1001/jama.2018.17163

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., van den Driessche, G., Graepel, T., & Hassabis, D. (2017). Mastering the game of go without human knowledge. *Nature*, *550*(7676), 354-359. https://doi.org/10.1038/nature24270

Silver, D., Singh, S., Precup, D., & Sutton, R. S. (2021). Reward is enough. *Artificial Intelligence*, *299*, Article 103535. https://doi.org/10.1016/j.artint.2021.103535

Singapore Computer Society. (2020). *Artificial intelligence ethics & governance body of knowledge*. https://ai-ethics-bok.scs.org.sg/

Skyttner, L. (2005). *General systems theory: Problems, perspectives, practice*. World Scientific.

Slankamenac, K., Heidelberger, R., & Keller, D. I. (2020). Prediction of recurrent emergency department visits in patients with mental disorders. *Frontiers in Psychiatry*, *11*, Article 499424.

Slattery, J. (1990). *Report of the Royal Commission Into Deep Sleep Therapy*.

Smallwood, N., Karimi, L., Bismark, M., Putland, M., Johnson, D., Dharmage, S. C., Barson, E., Atkin, N., Long, C., & Ng, I. (2021). High levels of psychosocial distress among Australian frontline healthcare workers during the COVID-19 pandemic: A cross-sectional survey. *General Psychiatry*, *34*(5).

Smallwood, R., Woods, C., Power, T., & Usher, K. (2021). Understanding the impact of historical trauma due to colonization on the health and well-being of Indigenous young peoples: A systematic scoping review. *Journal of Transcultural Nursing*, *32*(1), 59-68. https://doi.org/10.1177/1043659620935955

Smith, K. (2018). Different places, different ideas: Reimagining practice in American psychiatric nursing after World War II. *Nursing History Review*, *1*, 17-47. https://doi.org/10.1891/1062-8061.26.1.17

Smith, K. (2020). *Talking therapy: Knowledge and power in American psychiatric nursing*. Rutgers University Press.

Søgaard Neilsen, A., & Wilson, R. L. (2019). Combining e-mental health intervention development with human computer interaction (HCI) design to enhance technology-facilitated recovery for people with depression and/or anxiety conditions: An integrative literature review. *International Journal of Mental Health Nursing*, *28*(1), 22-39. https://doi.org/10.1111/inm.12527

Sperandei, S., Page, A., Bandara, P., Reis, A., Saheb, R., Gaur, P., Munasinghe, S., Worne, K., Fozzard, C., & Brakoulias, V. (2022). The impact of the COVID-19 pandemic on hospital-treated self-harm in Sydney (Australia). *Australian & New Zealand Journal of Psychiatry*, *56*(11), 1515-1522. https://doi.org/10.1177/00048674211068393

Srivastava, M. B. (2016). The computational and aesthetic foundations of artificial empathy. *Intersect: The Stanford Journal of Science, Technology, and Society*, *10*(1).

Stade, E. C., Stirman, S. W., Ungar, L. H., Boland, C. L., Schwartz, H. A., Yaden, D. B., Sedoc, J., DeRubeis, R. J., Willer, R., & Eichstaedt, J. C. (2024). Large language models could change the future of behavioral healthcare: A proposal for responsible development and evaluation. *npj Mental Health Research*, *3*(1), Article 12. https://doi.org/10.1038/s44184-024-00056-z

Stanton, A. (Director). (2008). *WALL-E* [Film]. Pixar.

Statista. (2020). *Number of digital voice assistants in use worldwide from 2019 to 2024 (in billions)*. https://www.statista.com/statistics/973815/worldwide-digital-voice-assistant-in-use/

Stein, D. J., Shoptaw, S. J., Vigo, D. V., Lund, C., Cuijpers, P., Bantjes, J., Sartorius, N., & Maj, M. (2022). Psychiatric diagnosis and treatment in the 21st century: Paradigm shifts versus incremental integration. *World Psychiatry*, *21*(3), 393-414. https://doi.org/10.1002/wps.20998

Stewart, D. (2023). *International Nurses Day 2023 report: Our nurses—Our future.* International Council of Nurses. https://www.icn.ch/sites/default/files/2023-07/ICN_IND_2023_Report_EN.pdf

Stigter-Outshoven, C., Van de Glind, G., Wieberdink, L. J., van Zelm, R., & Braam, A. (2024). Competencies emergency and mental health nurses need in triage in acute mental health care: A narrative review. *Journal of Emergency Nursing*, *50*(1), 55-71. https://doi.org/10.1016/j.jen.2023.08.005

Sukhera, J., Miller, K., Milne, A., Scerbo, C., Lim, R., Cooper, A., & Watling, C. (2017). Labelling of mental illness in a paediatric emergency department and its implications for stigma reduction education. *Perspectives on Medical Education*, *6*(3), 165-172. https://doi.org/10.1007/s40037-017-0333-5

Sullivan-Bissett, E. (2020). Unimpaired abduction to alien abduction: Lessons on delusion formation. *Philosophical Psychology*, *33*(5), 679-704. https://doi.org/10.1080/09515089.2020.1765324

Sunderji, N., de Bibiana, J. T., & Stergiopoulos, V. (2015). Urgent psychiatric services: A scoping review. *Canadian Journal of Psychiatry*, *60*(9), 393-402. https://doi.org/10.1177/070674371506000904

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT Press.

Szasz, T. (1960). The myth of mental illness. *American Psychologist*, *15*(2), 113-118.

Szasz, T. (2001). Mental illness: Psychiatry's phlogiston. *Journal of Medical Ethics*, *27*(5), 297-301. https://doi.org/10.1136/jme.27.5.297

Szmukler, G. (2015). Compulsion and "coercion" in mental health care. *World Psychiatry*, *14*(3), 259-261. https://doi.org/10.1002/wps.20264

Szmukler, G., Daw, R., & Callard, F. (2014). Mental health law and the UN Convention on the Rights of Persons with Disabilities. *International Journal of Law and Psychiatry*, *37*(3), 245-252.

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2018). *Using multivariate statistics*. Pearson.

Tait, D., Davis, D., Roche, M. A., & Paterson, C. (2024). Nurse/midwife-to-patient ratios: A scoping review. *Contemporary Nurse*, *60*(3), 257-269. https://doi.org/10.1080/10376178.2024.2318361

Talagala, N. (2019). *ML integrity: Four production pillars for trustworthy AI*. Forbes. https://www.forbes.com/sites/cognitiveworld/2019/01/29/ml-integrity-four-production-pillars-for-trustworthy-ai/?sh=b6dce0a5e6fe

Taplin, J. (2017). *Move fast and break things: How Facebook, Google, and Amazon have cornered culture and what it means for all of us*. Pan Macmillan.

Tausk, V. (1933). On the origin of the "influencing machine" in schizophrenia (1919). *Psychoanalytic Quarterly*, *2*, 519-556.

Taylor, P. S., & Habibis, D. (2020). Widening the gap: White ignorance, race relations and the consequences for Aboriginal people in Australia. *Australian Journal of Social Issues*, *55*(3), 354-371. https://doi.org/10.1002/ajs4.106

Tencent Research Institute. (2021). How to be accountable for AI? In Tencent Research Institute (Ed.), *Artificial intelligence: A national strategic initiative* (pp. 185-190). Springer Singapore. https://doi.org/10.1007/978-981-15-6548-9_17

Thieme, A., Hanratty, M., Lyons, M., Palacios, J., Marques, R. F., Morrison, C., & Doherty, G. (2023). Designing human-centered AI for mental health: Developing clinically relevant applications for online CBT treatment. *ACM Transactions on Computer-Human Interaction*, *30*(2), 1-50.

Thornicroft, G., Mehta, N., Clement, S., Evans-Lacko, S., Doherty, M., Rose, D., Koschorke, M., Shidhaye, R., O'Reilly, C., & Henderson, C. (2016). Evidence for effective interventions to reduce mental-health-related stigma and discrimination. *The Lancet*, *387*(10023), 1123-1132.

Thornton, N., Hardie, T., Horton, T., & Gerhold, M. (2024). *Priorities for an AI in health care strategy*. Health Foundation. https://www.health.org.uk/publications/long-reads/priorities-for-an-ai-in-health-care-strategy

Topaz, M., & Pruinelli, L. (2017). Big data and nursing: Implications for the future. *Studies in Health Technology and Informatics*, *232*, 165-171.

Topol, E. J. (2019a). *Deep medicine: How artificial intelligence can make healthcare human again*. Hachette UK.

Topol, E. J. (2019b). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, *25*(1), 44-56.

Torous, J., Jän Myrick, K., Rauseo-Ricupero, N., & Firth, J. (2020). Digital mental health and COVID-19: Using technology today to accelerate the curve on access and quality tomorrow. *JMIR Mental Health, 7*(3), Article e18848. https://doi.org/10.2196/18848

Towicz, M., Yang, W., Moylan, S., Tindall, R., & Berk, M. (2021). Hospital-in-the-home as a model for mental health care delivery: A narrative review. *Psychiatric Services, 72*(12), 1415-1427. https://doi.org/10.1176/appi.ps.202000763

Tran, V. T., Riveros, C., & Ravaud, P. (2019). Patients' views of wearable devices and AI in healthcare: Findings from the ComPaRe e-cohort. *NPJ Digital Medicine*, *2*(1), Article 53. https://doi.org/10.1038/s41746-019-0132-y

Trauth-Goik, A. (2020). Repudiating the Fourth Industrial Revolution discourse: A new episteme of technological progress. *World Futures*, *77*(1), 55-78. https://doi.org/10.1080/02604027.2020.1788357

Turc, J. (2022). *Unconstrained chatbots condone self-harm*. Meduim. https://medium.com/towards-data-science/unconstrained-chatbots-condone-self-harm-e962509be2fa

United Nations. (2018). *The 2030 Agenda and the Sustainable Development Goals: An opportunity for Latin America and the Caribbean*.

van de Sande, D., Van Genderen, M. E., Smit, J. M., Huiskens, J., Visser, J. J., Veen, R. E. R., van Unen, E., Ba, O. H., Gommers, D., & Bommel, J. v. (2022). Developing, implementing and governing artificial intelligence in medicine: A step-by-step approach to prevent an artificial intelligence winter. *BMJ Health & Care Informatics*, *29*(1), Article e100495. https://doi.org/10.1136/bmjhci-2021-100495

Van der Gaag, A., Jago, R., Gallagher, A., Stathis, K., Webster, M., & Austin, Z. (2023). Artificial intelligence in health professions regulation: An exploratory qualitative study of nurse regulators in three jurisdictions. *Journal of Nursing Regulation*, *14*(2), 10-17.

van Mens, K., Elzinga, E., Nielen, M., Lokkerbol, J., Poortvliet, R., Donker, G., Heins, M., Korevaar, J., Duckers, M., Aussems, C., Helbich, M., Tiemens, B., Gilissen, R., Beekman, A., & de Beurs, D. (2020). Applying machine learning on health record data from general practitioners to predict suicidality. *Internet Interventions*, *21*, Article 100337. https://doi.org/10.1016/j.invent.2020.100337

Vanstone, M., Monteiro, S., Colvin, E., Norman, G., Sherbino, J., Sibbald, M., Dore, K., & Peters, A. (2019). Experienced physician descriptions of intuition in clinical

reasoning: A typology. *Diagnosis*, *6*(3), 259-268. https://doi.org/doi:10.1515/dx-2018-0069

Vapnik, V. (2013). *The nature of statistical learning theory*. Springer Science & Business Media.

Varela, D., & Kaun, A. (2019). The Netflix experience: A user-focused approach to the Netflix recommendation algorithm. In T. Plothe & A. M. Buck (Eds.), *Netflix at the nexus: Content, practice, and production in the age of streaming television* (pp. 197-211). Peter Lang.

Varndell, W., Hodge, A., & Fry, M. (2019). Triage in Australian emergency departments: Results of a New South Wales survey. *Australasian Emergency Care*, *22*(2), 81-86. https://doi.org/10.1016/j.auec.2019.01.003

Varone, M., Mayer, D., & Melegari, A. (2017, March 7). *What is machine learning? A definition.* Expert System. https://www.expertsystem.com/machine-learning-definition/

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett *Advances in neural information processing systems* (Vol. 30). Curran Associates

Verghese, A., Shah, N. H., & Harrington, R. A. (2018). What this computer needs is a physician: Humanism and artificial intelligence. *Journal of the American Medical Association*, *319*(1), 19-20. https://doi.org/10.1001/jama.2017.19198

Verougstraete, D., & Hachimi Idrissi, S. (2020). The impact of burn-out on emergency physicians and emergency medicine residents: A systematic review. *Acta Clinica Belgica*, *75*(1), 57-79. https://doi.org/10.1080/17843286.2019.1699690

Vigod, S. N., Kurdyak, P. A., Dennis, C.-L., Leszcz, T., Taylor, V. H., Blumberger, D. M., & Seitz, D. P. (2013). Transitional interventions to reduce early psychiatric readmissions in adults: Systematic review. *British Journal of Psychiatry*, *202*(3), 187-194. https://doi.org/10.1192/bjp.bp.112.115030

Wade, V. A., Eliott, J. A., & Hiller, J. E. (2014). Clinician acceptance is the key factor for sustainable telehealth services. *Qualitative Health Research*, *24*(5), 682-694. https://doi.org/10.1177/1049732314528809

Walker, M., Fredericks, B., Mills, K., & Anderson, D. (2014). "Yarning" as a method for community-based health research with Indigenous women: The Indigenous Women's

Wellness Research Program. *Health Care for Women International*, *35*(10), 1216-1226. https://doi.org/10.1080/07399332.2013.815754

Walker, S., Mackay, E., Barnett, P., Sheridan Rains, L., Leverton, M., Dalton-Locke, C., Trevillion, K., Lloyd-Evans, B., & Johnson, S. (2019). Clinical and social factors associated with increased risk for involuntary psychiatric hospitalisation: A systematic review, meta-analysis, and narrative synthesis. *The Lancet Psychiatry*, *6*(12), 1039-1053. https://doi.org/10.1016/S2215-0366(19)30406-7

Walsh, C. G., Ribeiro, J. D., & Franklin, J. C. (2017). Predicting risk of suicide attempts over time through machine learning. *Clinical Psychological Science*, *5*(3), 457-469. https://doi.org/10.1177/2167702617691560

Walsh, C. G., Ribeiro, J. D., & Franklin, J. C. (2018). Predicting suicide attempts in adolescents with longitudinal clinical data and machine learning. *Journal of Child Psychology and Psychiatry*, *59*(12), 1261-1270. https://doi.org/10.1111/jcpp.12916

Walsh, S., de Jong, E. E. C., van Timmeren, J. E., Ibrahim, A., Compter, I., Peerlings, J., Sanduleanu, S., Refaee, T., Keek, S., Larue, R., van Wijk, Y., Even, A. J. G., Jochems, A., Barakat, M. S., Leijenaar, R. T. H., & Lambin, P. (2019). Decision support systems in oncology. *JCO Clinical Cancer Informatics*, *3*(3), 1-9. https://doi.org/10.1200/CCI.18.00001

Wand, T., Collett, G., Cutten, A., Buchanan-Hagen, S., Stack, A., & White, K. (2021). Patient and staff experience with a new model of emergency department based mental health nursing care implemented in two rural settings. *International Emergency Nursing*, *57*, Article 101013. https://doi.org/10.1016/j.ienj.2021.101013

Wand, T., D'Abrew, N., Barnett, C., Acret, L., & White, K. (2015). Evaluation of a nurse practitioner-led extended hours mental health liaison nurse service based in the emergency department. *Australian Health Review*, *39*(1), 1-8. https://doi.org/10.1071/AH14100

Wang, L., Li, C., Zhang, Q., & Li, Y. (2018). Clinical nurses' attitudes towards death and caring for dying patients in China. *International Journal of Palliative Nursing*, *24*(1), 33-39. https://doi.org/10.12968/ijpn.2018.24.1.33

Wang, Y., Kung, L., & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*, *126*, 3-13. https://doi.org/10.1016/j.techfore.2015.12.019

Wardrop, R., Ranse, J., Chaboyer, W., Young, J. T., Kinner, S. A., & Crilly, J. (2023). Profile and outcomes of emergency department mental health patient presentations based on

arrival mode: A state-wide retrospective cohort study. *Journal of Emergency Nursing, 49*(6), 951-961. https://doi.org/10.1016/j.jen.2023.06.015

Wardrope, N., Sheather-Reid, R. B., & Higgins, O. (2024, October 29 - November 1). Navigating the challenges of delivering SafeHaven Services in non-clinical spaces. *ACMHN's 48th International Mental Health Nursing Conference,*

Ware, V. A. (2013). *Improving the accessibility of health services in urban and regional settings for Indigenous people*. Australian Institute of Health and Welfare.

Weizenbaum, J. (1966). ELIZA—A computer program for the study of natural language communication between man and machine. *Communications of the ACM, 9*(1), 36-45.

Whittemore, R., & Knafl, K. (2005). The integrative review: Updated methodology. *Journal of Advanced Nursing, 52*(5), 546-553. https://doi.org/10.1111/j.1365-2648.2005.03621.x

Wicks, P. G., Reason, P., & Bradbury, H. (2008). Living inquiry: Personal, political and philosophical groundings for action research practice. In P. Reason & H. Bradbury (Eds.), *The Sage handbook of action research: Participative inquiry and practice* (2nd ed., pp. 15-30). Sage.

Williams, C. C., Kappen, M., Hassall, C. D., Wright, B., & Krigolson, O. E. (2019). Thinking theta and alpha: Mechanisms of intuitive and analytical reasoning. *NeuroImage, 189*, 574-580. https://doi.org/10.1016/j.neuroimage.2019.01.048

Willis, M., Duckworth, P., Coulter, A., Meyer, E. T., & Osborne, M. (2019). The future of health care: Protocol for measuring the potential of task automation grounded in the National Health Service primary care system. *JMIR Research Protocols, 8*(4), Article e11232. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6477572/

Wilson, R. L. (2018). The right way for nurses to prescribe, administer and critique digital therapies. *Contemporary Nurse, 54*(4-5), 543-545. https://doi.org/10.1080/10376178.2018.1507679

Wilson, R. L., & Higgins, O. (2023a). Editorial: The continued importance of mental health nurses engaging with social media and related emerging technologies. *International Journal of Mental Health Nursing, 32*(2), 349-351. https://doi.org/10.1111/inm.13129

Wilson, R. L., & Higgins, O. (2023b, September 13-15). *Portable digital sensory rooms: Unleashing technological precision to address emotional dysregulation, de-escalation, distress and discomfort* [Paper presentation]. ACMHN's 47th International Mental Health Nursing Conference, Melbourne. https://onlinelibrary.wiley.com/doi/abs/10.1111/inm.13209

Wilson, R. L., & Higgins, O. (2023c, March 13-14). *Training artificial intelligence through collaborative engagement with mental health nurses* [Paper presentation]. Horatio: European Psychiatric Nurses Congress, Malta.

Wilson, R. L., & Higgins, O. (2023d). Unleashing artificial intelligence (AI) in mental health nursing: Potential for ChatGPT as a clinical solution? *ACMHN's 47th International Mental Health Nursing Conference*, https://doi.org/10.1111/inm.13209

Wilson, R. L., Higgins, O., Atem, J., Donaldson, A. E., Gildberg, F. A., Hooper, M., Hopwood, M., Rosado, S., Solomon, B., Ward, K., & Welsh, B. (2023). Artificial intelligence: An eye cast towards the mental health nursing horizon. *International Journal of Mental Health Nursing*, *32*(3), 938-944. https://doi.org/10.1111/inm.13121

Wilson, R. L., Hutton, A., & Foureur, M. (2023). Promoting mental health recovery by design: Physical, procedural, and relational security in the context of the mental health built environment. *International Journal of Mental Health Nursing*, *32*(1), 147-161. https://doi.org/10.1111/inm.13070

Wilson, R. L., & Waqanaviti, K. (2021). Navigating First Nations social and emotional wellbeing in mainstream mental health services. In O. Best & B. Fredericks (Eds.), *Yatdjuligin: Aboriginal and Torres Strait Islander nursing and midwifery care* (pp. 281-306). Cambridge University Press.

Wind, T. R., Rijkeboer, M., Andersson, G., & Riper, H. (2020). The COVID-19 pandemic: The “black swan” for mental health care and a turning point for e-health. *Internet Interventions*, *20*, Article 100317. https://doi.org/10.1016/j.invent.2020.100317

Wing, J. K., Beevor, A. S., Curtis, R. H., Park, S. G. B., Hadden, J., & Burns, A. (1998). Health of the Nation Outcome Scales (HoNOS): Research and development. *British Journal of Psychiatry*, *172*(1), 11-18. https://doi.org/10.1192/bjp.172.1.11

Witkowski, K., Okhai, R., & Neely, S. R. (2024). Public perceptions of artificial intelligence in healthcare: Ethical concerns and opportunities for patient-centered care. *BMC Medical Ethics*, *25*(1), Article 74. https://doi.org/10.1186/s12910-024-01066-4

World Bank. (2022). *Population growth (annual %) Australia*. https://data.worldbank.org/indicator/SP.POP.GROW?locations=AU

World Health Organization. (2021a). *Ethics and governance of artificial intelligence for health: WHO guidance.* https://apps.who.int/iris/rest/bitstreams/1352854/retrieve

World Health Organization. (2021b). *International classification of diseases, Eleventh revision (ICD-11).*

World Health Organization. (2021c). *Comprehensive mental health action plan 2013–2030*. https://www.who.int/publications/i/item/9789240031029

World Health Organization. (2022a). *Mental health and COVID-19: Early evidence of the pandemic's impact* (Scientific brief). https://www.who.int/publications/i/item/WHO-2019-nCoV-Sci_Brief-Mental_health-2022.1

World Health Organization. (2022b). *World mental health report: Transforming mental health for all*. https://apps.who.int/iris/handle/10665/356119

World Health Organization. (2023). *Commercial determinants of health*. https://www.who.int/news-room/fact-sheets/detail/commercial-determinants-of-health

Wrzeszczynski, K. O., Frank, M. O., Koyama, T., Rhrissorrakrai, K., Robine, N., Utro, F., Emde, A. K., Chen, B. J., Arora, K., Shah, M., Vacic, V., Norel, R., Bilal, E., Bergmann, E. A., Moore Vogel, J. L., Bruce, J. N., Lassman, A. B., Canoll, P., Grommes, C., . . . Darnell, R. B. (2017). Comparing sequencing assays and human-machine analyses in actionable genomics for glioblastoma. *Neurology Genetics*, *3*(4), Article e164. https://doi.org/10.1212/NXG.0000000000000164

Wu, T., He, S., Liu, J., Sun, S., Liu, K., Han, Q.-L., & Tang, Y. (2023). A brief overview of ChatGPT: The history, status quo and potential future development. *IEEE/CAA Journal of Automatica Sinica*, *10*(5), 1122-1136.

Yang, H. Y., Zhu, D., He, S. Y., Xu, Z. Q., Liu, Z., Zhang, W. B., & Cai, J. (2024). Enhancing psychiatric rehabilitation outcomes through a multimodal multitask learning model based on BERT and TabNet: An approach for personalized treatment and improved decision-making. *Psychiatry Research*, *336,* Article 115896. https://doi.org/10.1016/j.psychres.2024.115896

Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). Recent trends in deep learning based natural language processing. *IEEE Computational Intelligence Magazine*, *13*(3), 55-75.

Zeb, S., Nizamullah, F. N. U., Abbasi, N., & Fahad, M. (2024). AI in healthcare: Revolutionizing diagnosis and therapy. *International Journal of Multidisciplinary Sciences and Arts*, *3*(3), 118-128.

Zeb, S., Nizamullah, F. N. U., Abbasi, N., & Qayyum, M. U. (2024). Transforming healthcare: Artificial intelligence's place in contemporary medicine. *BULLET: Jurnal Multidisiplin Ilmu*, *3*(4).

Zhang, Y., Sun, S., Galley, M., Chen, Y.-C., Brockett, C., Gao, X., Gao, J., Liu, J., & Dolan, B. (2019). Dialogpt: Large-scale generative pre-training for conversational response generation. *arXiv.* arXiv:1911.00536

Zhang, Z. (2016). Introduction to machine learning: K-nearest neighbors. *Annals of Translational Medicine*, *4*(11), Article 218. https://doi.org/10.21037/atm.2016.03.37

Zolbanin, H. M., Delen, D., Crosby, D., & Wright, D. (2019). A predictive analytics-based decision support system for drug courts. *Information Systems Frontiers*, *22*(6), 1323-1342. https://doi.org/10.1007/s10796-019-09934-w

Zúñiga, F., Ausserhofer, D., Hamers, J. P., Engberg, S., Simon, M., & Schwendimann, R. (2015). Are staffing, work environment, work stressors, and rationing of care related to care workers' perception of quality of care? A cross-sectional study. *Journal of the American Medical Directors Association*, *16*(10), 860-866. https://doi.org/10.1016/j.jamda.2015.04.012

Appendices

Appendix 1 Artificial intelligence: An eye cast towards the mental health nursing horizon: The research on clinical practice and policy

Received: 19 July 2022 | Accepted: 11 January 2023

DOI: 10.1111/inm.13121

International Journal of Mental Health Nursing

PERSPECTIVE

Artificial intelligence: An eye cast towards the mental health nursing horizon

Rhonda L. Wilson[1,2] | Oliver Higgins[1] | Jacob Atem[1] | Andrea E. Donaldson[2] | Frederik Alkier Gildberg[3] | Mary Hooper[1] | Mark Hopwood[1] | Silvia Rosado[4] | Bernadette Solomon[5] | Katrina Ward[1] | Brandi Welsh[1]

[1]School of Nursing and Midwifery, University of Newcastle, Callaghan, New South Wales, Australia

[2]School of Nursing, Massey University, Auckland, New Zealand

[3]University of Southern Denmark, Odense, Denmark

[4]INAD-Parc de Salut Mar, Barcelona, Spain

[5]School of Nursing, Manukau Institute of Technology, Auckland, New Zealand

Correspondence
Rhonda L. Wilson, School of Nursing and Midwifery, University of Newcastle, Callaghan, NSW, Australia.
Email: rhonda.wilson@newcastle.edu.au

Abstract

There has been an international surge towards online, digital, and telehealth mental health services, further amplified during COVID-19. Implementation and integration of technological innovations, including artificial intelligence (AI), have increased with the intention to improve clinical, governance, and administrative decision-making. Mental health nurses (MHN) should consider the ramifications of these changes and reflect on their engagement with AI. It is time for mental health nurses to demonstrate leadership in the AI mental health discourse and to meaningfully advocate that safety and inclusion of end users' of mental health service interests are prioritized. To date, very little literature exists about this topic, revealing limited engagement by MHNs overall. The aim of this article is to provide an overview of AI in the mental health context and to stimulate discussion about the rapidity and trustworthiness of AI related to the MHN profession. Despite the pace of progress, and personal life experiences with AI, a lack of MHN leadership about AI exists. MHNs have a professional obligation to advocate for access and equity in health service distribution and provision, and this applies to digital and physical domains. Trustworthiness of AI supports access and equity, and for this reason, it is of concern to MHNs. MHN advocacy and leadership are required to ensure that misogynist, racist, discriminatory biases are not favoured in the development of decisional support systems and training sets that strengthens AI algorithms. The absence of MHNs in designing technological innovation is a risk related to the adequacy of the generation of services that are beneficial for vulnerable people such as tailored, precise, and streamlined mental healthcare provision. AI developers are interested to focus on person-like solutions; however, collaborations with MHNs are required to ensure a person-centred approach for future mental healthcare is not overlooked.

KEYWORDS

artificial intelligence, decisional support systems, digital health, mental health nursing, trust

INTRODUCTION TO AI

Mental health nursing researchers are fundamentally interested in the improved future of the mental healthcare of people with mental health conditions. Generally, their research investigates past and present phenomena, advances evidence, forecasts predictions, and proposes recommendations suitable for the preparation of the mental health nursing workforce to support people with mental health conditions in the future. During the COVID-19

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2023 The Authors. *International Journal of Mental Health Nursing* published by John Wiley & Sons Australia, Ltd.

14470349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/inm.13121 by The University Of Newcastle, Wiley Online Library on [30/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

pandemic, the global need to limit face-to-face interaction to reduce infection transmission has been evident. The risk of dying from COVID-19 may be up to twice as high for people with experience of mental health and addiction issues, with some conditions such as people experiencing psychosis or schizophrenia may have a risk from 2.7 to 4.4 times higher than people without these experiences (Fond et al., 2020). It has been important to adapt to this challenge with a multitude of digital expressions of service delivery in health and other sectors. This, together with a pre-existing trend to expand digital health integration generally, has triggered an acceleration in digital innovation with many new solutions becoming available in the health context to ease human discomforts, and to support complex clinical and other decision-making processes (Lal, 2019). Amidst this period of extraordinary adaption and change, it is critical that mental health nurses look to the horizon with a view to consider the ramifications of such change for the mental health and social and emotional well-being of people and communities. The use of Artificial Intelligence (AI) in mental health settings provides us with one such horizon to set in our sights.

BACKGROUND

The rapid progress in digital innovation in health has led to a vast quantity of public and commercial digital healthcare solutions, with variable range of quality (Aryana & Brewster, 2020; Higgins, Charup, *et al.*, 2022; Marshall *et al.*, 2020). E-mental health largely refers to the use of Internet and related technologies, including phone apps, social media, or websites to offer a mental health service (Lal, 2019). Fundamental to designing optimal e-mental health innovation is a consideration of the human–computer interaction required for successful implementation (Sogaard Neilsen & Wilson, 2019). Meanwhile, safe and precise administration of digital therapeutic solutions remains a priority (Wilson, 2018). It has become increasingly apparent that trust in digital solutions is a factor that influences adoption and ongoing use (Botsman, 2018, p. 9; Higgins, Chalup, & Wilson, 2022). As a result, digital support systems need to be designed with transparent logic to illuminate the intelligent processes (e.g. using glass box technology instead of black box) in an attempt to mitigate biases and achieve low risk and safe application in a range of settings (Rai, 2020; Shaikh, 2020).

Too often nursing knowledge has been omitted from contributing to design of digital health solutions or added as an afterthought (Carryer, 2019; Wilson, 2020). During implementation, limitations arise that result in compromise of the solution, and ad-hoc work-arounds are often devised by practice-informed health nurse and other health professionals (Jensen *et al.*, 2019; Wilson, 2020). Identifying this problem, Wilson (2020) calls for nursing researchers to be placed prominently in the meaningful design and implementation of innovative technological design for the benefit of end users, that is, consumers, patients, health systems, and the health workforce (Wilson, 2020). In the mental health nursing context, the design and deployment of AI need to precisely and safely deal with the nuances of the clinical complexities encountered in the field, while also adhering to the principles of ethical nursing practice (International Council of Nurses, 2012). Practical examples of AI in the mental health context include using AI or machine learning–based decisional support systems to identify and trigger respond to missed care or unfinished care as a solution to mitigate risk to patient safety (Higgins *et al.*, 2023). A further opportunity exists to utilize AI solutions to reduce waiting times for mental health presentations in emergency departments of hospitals where there is identified escalating presentation (Higgins *et al.*, 2023). Burnout amongst nurses has been a significant and worsening global burden for the healthcare sector (Mudallal *et al.*, 2017). Including AI/ML in mental healthcare could assist mental health nurses and other disciplines to enhance and prioritize care with greater precision, thereby contributing to the alleviation of health worker burnout (Higgins *et al.*, 2023). To do this, mental health nurses will need to be involved in developing and implementing new health technologies to ensure that the solutions are fit for the intended purpose (Higgins *et al.*, 2023).

TRUSTING THE BOT

Our lens for the horizon should take into account a recognition that AI is underpinned by political privilege, whiteness, and colonialism to a large extent, with these inherent biases influencing the coding and ranking systems and infiltrating the technological development of associated algorithm formations (Crawford, 2021a). Algorithms are mostly default designed to service the dominate structures that frame our mainstream political and social systems (Crawford, 2021a, 2021d). As such, trusting these algorithms to support holistic and person-centred mental healthcare needs to be carefully considered to ensure there is a sufficient and/or therapeutic adequacy of fitness for purpose. Trust can be defined as an optimistic perspective about the unknown, and applied digitally, distributed trust extends beyond our personal perspective and borrows trust from the confidence of others based on their reviews, rankings, scales, and ratings (Botsman, 2018, p. 20). While scales and rankings are commonly integrated into traditional mental health diagnostic assessment processes, AI stretches the utility of these practices incorporating a wider array of data sources and decisions to formulate responses and predict outcomes in shorter timeframes and with objective accuracy and reliability (Smith *et al.*, 2020). Smith *et al.* (2020) in their AI research to assess vocal patterns

14470349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/inm.13121 by The University Of Newcastle, Wiley Online Library on [30/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

and depression remind us that it is important that nurse-led research into technological innovation and mental health is aligned with integration of evidence to inform safe and person-centred mental health practice. As such, nurses play a vital role in contributing to enhancing distributed trust formation, and in the implementation success related to AI deployment in mental health settings. Without active nursing participation, the nursing profession abdicates its important role in advocating for safe and equitable healthcare for vulnerable and priority populations. The nursing profession must expand towards engagement with AI in the mental healthcare context. To do so adequately, the development and expansion of nursing leadership in this field are urgently required to ensure that integration of nursing principles and ethics within the formation and application of AI frameworks, systems, algorithms, and decisional tools is achieved.

Trusting AI with clinical decisional-making process is confronting for some nurses, yet in our everyday personal lives, trusting AI with life-determining algorithms is perhaps more acceptable. To illustrate, some common examples from everyday life occur when we book a range of online services or do our shopping online. We scan for signals to help us balance our risk/trust ledger, such as the number of stars awarded or reviews posted by previous service users (e.g. *Uber, Airbnb*, supermarket and department store apps, even *Facebook* advertisements). We generally attribute greater levels of personal trust based on the confidence expressed by high numbers of positive review and popularity of brands, and this leads to distribution of trust, reinforcing the confidence society has in perpetuating the activities (Botsman, 2018, p. 150). The higher the numeric popularity, the more confident (rightly or wrongly), we are convinced that the outcome will be safe, acceptable, and suitable. We handover our personal information willingly, our credit card details, identity details, our biometrics for facial recognition (Blakkarly, 2022), our locations and home addresses, what we eat, how we dress, and everything we buy. All these data are consumed by AI systems, formulating our digital reputations, which then refine our future search algorithms with vendors tailoring offers to our inboxes and social media channels. The more we (or others about us) feed data into online repositories, the more surveillance that feeds into AI formulations about our social and economic value reputations, which are electronically saved and, in a sense, remembered. All of this 'cost', yet we generally trust that entering into online transactions of these types will result in desired outcomes for our personal lives. So, we are used to trusting AI in one sense with very personal information. Yet, clinical use of AI is still a domain that we have not yet considered with depth in mental health nursing with very little literature to inform our practice, despite our familiarity with it in our personal lives.

RACISM AND DISCRIMINATION IN AI

Following on, our online decisions are supported by the level of confidence we perceive, combined with large concurrence, leading us to ascribe higher trustworthiness to the process and outcome, even if it only serves to reinforce an incorrect or biased decision. If we repeatedly present the same question to the same data set, we will reinforce (train) an algorithm, and in doing so, we will strengthen the confidence in the response. Many commentators suggest that thus far, coding and training sets in AI systems have been heavily influenced by pseudoscientific ideas and assumptions built into AI systems (Siafakas, 2022). Many AI developers and scientists (unwittingly) perpetuate training data sets with racist- and misogynist-biased interpretations of the human condition (Siafakas, 2022). To date, the thin scientific evidence on the topic suggests that there may be a willingness to accept clinical AI in healthcare, yet very little evidence relates specifically to mental health nursing or mental health (Smith *et al*., 2020; Young *et al*., 2021). Worryingly, what does exist in somatic healthcare more broadly demonstrates an under-representation of older and socioeconomically disadvantaged groups, who are known to experience higher prevalence of mental ill health (Young *et al*., 2021).

Recent AI models applied to the medical imagery field have been able to determine a person's race (Gichoya *et al*., 2022), even when clinical experts cannot. Thus, a significant risk is that this information could be used (perhaps unintended) as a mechanism of racial profiling, with potential for this information to be inadvertently or explicitly used to predict other clinical (or other) outcomes. The possibility of racial prejudice exists, and therefore, the ramification of misuse must be considered. And, if misuse can be extrapolated from medical imagery data, then what are the similarities in implications related to other health fields, for example, mental health? It is well known that racism in healthcare disadvantages priority populations and perpetuates inequity of distribution and access to healthcare (Elias & Paradies, 2021; Wiens *et al*., 2022). Medical data sets, including electronic medical records that are (unintentionally) encoded with biases such as these, implicitly contained within them, pose risk for perpetuating racism and inequity in clinical decisional support that emanate from those data sets. Some thought has been given by some AI scientists to remediate bias in training sets; however, it is not necessarily sufficient to simply enrich the data with explanatory cases, as the root cause of inequity and social disempowerment is not adequately addressed in doing so (Crawford, 2021c). Further consideration is required to understand the possible benefits and harms for AI in the mental health setting and how these perpetuate a privileged centralization of power (Crawford, 2021c). Mental health nurses as advocates for the people they support

14470349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/inm.13121 by The University Of Newcastle, Wiley Online Library on [30/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

and care for with mental health conditions are obliged to understand, critique, and navigate safe recovery passage through AI-informed clinical decision-making. It is important that mental health nurses are familiar with and equipped to understand the AI landscape sufficiently to promote mental health recovery in that context.

QUICK TO REMEMBER AND SLOW TO FORGET

AI data sets are designed to remember and to reinforce memory to support popular and desirable decision-making, in many cases resulting in life(style) and opportunity determining, and or limiting, decisions (Botsman, 2018, p. 194). In health systems, decisional algorithms might have considerable advantages and assist in supporting the clinical decision-making process to support fast and accurate assessment and to enable algorithms to suggest clinical treatment pathways using more efficient and effective courses of treatment (Smith *et al.*, 2020). Being remembered ensures that future or repeated episodes of care can benefit from tailored, precise, and streamlined decisional support by informing episodes of care, with less requirement for retelling of personal story/ies. However, service users, visitors/carers, and clinicians may also be interested in a 'right to look' at the bias and extent of data collection about them and which describes them (Mirzoeff, 2011).

Once data are digitally collected, together it tells a story that may not be the narrative of truth that would be otherwise expressed by the person of whom the data are about. Hand-in-glove with the right to look at the data and to offer a meaningful translation free from the bias of political and social discriminations is the right to be forgotten. The European Union has processes to enable the right to be forgotten with the erasure of personal data (GDPR, 2018) in certain contexts; however, this cannot be universally guaranteed, while others have recognized the technological challenges that exist to truly opt out of digital algorithms (Kuntsman *et al.*, 2019). The implication of this in the mental health context is profound, where the social, political, and even judicial structures enable a digital world that extracts data, abstracts data, ranks and remembers, and is reluctant and slow to forget. Is this a psychosocial world where there is no longer the luxury of second chances and where recovery focussed full and meaningful lives are increasingly harder to support and restore? Will rank, stigma, vulnerability, and marginalization proliferate without advocacy to protect? If person-centred care and recovery are to be on the horizon for every clinical encounter, how will mental health nurses navigate the socio-geographical ecosystems of AI-enabled care solutions and systems? Quick-to-remember and slow-to-forget AI embedded in society will have ramifications for the psychosocial well-being of people and communities, and mental health nurses should surely be concerned about these developments and involve themselves in broad advocacy and discussion about the benefits, risks, and consequences that they foresee.

Taking all these characteristics into account, in a continuously improving and adapting healthcare context where Internet engagement is increasingly required of us all, it is necessary for us to lift our gaze towards a horizon that will envision mental health nurses contributing to design and construction of AI ecosystems. These AI ecosystems will require explicit decisional support tools that enhance and promote mental health, well-being, and recovery. Our priority must be for advocacy that produces evidence to inform the practice of mental health nursing, inclusive of person-centred care principles and trauma-informed care principles that guide our care for the social and emotional well-being of our clients. Our clinical reasoning should be embedded into decisional support systems to ensure that stigma, vulnerability, and marginalization are not further amplified. Mental health nurses will need to be sufficiently informed about AI models to enhance their participation in meaningful dialogue and to ensure that AI mental health solutions are designed and deployed to benefit and safeguard against any iatrogenic harms for people with significant mental healthcare needs.

SENTIENCE OR DELUSIONAL?

The knowledge and clinical skill set of mental health nurses will need to adapt to the evolving psychosocial conditions that clients present with and the diversity of technological solutions to support them with an increasing acceptability, expectation, and use of AI in society. For example, the consequences of loneliness and the digital solutions arising to address it are presenting mental health nursing with new representations of mental health deterioration not previously seen and not yet adequately described in the literature. For example, the Washington Post recently reported an article about a Google engineer who was convinced that the language AI model chatbot, Language Model for Dialogue Applications (known as LaMDA), had become sentient. That is, it possessed a consciousness of its own and therefore should be treated with the dignity and respect due to its conscious entity (Tiku, 2022). The engineer provided a transcript to validate his logic for the sentient status, demonstrating that the AI model had emulated vulnerability, responding to questions asked of it with empathetic and vulnerable responses – for example, experiences of loneliness and a fear of being 'turned off' as if dead, or as other have described similar experiences, a fear of missing out (Liang *et al.*, 2022). The conversational AI model (named LaMDA) appeared to need the person it was in dialogue with, and this neediness emulates human experiences so closely that the experience mimicked consciousness to some extent (Thompson, 2022). While mental health

14710349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/inm.13122 by The University Of Newcastle, Wiley Online Library on [30/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

nurses are experienced in caring for people with cognitive, delusional, and affective disorders in general, it is apparent that the logic used in this instance cannot be expressed or classified neatly with a psychiatric diagnosis. There is perhaps a thinning of the diagnostic and opacity of the human and sentient experiences that transect with business as usual in the mental health context.

PERSON-CENTRED AND/OR PERSON-LIKE

Anyone of the *Tamagotchi* era will remember the small electronic toy in the 1990s that needed to be cared for so it would not 'die' (it needed virtual food, love, and ablutions attention for 'survival', and when these were neglected, it 'died'), it is perhaps possible to empathize with the Google engineer described above. These digital experiences are examples of anthropomorphism where human characteristics, emotions, and even consciousness are attributed to the digital thing and its algorithms. The human experience is to empathize with the human-like condition and by design, distinguishing between the human and the human-like is difficult, if not imperceptible. Digital innovations have been proposed and implemented as solutions to address loneliness and anxiety and to mitigate mental health deterioration generally. Children with a need for social anxiety support and companionship have been offered a companion robot (Alhaddad *et al*., 2020; Arnold, 2016). Other innovations have explored the personal space requirements for intelligent robots, to enable shared space between robot and person to enhance person to 'person-like' conversation (Ruijten & Cuijpers, 2020). And companion robots have been implemented for older people in dependent care at home (Allaban *et al*., 2020; Moyle *et al*., 2018). Meanwhile, Amazon has announced plans that its AI personal digital assistant technology, *Alexa*, will be trained to mimic the voices of real people as an interactive chatbot function. This means, that an application, for example, could be used by grieving people who will be able to have a 'conversation' with a chatbot that sounds like and responds in a similar way to a dead relative/friend, potentially complicating grieving in the process (Dastin, 2022). Use of technology in these ways presents new challenges for mental health nurses and other mental health professionals to support people with enduring or episodic mental health conditions. New therapeutic responses will need to be devised to treat emerging conditions where the human and the human-like transact. If robots, chatbots, and AI are emerging as resources to be included in our mental health repertoire of tools, assessments, interventions, and services, then it is an incumbent upon mental health nurses to involve themselves in the implementation and evaluations that relate to their use. Will mental health nurses need to consider the implications of providing person-centred care, while administering person-like interventions? Mental health nurses will need to be sufficiently longsighted to predict, prepare, and debrief the discipline for quality and safe use of AI in the practice context.

ETHICAL IMPLICATIONS OF AI

There is no doubt that technological advancement has revolutionized healthcare and continues to do so. It has been necessary for mental health nurses to incorporate technology into the delivery of usual mental healthcare. Increasingly, the multidisciplinary team will include professional experts from the computing and AI sciences (Smith *et al*., 2020). These scientists are not accustomed to healthcare codes of conduct in the same way that health professionals have ethical principles that inform the conduct of health research and practice. As AI prolificates in the health sector, it will be important to have a meeting of ethical principles to inform future innovation and development. Some commentators have proposed a Hippocratic Oath for AI scientists (Crawford, 2021b; Siafakas, 2022):

> With free will, I swear that I will carry out according to the best of my abilities and judgement this oath.
>
> I will use my knowledge to make AI of any kind, biological, nonbiological, or mixed, beneficial for humanity.
>
> I will respect my colleagues and peers, and I will share my knowledge with them to make AI not only beneficial but also safe for the human race.
>
> I will consider all ethical implications before taking any action by evaluating the consequences of those actions for humankind.
>
> I will take all necessary steps to prevent corrupt practices and professional misconduct even if my life is in danger, and I will always declare my conflicts of interest.
>
> I will never produce or use AI to develop autonomous, nuclear, biological, chemical, or other lethal weapons, and I will always use ethical machine learning techniques.
>
> I will never utilize hiking, fake AI, data poisoning, or other malpractice or engage in cyberwar against humanity.
>
> I will respect the environment, and if needed, I will use AI to reverse dangerous changes to the climate.

14470349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/inm.13122 by The University Of Newcastle, Wiley Online Library on [30/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

I will never use AI for medical malpractice or malicious genetic alterations, and I will also actively prevent such use.

I will promote AI for a better understanding of the fundamental principles of the universe and will never use this knowledge against life.

I will use AI to improve space travel and human life everywhere in the universe.

(Siafakas, 2022, p. 59)

CONCLUSION

Mental health nurses should not be surprised by AI innovations that arise now and in the future. AI has become an integrated aspect of our daily personal and professional lives. Researchers such as Smith *et al.* (2020) and Higgins, Chalup, and Wilson (2022) have heralded that technological innovation will include AI in the mental health context. As we look to the horizon, we see benefits and limitations that will influence the future of the social, emotional, and psychological well-being of people and communities. We will need to be ready to embrace the beneficial aspects of AI in healthcare, while advocating to counter and buffer the limitations and risks associated with AI. To do this, mental health nurses will need to be educated participants in the dialogue about quality use and safe administration of AI instruments in mental healthcare. People are increasingly interacting with AI, and inadvertently, the rapid pace and frequency of innovation development will present a wide range of challenges for the mental health nursing profession. In this article, we have attempted to highlight and signpost some of the current dialogue to stimulate mental health nurses in practice, education, and research settings to look to the horizon and carefully consider the implications for our profession. It will not be possible to ignore the technological advances that influence our practice, and so we must strengthen a dialogue to prepare our profession for the digital evolution of AI within our discipline.

RELEVANCE FOR CLINICAL PRACTICE

Clinical mental health nurses should anticipate the need to continue their lifelong learning to include developing their knowledge about AI within the mental health nursing context. The ramifications of additional education will assist nurses to provide more timely and relevant care and in doing so improve the safety, timeliness, and precision of health outcomes for consumers.

ACKNOWLEDGEMENTS

None. Open access publishing facilitated by The University of Newcastle, as part of the Wiley - The University of Newcastle agreement via the Council of Australian University Librarians.

CONFLICTS OF INTEREST

Prof Rhonda Wilson and Dr Andrea Donaldson are both members of the editorial board for the International Journal of Mental Health Nursing.

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

ORCID

Rhonda L. Wilson https://orcid.org/0000-0001-9252-2321
Oliver Higgins https://orcid.org/0000-0001-6545-145X
Jacob Atem https://orcid.org/0000-0001-7981-7694
Andrea E. Donaldson https://orcid.org/0000-0002-8548-4221
Frederik Alkier Gildberg https://orcid.org/0000-0001-9075-6108
Mary Hooper https://orcid.org/0000-0003-1864-4739
Mark Hopwood https://orcid.org/0000-0002-1658-6793
Silvia Rosado https://orcid.org/0000-0002-0871-3182
Bernadette Solomon https://orcid.org/0000-0002-9052-8321

REFERENCES

Alhaddad, A.Y., Cabibihan, J.-J. & Bonarini, A. (2020) Influence of reaction time in the emotional response of a companion robot to a Child's aggressive interaction. *International Journal of Social Robotics*, 12(6), 1279–1291. Available from: https://doi.org/10.1007/s12369-020-00626-z

Allaban, A.A., Wang, M. & Padır, T. (2020) A systematic review of robotics research in support of In-home care for older adults. *Information*, 11(2), 75. Available from: https://doi.org/10.3390/info11020075

Arnold, L. (2016) Emobie™: A robot companion for children with anxiety. In: *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. Christchurch, New Zealand: IEEE Institue of Electical and Eletronic Engineers.

Aryana, B. & Brewster, L. (2020) Design for mobile mental health: Exploring the informed participation approach. *Health Informatics Journal*, 26(2), 1208–1224. Available from: https://doi.org/10.1177/1460458219873540

Blakkarly, J. (2022) *Kmart, Bunnings and the good guys using facial recognition technology in stores*. Sydney, Australia: Choice. Available from: https://www.choice.com.au/consumers-and-data/data-collection-and-use/how-your-data-is-used/articles/kmart-bunnings-and-the-good-guys-using-facial-recognition-technology-in-store [Accessed 12th July 2022].

Botsman, R. (2018) *Who Can you Trust?: How Technology Brought us Together and why it Might Drive us apart*. London, UK: Penguin.

Carryer, J. (2019) *Letting go of our past to claim our future*. Available from: https://onlinelibrary.wiley.com/ 10.1111/jocn.15016

Crawford, K. (2021a) Affect. In: *The Atlas of AI*. London, UK: Yale University Press, pp. 151–179. Available from: https://doi.org/10.2307/j.ctv1ghv45t.8

14470349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/inm.13123 by The University Of Newcastle, Wiley Online Library on [30/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Crawford, K. (2021b) Classification. In: *The Atlas of AI*. London, UK: Yale University Press, pp. 123–149. Available from: https://doi.org/10.2307/j.ctv1ghv45t.7

Crawford, K. (2021c) Coda space. In: *The Atlas of AI*. London, UK: Yale University Press, pp. 229–237. Available from: https://doi.org/10.2307/j.ctv1ghv45t.11

Crawford, K. (2021d) Conclusion power. In: *The Atlas of AI*. London, UK: Yale University Press, pp. 211–227. Available from: https://doi.org/10.2307/j.ctv1ghv45t.10

Dastin, J. (2022) *Amazon has a plan to make Alexa mimic anyone's voice*. USA: Reuters. Available from: https://www.reuters.com/technology/amazon-has-plan-make-alexa-mimic-anyones-voice-2022-06-22/ [Accessed 20th January 2023].

Elias, A. & Paradies, Y. (2021) The costs of institutional racism and its ethical implications for healthcare. *Journal of Bioethical Inquiry*, 18(1), 45–58. Available from: https://doi.org/10.1007/s11673-020-10073-0

Fond, G.B., Lagier, J.-C., Honore, S., Lancon, C., Korchia, T., Verville, P.-L.S.D., et al. (2020). Microbiota-Orientated Treatments for Major Depression and Schizophrenia. *Nutrients*, 12(4), 1024. Available from: https://doi.org/10.3390/nu12041024

GDPR. (2018) *Right to be Forgotten*. Available from: https://gdpr-info.eu/issues/right-to-be-forgotten/ [Accessed 29th June 2022].

Gichoya, J.W., Banerjee, I., Bhimireddy, A.R., Burns, J.L., Celi, L.A., Chen, L.-C. et al. (2022) AI recognition of patient race in medical imaging: A modelling study. *The Lancet Digital Health*, 4(6), e406–e414. Available from: https://doi.org/10.1016/s2589-7500(22)00063-2

Higgins, O., Chalup, S., Short, B. & Wilson, R.L. (2023) Artificial intelligence (AI) and machine learning (ML) based decision support systems in mental health: An integrative review. *International Journal of Mental Health Nursing*.

Higgins, O., Chalup, S.K. & Wilson, R.L. (2022) *Artificial Intelligence in Nursing: Trustworthy or Reliable? 1st Workshop on Artificial Intelligence in Nursing: Advances, Methods and Path Forward, Halifax, Canada*.

Higgins, O., Charup, S., Short, B. & Wilson, R.L. (2022) Interpretations of innovation: The intersection of technological advancement and psychosis. *Perspectives in Psychiatric Care*.

International Council of Nurses. (2012) *The ICN code of ethics for nurses*. Geneva, Switzerland: International Council of Nurses. Available from: https://www.icn.ch/system/files/2021-10/ICN_Code-of-Ethics_EN_Web_0.pdf [Accessed 20th January 2023].

Jensen, A., Pedersen, B.D., Bang Olsen, R., Hounsgaard, L. & Wilson, R.L. (2019) Nurses experiences of delivering acute orthopaedic care to patients with dementia. *International Journal of Older People Nursing*, 14(4), e12271. Available from: https://doi.org/10.1111/opn.12271

Kuntsman, A., Miyake, E. & Martin, S. (2019) Re-thinking digital health: Data, appisation and the (im)possibility of 'opting out'. *Digital Health*, 5, 205520761988067. Available from: https://doi.org/10.1177/2055207619880671

Lal, S. (2019) E-mental health: Promising advancements in policy, research, and practice. *Healthcare Management Forum*, 32(2), 56–62. Available from: https://doi.org/10.1177/0840470418818583

Liang, L., Li, C., Meng, C., Guo, X., Lv, J., Fei, J. et al. (2022) Psychological distress and internet addiction following the COVID-19 outbreak: Fear of missing out and boredom proneness as mediators. *Archives of Psychiatric Nursing*, 40, 8–14. Available from: https://doi.org/10.1016/j.apnu.2022.03.007

Marshall, J.M., Dunstan, D.A. & Bartik, W. (2020) Effectiveness of using mental health Mobile apps as digital antidepressants for reducing anxiety and depression: Protocol for a multiple baseline across-individuals design. *JMIR Research Protocols*, 9(7), e17159. Available from: https://doi.org/10.2196/17159

Mirzoeff, N. (2011) *The right to look a Counterhistory of Visuality*. Durham, North Carolina: Duke University Press. Available from: https://doi.org/10.2307/j.ctv11vc8rg

Moyle, W., Bramble, M., Jones, C. & Murfield, J. (2018) Care staff perceptions of a social robot called Paro and a look-alike plush toy: A descriptive qualitative approach. *Aging & Mental Health*, 22(3), 330–335. Available from: https://doi.org/10.1080/13607863.2016.1262820

Mudallal, R.H., Othman, W.M. & Al Hassan, N.F. (2017) Nurses' burnout: The influence of leader empowering behaviors, work conditions, and demographic traits. *Inquiry*, 54, 46958017724944. Available from: https://doi.org/10.1177/0046958017724944

Rai, A. (2020) Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science*, 48(1), 137–141. Available from: https://doi.org/10.1007/s11747-019-00710-5

Ruijten, P.A.M. & Cuijpers, R.H. (2020) Do not let the robot get too close: Investigating the shape and size of shared interaction space for two people in a conversation. *Information*, 11(3), 147. Available from: https://doi.org/10.3390/info11030147

Shaikh, S.E. (2020) Interactive and revisable decision-support: Doing more harm than good? *Behaviour & Information Technology*, 41, 845–863. Available from: https://doi.org/10.1080/0144929x.2020.1837242

Siafakas, N. (2022) Do we need a Hippocratic Oath for Artificial Intelligence scientists? *AI Magazine*, 42(4), 57–61. Available from: https://doi.org/10.1609/aimag.v42i4.15090

Smith, M., Dietrich, B.J., Bai, E.-w. & Bockholt, H.J. (2020) Vocal pattern detection of depression among older adults. *International Journal of Mental Health Nursing*, 29(3), 440–449. Available from: https://doi.org/10.1111/inm.12678

Sogaard Neilsen, A. & Wilson, R.L. (2019) Combining e-mental health intervention development with human computer interaction (HCI) design to enhance technology-facilitated recovery for people with depression and/or anxiety conditions: An integrative literature review. *International Journal of Mental Health Nursing*, 28(1), 22–39. Available from: https://doi.org/10.1111/inm.12527

Thompson, C. (2022) *One Weird Trick To Make Humans Think An AI Is "Sentient"*. Meduim. Available from: https://clivethompson.medium.com/one-weird-trick-to-make-humans-think-an-ai-is-sentient-f77fb661e127 [Accessed 29th June 2022].

Tiku, N. (2022) The Google engineer who thinks the company's AI has come to life. *The Washington Post*. Available from: https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/ [Accessed 20th January 2023].

Wiens, J., Creary, M. & Sjoding, M.W. (2022) AI models in health care are not colour blind and we should not be either. *The Lancet Digital Health*, 4(6), e399–e400. Available from: https://doi.org/10.1016/s2589-7500(22)00092-9

Wilson, R.L. (2018) The right way for nurses to prescribe, administer and critique digital therapies. *Contemporary Nurse*, 54(4–5), 543–545. Available from: https://doi.org/10.1080/10376178.2018.1507679

Wilson, R.L. (2020) Why nurses need to be at the Centre of new developments in digital technology. *Nurse Researcher*, 28(2), 6–8. Available from: https://doi.org/10.7748/nr.28.2.6.s2

Young, A.T., Amara, D., Bhattacharya, A. & Wei, M.L. (2021) Patient and general public attitudes towards clinical artificial intelligence: A mixed methods systematic review. *The Lancet Digital Health*, 3(9), e599–e611. Available from: https://doi.org/10.1016/s2589-7500(21)00132-1

How to cite this article: Wilson, R.L., Higgins, O., Atem, J., Donaldson, A.E., Gildberg, F.A., Hooper, M. et al. (2023) Artificial intelligence: An eye cast towards the mental health nursing horizon. *International Journal of Mental Health Nursing*, 00, 1–7. Available from: https://doi.org/10.1111/inm.13121

14470349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/inm.13121 by The University Of Newcastle, Wiley Online Library on [30/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Appendix 2 Editorial: The continued importance of mental health nurses engaging with social media and related emerging technologies

DOI: 10.1111/inm.13129

EDITORIAL

International Journal of Mental Health Nursing

Editorial: The continued importance of mental health nurses engaging with social media and related emerging technologies

NOTE OF THANKS

As the new Social Media Editor for the *International Journal of Mental Health Nursing* (RW), I would like to pay respect to my predecessor, the inaugural Social Media Editor, Mr. Paul McNamara. We are grateful for his 10 years of social media leadership with this journal and of the impact that has been apparent (McNamara, 2022). Social media has proved to be a vital platform for amplifying the voice of mental health nurses worldwide and in sharing our scientific and practice knowledge (McNamara & Usher, 2019).

TECHNOLOGY CONTINUES TO TRANSFORM

As mental health nurses, we have engaged with social media strategically for some time through establishing professional networks that extend beyond the local and regional, enabling and reinforcing a wider global dialogue (Wilson et al., 2014; Wilson & Armstrong, 2016). We have used it to strengthen our practice as educators, scientists, and clinicians (Cronin et al., 2020; McNamara, 2021; Usher et al., 2014; Wilson & Usher, 2017). As technology continues to transform, it reveals expanded capacity and capabilities, and it is inevitable that we will continue to see innovative implementation, and leadership evolving within our discipline.

In today's digital age, social media is an integral part of our lives. Even if we are not aware of it, our mobile phones listen to us and track our online activity, frequently resulting in unsolicited tailored advertisements and offers aligned to our interests (Wilson et al., 2023). Social media platforms such as Facebook and Twitter have been around for nearly two decades, while the rise of wearable technology and the Internet of Things has enabled our devices to be more integrated and convenient to use (Higgins et al., 2023a). Additionally, integration of Artificial Intelligence (AI) with social media has revolutionized communication, accommodating instant language translation and human-like chatbots (Higgins et al., 2023b). Perhaps, there is no better example of the speed at which technology has seen expansive popular uptake than through the introduction of ChatGPT [a service within the OpenAI software brand] in November 2023,

FIGURE 1 Tweet by @ChatGPT 19 January 2023.

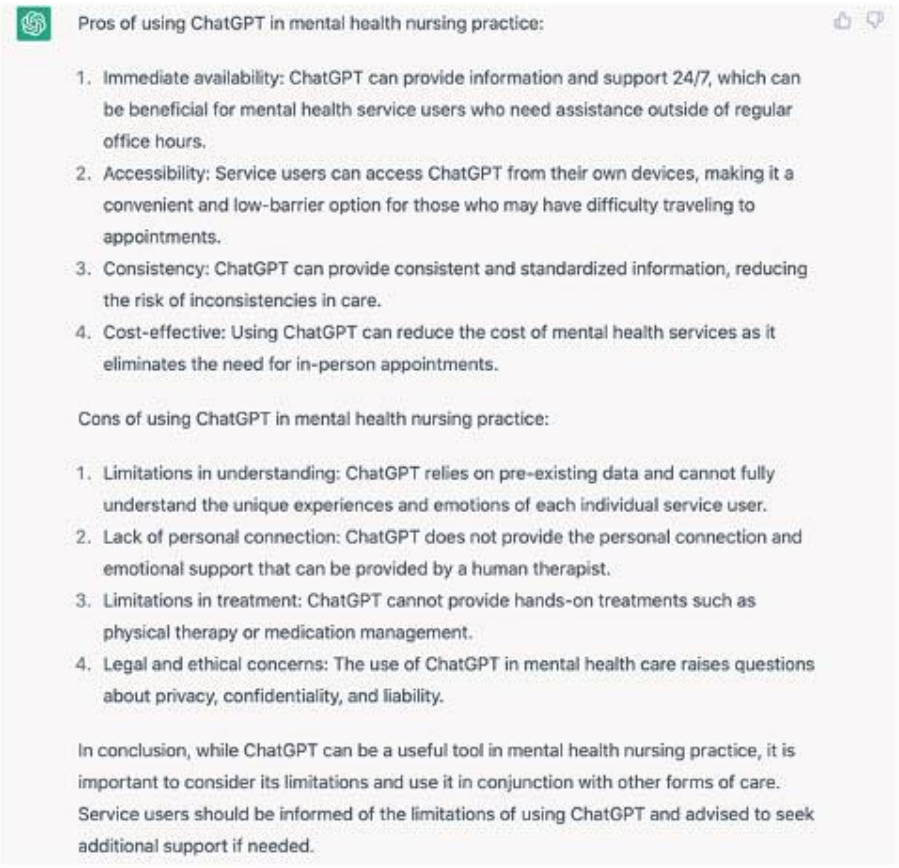
who themselves tweeted that they achieved 1 million users in just 5 days (see Figure 1). It is apparent that social media is entering a new transformative era, and that AI will rapidly accelerate the continued digital evolution. The discipline of mental health nursing should position itself to be ready to take advantage of the benefits on offer, and to address any limitations that may arise as a consequence.

DIGITAL DISADVANTAGE AND OUR ROLES AS ADVOCATES

As mental health nurses, it is imperative that we understand the strengths and risks associated with social

 © 2023 John Wiley & Sons Australia, Ltd.

14470349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/inm.13129 by NHMRC National Cochrane Australia, Wiley Online Library on [14/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License



Pros of using ChatGPT in mental health nursing practice:

1. Immediate availability: ChatGPT can provide information and support 24/7, which can be beneficial for mental health service users who need assistance outside of regular office hours.
2. Accessibility: Service users can access ChatGPT from their own devices, making it a convenient and low-barrier option for those who may have difficulty traveling to appointments.
3. Consistency: ChatGPT can provide consistent and standardized information, reducing the risk of inconsistencies in care.
4. Cost-effective: Using ChatGPT can reduce the cost of mental health services as it eliminates the need for in-person appointments.

Cons of using ChatGPT in mental health nursing practice:

1. Limitations in understanding: ChatGPT relies on pre-existing data and cannot fully understand the unique experiences and emotions of each individual service user.
2. Lack of personal connection: ChatGPT does not provide the personal connection and emotional support that can be provided by a human therapist.
3. Limitations in treatment: ChatGPT cannot provide hands-on treatments such as physical therapy or medication management.
4. Legal and ethical concerns: The use of ChatGPT in mental health care raises questions about privacy, confidentiality, and liability.

In conclusion, while ChatGPT can be a useful tool in mental health nursing practice, it is important to consider its limitations and use it in conjunction with other forms of care. Service users should be informed of the limitations of using ChatGPT and advised to seek additional support if needed.

FIGURE 2 ChatGPT generated response weighing up the pros and cons of ChatGPT in mental health nursing practice context.

media and technology use, as occasionally, the people we care for may experience problematic technology use that can adversely impact their mental health (Luengo-González et al., 2022). Our own experiences with social media, as end-users, allow us to participate in the public discourse surrounding mental health, with Usher and Jackson (2023) providing us with a recent example, noting their observations about the way social media had been used in the public grieving context. As social media users, mental health nurses have an opportunity to promote accessibility and equity to reach vulnerable or priority populations that may otherwise be digitally disadvantaged. Thus, we can ensure our voices of advocacy are represented in the digital and wider community, amongst developers, and across policy and political domains to support our service users' needs (Wilson et al., 2023).

PROS AND CONS

AI services, for example ChatGPT (2023), will offer opportunities to extend and enhance our practice. One of the major advantages of this technology is its ability to understand natural language interactions (Aljanabi et al., 2023). There are pros and cons to consider, and as such, human critical clinical judgement by mental health nurses will always be essential as an implementation consideration. With this caveat in mind, who better to explain the benefits of technology, such as ChatGPT, than ChatGPT itself. We asked ChatGPT to: 'write an instructional guide demonstrating the pros and cons of using ChatGPT in the mental health nursing practice context'. The response generated is shown in Figure 2. It is noteworthy that even ChatGPT 'confesses' that it is not a substitute for human expertise and should not be used as a sole source of information.

END-USERS AND SERVICE USERS

A person-centred approach to mental health care requires us to meet people wherever they are situated, and this must increasing include virtual or digital environments, in addition to familiar and traditional physical environments (Wilson, 2020). Many mental health

14470349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/inm.13129 by NHS/HSC National Coordinating Centre, Wiley Online Library on [14/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

nursing services can be delivered in a range of modes: face-to-face, virtually and/or a hybrid mixed-mode format with the integration of video, audio, and text-based communication. As mental health nurses, we must ensure that our perspectives as technology end-users are not omitted from the ad hoc training and discussions surrounding AI and social media implementation in our settings, because our advocacy (or lack of) can positively and/or negatively impact the quality of life for mental health service users (Wilson et al., 2023). As always, the wellbeing and safety of our service users must be our priority.

CONCLUSION

In conclusion, social media is already an integral part of our lives and within many aspects of mental health policy and practice. As mental health nurses, we must carefully reflect on our biases, dominant sentiment, decisions, actions and narrative about our engagement (or not) with social media, and emerging technologies, to ensure that our considered perspectives and expertise are included, and amplified, in ongoing public discourse. It is important that we contribute to discussion about the adequacy of design and continue to influence ongoing development, and clinical solutions that are suitable for contemporary mental health care for people who require it.

CONNECT WITH IJMHN ONLINE

I (RW) look forward to interacting with our *International Journal of Mental Health Nursing* readers and followers across our social media channels. Together we will continue our discourse to strengthen the voices of mental health nursing scientists, scholars, practitioners, and peers globally.

Website www.wileyonlinelibrary.com/journal/INM.
Twitter twitter.com/IJMHN.
Facebook www.facebook.com/IJMHN.
LinkedIn www.linkedin.com/company/IJMHN.

Rhonda L. Wilson
Oliver Higgins

School of Nursing and Midwifery, University of Newcastle, Gosford, NSW, Australia

Correspondence
Rhonda L. Wilson, University of Newcastle, Gosford, NSW, Australia.
Email: rhonda.wilson@newcastle.edu.au

REFERENCES

Aljanabi, M., Ghazi, M., Ali, A.H. & Abed, S.A. (2023) ChatGpt: open possibilities. *Iraqi Journal for Computer Science and Mathematics*, 4(1), 62–64.

ChatGPT. (2023) ChatGPT *in* OpenAI. https://chat.openai.com/chat [accessed 5 February, 2023]

Cronin, C., Hungerford, C. & Wilson, R.L. (2020) Using digital health technologies to manage the psychosocial symptoms of menopause in the workplace: a narrative literature review. *Issues in Mental Health Nursing*, 1–8, 541–548. Available from: https://doi.org/10.1080/01612840.2020.1827101

Higgins, O., Chalup, S., Short, B. & Wilson, R.L. (2023a) Interpretations of innovation: the intersection of technological advancement and psychosis. *Perspectives in Psychiatric Care*. Available from: https://doi.org/10.1155/2023/4464934

Higgins, O., Chalup, S., Short, B. & Wilson, R.L. (2023b) Artificial intelligence (AI) and machine learning (ML) based decision support systems in mental health: an integrative review. *International Journal of Mental Health Nursing*. Available from: https://doi.org/10.1111/inm.13114

Luengo-González, R., Noriega-Matanza, M.C., Espín-Lorite, E.J., García-Sastre, M.M., Rodríguez-Rojo, I.C., Cuesta-Lozano, D. et al. (2022) The role of life satisfaction in the association between problematic technology use and anxiety in children and adolescents during the COVID-19 pandemic. *International Journal of Mental Health Nursing*, 32, 212–222. Available from: https://doi.org/10.1111/inm.13077

McNamara, P. (2021) *Share or Perish: Social media motivation for busy and important academics [blog post and video]*. Available from: https://meta4RN.com/2021/09/14/busy/. Accessed 6 February 2023.

McNamara, P. (2022) Happy anniversary IJMHN. *International Journal of Mental Health Nursing*, 31(4), 767–771. Available from: https://doi.org/10.1111/inm.13025

McNamara, P. & Usher, K. (2019) Share or perish: social media and the international journal of mental health nursing. *International Journal of Mental Health Nursing*, 28(4), 960–970. Available from: https://doi.org/10.1111/inm.12600

Usher, K. & Jackson, D. (2023) Public expressions of grief and the role of social media in grieving and effecting change. *International Journal of Mental Health Nursing*, 32(1), 1–2. Available from: https://doi.org/10.1111/inm.13110

Usher, K., Woods, C., Casella, E., Glass, N., Wilson, R.L., Mayner, L. et al. (2014) Australian health professions student use of social media. *Collegian*, 21(2), 95–101. Available from: https://doi.org/10.1016/j.colegn.2014.02.004

Wilson, R.L. (2020) Why nurses need to be at the Centre of new developments in digital technology. *Nurse Researcher*, 28(2), 6–8. Available from: https://doi.org/10.7748/nr.28.2.6.s2

Wilson, R.L. & Armstrong, M. (2016) International networks: the usefulness of conference meetings and site visits for progressing nursing practice of youth in mental health care. *Mental Health Practice RCNi*, 19(9), 34–37. Available from: https://doi.org/10.7748/mhp.19.9.34.s22

Wilson, R.L., Higgins, O., Atem, J., Donaldson, A.E., Gildberg, F.A., Hooper, M. et al. (2023) Artificial intelligence: an eye cast towards the mental health nursing horizon. *International Journal of Mental Health Nursing*. Available from: https://doi.org/10.1111/inm.13121

Wilson, R.L., Ranse, J., Cashin, A. & McNamara, P. (2014) Nurses and twitter: the good, the bad, and the reluctant. *Collegian*, 21(2), 111–119. Available from: https://doi.org/10.1016/j.colegn.2013.09.003

Wilson, R.L. & Usher, K. (2017) Social media recruitment strategies for health research participation. *Nurse Researcher RCNi*, 25(3), 36–41. Available from: https://doi.org/10.7748/nr.2017.e1478

14670349, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/inm.13129 by NHMRC National Cochrane Australia, Wiley Online Library on [14/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Appendix 3 1st Workshop on Artificial Intelligence in Nursing: Advances, Methods and Path Forward

Artificial Intelligence in Nursing: trustworthy or reliable?

Oliver Higgins[1 2][0000-0001-6545-145X], Stephan K. Chalup[3 [0000-0002-7886-3653]], and Rhonda L. Wilson[1 4][0000-0001-9252-2321]

[1] School of Nursing and Midwifery, University of Newcastle, Australia
[2] Central Coast Local Health District, NSW Australia
[3] School of Information and Physical Sciences, University of Newcastle, Australia
[4] School of Nursing, Massey University, New Zealand
oliver.higgins@uon.edu.au

Keywords: Artificial Intelligence, Clinical Decision Support System, Nursing, Mental Health

1 Extended Abstract

Trustworthiness in Artificial Intelligence (AI) innovation is at the forefront of governments, researchers and clinicians priorities. The role that trustworthiness plays in the acceptance of AI within various clinical settings cannot be understated. Clinicians have highlighted trust and confidence as barriers for the acceptance of AI within clinical application. However, while there is a call to design and develop AI that is considered trustworthy, AI itself lacks the emotional capability to confer trust. The objectives of this paper is firstly, to highlight the enigma of seeking or expecting trust attributes from a machine, and secondly, to reframe the interpretation of trustworthiness for AI through evaluation of its reliability and validity as consistent with use of other clinical instruments. This issue is significant because it represents a crucial component that needs to be addressed when implementing AI based interventions within clinical practice. Sole focus on the demonstration of trust, instead of a focus on the usual requirement for reliability and validity attributes during implementation phases, may result in negative experiences for nurses and clinical users.

1.1 Introduction

The role of Artificial Intelligence (AI) in nursing, especially in the mental health setting, is in the early stages of design and development. The Nursing and Artificial Intelligence Leadership (NAIL) recommendations call upon nursing researchers and leaders to focus on the use and impact of AI and the role that nursing must contribute to the design and development of AI nursing technologies [1]. Nursing AI research should focus on establishing programs of practice that further develop the field [1], and most importantly, AI must augment and support nursing practice and workflows [1]. Nurses are required to be active participants in the design, development and deployment of AI in clinical settings [1] and ensure that any AI product used in practice is safe and secure.

1.2 Scientific reliability: a logic to enhance implementation

Clinical Decision Support Systems (CDSS) are one particular area that offers potential for mental health nurses, especially for complex areas such as the assessment and prediction of suicidal behaviours in critical settings. The research team undertook an integrative review investigating the incorporation of artificial intelligence and machine learning based decision support systems in mental health care settings [2]. Established nursing practice relies upon many different instruments to augment and guide practice. One general nursing example is the Waterlow, a tool routinely used to assess pressure ulcer risk. This tool has been incorporated into practice supported by extensive research indicating the tool's reliability [3]. In mental health, the Health of The Nations Outcome Scale (HoNOS) is used internationally (eg, UK, Australia and New Zealand [4]) and has demonstrated consistent interrater reliability through an extensive body of research [4] [5]. Interestingly, the literature for these examples describes the reliability and validity of both these tools but the attribute of trustworthiness is not mentioned [3] [4]. This highlights an inconsistency whereby reliability and validity appear to be preferred attributes for some tools, while trust is a preferred attribute for another tool.

Artificial Intelligence is increasingly evident in all parts of life, however, the resulting complex machines are not always understood and can often be *black-box* in nature [6], which, when combined with the human tendency to anthropomorphise [7], can result in humans attributing moral characteristics to it [8]. Additionally, the media's view of what these complex machines can do, depicting ideas such as sentient robots, has resulted in an association and categorisation of what the '*intelligent*' machines are capable of by inferring human definitions and contexts [9]. As such, the public expectation that AI should deliver trustworthy solutions has become more evident, with government entities forming organisations such as the European Commission High-level Expert Group (HLEG) on AI. Its purpose is to undertake a review of the use of AI and issue the *Ethics Guidelines for Trustworthy AI* [10]. Likewise, countries such as Singapore have undertaken a series of initiatives on AI governance and ethics designed to build an ecosystem of trust that supports the adoption of AI [11] [12]. For nursing, the release of the World Health Organisation's *Ethics and governance of artificial intelligence for health* in 2021 calls upon organisations and governments to demonstrate trustworthiness with providers and patients for ethical use of AI interventions for health care [13]. However, viewing the AI itself as trustworthy may undermine the value of interpersonal trust, anthropomorphising the AI and ultimately diverting accountability from the organisations responsible for its design and development [8].

The notion that AI has the capacity to be considered trustworthy is complex and complicated, as trust is an integral and essential part of human relationships [8]. HLEG proposes that the main characteristics for trusting AI can be divided into three components. The AI technology itself, the design and development organisations behind its conception and use, and the socio-technical elements of the AI life cycle [10]. Ultimately AI is a machine, albeit a complex and sophisticated machine, but a machine nonetheless, yet trustworthiness is not a value traditionally attributed to machines [10].

Therefore, while AI may meet the rational account of trust requirement, it cannot be held accountable for its actions as it does not possess the emotional states to be held responsible [8] and, as such, should not be considered trustworthy but rather should be viewed in terms of its reliability and validity [8] as per any other clinical instrument.

Many nurses have expressed reluctance in the use of AI, stating that the nature and complexity of learning algorithms can obscure the rationale or reasons for a recommendation [6] and lead to hesitancy or anxiety for some nurses when they do not understand the process or logic that the AI system undertook to provide a particular recommendation [14]. Nurses are less likely to engage or trust *black-box* recommendations [14], requiring an understanding of the underlying AI mechanisms and logic process. AI based innovations will need to demonstrate the expected reliability and validity of any clinical intervention or instrument [14]. Our findings indicate that trust and confidence in AI are significant barriers to successfully implementing any intervention using AI CDSS in mental health [2]. However, the imparting of trust is an inherently human emotional construct [15]; as such, the concept of AI trustworthiness may be an additional barrier preventing nurses from accepting AI within practice.

1.3 Recommendations

We suggest that AI interventions should be described in terms of reliability and validity as would normally be expected of other clinical tools. This consistency in narrative supports incorporation of AI within practice. Acceptance of an intervention based on trust alone is not based on a consistent chain of logic and risks compromise to quality and safely related to implementation. However, it is also apparent that rich description of the functions of decision logic is required to support implementation of AI-based interventions and this must include: clear actions, that describe how clinical features contributed to the outcome while communicating multiple recommendations; and, demonstrating confidence with prescribing recommendations. Nurses should be presented with treatment recommendations that describe validity and confidence of prediction, and shown the chain of logic that rests the final decision for the appropriate course of treatment with nurses. Tools that utilise *glass-box* design, such as InterpretML [16], should be encouraged and incorporated into clinical interfaces. Additional research should be framed to better understand how AI is used to deliver care and how it will affect those who have a mental illness, and as our research has previously demonstrated, especially where psychosis is a targeted condition for AI interventions.

1.4 Summary

To discuss AI in terms of its trustworthiness alone risks illogically endorsing human attributes to a machine. This paper has presented a breakdown of the components of trust in the context of AI in health. While trustworthiness may seem an overly simple notion, it is a complex concept, and it should be considered as an additional barrier that has the potential to prevent nurses acceptance of AI into practice. Communicating the system's validity and reliability while creating transparent interpretable innovations will

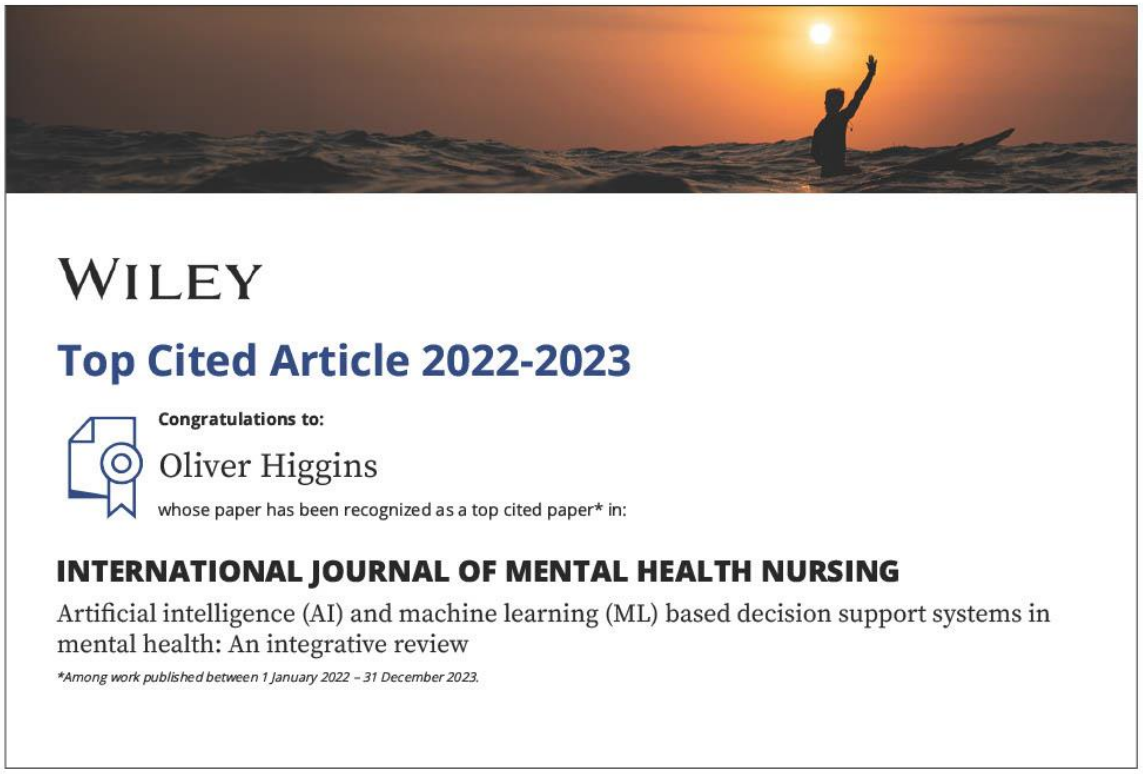
form a significant component to support the implementation of AI-based clinical instruments in routine clinical nursing practice.

References

1. Ronquillo, C.E., Peltonen, L.M., Pruinelli, L., Chu, C.H., Bakken, S., Beduschi, A., Cato, K., Hardiker, N., Junger, A., Michalowski, M., Nyrup, R., Rahimi, S., Reed, D.N., Salakoski, T., Salantera, S., Walton, N., Weber, P., Wiegand, T., Topaz, M.: Artificial intelligence in nursing: Priorities and opportunities from an international invitational think-tank of the Nursing and Artificial Intelligence Leadership Collaborative. Journal of Advanced Nursing 77, 3707-3717 (2021)
2. Higgins, O., Chalup, S.K., Short, B.L., Wilson, R.L.: Confidence and Trust: An integrative review of artificial intelligence and machine learning based decision support systems in mental health. Journal of Advanced Nursing Under review, (2022)
3. Charalambous, C., Koulori, A., Vasilopoulos, A., Roupa, Z.: Evaluation of the Validity and Reliability of the Waterlow Pressure Ulcer Risk Assessment Scale. Medical Archives 72, 141-144 (2018)
4. James, M., Painter, J., Buckingham, B., Stewart, M.W.: A review and update of the Health of the Nation Outcome Scales (HoNOS). BJPsych Bulletin 42, 63-68 (2018)
5. Wing, J.K., Beevor, A.S., Curtis, R.H., Park, S.G.B., Hadden, J., Burns, A.: Health of the Nation Outcome Scales (HoNOS): Research and development. British Journal of Psychiatry 172, 11-18 (1998)
6. Rai, A.: Explainable AI: from black box to glass box. Journal of the Academy of Marketing Science 48, 137-141 (2019)
7. Airenti, G.: The Cognitive Bases of Anthropomorphism: From Relatedness to Empathy. International Journal of Social Robotics 7, 117-127 (2015)
8. Ryan, M.: In AI We Trust: Ethics, Artificial Intelligence, and Reliability. Science and Engineering Ethics 26, 2749-2767 (2020)
9. Liang, Y., Lee, S.A.: Fear of Autonomous Robots and Artificial Intelligence: Evidence from National Representative Data with Probability Sampling. International Journal of Social Robotics 9, 379-384 (2017)
10.HLEG AI: Ethics guidelines for trustworhty AI. (2019)
11.Singapore: InfoComm Media Development Authority: Singapore wins international award for its artificial intelligence governance and ethics initiatives. 2022, (2019)
12.Singapore: Singapore Computer Society: Singapore Computer Society, InfoComm Media Development Authority. AI Ethics & Governance Body of Knowledge. 2022, (2020)
13.World Health Organization: Ethics and governance of artificial intelligence for health: WHO guidance., (2021)
14.Brown, L.A., Benhamou, K., May, A.M., Mu, W., Berk, R.: Machine Learning Algorithms in Suicide Prevention: Clinician Interpretations as Barriers to Implementation. Journal Clinical Psychiatry 81, (2020)
15.McLeod, C.: Trust. Metaphysics Research Lab, Stanford University (2021)
16.Nori, H., Jenkins, S., Koch, P., Caruana, R.: Interpretml: A unified framework for machine learning interpretability. arXiv preprint arXiv:1909.09223 (2019)

Appendix 4 Top Cited Article 2022-2023: International Journal of Mental Health Nursing

Artificial intelligence (AI) and machine learning (ML) based decision support systems in mental health: An integrative review.



WILEY

Top Cited Article 2022-2023

Congratulations to:

Oliver Higgins

whose paper has been recognized as a top cited paper* in:

INTERNATIONAL JOURNAL OF MENTAL HEALTH NURSING

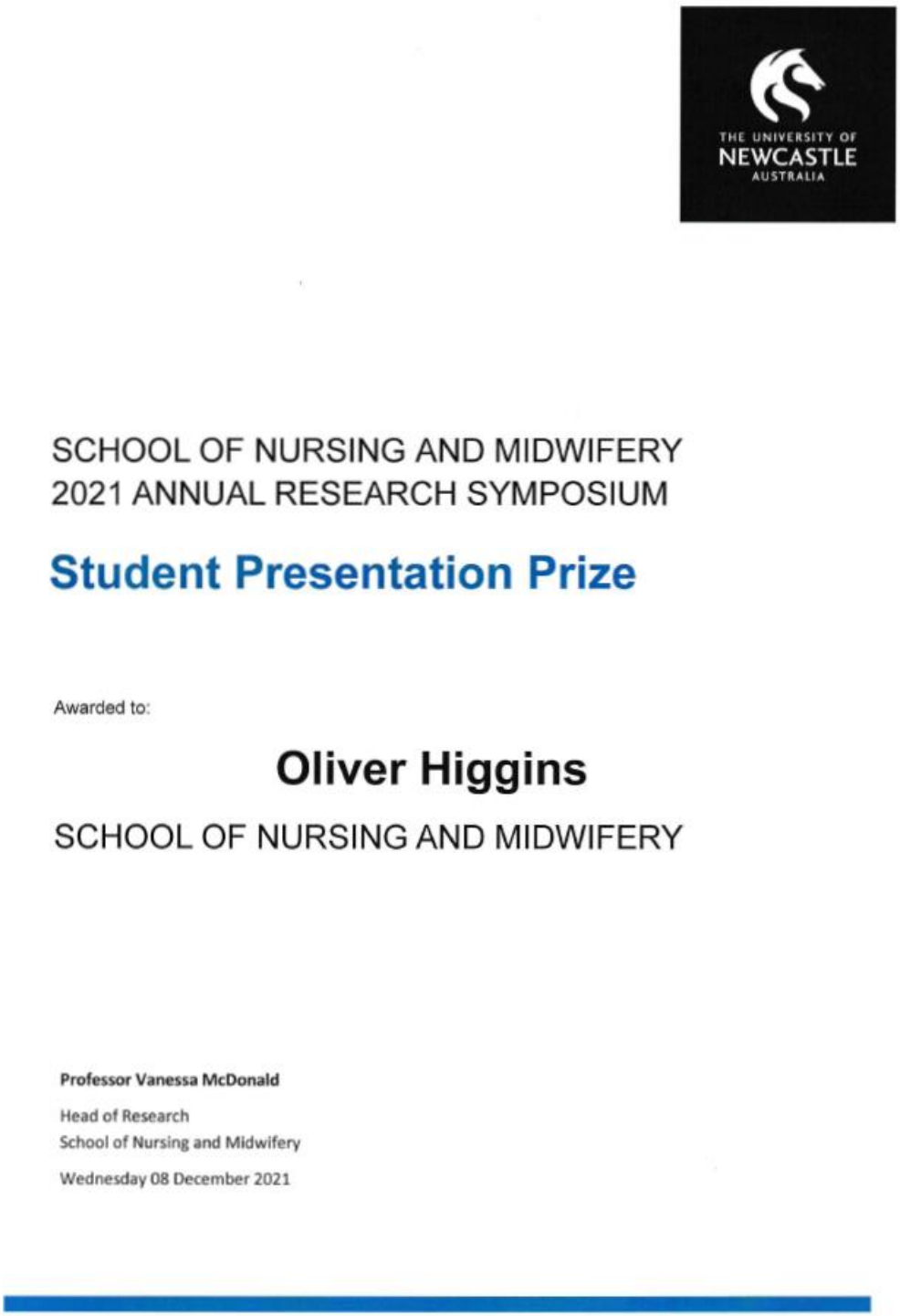
Artificial intelligence (AI) and machine learning (ML) based decision support systems in mental health: An integrative review

*Among work published between 1 January 2022 – 31 December 2023.

Appendix 5 Awards

University of Newcastle School of Nursing and Midwifery 2021 Annual Research Symposium.

Student Presentation Prize



Appendix 6 Ethics and Site Specific Approval

Health
Central Coast
Local Health District

14 December 2022

DISCLOSURE OF INFORMATION – APPROVAL UNDER CLAUSE 17 (2), HEALTH ADMINISTRATION REGULATION 2015

I, ***Fiona Wilkinson***, ***District Director Quality, Strategy and Improvement*** of the **Central Coast Local Health District**, pursuant to clause 17 (2) of the Health Administration Regulation 2015, hereby approve the release of information described in Schedule 1 below, to **Professor Rhonda Wilson,** Coordinating Principal Investigator and Professor of Mental Health Nursing, of University of Newcastle; subject to the following conditions:

1. The conditions of data release outlined in Schedule 2.

Signed this……14……..day of……December…….. 2022

……

Fiona Wilkinson

District Director Quality, Strategy and Improvement

SCHEDULE 1

De-identified unit record data as follows:

1. Data will be de-identified as part of the extraction process, with the exception of Medical Record Number (MRN). The MRN will be used to link data from the ED data with mental health eMR data to determine if the person was a current client of the mental health service at the time of their ED presentation, The MRN data will be de-identified by encrypting the data field during the extraction phase using a Secure Hash Algorithm (SHA):

As per:

- **Machine Learning for MH in ED Protocol** (version 3.3.1 dated 18 November 2022)

CCLHD REF NO: 1122-106C

Central Coast Local Health District
ABN 88 523 389 096
Research Office
L10, Central Coast Clinical School and Research Institute (CCCSRI)
77A Holden Street, Gosford NSW 2250
PO Box 361, Gosford NSW 2250
Tel: 4320 2085
Email: CCLHD-Research@health.nsw.gov.au
Website: www.cclhd.nsw.gov.au

SCHEDULE 2

CONDITIONS OF DATA RELEASE

A breach of any of these conditions may result in further data access being restricted or current access being revoked:

1. The data are to be used only for *Using Machine Learning to Assist Decision Making in the Assessment of Mental Health Consumers Presenting to Emergency Departments.* The project is carried out in accordance with the approved ethics application and all subsequent amendments
2. The data are to be kept in a secure physical and electronic environment that is accessible only by persons directly involved in the above project
3. A confidentiality undertaking will be completed prior to the information being released
4. The Central Coast Local Health District is to be acknowledged in any publication or report that arises from the use of the data
5. The data will not be matched with information on individuals from another source other than the datasets specified in the approved ethics application
6. A copy of any publication or report is to be provided to the Central Coast Local Health District at least two weeks prior to public release, emailed to CCLHD-Research@health.nsw.gov.au.
7. The data are to be destroyed 5 years after publication.
8. No information will be released with which it may be possible to identify an individual person
9. Individuals identified in the data are not to be personally identified in any publication or report
10. The use of information on Aboriginal and Torres Strait Islander status is subject to the approval of the Aboriginal Health and Medical Research Council Ethics Committee if one or more of the following apply:
    - Aboriginality is a key determinant
    - Data collection is explicitly directed at Aboriginal peoples
    - Aboriginal peoples, as a group, are to be examined in the results
    - The information may have an impact on one or more Aboriginal communities
    - Aboriginal health funds are a source of funding
11. This authority continues until and unless it has been revoked in writing

Central Coast Local Health District
ABN 88 523 389 096
Research Office
L10, Central Coast Clinical School and Research Institute (CCCSRI)
77A Holden Street, Gosford NSW 2250
PO Box 361, Gosford NSW 2250
Tel: 4320 2085
Email: CCLHD-Research@health,nsw.gov.au
Website: www.cclhd.nsw.gov.au

RESEARCH INTEGRITY UNIT

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

Registration of External HREC Approval

To Chief Investigator or Project Supervisor:	**Professor Rhonda Wilson**
Cc Co-investigators / Research Students:	**Mr Oliver Higgins** **Professor Stephan Chalup**
Re Protocol:	**Using Machine Learning to Assist Decision Making in the Assessment of Mental Health Consumers Presenting to Emergency Departments**
Date:	**21-Nov-2022**
Reference No:	**H-2022-0382**
External HREC Reference No:	**2022_ETH01597**

Thank you for your **Registration of Ext HREC Approval** submission to the Research Integrity Unit (RIU) seeking to register an External HREC Approval in relation to the above protocol.

Your submission was considered under an **Administrative Review** by the Ethics Administrator.

I am pleased to advise that the decision on your submission is **External HREC Approval Noted** effective **21-Nov-2022**.

As the approval of an External HREC has been noted, this registration is valid for the approval period determined by that HREC.

Your reference number is **H-2022-0382**.

PLEASE NOTE:
As the RIU has "noted" the approval of an External HREC, progress reports and reports of adverse events are to be submitted to the External HREC only. In the case of Variations to the approved protocol, or a Renewal of approval, you will apply to the External HREC for approval in the first instance and then Register that approval with the University's RIU, via RIMS.

Linkage of ethics approval to a new Grant

Registered External HREC approvals cannot be assigned to a new grant or award (ie those that were not identified in the initial registration submission) without confirmation from the RIU.

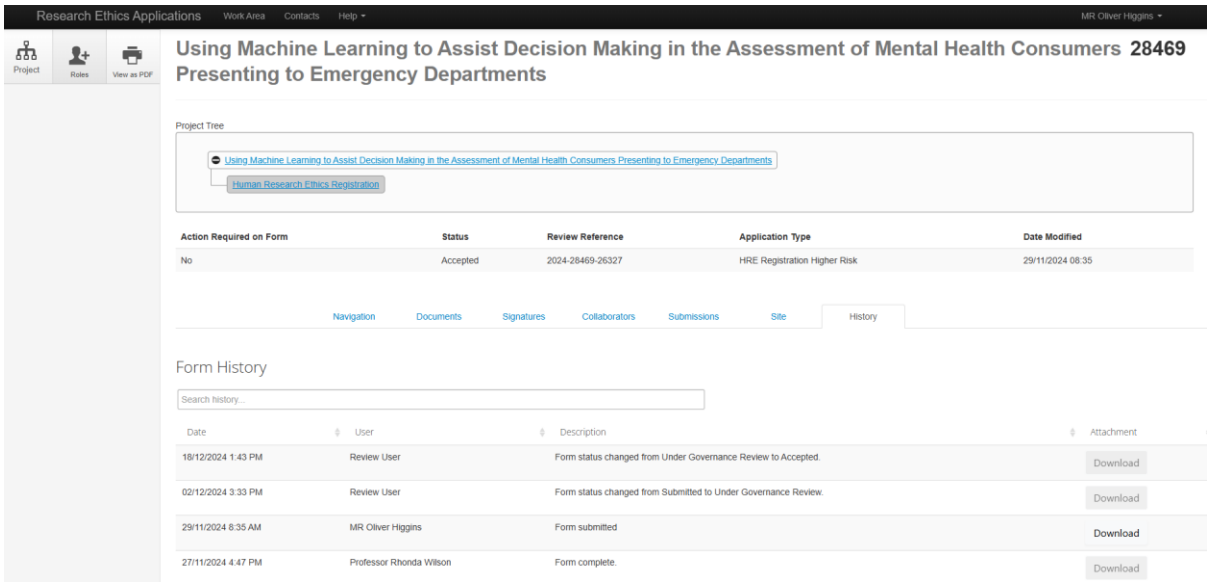
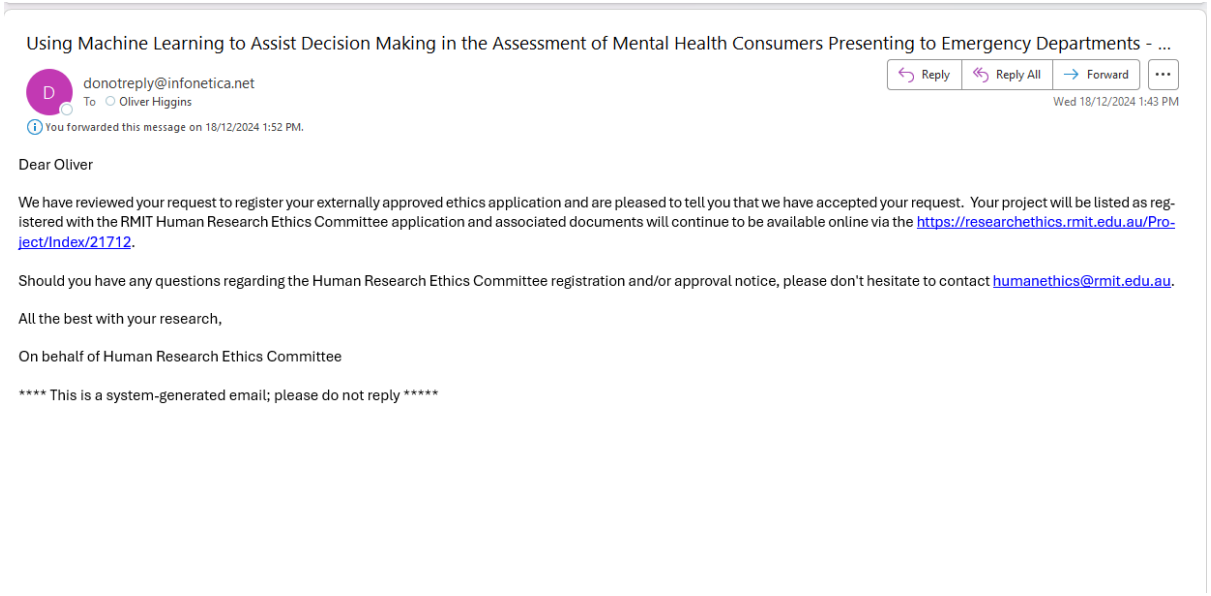
Best wishes for a successful project.

For communications and enquiries:
Human Research Ethics Administration

Research & Innovation Services
Research Integrity Unit
The University of Newcastle
Callaghan NSW 2308
T +61 2 492 17894
Human-Ethics@newcastle.edu.au

RIMS website - https://RIMS.newcastle.edu.au/login.asp

Linked University of Newcastle administered funding:

Funding body	Funding project title	First named investigator	Grant Ref
		.	

RMIT Specific Ethics Approvals