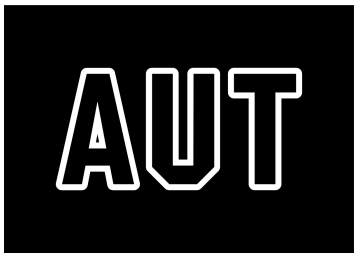


Advancing Asthma Care: Predictive Modelling and Community Perspectives in New Zealand

Widana Kankanamge Darsha Jayamini

School of Engineering, Computer & Mathematical Sciences
Auckland University of Technology

A thesis submitted to Auckland University of Technology
in fulfilment of the requirements for the degree of
Doctor of Philosophy (PhD)

July 2026

Abstract

Asthma is a chronic respiratory disease that affects people of all ages worldwide. This medical condition places a significant burden on individuals, healthcare systems and communities. Over 300 million people are affected globally by this respiratory condition. Among regions with a high prevalence of asthma, the Western Pacific region ranks among the highest. New Zealand (NZ), which lies within this region, has demonstrated a high prevalence of asthma, particularly among different ethnic groups, including the Indigenous Māori population and Pacific Peoples (also known as Pasifika). Worsening of asthma symptoms may lead to asthma attacks, resulting in emergency department visits and hospital admissions. In some cases, it may also be fatal. Globally, asthma causes approximately 450,000 deaths each year, while 101 people have died in the year 2019 from asthma in NZ. Therefore, asthma management is critical. Supporting earlier identification of individuals at risk of severe asthma outcomes is therefore critical for improving asthma management and reducing preventable harm.

Recent advances in Machine Learning (ML) offer opportunities to improve asthma management through predictive analytics. While these technological developments have the potential to enhance the quality of life and reduce healthcare costs, existing research remains limited in several key areas, including the development of robust and generalisable prediction models for imbalanced asthma data, the prediction of hospital length of stay following asthma admissions, and the consideration of how such technologies are perceived by the most vulnerable communities in NZ. Addressing these gaps is essential to ensure that predictive tools are not only technically sound but also meaningful and acceptable in real-world healthcare settings. Accordingly, this thesis adopts a mixed-methods approach to advance data-driven, equitable, and community-informed asthma care.

The quantitative components of the thesis followed the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology to develop the ML-based prediction models. The findings demonstrated that ML models can effectively predict both future asthma attacks and hospital length of stay using routinely collected hospital and pharmaceutical data in NZ, even in the presence of substantial data imbalance. Across these analyses, prior asthma attacks and medication use consistently emerged as key predictors of impending asthma attacks, showing that reliable predictions can be achieved using a smaller number of clinically meaningful

features. These findings contribute to improved understanding of predictive performance, generalisability, and efficiency in asthma attack prediction modelling. Furthermore, it provided new insights into asthma hospitalisation patterns in NZ. The results showed that demographic factors and admission time affect the length of stay for asthma.

Complementing the predictive modelling, the qualitative components were conducted using the inductive thematic analysis methodology through semi-structured interviews. The thesis explored the perspectives of different ethnic groups in NZ, including the Indigenous Māori population and Pasifika, on the use of Artificial Intelligence and digital technology for asthma management, specifically for predicting asthma attack risk. The findings highlighted openness to the potential benefits, alongside key concerns related to trust, data privacy, accessibility, and cultural alignment. The results showed that technical accuracy alone is insufficient for successful adoption, and that cultural values and expectations of these technology-based solutions play a central role in enhancing the acceptability. The thesis also provides a framework for Pasifika asthma care, serving as a guide for enhancing asthma awareness and designing future technology-based solutions for the community. The thesis further identifies and reflects on future research directions informed by the results and their broader implications.

Table of contents

List of figures

List of tables

Attestation of Authorship

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the qualification of any other degree or diploma of a university or other institution of higher learning.

Widana Kankanamge Darsha Jayamini
July 2026

Co-authorship Contribution

STUDENT AND SUPERVISOR APPROVALS

By signing you are confirming that the co-author contributions stated in the table(s) below are accurate.

Student Name	W.K. Darsha Jayamini	**Signature**		**Date**	20/07/2026
Supervisor Name	Farhaan Mirza	**Signature**		**Date**	20/07/2026

Chapter Number:	**Chapter 2**
Manuscript Title:	**Investigating Machine Learning Techniques for Predicting Risk of Asthma Exacerbations: A Systematic Review**
Publication Status:	**Published**
Reference if published:	**https://doi.org/10.1007/s10916-024-02061-3.**

AUTHOR SURNAME: (order as per manuscript)	CONTRIBUTION (May copy from the guidelines above)
W.K.D. Jayamini (W.K.D.J.)	**Conceptualisation: W.K.D.J., F.M.; Methodology: W.K.D.J., F.M.; Discussion: W.K.D.J.; Implementation: W.K.D.J., Validation: W.K.D.J., F.M., M.A.N., A.H.Y.C; Writing - original draft preparation: W.K.D.J.; Writing - review and editing: W.K.D.J., F.M., M.A.N., A.H.Y.C.; Project supervision - F.M., M.A.N., A.H.Y.C.**
F. Mirza (F.M.)	
M.A. Naeem (M.A.N.)	
A.H.Y. Chan (A.H.Y.C.)	

Chapter Number:	**Chapter 3**
Manuscript Title:	**Predicting the Risk of Asthma Attacks in New Zealand Using Machine Learning**
Publication Status:	**Published**
Reference if published:	**https://doi.org/10.24251/HICSS.2025.448**

AUTHOR SURNAME: (order as per manuscript)	CONTRIBUTION (May copy from the guidelines above)
W.K.D. Jayamini (W.K.D.J.)	**Conceptualisation: W.K.D.J.; Data Acquisition: A.H.Y.C, A.T., H.T., K.A.B.; Methodology: W.K.D.J.; Discussion: W.K.D.J.; Implementation: W.K.D.J., Validation: W.K.D.J.; Writing - original draft preparation: W.K.D.J.; Writing - review and editing: W.K.D.J., F.M., M.A.N., A.H.Y.C., A.T., H.T., K.A.B. ; Project supervision - F.M., M.A.N., A.H.Y.C.**
F. Mirza (F.M.)	
M.A. Naeem (M.A.N.)	
A.H.Y. Chan (A.H.Y.C.)	
A. Tomlin (A.T.)	
H. Tibble (H.T.)	
K.A. Beyene (K.A.B.)	

Chapter Number:	**Chapter 4**
Manuscript Title:	**Data-Efficient Machine Learning Approach for Predicting Asthma Attack Risk**
Publication Status:	**Published**
Reference if published:	**https://doi.org/10.1038/s41598-026-62302-y**
AUTHOR SURNAME: (order as per manuscript)	**CONTRIBUTION** (May copy from the guidelines above)
W.K.D. Jayamini (W.K.D.J.)	**Conceptualisation: W.K.D.J.; Data Acquisition: A.H.Y.C, A.T., H.T., K.A.B.; Methodology: W.K.D.J.; Discussion: W.K.D.J.; Implementation: W.K.D.J., Validation: W.K.D.J.; Writing - original draft preparation: W.K.D.J.; Writing - review and editing: W.K.D.J., F.M., M.A.N., A.H.Y.C., A.T., H.T., K.A.B. ; Project supervision - F.M., M.A.N., A.H.Y.C.**
F. Mirza (F.M.)	
M.A. Naeem (M.A.N.)	
A.H.Y. Chan (A.H.Y.C.)	
A. Tomlin (A.T.)	
H. Tibble (H.T.)	
K.A. Beyene (K.A.B.)	

Chapter Number:	**Chapter 5**
Manuscript Title:	**State of Asthma-Related Hospital Admissions in New Zealand and Predicting Length of Stay Using Machine Learning**
Publication Status:	**Published**
Reference if published:	**https://doi.org/10.3390/app12199890**
AUTHOR SURNAME: (order as per manuscript)	**CONTRIBUTION** (May copy from the guidelines above)
W.K.D. Jayamini (W.K.D.J.)	**Conceptualisation: W.K.D.J., F.M.; Data Acquisition: A.H.Y.C.; Methodology: W.K.D.J.; Discussion: W.K.D.J.; Implementation: W.K.D.J., Validation: W.K.D.J.; Writing - original draft preparation: W.K.D.J.; Writing - review and editing: W.K.D.J., F.M., M.A.N., A.H.Y.C.; Project supervision - F.M., M.A.N., A.H.Y.C.**
F. Mirza (F.M.)	
M.A. Naeem (M.A.N.)	
A.H.Y. Chan (A.H.Y.C.)	

Chapter Number:	**Chapter 6**
Manuscript Title:	**Perceptions Towards Using Artificial Intelligence and Technology for Asthma Attack Risk Prediction: A Qualitative Exploration of Māori Views**
Publication Status:	**Published**
Reference if published:	**https://doi.org/10.2196/59811**

AUTHOR SURNAME: (order as per manuscript)	CONTRIBUTION (May copy from the guidelines above)
W.K.D. Jayamini (W.K.D.J.)	Conceptualisation: W.K.D.J., F.M., A.H.Y.C.; Data Acquisition: W.K.D.J., M.C.B.P.; Methodology: W.K.D.J.; Discussion: W.K.D.J.; Implementation: W.K.D.J., Validation: W.K.D.J.; Writing - original draft preparation: W.K.D.J.; Writing - review and editing: W.K.D.J., F.M., M.A.N., A.H.Y.C., M.C.B.P.; Project supervision - F.M., M.A.N., A.H.Y.C.
F. Mirza (F.M.)	
M. C. Bidois-Putt (M.C.B.P.)	
M.A. Naeem (M.A.N.)	
A.H.Y. Chan (A.H.Y.C.)	

Chapter Number:	**Chapter 7**
Manuscript Title:	**Leveraging AI and Digital Technology for Holistic Asthma Management in the Pasifika Community**
Publication Status:	**Published**
Reference if published:	**https://doi.org/10.1016/j.teler.2026.100302**

AUTHOR SURNAME: (order as per manuscript)	CONTRIBUTION (May copy from the guidelines above)
F. Mirza (F.M.)	Conceptualisation: W.K.D.J., F.M., A.H.Y.C., R.L.; Methodology: W.K.D.J.; Discussion: W.K.D.J.; Implementation: W.K.D.J., A.M.I., K.P., A.F.M.; Validation: W.K.D.J.; Writing - original draft preparation: W.K.D.J.; Writing - review and editing: W.K.D.J., F.M., A.H.Y.C.; Project supervision - F.M., A.H.Y.C.
W.K.D. Jayamini (W.K.D.J.)	
R. Lutui (R.L.)	
K. Pole (K.P.)	
A.M. Ikihele (A.M.I.)	
A.F. Mu'aulama (A.F.M.)	
A.H.Y. Chan (A.H.Y.C.)	

Acknowledgements

First and foremost, I extend my deepest appreciation to my supervisory team. I am profoundly grateful to my primary supervisor, Associate Professor Farhaan Mirza, for his consistent guidance, intellectual leadership, and unwavering support throughout this research journey. His expertise, thoughtful feedback, and encouragement were instrumental in shaping the direction and quality of this thesis. I am also sincerely thankful to my co-supervisors, Professor Asif Naeem, for his technical expertise and guidance on the methodological and analytical aspects of this research, and Associate Professor Amy Chan, for her clinical insight, domain expertise, and valuable contributions that strengthened the practical and clinical relevance of the study.

A special note of thanks goes to the research collaborators and clinical advisors, including Dr Andrew Tomlin, Dr Holly Tibble, and Assistant Professor Kebede Beyene, for their generous support, advice, and contributions to this research. Their assistance with data access, methodological discussions, and critical feedback played an important role in strengthening this work. I am deeply grateful to the cultural and community experts and advisors who supported this research. I sincerely thank Marie-Claire Bidois-Putt, Māori advisor, for sharing her knowledge, guidance, and insights, which were essential in ensuring that the research was conducted in a culturally informed and respectful manner. I also acknowledge Arieta Fa'apesolo Mu'aulama, a Pasifika clinical advisor, and Moana Connect, a Pacific-led research and community organisation, for their invaluable support in facilitating community engagement and strengthening the connection between this research and the communities it seeks to serve.

I also extend my appreciation to the University of Kelaniya, Sri Lanka and the Science and Technology Human Resource Development Project of the Ministry of Education, Sri Lanka, for awarding me a full scholarship that made this PhD possible. I am also thankful for the financial support that facilitated the conduct of this research, including the New Zealand Health Research Council Summer Studentship and the DCT Faculty Contestable Funds, which provided resources essential to the successful completion of this work.

This research would not have been possible without the support of the School of Engineering, Computer and Mathematical Sciences at Auckland University of Technology

for providing a supportive academic environment, access to resources, and administrative assistance throughout my doctoral studies.

I am also grateful to my friends for their encouragement, companionship, and moral support during this journey. In particular, I would like to thank Uvini Perera, Nimesha Weerasinghe, and Avy Sheina, whose friendship, encouragement, and understanding helped me navigate both the challenges and milestones of this PhD.

Above all, I am deeply thankful to my family: mother, father, sister, and brother-in-law, for their unconditional love, patience, and unwavering belief in me. Their constant encouragement and reassurance sustained my motivation through challenging times and made the completion of this work possible.

This thesis is the result of the collective support, guidance, and encouragement of many individuals and institutions, and I am sincerely grateful to all who contributed to this journey.

I dedicate this thesis to my loving parents and to my sister and brother-in-law, with heartfelt gratitude for their unwavering support and encouragement.

Publications

Peer-Reviewed Full Paper Academic Publications

1. W. K. D. Jayamini, F. Mirza, M. A. Naeem, and A. H. Y. Chan, "Investigating Machine Learning Approaches for Detecting and Predicting Asthma: A Systematic Review". *PROSPERO*, 2024, https://www.crd.york.ac.uk/PROSPERO/view/402178

2. W. K. D. Jayamini, F. Mirza, M. A. Naeem, and A. H. Y. Chan, "Investigating Machine Learning Techniques for Predicting Risk of Asthma Exacerbations: A Systematic Review," *Journal of Medical Systems*, vol. 48, art. no. 49, May 2024, doi: https://doi.org/10.1007/s10916-024-02061-3. (**IF=5.7, Q1**)

3. W. K. D. Jayamini, F. Mirza, M. A. Naeem, A. Tomlin, H. Tibble, K. Beyene, and A. H. Y. Chan, "Predicting the Risk of Asthma Attacks in New Zealand Using Machine Learning," in *Proc. 58th Hawaii Int. Conf. System Sciences (HICSS)*, Jan. 2025, pp. 3732–3741, doi: https://doi.org/10.24251/hicss.2025.448. (**Rank A**)

4. W. K. D. Jayamini, F. Mirza, M. A. Naeem, A. Tomlin, H. Tibble, K. A. Beyene, and A. H. Y. Chan, "Data-Efficient Machine Learning Approach for Predicting Asthma Attack Risk," *Scientific Reports*, Jul. 2026, doi:https://doi.org/10.1038/s41598-026-62302-y (**IF=3.9, Q1**)

5. W. K. D. Jayamini, F. Mirza, M. A. Naeem, and A. H. Y. Chan, "State of Asthma-Related Hospital Admissions in New Zealand and Predicting Length of Stay Using Machine Learning," *Applied Sciences*, vol. 12, no. 19, p. 9890, Oct. 2022, doi: https://doi.org/10.3390/app12199890. (**IF=2.5, Q2**)

6. W. K. D. Jayamini, F. Mirza, M. C. Bidois-Putt, M. A. Naeem, and A. H. Y. Chan, "Perceptions Toward Using Artificial Intelligence and Technology for Asthma Attack Risk Prediction: Qualitative Exploration of Māori Views," *JMIR Formative Research*, vol. 8, e59811, 2024, doi: https://doi.org/10.2196/59811. (**IF=2.1, Q2**)

7. F. Mirza, **W. K. D. Jayamini**, R. Lutui, K. Pole, A. M. Ikihele, A. F. Mu'aulama, and A. H. Y. Chan, "Leveraging AI and Technology for Holistic Asthma Management in the Pasifika Community," *Telematics and Informatics Reports*, p.100302, Feb. 2026, doi: https://doi.org/10.1016/j.teler.2026.100302. (**IF=4.7, Q1**)

Abstract Publications

1. W. K. D. Jayamini, F. Mirza, M. A. Naeem, A. H. Y. Chan, A. Tomlin, H. Tibble, and K. Beyene, "Predicting asthma attacks in New Zealand using machine learning," *European Respiratory Journal*, vol. 64, suppl. 68, p. PA783, Sep. 2024, doi: https://doi.org/10.1183/13993003.congress-2024.PA783. (**IF=21.2, Q1**)

2. W. K. D. Jayamini, F. Mirza, M. C. Bidois-Putt, M. A. Naeem, and A. H. Y. Chan, "Perceptions towards using artificial intelligence and technology for asthma – A qualitative exploration of Māori views," *European Respiratory Journal*, vol. 64, suppl. 68, p. PA405, Sep. 2024, doi: https://doi.org/10.1183/13993003.congress-2024.PA405. (**IF=21.2, Q1**)

Awards and Achievements

1. Fully funded PhD scholarship awarded by the Science and Technology Human Resource Development Project (STHRD) held by the Ministry of Education, Sri Lanka, 2021 (Reference ID: STHRD/KE/LT-05).

2. Ethics Summer Studentship, awarded by the Health Research Council of New Zealand, 2022. (Reference ID: 22/925).

3. Outstanding poster award for the poster titled "*State of Asthma-Related Hospital Admissions in New Zealand and Predicting Length of Stay Using Machine Learning*" in the Data Science Research Center Poster Competition, AUT Research Week, 2023.

4. Nominated participant in the Doctoral Consortium at the Australasian Conference on Information Systems (ACIS) held in Wellington, New Zealand, 2023.

5. Faculty Contestable Research Funding awarded by the Faculty of Design and Creative Technologies, Auckland University of Technology (AUT), New Zealand, 2024.

6. Contestable conference funding award to participate in the European Respiratory Society (ERS) Congress held in Vienna, Austria, 2024.

7. Presenting the poster titled "*Predicting the Risk of Asthma Attacks in New Zealand Using Machine Learning*" at the New Zealand IT Project Management Research Symposium, 2025.

8. AUT Research Center of Computer Information and Sciences (RCCIS) award for conference attendance at the ACIS Conference held in Brisbane, Australia, 2025.

Nomenclature

Acronyms / Abbreviations

ACT	Asthma Control Test
AI	Artificial Intelligence
AQI	Air Quality Index
AUPRC	Area Under the Precision-Recall Curve
AUROC	Area Under the Receiver Operating Curve
BMI	Body Mass Index
CCI	Charlson Comorbidity Index
COREQ	Consolidated Criteria for Reporting Qualitative Research
CRISP-DM	Cross-Industry Standard Process for Data Mining
CV	Cross-Validation
DHB	District Health Board
DIHT	Data Imbalance Handling Technique
DT	Decision Trees
ED	Emergency Department
EHR	Electronic Health Records
FEV1	Forced Expiratory Volume in 1 second

FVC	Forced Vital Capacity
GBM	Gradient Boosting Machines
ICS	Inhaled Corticosteroids
ICU	Intensive Care Unit
KNN	K-Nearest Neighbour
LASSO	Least Absolute Shrinkage and Selection Operator
LoS	Length of Stay
LR	Logistic Regression
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
ML	Machine Learning
NN	Neural Networks
NPV	Negative Predictive Values
NZ	New Zealand
OHE	One Hot Encoding
PCA	Principal Component Analysis
PEFR	Peak Expiratory Flow Rate
PI	Pacific Islander
PPV	Positive Predictive Value
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PSI	Population Stability Index
RF	Random Forest

RFE	Recursive Feature Elimination
RFECV	Recursive Feature Elimination with Cross-Validation
RMSE	Root Mean Squared Error
ROS	Random Over-Sampling
RUS	Random Under-Sampling
SABA	Short-Acting Beta-Agonists
SLR	Systematic Literature Review
SMOTE	Synthetic Minority Oversampling Technique
SMOTE-NC	Synthetic Minority Oversampling Technique for Nominal and Continuous
SNP	Single Nucleotide Polymorphism
SVM	Support Vector Machines
SVR	Support Vector Regression
XGB	Extreme Gradient Boosting

Chapter 1

Introduction

This chapter provides background on the chronic respiratory disease, asthma and the importance of controlling and managing it. Then it discusses various strategies to support asthma management. The chapter outlines the rationale and significance of this thesis from both healthcare and scientific perspectives. Additionally, it presents the research gaps and questions, followed by an explanation of the research methodology. Finally, this chapter outlines the overall structure of the thesis by publication.

1.1 Research Background

1.1.1 Asthma: Definition and Global Prevalence

Respiratory diseases are a major global health burden affecting millions of individuals and imposing substantial economic costs on healthcare systems worldwide. Among these conditions, asthma is among the most prevalent chronic respiratory diseases. This condition is characterised by inflammation of the airways, causing intermittent airflow obstructions with symptoms including wheezing, coughing, shortness of breath and chest tightness [1, 2]. Often beginning in childhood, asthma affects all ages [3]. According to the Global Initiative for Asthma (GINA) report in 2024, asthma affects approximately 300 million people worldwide and causes around 1,000 deaths per day [2]. A study found that there were more than 260 million prevalent cases globally in 2021 [4]. A study examining global subgroups in 2019 reported continent-specific prevalence rates of 3.44% in Asia, 3.67% in Africa, 4.90% in South America, 5.69% in Europe, 8.29% in North America and 8.33% in Oceania [5]. In addition, a systematic analysis and modelling study conducted on global, regional and national prevalence of asthma in 2019 showed that the Western Pacific region had the most cases of ever-asthma at 156.55 (95% CI=124.2-196.14) million [6]. They have

categorised the countries into regions according to the World Health Organisation (WHO). These highlight the fact that asthma has a significant impact in the Western Pacific Region, which includes China, Hong Kong SAR, China, Macao SAR, China, Taiwan Province, Japan, the Republic of Korea, Brunei Darussalam, Singapore, Australia, New Zealand (NZ), New Caledonia, and Guam. Asthma exhibits a clear sex-based disparity. While its prevalence is high among boys under the age of 13 (nearly 65%), this pattern reverses in adulthood, with women showing a similarly higher prevalence compared to men in developed countries [7].

1.1.2 Asthma Prevalence in New Zealand

Asthma represents a significant health burden in NZ, with prevalence rates among the highest globally. Over 615,000 people in NZ use asthma medication, representing approximately one in eight [8]. This health condition has been identified as the third most common cause of death in NZ [8]. The Asthma and Respiratory Foundation NZ report in 2023 showed that approximately 12.4% of children under 15 years were receiving asthma medication [9]. Around 11.4% of adults were reported to be using asthma treatment [9]. Further, ethnic disparities were also reported in the NZ community. Māori and Pasifika represent approximately 17% and 8% of the NZ population, respectively. In the most recent surveillance report, Māori children, the Indigenous population of NZ, aged 2–14 years, had the highest prevalence of medicated asthma at 19.7%, followed by Pacific children at 10.3%, compared with European/Other children at 12.7% and Asian children at 7.3% [10].

As discussed above, the high asthma burden in NZ made it a highly relevant setting for this thesis. The fact that most members of the research team were based in NZ further supported this choice, as it enables practical access to the datasets, clinical insights and advisory support throughout the work. In addition, the higher asthma prevalence among different ethnic groups in NZ, combined with persistent health inequities across communities, provided an important context for examining asthma outcomes and for informing the development of more equitable approaches to care.

1.1.3 What is an Asthma Attack?

Asthma attacks, also called asthma exacerbations, occur when the airways suddenly become more inflamed and narrowed, resulting in a rapid worsening of the aforementioned symptoms, which makes breathing harder [11, 12]. During an attack, the airway muscles tighten, and excess mucus is produced, making breathing difficult. Asthma symptoms can be triggered by several factors, including respiratory infections, exercise, weather changes, environmental pollutants, exposure to pets and animals, and smoke [13]. If these episodes are not managed

promptly with appropriate treatments and medication, they can escalate to a medical emergency, which may lead to hospitalisation [14]. Asthma attacks are considered life-threatening as their severity and inability to be managed properly can cause death [14, 15]. Our work represents a step forward in the early escalation of such medical events, with the potential to save lives and reduce risks to individuals [16].

1.1.4 Importance of Asthma Management

Asthma symptoms are generally intermittent and reversible with proper treatment [17]. If asthma is not appropriately managed, then it might result in a life-threatening event. In addition to saving patients' lives and improving their quality of life, effective asthma management would significantly reduce national health care costs. Asthma imposes a significant economic burden globally and in NZ, as discussed above. Recent estimates indicate that the worldwide cost of asthma will reach USD 1.56 trillion by 2050 [18]. High-income countries continue to bear a significant portion of this burden, with their costs increasing from USD 130.29 billion to USD 133.37 billion over the same period [18]. In NZ, asthma also places a significant strain on the health system and society, with an annual cost of approximately NZD 1.20 billion (about USD 690 million), including NZD 276.00 million (about USD 159 million) in direct healthcare costs and NZD 918.90 million (about USD 528.3 million) in indirect costs such as lost productivity and early mortality [8]. With a population of about 5 million, these costs are substantial for NZ. Therefore, asthma management is critical from both individual and national perspectives.

Asthma management strategies can be divided into two main phases: *proactive* and *reactive*. We conceptualise these strategies as shown in Figure 1.1 from a computing technology perspective. It captures the types of strategies during different asthma attack episodes *prevention*, *prediction* and *post-care*. The following sections explain these different strategies and their subcategories.

1.1.4.1 Reactive Strategies

Acute asthma episodes are not always preventable. In such scenarios, there are some strategies that can assist with aftercare. *Reactive strategies* can be followed up as subsequent actions proceeding an acute episode. Digital technologies can support reactive asthma management by enabling rapid assessment and timely intervention during symptom worsening. Smart inhalers can automatically record rescue-medication use, allowing patients and clinicians to detect acute changes in control [19]. In addition, mobile health applications could also provide instant decision support during exacerbations, guiding the users through personalised

action plan steps [20, 21]. As previously explained, depending on the severity, asthma attacks could result in hospital admissions. In such scenarios, patients may need to stay in the hospital for several days to receive appropriate medical treatment and care. The number of days a patient stays in the hospital is referred to as the Length of Stay (LoS). If these hospital stays can be forecasted earlier, it would enable hospital staff to be prepared with the required resources to manage patients with asthma before the actual event occurs. This may include bed allocation and occupancy management, staff scheduling and Intensive Care Unit (ICU) and ward planning by identifying high-risk patients. For this, Machine Learning (ML) techniques can be used to predict a patient's LoS after admission. These predictions would help hospital management staff allocate existing resources effectively and efficiently.

Reactive strategies are necessary to manage health events once they occur; however, early action is better than reacting later. *Proactive strategies*, on the other hand, enable early action, reducing the risk of major complications and easing the burden on clinical systems while avoiding significant health events.

1.1.4.2 Proactive Strategies

Proactive strategies are strategies that can be implemented in order to prevent asthma attacks or minimise their impact and maintain control. As demonstrated in Figure 1.1, the *proactive strategies* can be subdivided into two: asthma attack *prevention* and *prediction*.

Prevention strategies can be tailored to demographics. Ethnic communities, gender, age, and regions may have specific considerations that enhance the effectiveness of prevention strategies. For instance, the health education needs of one ethnic group may differ from those of another. Therefore, appropriate community-based strategies need to be implemented to raise awareness of this health condition and its management.

Digital technology-based solutions, such as *smart devices*, are available on the market to assist asthma patients in managing and controlling their respiratory disorder. One potential solution is *digital inhalers*. These devices contain sensors that record details such as dates and timestamps of actuation, frequency and inhaler technique of using the device (for example, the orientation of using the device) that will appear on the patient's mobile application or the clinician dashboard, helping decision making. This allows the patient and clinician to understand if the medication is effective and accurate. Also, these outcomes will help clinicians observe any irregularities and changes in inspiratory patterns that might lead to an acute episode. Similarly, *smart peak flow* meters can store measurements and generate alerts to remind users to take regular readings through the corresponding apps. These solutions have improved the controller medication adherence and asthma control [19, 22]. Also, *mobile health (mHealth) applications* are designed to manage asthma by enabling users to routinely

track symptoms, medication use and lung function. These tools also provide reminders, education and personalised feedback which helps maintain consistent control [22].

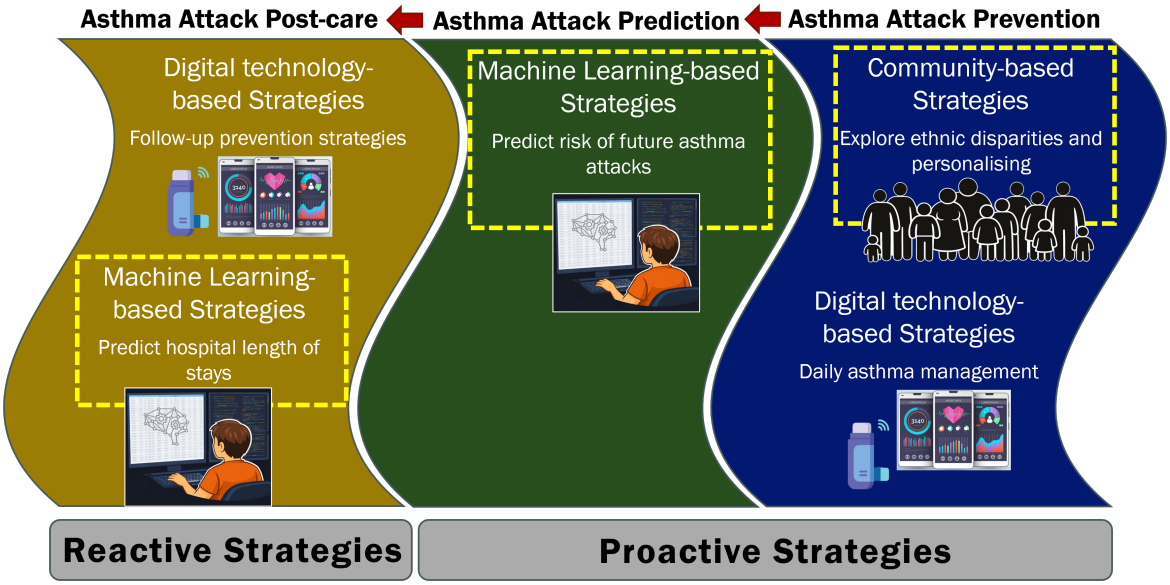


Fig. 1.1 Strategies for facilitating asthma management.

In addition to *prevention*, *risk prediction* is another key approach within *proactive strategies* to facilitate asthma management. This can help the patients be aware of any possible impending asthma attack. Hence, taking necessary actions such as optimising treatment strategies, improving medication adherence, increasing inhaled medication use, reducing exposure to triggers, while enhancing patient self-management would mitigate severe outcomes.

ML is an emerging technique used in supporting risk prediction [23, 24, 25, 26, 27]. The ability of ML algorithms to learn patterns and associations among variables and produce accurate outcomes has led to their popularity in risk prediction, including asthma attacks. There are several ML algorithms used in asthma attack risk prediction, such as Decision Trees (DT), Support Vector Machines (SVM), tree-based ensemble methods like Random Forest (RF), Extreme Gradient Boosting (XGB), Neural Networks (NN) and deep learning approaches. For developing prediction models, asthma-related outcomes can be determined at a healthcare encounter or treatment event. For instance, asthma-related Emergency Department (ED) visits or hospitalisations identified by International Classification of Diseases (ICD) codes J45/J46 [28] are considered as asthma-related admissions. Furthermore, the prescription of systemic oral corticosteroids is considered a primary marker of an exacerbation, either alone or as part of a composite outcome that includes an ED visit or hospitalisation [29]. Moreover,

if a person stays in the hospital for a period of time preceding an asthma attack, the LoS is another important outcome.

1.2 Rationale and Significance

Asthma is a major health condition, and its effective management is critical to prevent or mitigate life-threatening events among individuals living with asthma. As discussed in the previous sections, the management of this condition remains challenging due to its clinical complexity and variability across populations. A range of strategies can be implemented to address this health challenge, and this thesis contributes new knowledge while extending the boundaries of precision health.

In addition, this work expands current understanding of the application of ML and other computational techniques, such as feature selection and data imbalance handling, which are particularly relevant when working with real-world health data within this research scope. This thesis provides valuable insights into how limited and imbalanced datasets can be effectively utilised for predicting asthma-related outcomes.

Furthermore, this thesis offers empirical insights into the perceptions of different asthma-vulnerable communities regarding the use of ML and digital technologies to support asthma management, with a specific focus on asthma attack risk prediction. By incorporating these perspectives, the thesis highlights the importance of contextual and community-specific considerations when developing and deploying predictive health technologies.

In this regard, the study contributes to both scientific and applied domains, particularly within healthcare, for several important reasons, as discussed below.

1.2.1 Significance from a Healthcare Perspective

Asthma leads to deaths, prolonged suffering, and considerable physical, emotional, and social burdens for patients and their families. Therefore, the impact of asthma is substantial from both individual and societal perspectives. Accordingly, research in this area will be significant to personal and population healthcare management. In the following text, we present three reasons why this study is significant from a healthcare perspective.

Firstly, as discussed in the asthma management section 1.1.4 above, this medical condition generates a *huge cost* locally and globally. Asthma continues to impose a substantial economic burden globally, including both direct medical costs and wider indirect expenses, such as lost workdays, reduced productivity and premature mortality [30]. According to the Asthma Foundation NZ Update in 2023, all respiratory diseases, including asthma, chronic

obstructive pulmonary disease, and pneumonia, cost NZ about NZD 8.44 billion per year. They also reported that 1 in 11 hospital stays in NZ is due to respiratory diseases, including asthma. Therefore, predicting asthma attacks can reduce costs for both the healthcare system and individuals while enhancing quality of life.

Secondly, asthma requires *intensive regular monitoring*. Asthma is a heterogeneous condition that often remains manageable for patients when symptoms are absent or mild. However, asthma can worsen rapidly, leading to hospitalisation or death [31]. Asthma-related hospitalisations result in high morbidity and mortality. Therefore, early detection of changes or exacerbations is essential to ensure rapid mitigation. Asthma prediction is challenging and complex due to this heterogeneity in presentation and diverse multifactorial triggers unique to individuals [32, 33]. A study found that hospitalisation due to COVID-19 has increased among children (5–17 years) with poorly controlled asthma compared to children with well-controlled or without asthma in Scotland [34]. This highlights the risks asthma patients face, which might be affected by other medical conditions. It emphasises that this respiratory condition requires regular medical treatments and proper management because the disease's severity varies with other factors, including environmental and meteorological conditions [35, 36, 37], workplace conditions, and severe adverse life events [38]. If asthma is not adequately managed, it can lead to sudden exacerbations that require emergency health care and hospital admission for treatment and monitoring [35]. Therefore, asthma patients are at a high risk of needing medical care, which could be prolonged and costly to the health system. Hence, our research outcomes will support physicians and asthma patients in managing the condition.

Thirdly, there exist *health disparities* within *asthma-vulnerable communities* related to asthma management and treatment. Diversity within communities creates a need to investigate how the unique practices and beliefs of these communities can be valued and integrated into health-supporting systems and tools developed using Artificial Intelligence (AI) and digital technologies. Also, understanding high-risk asthma communities is critical for developing such tools to enhance the acceptability, enabling asthma management through lifestyle adjustments.

Finally, in many real-world settings, complete clinical data may be unavailable or accessible only in limited form. Therefore, a *generalised and optimised* computer-based system capable of predicting asthma-attack risk using only the most essential variables would be widely applicable across countries and regions. Such a tool has the potential to substantially reduce both regional and global economic burden.

1.2.2 Significance from a Scientific Perspective

From a scientific perspective, this research is significant because it addresses several existing challenges in applying ML to real-world healthcare, particularly in the context of a chronic respiratory condition, asthma. In real-world contexts, asthma outcomes are *highly imbalanced*, with severe events occurring less frequently than stable outcomes. Despite rapid growth in research on predictive modelling, there remains limited empirical understanding of how ML techniques perform when applied to large-scale, routinely collected, complex, and often imbalanced health data. Further, this gap limits the reliability and clinical relevance of many existing predictive models. Therefore, the thesis is important for gaining insights into ML applications in this research scope, especially in handling data imbalance.

Existing studies have identified a wide range of clinical, biological, hospital and environmental factors associated with asthma attacks and developed predictive models. However, those models often used specific datasets with several predictors, without focusing on the *generalisability* of the models *through feature optimisation and model simplification*, thereby enabling applicability across diverse populations and healthcare settings. As a result, there is a limitation in determining which predictors are important for asthma attack risk prediction in a generalisable and simplified model. Without a systematic evaluation of these aspects, predictive models may become overly fit to specific datasets or contexts, making them less reliable and interpretable when applied in broader clinical settings. Addressing this gap is essential for moving ML toward systems and tools that can be trusted and used across diverse populations and clinical contexts.

Furthermore, there is a lack of research that explicitly considers equity and population diversity, specifically in the NZ context, particularly for asthma-vulnerable communities. To our knowledge, few studies have explored health disparities among high-risk asthma communities in relation to asthma management and technological support. This thesis is therefore significant in responding to the need for more scientifically rigorous approaches that account for asthma-vulnerable population heterogeneity, particularly within the context of NZ.

Overall, the thesis is significant because it strengthens the applied methodological foundations of ML in the field of health risk prediction and prevention, particularly asthma. The thesis supports the creation of predictive methods that are simplified and appropriate for practical clinical use by addressing fundamental issues in data quality, imbalance, interpretability, and generalisability.

1.3 Research Gaps

To enhance proactive asthma management, we identified the following research gaps to be addressed in this thesis.

RG1 - Limited studies on developing robust ML models for imbalanced asthma attack data and identifying key predictors required for inclusion in a generalisable prediction model.

RG2 - Limited studies on predicting LoS using ML methods for asthma hospital stays.

RG3 - Limited studies exploring the perceptions of asthma prediction models in asthma high-risk communities in NZ.

1.4 Research Questions

The following research questions were formulated to address the research gaps identified above.

RQ1 - How do ML models perform in predicting asthma attacks with imbalanced data, and what are the key risk factors associated with future asthma attacks?

RQ2 - How accurately can we predict the asthma LoS using ML algorithms?

RQ3 - How do high-risk asthma communities in NZ perceive the use of AI and digital technology for predicting the risk of asthma attacks?
To explore this, we consider the asthma-vulnerable communities, Māori and Pacific Peoples in NZ.

1.5 Research Contributions

To address the above research gaps and questions, this thesis makes research contributions from both clinical and scientific perspectives, as outlined below.

RC1 - Demonstrating the potential of ML techniques in asthma-related predictive modelling in the NZ context.

RC2 - Identifying effective strategies for data imbalance handling in asthma predictive modelling.

RC3 - Identifying the most important asthma attack risk predictors.

RC4 - Providing a comprehensive performance analysis of ML models, highlighting their predictive accuracy and efficiency.

RC5 - Providing insights into which feature combinations contribute to model generalisation through comparative performance analysis.

RC6 - Providing insights into asthma hospitalisation patterns in NZ through exploratory analysis of hospitalisation data.

RC7 - Evaluating the predictive performance of ML models for asthma *LoS* prediction.

RC8 - Investigating perceptions of different asthma-vulnerable communities, including Māori and Pasifika, regarding the use of AI and digital technology for asthma risk prediction.

RC9 - Generating themes on using AI and digital technology for asthma attack risk prediction for Māori and Pasifika communities.

RC10 - Postulating a framework to strengthen asthma-related knowledge and inform future development of AI and digital technology-based solutions for Pasifika asthma care.

Figure 1.2 provides an overview of how the research gaps inform the research questions, which in turn lead to the research contributions and related publications.

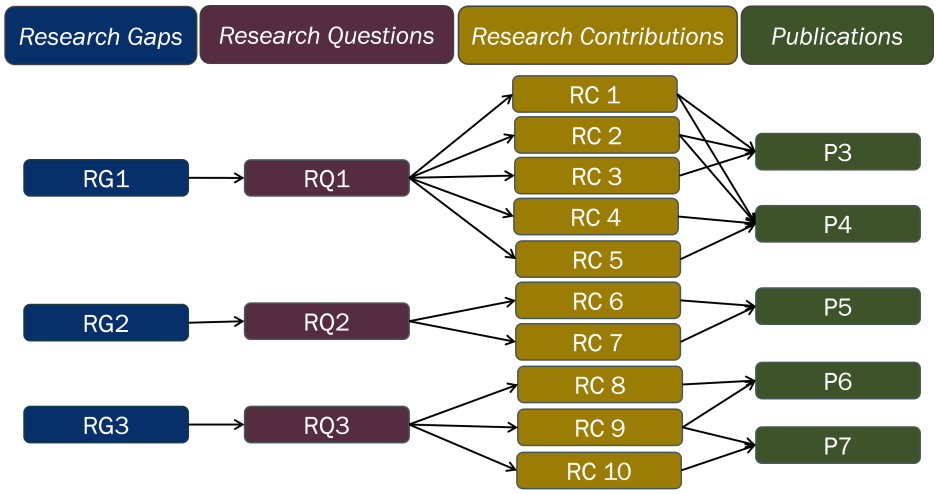


Fig. 1.2 Overview of the alignment between research gaps, research questions, research contributions, and the related research publications.

1.6 Research Methodology

This thesis adopts a comprehensive approach to investigate asthma management and risk prediction, and examines how these approaches can be, or need to be, effectively applied within Māori and Pasifika communities. To achieve this objective, there is engagement of clinical partners and advisors, hospitals and government bodies, such as the NZ Ministry of Health, providing suitable data to support the analysis and modelling. The data collected for this thesis includes hospital, pharmaceutical and demographic data. In addition, the thesis generated and utilised empirical qualitative data through interviews conducted within communities with the involvement of experts.

We have been fortunate to have the support of the Auckland Hospital, including a wider collaboration with clinical researchers and advisors as a part of this thesis. As this thesis involves health data and human participation, we had to adhere to the ethical standards of the National Ethical Standards from the National Ethics Advisory Committee (NEAC) and FAIR and CARE principles in careful management of real patient data to ensure the privacy and confidentiality of the data [39]. Accordingly, we obtained four separate ethics approvals and amendments to conduct this work (refer to Appendices H, D, N, T). To best illustrate the work we have done, Figure 1.3 provides a comprehensive overview of the thesis. The thesis utilises a mixed-methods approach including qualitative and quantitative methodologies such as semi-structured interviews, data analysis and ML.

To simplify the presentation of research methods that have been applied, we categorise the mixed-methods approach into two main parts. The first part relates to the quantitative

approach we followed using the Cross-Industry Standard Process for Data Mining (CRISP-DM) computational methodology. The second part deals with our engagement with asthma patients in the community. It presents the qualitative approach we followed, mainly through conducting interviews. The finer details of each method, the data involved, and the findings are presented in the publications (Chapters 2-7). Figure 1.3 presents research gaps, research questions, research contributions, and corresponding publications.

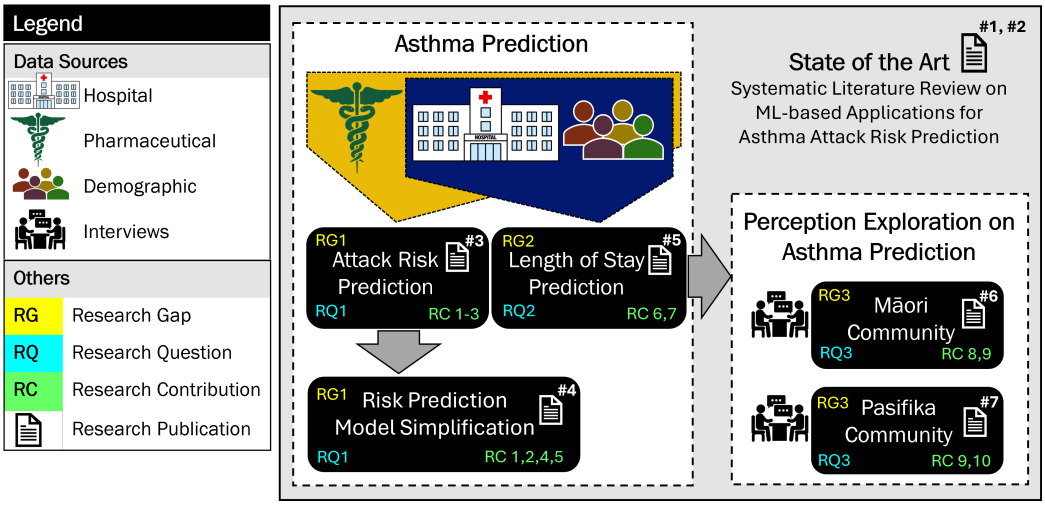


Fig. 1.3 Comprehensive overview of the thesis.

The *asthma prediction* phase of the thesis follows the CRISP-DM approach [40], which provides a structured methodology for predictive modelling. It addresses the research gaps RG1 and RG2 while answering the research questions RQ1 and RQ2. This methodology is demonstrated in Figure 1.4. This includes six phases. In the first and second phases, we iteratively try to understand the asthma-related hospital datasets and clinical context while identifying clinically meaningful outcomes. Next, we prepare and clean the data, followed by developing prediction models using different ML algorithms for generating the corresponding predictions. While developing algorithms, it may be necessary to revert to the previous step and conduct data pre-processing accordingly. Once the models are developed, their performance is evaluated using various evaluation metrics. Whilst the model performance is evaluated based on the metrics, we will reflect on our understanding of the scope. The final phase of this methodology involves providing different options with respect to different outcomes and feature sets that can be taken into account based on the requirement and the context. However, prior to real-world application, the models could be validated and configured according to the relevant context. The Chapters 3, 4, and 5 provide detailed information about this quantitative phase.

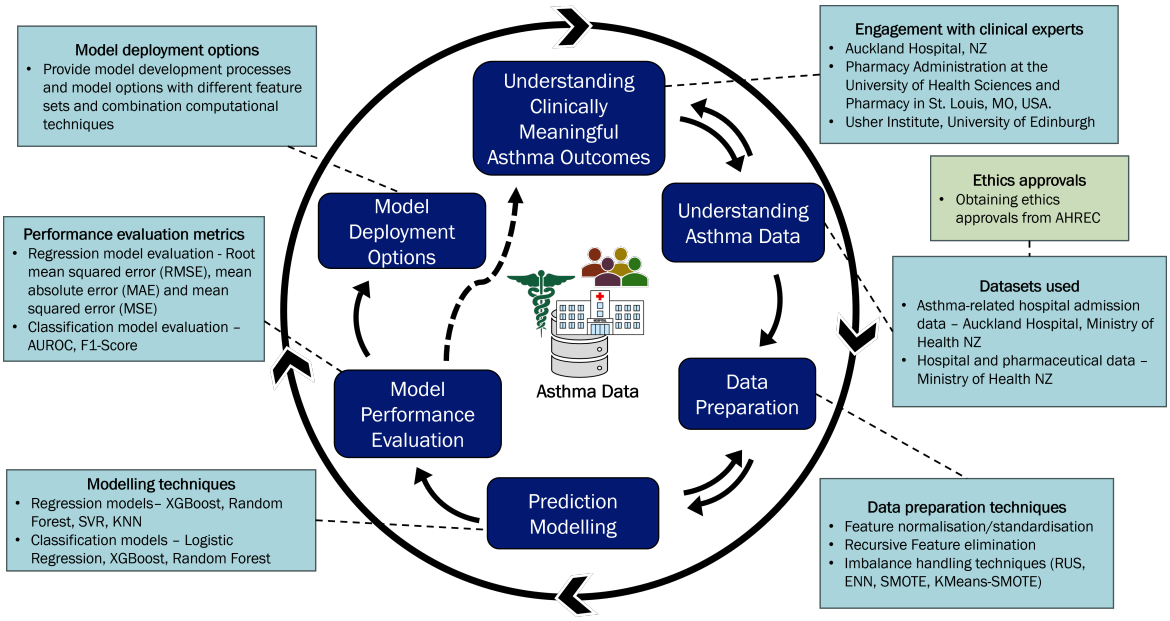


Fig. 1.4 CRISP-DM approach followed in the quantitative studies.

The qualitative research phase of the thesis follows a community-informed, exploratory design. This addresses the research gap RG3 and answers the research question RQ3. Here, semi-structured interviews and focus groups are conducted with Māori and Pasifika participants to explore perceptions of AI and digital technologies for asthma management. Inductive thematic analysis method by Braun and Clarke [41] is used to identify key patterns, values, concerns, and expectations, ensuring that the findings reflect the voices and perspectives of the communities most affected. The methodology used in the qualitative data analysis is demonstrated in Figure 1.5.

Initially, relevant documents, including an information sheet, interview guide, consent form, and a flyer, are prepared. Next, ethics approval is sought and obtained. Following this, the flyer is used to advertise the research and recruit participants for interviews. In addition, participants are recruited through the research team's network. During the recruitment, participants are provided with the information sheet, and the interviews will be scheduled at a convenient time and location, either online or offline. Interviews are then conducted by the interviewers from the research team, who are fluent in the preferred language of the interviewee. Interviews are recorded and stored anonymously and carefully in a password-secured location. Interview recordings are then transcribed into transcripts for the purpose of analysis. This is an iterative process done by the team, ensuring the correctness of the transcribing process. Subsequently, coding is conducted inductively, meaning that the codes are derived from the data itself rather than from pre-existing theoretical frameworks. Next, by examining the initial codes, we identify the meaningful and potential themes. Related codes

are grouped together, and preliminary themes and subthemes are constructed to represent shared meanings across the data. This step marks the transition from coding to thematic interpretation. The candidate themes are refined and evaluated in relation to both the coded data and the dataset as a whole through discussion within the team. This ensures that the themes accurately reflect meaningful patterns in the data. The final step involves writing up the findings of the analysis. The Chapters 6 and 7 in the thesis provide further details of this part of the work.

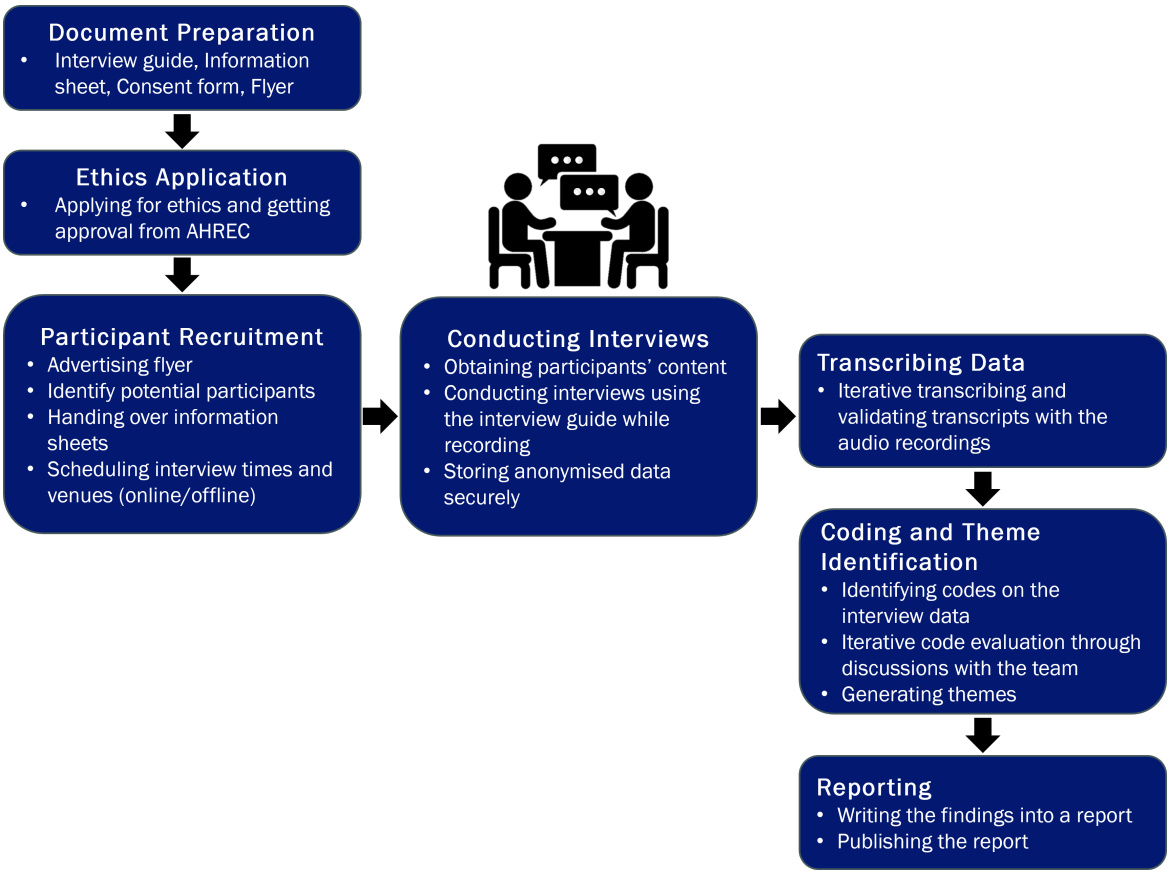


Fig. 1.5 Inductive thematic analysis methodology for qualitative studies.

1.7 Structure of the Thesis (Thesis by Publication)

The schematic representation of this PhD thesis structure is shown in Figure 1.6. This is a thesis by publication and is organised as follows. Chapter 2 provides a systematic literature review focusing on systematically reviewing recent applications of ML techniques in predicting the risk of asthma attacks to assist asthma control and management. Chapters 3 to 7 present the research publications that address RQ1, RQ2, and RQ3. These chapters comprehensively discuss the methodologies and findings. Chapter 8 provides a conclusion summarising the contributions, research findings, and future research directions. Finally, the Appendices present the related additional documents and information.

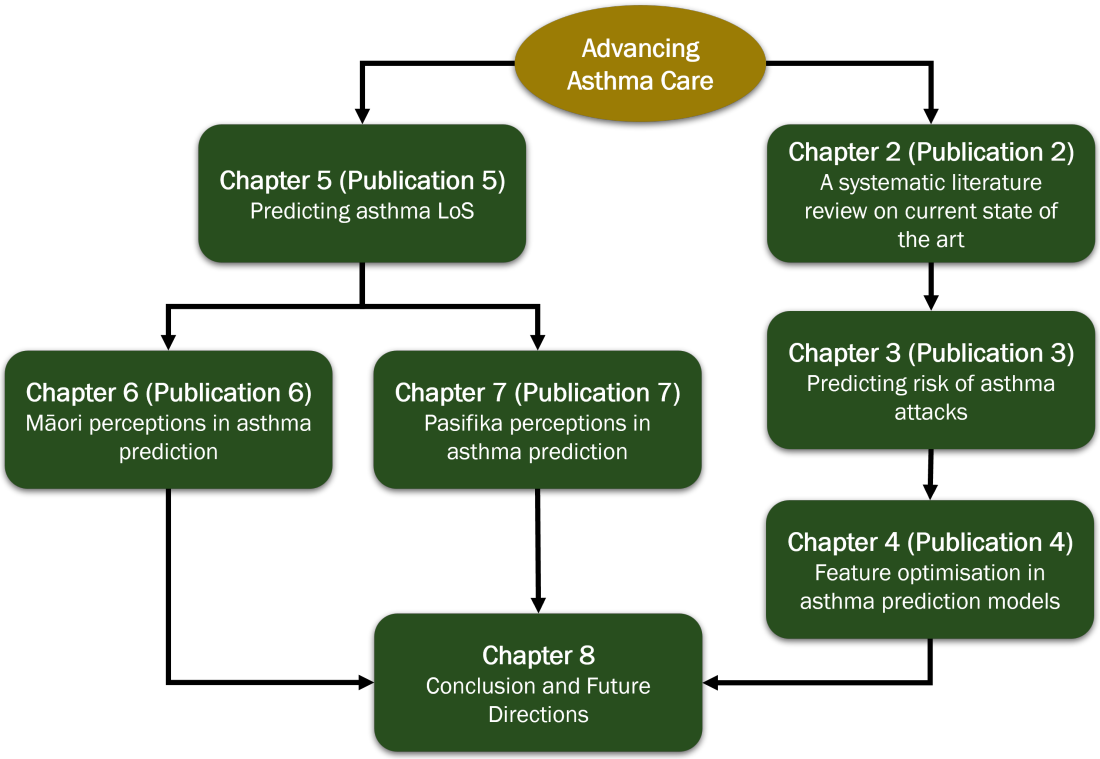


Fig. 1.6 Structure of the thesis (thesis by publication).

Chapter 2

Investigating Machine Learning Techniques for Predicting Risk of Asthma Exacerbations: A Systematic Review (Publication #2)

Citation: W. K. D. Jayamini, F. Mirza, M. A. Naeem, and A. H. Y. Chan, "Investigating Machine Learning Techniques for Predicting Risk of Asthma Exacerbations: A Systematic Review," *Journal of Medical Systems*, vol. 48, art. no. 49, May 2024, doi: https://doi.org/10.1007/s10916-024-02061-3 (**IF=5.7, Q1**).

2.1 Publication #2 Prelude

In asthma management, ML excels at well-defined tasks, such as diagnosis, prediction, medication selection, and management. However, there remain uncertainties about how ML can be applied to predict asthma attacks. An SLR was conducted to understand the state of the art in ML-based approaches for predicting the risk of asthma attacks. This chapter reviews previous studies in this area, focusing on several aspects of each. We compare the performance of ML-based prediction models for asthma attacks. Also, we categorise the studies by the nature of the outcome and discuss each. In addition, we identify the various data sources used in past research within this scope of study. Aggregating the findings from the review, we propose a conceptual model that summarises how ML and available data sources can be leveraged to develop effective models for the early detection of asthma attacks. We also present a list of data sources that other researchers may use in similar work in the

future. Furthermore, we present opportunities for further research and the limitations of the preceding studies.

2.2 Introduction

Asthma, a common long-term lung condition, is caused by inflammation in the airways of the respiratory system. Asthma affected 262 million people in 2019, and it causes an average of 461,000 deaths each year [14]. Often beginning in childhood, asthma affects all ages [3]. A diagnosis of asthma is based on clinical signs and symptoms, including wheezing, breathlessness, chest tightness, and coughing [2]. Asthma exacerbations could arise in patients due to multiple factors, including genetic, clinical, and environmental factors. For instance, if asthma patients are exposed to atmospheric changes such as dust, fumes, and pollen, their asthma can worsen, and it can lead to asthma attacks.

The reasons for asthma exacerbations are multifactorial, and worsening of asthma control may go unnoticed in patients if their symptoms are mild. However, asthma can worsen rapidly, leading to hospitalisation or death [31]. The National Review of Asthma Deaths (NRAD), conducted in the United Kingdom, found that 58% of patients with asthma who died were diagnosed as having mild or moderately severe asthma, with almost half having had no asthma review in the previous 12 months [42]. This highlights the need for innovative approaches to managing asthma and responding to attacks to ensure that care is delivered promptly. Another study found that hospitalisation due to COVID-19 has increased among children (5–17 years) with poorly controlled asthma, as compared to children with well-controlled or without asthma [34]. This highlights the increased risk that asthma patients face and emphasises the need for regular medical review, treatment, and proper management. This is important since asthma control can be affected by other factors, including environmental and meteorological conditions [35, 36, 37], workplace conditions, and severe adverse life events [38]. Therefore, early detection of asthma symptoms or exacerbations is essential to ensure rapid mitigation.

Even so, asthma prediction is challenging and complex due to its heterogeneous nature and the diverse multifactorial triggers unique to individuals [33, 43]. Patients with asthma are at a high risk of needing unscheduled medical care, which could be prolonged and costly to the health system. Any approach to predicting asthma attacks will enhance asthma management, reduce costs and increase quality of life. Whilst there has been a range of approaches to predicting the risk for asthma attacks that have been investigated in prior literature [44], the recent development of novel AI techniques has seen a rapid rise in efforts to predict asthma exacerbations. However, the clinical relevance of these techniques and the

potential to improve the accuracy and reliability of risk prediction is not yet known. There is a need to synthesise the available ML models that have been developed to a) identify the performance of these models in risk prediction and b) determine the range of clinical factors that have been included in these models. Understanding these factors is key if these models are to be used in routine practice for triaging patients based on their level of attack risk, or for informing changes in their treatment as their risk evolves, as part of clinical decision-making.

ML techniques that have emerged to manage and treat various diseases are increasingly being applied, particularly in terms of risk prediction. ML is developed on the foundations of mathematics, logic, probability, neuroscience, and decision theory. Using these concepts, various algorithms are developed to store important information from data elements through continuous training sessions. Consequently, the models built upon these algorithms have the ability to identify patterns and generate outcomes from complex data without any human-generated explicit programming code [45]. These computational methodologies have been found to outperform traditional statistical approaches, facilitating personalised medicine by addressing individual patient characteristics while processing vast amounts of data [46, 47]. Previous studies have shown that ML models can support asthma monitoring, prediction, diagnosis, and control in children and adults. These techniques complement clinicians if expertise or resources are limited [48], helping to prevent or mitigate tragic circumstances. This may be particularly true for exacerbation prediction, as it will be difficult for health practitioners to consider a variety of factors beyond medical and clinical ones when determining future attacks. Computer-based assistance can help predict impending attacks by considering a range of factors and assist clinicians in making appropriate decisions for the patient. The existence of a risk prediction model will help physicians monitor disease control. It can also help patients, as they will be made aware of any possible upcoming attacks.

There are three categories of ML algorithms: supervised, unsupervised and reinforcement. Supervised algorithms are used to solve classification and regression problems. Typical applications of these algorithms are seen in spam filters, fraud detection systems, recommendation engines, image recognition systems, and so on. Figure 2.1 shows some of the commonly used ML algorithms for making predictions.

This systematic review explores the application of ML techniques to predict the risk of asthma attacks. The review question was "How do machine learning techniques perform in predicting asthma attacks?" Specifically, we identify the ML techniques used, categorise the predictive models based on the outcome, and identify the suitability of ML approaches as they relate to the scenarios we analysed. We also identify the best-performing ML algorithms and provide prospective research directions.

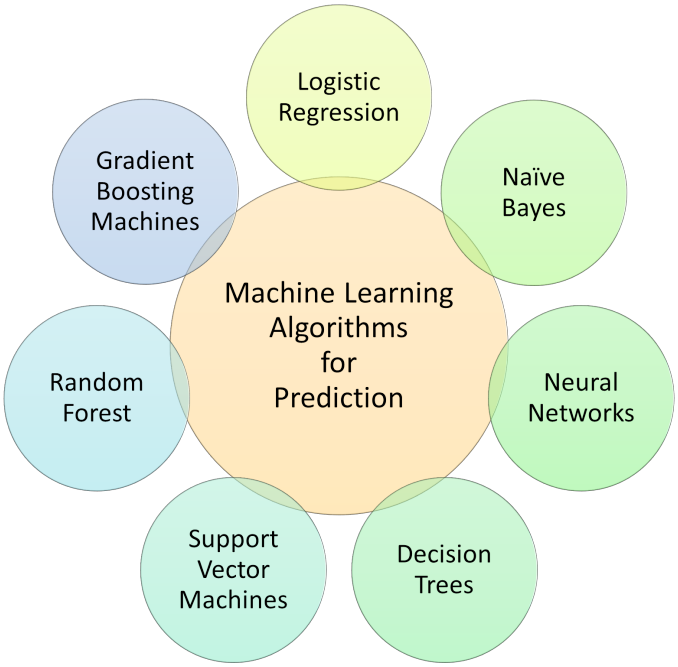


Fig. 2.1 Commonly used ML algorithms for making a prediction.

The subsequent sections of this chapter are organised as follows. The *Methodology* section explains the study design and methods employed in this investigation, while the *Results* section articulates the findings derived from the study. The *Discussion* section provides an in-depth discussion of these findings.

2.3 Methodology

This review adopted the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) reporting guidelines [49]. The protocol has been registered with the PROSPERO International Prospective Register of Systematic Reviews (Publication #1) (PROSPERO CRD42023402178). The PRISMA checklist is provided in Appendix B.

2.3.1 Inclusion Criteria

This study included primary studies of ML-based solutions for asthma in adults and children published in English from January 2010 to February 2023. No limitations were placed on the study design, setting, or minimum follow-up, but studies were limited to the English language due to the language capabilities of the research team.

2.3.2 Exclusion Criteria

Studies focusing on the diagnosis/prediction of asthma itself (as a condition), rather than asthma exacerbations, were excluded from the current study. Any studies that have followed a non-ML approach were excluded. Additionally, studies focusing on predicting asthma symptoms/control/severity level, Peak Expiratory Flow Rate (PEFR) and ED visits due to other diseases were eliminated from the study pool. Further, we excluded research work published as reviews, systematic reviews, editorials, letters, comments to the editor, book chapters, abstracts, conceptual papers, opinions, unavailable sources, protocols, commentary, and unpublished full-text documents.

2.3.3 Data Sources and Search Strategy

We searched the databases Medline (via PubMed), Cochrane (via Wiley), Embase, IEEE Xplore, and Google Scholar using the search terms Asthma AND (attack* OR exacerbat* OR control* OR symptom*) AND (detect* OR predict* OR diagnos* OR manag*) AND ("artificial intelligence" OR AI OR "machine learning" OR "deep learning" OR "neural network" OR computer-based OR "computer based" OR computer-assisted OR "computer assisted" OR "computer technology" OR technology).

2.3.4 Selection Process

The database search identified 860 eligible records. We used the tool RAYYAN for the study selection process [50]. After removing 122 duplicates, the titles and the abstracts of the remaining studies were screened. At the end of this step, we eliminated 667 records, and the rest of the studies were sent through full-text screening, which resulted in the selection of 20 studies to be included in our review. These were blindly reviewed by two different researchers. This study selection process is illustrated in Figure 2.2.

2.3.5 Data Extraction

A data extraction table was created to record the specific data extracted from the selected studies. The design of the table structure was informed by previous research [51, 52]. We extracted the year, country, title, aim, data source/s, sample size, research techniques, findings, evaluation metrics, model performance, and the limitations and strengths of each study. The effect measures extracted included the Area Under the Receiver Operating Curve (AUROC), accuracy, sensitivity, specificity, precision, recall, and F1-score, where available.

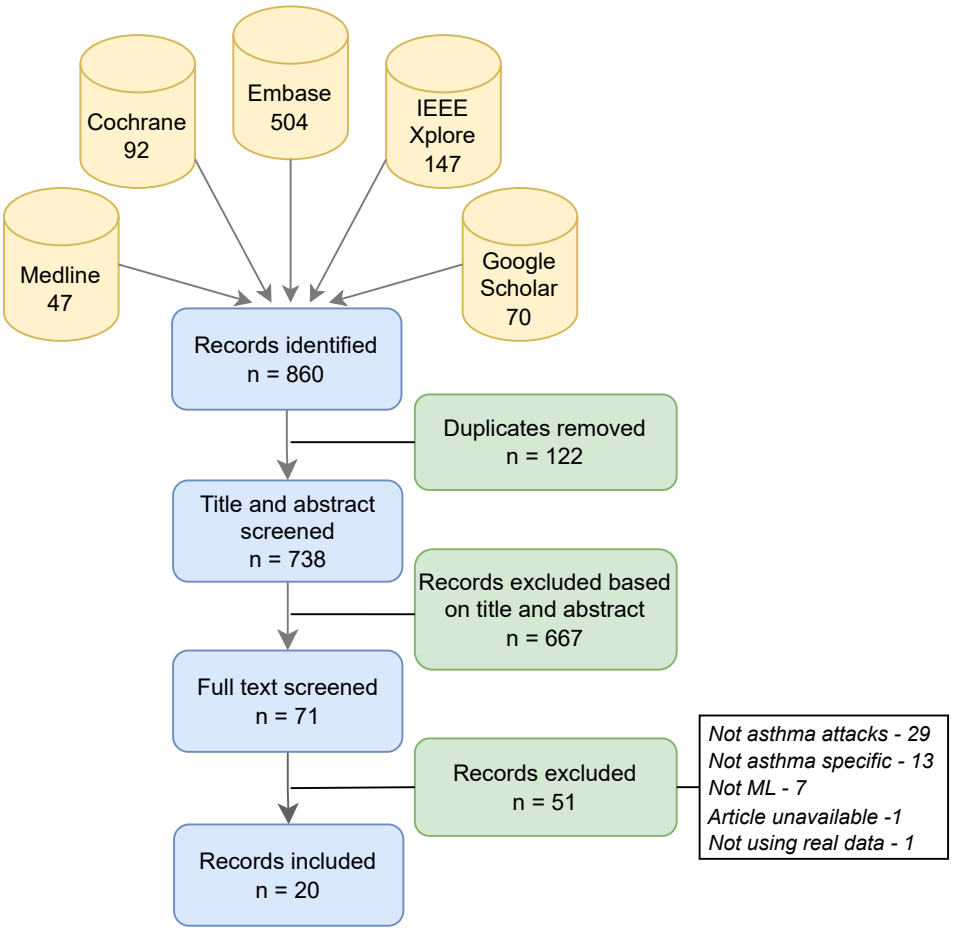


Fig. 2.2 Study inclusion process via PRISMA.

2.3.6 Risk of Bias Assessment

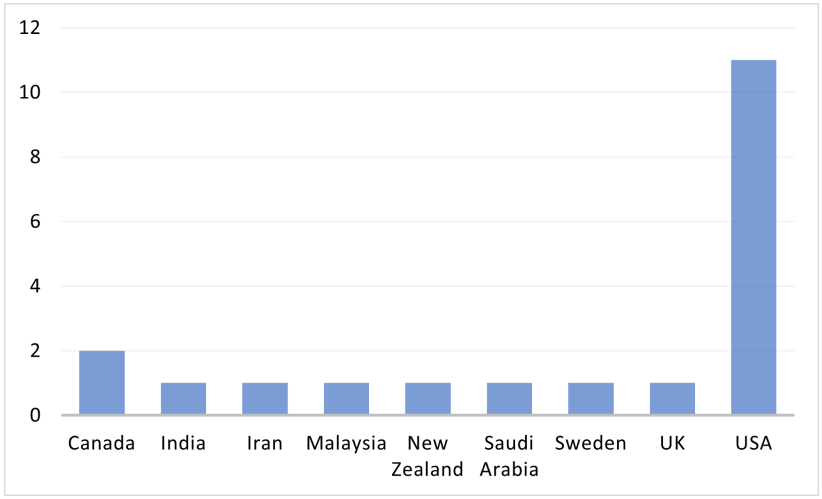
A risk of bias assessment of the selected studies was conducted using the Critical Appraisal Skills Programme (CASP) for Clinical Prediction Rule [53]. The CASP checklist was independently completed by two authors for each study. Any disagreements were resolved by consensus discussion. The Table A.1 shows the quality of each study.

2.4 Results

2.4.1 Characteristics of Included Studies

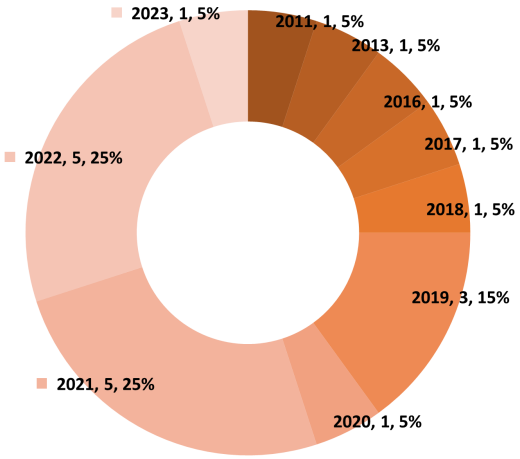
Characteristics of the 20 studies included in this review are elaborated in Figure 2.3. The included studies were conducted in a range of different countries, represented in Figure 2.3a. A study that used an international data set from multiple countries is not included in the

figure [54]. Importantly, as can be seen from Figure 2.3a, the United States of America is the main country of origin for many of these studies predicting asthma attack risk using ML techniques. As shown in Figure 2.3b, 75% of the studies were published as journal articles and the rest as conference papers. Figure 2.3c shows the distribution of the studies by year of publication. The majority of the studies have been published after 2020. The data sets employed in these studies have different sample sizes according to the number of participants or records (data instances). These details are presented in Table 2.1, showing the absolute values and percentages. Table 2.1 highlights the distribution of the two classes of the target variable: asthma attacks, both absent and present, as well as the portions of data used for training and testing the prediction models.



(a) Asthma risk prediction study count based on country.

(b) Publication formats of the included studies.



(c) Publication years of the included studies.

Fig. 2.3 Characteristics of the studies included in the review.

In this review, we identify the data sources that the previous studies incorporated to predict the risk of asthma attacks. Table 2.2 shows the different data domains, including biological, clinical, environmental and meteorological, hospital and medical, and socio-demographic, that have been used to develop the asthma risk prediction models. Clinical data may include asthma symptoms, PEFR, and inhalations, while prescribed medications and treatments come under medical data. Hospital data includes hospital admissions, prior attacks, comorbidities, ED visits, etc.

Table 2.1 Characteristics of datasets used in previous studies.

Study	Data sample size (total)				Training	Testing
	#Patients (P)	#Records (R)	Asthma attack absent	Asthma attack present	data	data
[55]	266	132,972	179 P (67.3%)	87 P (32.7%)	165 P (62%)	101 P (38%)
[56]		3,470	3,350 R (96.5%)	120 R (3.5%)	3,473 R (70%)	997 R (30%)
[57]		2435	2,287 R (93.9%)	148 R (6.1%)	1,452 R (70%)	983 R (30%)
[58]	29,396				17,638 P (60%)	11,758 P (40%)
[59]	298		234 P (78.5%)	64 P (21.5%)	184 P (60%)	60 P (20%)
[54]		728,535	727,959 R (99.9%)	576 R (0.1%)	582,828 R (80%)	145,707 R (20%)
[60]	5,982	17,907	12,862 R (71.8%)	5,045 R (28.2%)	4,008 P (67%)	1,974 P (33%)
[61]	3,057				2,447 P (80%)	610 P (20%)
[62]		334,564	322,420 R (96.4%)	12,144 R (3.6%)	315,308 R (94%)	19,256 R (6%)
[63]	3,678		2,855 P (77.6%)	823 P (22.4%)	2,942 P (80%)	756 P (20%)
[64]	109,488				98,823 P (90%)	10,665 P (10%)
[65]	21	655			524 R (80%)	131 R (20%)
[66]	240		106 P (45.4%)	134 P (54.6%)	216 P (90%)	24 P (10%)
[67]	10					
[68]		12,000			9,000 R (75%)	3,000 R (25%)
[69]		29,392	24,435 R (83%)	4957 R (17%)	23,514 R (80%)	5,878 R (20%)
[70]	581		404 P (69.5%)	177 P (30.5%)	417 P (72%)	164 P (28%)
[71]	60,302				48,242 P (80%)	12,060 P (20%)
[72]	178	3,970			3,000 R (75%)	970 R (25%)
[47]	31,433		29,171 P (92.8%)	2,262 P (7.2%)	25,146 P (80%)	6,287 P (20%)

Table 2.2 Data sources used in previous studies.

Data source	Previous studies
Biological	[70]
Clinical	[47, 54, 55, 56, 57, 59, 62, 63, 65, 68, 70, 71]
Environmental and meteorological	[65, 67, 69, 70, 72]
Hospital and medical	[47, 58, 59, 61, 62, 64, 66, 70, 71, 72]
Socio-demographic	[47, 58, 59, 60, 62, 63, 64, 67, 69, 70, 71]

2.4.2 Application of ML Models

The 20 studies used different ML techniques to predict the risk of asthma attacks. The outcome, asthma exacerbation, was treated either as a categorical variable or as a continuous variable in the form of a probability. Therefore, the studies can be categorised into 2 groups: 1) studies that predict the risk of asthma attacks as a classification (n=18) and 2) studies that predict the risk of asthma attacks as a probability (n=2). The classification group was further divided into 2 groups: studies with (n=11) and without (n=7) a prediction window. The studies with a prediction window can again be subdivided based on the window size: less than (n=6) and more than (n=5) a month. This categorisation of the studies is illustrated in Figure 2.4.

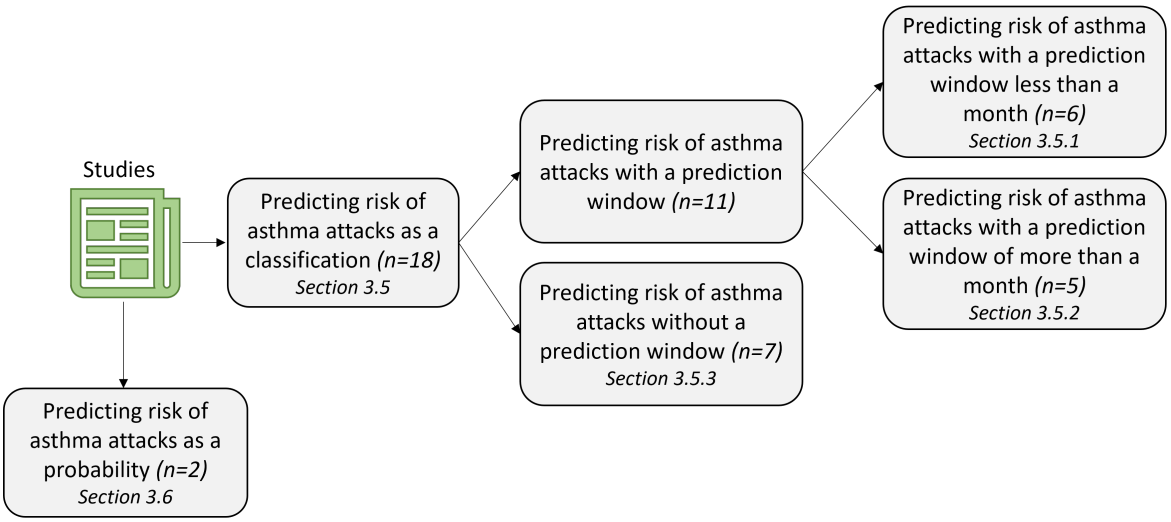


Fig. 2.4 Presentation of the results of the review.

In the literature, many studies predicted the risk of asthma exacerbations without considering the temporal effect. For instance, weather data from the previous day or a few days ago might have triggered asthma symptoms. Therefore, it is critical to consider the impact of the different factors from previous days (lags) in forecasting the risk of asthma attacks. Further, instead of just making a prediction, a group of studies constructed models to predict asthma attacks for a specific time (prediction window), such as the coming 3 days, 7 days, 3 months, 1 year, and so on. Details about these models are presented in the following sections.

Among the ML algorithms employed in these studies, Logistic Regression (LR), DT, RF, Gradient Boosting Machines (GBM), XGB, SVM, and NN algorithms were used most often. Most of the studies employed the k-fold Cross-Validation (CV) technique to validate the model on the training data. Different studies chose different k values such as 3 [69], 4 [58, 59], 5 [47, 54, 55, 63, 65] and 10 [61, 64, 66].

2.4.3 Hyperparameter Tuning

Hyperparameters in ML models are external configuration settings that are not learned from the data but are set prior to the training process. They influence the overall behaviour of the model and affect its performance. Hyperparameters can be optimised using different techniques. Among the studies included in this review, only a very few conducted hyperparameter tuning [55, 65, 69, 71]. The grid search technique was applied in two of these studies [55, 65] while the randomised search technique was applied in another one [69]. There are not enough details regarding the hyperparameter tuning process in [71]. Table 2.3 presents the details of the hyperparameter tuning conducted by past studies. It shows the various hyperparameters tuned with different values and techniques.

2.4.4 Model Performance

The studies used different evaluation metrics to evaluate and compare the performance of the models, as shown in Table 2.4. These metrics were predominantly accuracy, AUROC, specificity, sensitivity, Positive Predictive Value (PPV), and Negative Predictive Value (NPV). Accuracy is the ratio of correctly predicted outcomes and the total number of samples, simply the overall correctness of the model. The AUROC value represents the capability of the model to distinguish between the classes. Sensitivity, also called recall, measures the completeness of positive predictions, while specificity measures the completeness of negative predictions. PPV, also called precision, is the accuracy of positive predictions, while NPV is the accuracy of negative predictions.

2.4.5 Predicting Risk of Asthma Attacks as a Classification

Nineteen of the previous studies predicted the risk of an asthma attack as a binary classification, and one study [68] treated asthma attacks as a multi-class classification problem. Most studies in the binary category predicted the presence or absence of an asthma attack, while others considered different levels of asthma attack, such as mild, moderate, or severe. This section discusses studies that predicted the risk of asthma exacerbation as a category. Furthermore, imminent attacks could be predicted for a future time frame, for instance, the possibility of an attack in the next 3 days. While some studies constructed models by taking prediction windows into account, others did not. The following sub-section describes these groups.

Table 2.3 Hyperparameter optimisation details of the previous studies.

Study	Model	Hyperparameters	Values	Technique
[65]	DT	tree depth	2, 4, 6, 8 ,10, 12	grid Search
		split criterion	gini, entropy	
	GBM	maximum depth	3,7,9,11	grid Search
		subsample	0.5, 0.7, 1	
	LR	solver	newton-cg, lbfgs, liblinear	grid Search
		C (regularization strength inverse)	0.01, 0.1, 1, 10, 100	
	RF	subsample	10, 100, 1000	grid Search
		maximum features	sqrt, log2	
	SVM	C (regularization strength inverse)	0.1, 1, 10, 100, 1000	grid Search
		kernel	radial basis function, polynomial, sigmoid, linear	
[55]	XGB	number of trees	25, 100, 200	
		maximum depth	1, 3, 5, 7, 8, 9	grid Search
		learning rate	0.1, 0.3, 0.5, 0.7, 0.9	
	one class SVM	nu	0.001, 0.0015, 0.002, 0.004, 0.006, 0.008, 0.01	grid Search
		gamma	0.001, 0.01, 0.1, 1	
	LR	penalty	0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9	grid Search

2.4.5.1 Predicting Risk of Asthma Attacks with a Prediction Window of Less Than a Month

In the literature, many studies predicted the risk of asthma exacerbations without considering the temporal effect. For instance, the impact of weather data from the previous day or a few days ago might have triggered the symptoms of asthma patients. Therefore, it is critical to consider the impact of the different factors from previous days (lags) in forecasting the risk of asthma attacks. Further, instead of just making a prediction, a group of studies constructed models to predict asthma attacks for a specific time (prediction window), such as the coming 3 days, 7 days, 3 months, 1 year and so on. Figure 2.5 depicts the association between prediction window size and the model's performance. The figure highlights that the shorter the prediction window, the higher the model's performance. The following section discusses those studies. We synthesised these studies into two categories according to the size of the prediction window as follows. This section represents the studies that classified the risk of asthma attacks using prediction windows for less than a one-month period. Table A.2 shows a summary of the studies that developed ML models to predict asthma risk as a category using the prediction window concept.

Six studies developed models for short-term prediction of severe asthma exacerbations [54, 55, 56, 57, 58, 59]. Five studies [54, 55, 56, 57, 59] kept the prediction window at less than a week, while one study [58] used more than 2 weeks (15 days) for the prediction window. In training the ML models, the authors used data from several previous days, which they defined as a lookback window. Four of the studies [55, 56, 57, 58] applied

Table 2.4 Performance of the ML models developed for asthma risk prediction.

Application type	Study	Best model (overall)	AUROC[1]	AUPRC[2]	Accuracy	Specificity	Sensitivity	NPV[3]	PPV[4]
Classification Without prediction window	[65]	GBM	0.97	-	97%	-	0.98	-	-
	[66]	NB	-	-	70.7%	0.73	0.70	0.53	0.85
	[67]	NN	-	-	94%	-	-	-	-
	[68]	BPNN	-	-	86%	0.86	0.85	-	-
	[69]	XGB	0.84	-	-	-	-	-	-
	[70]	RF	0.50	-	-	-	-	-	-
	[71]	XGB	Non-severe: 0.71 ED visits: 0.88 Hospitalisation: 0.85	-	0.67 0.76 0.73	0.64 0.84 0.86	0.78 0.99 1.0	0.51 0.12 0.05	-
Classification With prediction window	[55]	LR	0.88	-	-	-	-	-	-
	[56]	CART	-	-	80.9%	0.97	0.65	-	-
	[57]	ABN	-	-	100%	1.00	1.00	-	-
	[60]	XGB	0.76	-	-	-	-	-	-
	[61]	Primary outcome: XGB Secondary outcome: RF	0.74 0.68	-	-	-	-	-	-
	[64]	Elastic-net LR	0.70	-	-	0.57	0.72	-	-
	[58]	Patients w/o comorbidities: XGB All patients: LightGBM	0.90 0.90	0.007 0.01	-	-	0.007 0.01	-	0.03 0.03
	[62]	XGB	0.86	-	-	-	-	-	-
	[59]	XGB	0.83	-	-	-	-	-	-
	[63]	RF	-	-	66.05%	0.68	0.65	-	-
	[54]	LR	0.83	-	-	0.83	0.90	-	-
Probability prediction	[72]	LSTM	RMSE[5] = 0.093	-	-	-	-	-	-
	[47]	TSANN	0.69	-	-	-	-	-	-

[1] Area under the receiver operating characteristic curve
[2] Area under the precision-recall curve
[3] Negative predictive value
[4] Positive predictive value
[5] Root mean square error

the lookback window concept, with the size of the lookback window ranging from 5 to 365 days, for prediction. However, the study [58] included inputs such as the count of events for multiple lookback window sizes - 10, 30, 60, 90, and 365 days. With the aim of exploring telemonitoring data for asthma risk prediction, and using the Minimum Description Length (MDL) principle, [57] found that the telemonitoring alert (out of four zones) on day 7 has higher importance in predicting the asthma risk on day 8. Comorbidity burden and previous exacerbations were important predictors, identified through collinearity [58]. Even though they have implemented Principal Component Analysis (PCA) and Recursive Feature

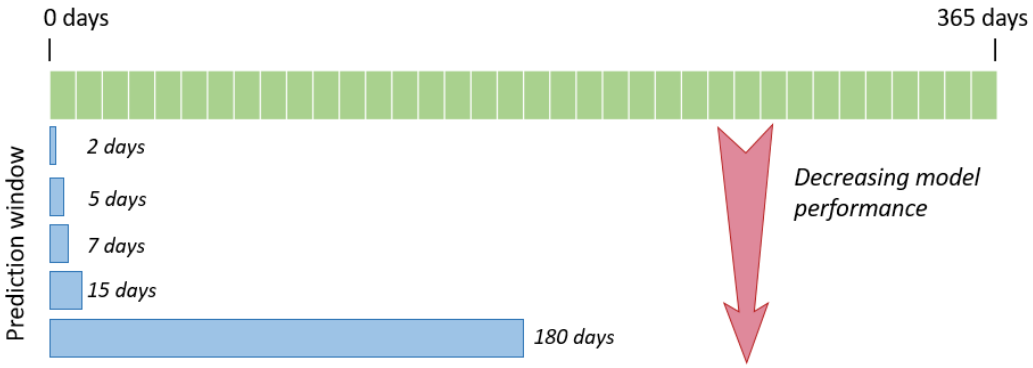


Fig. 2.5 Association between prediction window size and model performance.

Elimination (RFE) techniques to identify important features, these are not clearly stated in the article. Most works used tree-based algorithms such as DT [54], RF [58, 59], XGB [55, 58, 59], and CART [56]. Studies also developed models using LR [54, 55, 58, 59], SVM [55, 57], and NN [54, 58] algorithms. Only two studies applied Data Imbalance Handling Techniques (DIHTs)- Random Under-Sampling (RUS) [54, 58], Random Over-Sampling (ROS) [54], and Synthetic Minority Oversampling Technique (SMOTE) [54]. There is no clear data available for data imbalance handling in other works.

2.4.5.2 Predicting Risk of Asthma Attacks with a Prediction Window of More Than a Month

A set of studies defined their prediction window size as greater than or equal to a one-month period. One study [60] used a 1-month period while the other [61] used a 6-month period as the prediction window size. All of the other studies [62, 63, 64] kept the prediction window size to 1 year. While the studies [60, 61, 63] considered lookback windows similar in size to prediction windows, no clear details of lookback windows are provided in two studies [62, 64]. One study identified clinical factors such as obesity, atopy, medication, asthma controller plan and patient service utilisation history as important asthma risk predictors [60]. Asthma medication also played an important role [64]. Further, previous asthma exacerbations and length of treatment with biologics were key predictors of asthma risk [61]. Meanwhile, age, hospital stay, blight prevalence, and neighbourhood inequality are important predictors, according to another study [63]. In developing prediction models, the most common ML algorithms utilised by these studies are LR [60, 61, 64], RF [60, 61, 63, 64], and XGB [60, 62, 64]. Only one study in this category applied RUS to handle data imbalance [63].

2.4.5.3 Predicting Risk of Asthma Attacks Without a Prediction Window

This section focuses on the studies that developed ML models to predict the risk of asthma attacks without considering a prediction window. Table A.3 represents the study summary for these studies. Seven studies [65, 66, 67, 68, 69, 70, 71] are included in this category, and six of them considered this to be a binary classification problem. The seventh study [68] defined three target classes for risk as low, medium and high. A few studies specifically chose hospitalisations [69, 71] or ED visits [71], due to asthma attack being the target variable. Only two studies exercised feature selection. While the study [65] used the Least Absolute Shrinkage and Selection Operator (LASSO) regression technique, the other study [70] did feature selection with RF. With the aim of utilising environmental triggers and bio-signals for asthma risk prediction, one study [65] identified temperature, wind speed, relative humidity, Air Quality Index (AQI), smoking status, allergies, normal PEFR and daily readings, asthma symptoms of the last 4 weeks, and number of asthma attacks for the last 4 weeks as important predictors. Incorporating other data with community viral load, the study [69] determined the key predictors as patient vital signs, acuity, age, weight, socio-economic status, dry bulb temperature, wind speed, and station pressure. Another work mentioned the key features as age, long-acting beta-agonist, high dose inhaled glucocorticoid or chronic oral glucocorticoid therapy, Forced Expiratory Volume in 1 second (FEV1), and FEV1 to Forced Vital Capacity (FVC) ratio [71]. For developing the prediction models, commonly used ML algorithms were DT [65, 66, 69], RF [65, 69, 70, 71], GBM variations [65, 69, 71], LR [65, 69, 71], and SVM [65, 66, 68]. Additionally, some authors experimented with Naive Bayes (NB) [66] and NN extensions [67, 68]. Only one of the studies [66] did not possess a high imbalance data set. However, all the other works had highly imbalanced data sets, and they have not tried to handle it, other than the study [65], which applied SMOTE with the SVM technique to handle data imbalance.

2.4.6 Predicting Risk of Asthma Attacks as a Probability

Most of the previous studies on predicting the risk of asthma attacks addressed this risk as a binary or multi-class classification scenario. However, a prediction generated as a probability gives more meaning to the output. Rather than saying yes or no, a probability indicates the likelihood of something happening. Furthermore, a prediction presented as a risk score will help the medical practitioner better understand the status of the asthma patient. Accordingly, in the current review, we identified a few studies that used a probability or a score as the prediction outcome, which is clarified in this section. Table A.4 summarises the studies that predict the risk of asthma attacks as a probability.

One study presents a Deep Q-learning Network (DQN) framework to predict asthma attacks using the Q-learning method and Long Short-Term Memory (LSTM) to consider the personal risk scores of triggers [72]. The authors calculated a relative risk score as the probability of an event in the exposed group divided by the probability of an event in the non-exposed group. The authors used this separately with vital signs (respiratory rate, temperature, pulse, blood pressure and blood oxygen saturation) and environmental time series data, and they updated each record with a calculated personalised relative risk, presented as a numeric value. The Q-learning reward function was augmented with personalised risk scores, allowing the agent to favour actions associated with high-risk triggers and avoid non-trigger states. The deep learning algorithm, LSTM, was employed to predict the relative risk of asthma patients, which is then utilised in the Q-learning algorithm. Similarly, another study was conducted to predict the risk of asthma exacerbations while exploring the potential risk factors using LSTM [47]. They proposed a Time-Sensitive Attentive Neural Network (TSANN) by applying an attention mechanism on both code-level and visit-level variables, which eased the model's interpretability. They also added elapsed time embeddings into the model to indicate the relative time interval between each visit date and the prediction date. They identified gender, race, a few diagnoses (such as wheezing, chronic airway obstruction, inverse of regularization strength, nausea, chest pain, migraine) and medications as important predictors of asthma attacks [47].

2.5 Discussion

2.5.1 Data for Predicting Risk of Asthma Attacks

A conceptual model for predicting the risk of asthma attacks using ML techniques is presented in Figure 2.6. As asthma is a health condition affected by many factors, data from heterogeneous domains have been integrated to predict the risk of asthma exacerbations listed in Table 2.2. Asthma is affected by socio-demographic factors, including age, sex, education level, socio-economic status, and marital status [73]. Socio-demographic features, including age, gender, race, insurance type, and area of residence [47, 60, 61, 63, 64, 67, 70, 71] have been used with clinical factors such as asthma symptoms, PEFR, reliever medications, and inhalations [55, 56, 59, 65, 71]. Diagnoses, Body Mass Index (BMI), smoking status, previous asthma attacks, dispensed medications, Charlson Comorbidity Index (CCI), comorbidities, ED visits, and hospital visits were embedded as hospital and medical factors [58, 60, 61, 64, 65, 66, 72].

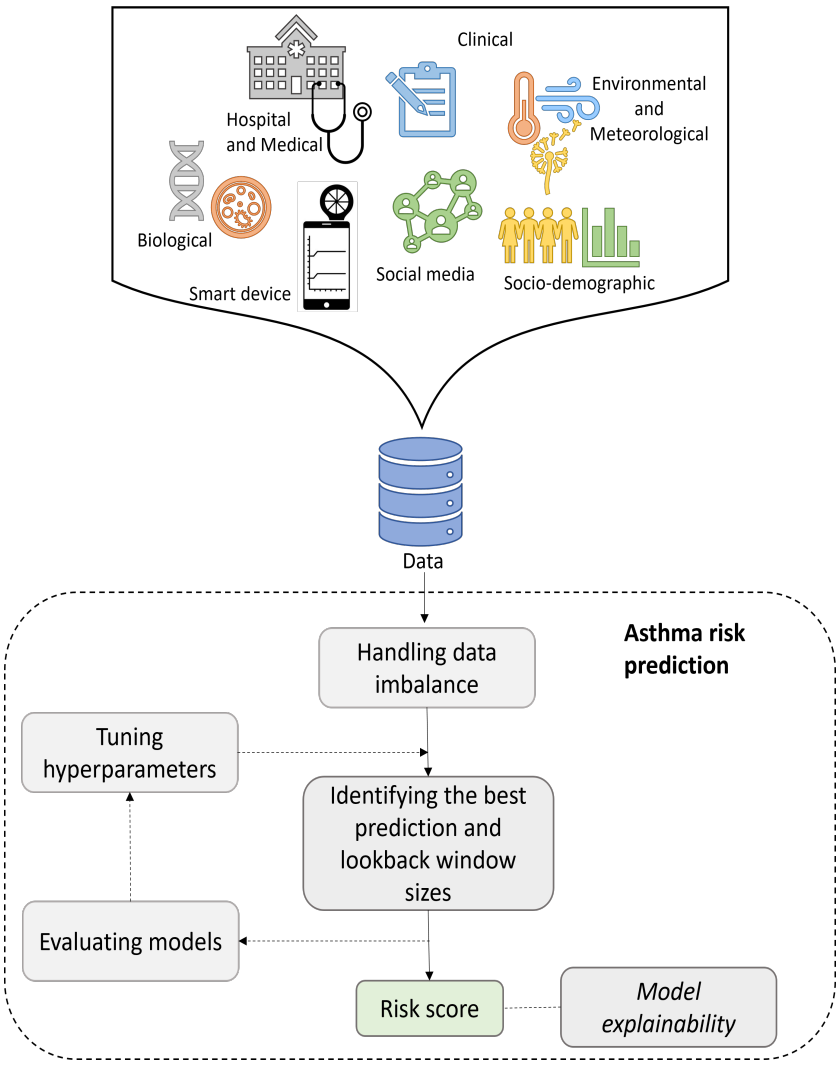


Fig. 2.6 Conceptual overview for asthma risk prediction using ML.

Environmental and meteorological data can be used for health forecasting with respect to respiratory diseases [74]. Meteorological data, such as temperature, humidity, air pressure, and wind speed, have impacts on asthma [35, 36]. Air pollutants can have chronic effects on humans and thereby pose a greater risk to a larger population. These factors are useful for predicting asthma since they could lead to a high number of asthma-related hospital admissions. Asthma triggers, including temperature, relative humidity, wind speed, AQI, air pressure, $PM_{2.5}$, pollen and gases like CO, NO_2, SO_2 were the environmental and meteorological factors investigated in the predictor list by [60, 65, 67, 68, 69, 72]. Additionally, we found one study using the biological factor, Single Nucleotide Polymorphism (SNP) [70].

Among several factors collected from multiple domains, previous studies have found that smoking status [65], normal PEFR and daily readings, previous asthma attacks, asthma

symptoms, medication usage [47, 58, 59, 60, 64, 65], comorbidities [58, 59, 60], and hospital stays [63] are important predictors of future asthma attacks. Socio-economic factors, including socio-economic status [69], age [63, 69, 71], gender, and race [47], have been shown to be important in predicting asthma attacks. Weather changes highly affect the control level of asthma patients, and accordingly, past studies show that temperature, wind speed, relative humidity [65, 69], and AQI [65] have strong predictive power.

These results show the importance of including multiple factors from different domains (see Figure 2.6). However, only a few studies have utilised predictors from more than three of the above heterogeneous domains. It seems important to incorporate different factors from multiple domains to forecast impending asthma attacks, as previous works have identified important predictors from each domain. Further, with the vast development of technology, smart devices such as smart peak flow meters and smart inhalers are in the hands of asthma patients. Utilising data from these devices will be more accurate than using data from daily diaries because manually entered data is prone to human errors and recall bias. Finally, social media usage has become a popular place where participants share their feelings and opinions. Therefore, it is worth investigating the use of social media data streams in asthma risk prediction.

2.5.2 Handling Data Imbalance

In real-world data, especially health data, the existence of positive cases is very low compared to the negatives. For instance, the proportion of asthma patients who are admitted to hospitals due to asthma attacks is very low compared to the number of patients who are not admitted to hospitals. This kind of research aims to identify the minority class over the majority. Hence, class imbalance needs to be carefully addressed when applying ML models that highly depend on data. Otherwise, the model will not be able to learn the patterns of positive cases and thereby produce a lower performance for the minority class.

There are several ways to handle data imbalance. The most commonly used techniques are RUS [54, 58, 63] and ROS [54]. In RUS, the number of records in the majority class will be reduced to match the number of records in the minority class, and vice versa for ROS. Additionally, SMOTE is a more advanced technique that is used to handle data imbalance. An alternative to the ROS technique mentioned above, SMOTE uses a more robust approach to generate new points based on the existing data. Further, the Edited Nearest Neighbours (ENN) and Tomek link are under-sampling techniques that can be used for down-sampling data.

Data imbalance should be handled; otherwise, classifiers would be biased or generate misleading results [35, 58, 75]. However, applying a data balancing technique is not necessary,

provided that the data set has considerable balance among target values. It should be noted that these data balancing techniques need to be applied only on the training set; otherwise, a data leakage issue arises. These techniques may improve the predictions in some cases by artificially eliminating or simplifying the records in the data set. The prediction model using the random under-sampling technique showed the best performance with higher specificity and sensitivity values compared to the other models (see Table 2.3). This highlights the importance of addressing data imbalance prior to developing the ML models.

2.5.3 ML Techniques in Asthma Risk Prediction

Clinical practices are mostly dependent on the professional skills that individuals acquire with practice and experience. However, the human brain has limited abilities to identify longitudinal patterns from massive amounts of data [45]. ML-based models can assist in analysing the growing amount of patient data, achieving patient-oriented decisions [76]. ML techniques can discover undisclosed patterns and find new information from the data. Costs incurred in healthcare could be reduced with the support of ML. Even though the knowledge and experience of a physician are irreplaceable, computer-aided support could increase the efficiency and accuracy of the diagnostic process. Therefore, ML could play a significant role in diagnosing and predicting asthma.

There are several ML algorithms that previous studies have used to predict imminent asthma attacks, as mentioned in Tables A.2 and A.3. In terms of classifying a patient to have or not have an asthma attack in the future, XGB, RF, and LR algorithms have achieved the best performance compared to other classifiers, regardless of the presence of a prediction window or not (see Table 2.4). On the other hand, both studies that predict the risk probability have used extensions of NN. In contrast to predicting the risk of asthma attacks as a category (yes or no), the probability of having an asthma attack will deliver more information to the health practitioner. The limited number of studies in this review that utilised probability prediction (see Table A.4) emphasise that further research is required to understand the applicability of ML techniques in predicting the risk of asthma attacks as a probability.

2.5.4 Prediction and Lookback Window

When anticipating the risk of attacks, it is more useful if the prediction can be made within a specified time window. As presented previously, researchers set the prediction window from 1 day [56, 57] to 1 year [62, 63, 64]. However, none of the studies mentioned a specific reason for choosing the window size. They used either a random value or chose the window size according to the nature of the dataset. From the analysis of the results of those

models, we identified that short-term predictions achieved better performance compared to long-term predictions (refer to Figure 2.5). In parallel, the researchers fed the past data into the prediction models within a lookback window, which is also called lag (i.e. past data for a given period of time). Incorporating historical data is crucial, especially when considering more common triggers, such as weather factors. While most of the studies have considered lookback windows, there is no clear information in defining the window size. Also, it seems that they have defined the window size according to the availability of data. There is a lack of knowledge about the optimum size of the lookback and prediction windows in predicting the risk of asthma attacks. Hence, it is important to explore and determine these windows for an accurate prediction.

2.5.5 Generating an Explainable Risk Score

Different clinical tools have been developed for monitoring and managing asthma. Global Initiative for Asthma (GINA) guidelines have been formulated based on the level of symptoms in the last four weeks to differentiate between controlled, partially controlled, and poorly controlled asthma. The Asthma Control Test (ACT) [77] and Asthma Control Questionnaire (ACQ) are the most common clinical scores used for asthma control. A study predicted the Pediatric Asthma Risk Score (PARS) with a logistic regression model by assigning weights to the predictors using the rounded odds ratios [78]. Many studies have used multivariate logistic regression to develop a risk score [79, 80, 81]. Accordingly, past models have considered a linear relationship between the predictors in generating the risk score or used the manually calculated scores as the target outcome of ML models. However, when the number of features becomes high, it is unreasonable to assume that they always have linear relationships. Therefore, developing a new asthma risk score requires more advanced technologies, such as ML, which can handle non-linearity among the predictors. While using these techniques, the prediction should be explainable to the clinicians in order for them to trust the score. This transparency is essential in the healthcare industry. The existing asthma risk scores are derived mainly from demographic and clinical factors. Instead of considering only the subjective factors, such as the prevalence of symptoms for a period, it is important to examine objective factors, such as peak flow and inhaler data, in deriving a control score. Asthma is a multi-factorial disease; it triggers patients' disease levels based on individual factors. Therefore, a more objective continuous score is necessary to predict the asthma control level. Further, rather than keeping the model a "black box", adding the reason behind the generated value would add more value and transparency to the prediction. It adds interpretability to the model, which will be clinically important.

2.5.6 Hyperparameter Tuning

ML algorithms comprise diverse parameters assigned varying values to optimise performance. Among these, model parameters, internal in nature, undergo configuration during the learning process as the model adapts to the data provided to it. A prime example is the weights of a neural network, acquired during the training phase. Conversely, hyperparameters, external in nature, necessitate the assignment of values before the training stage. Identification of hyperparameter values, termed hyperparameter optimisation or tuning, precedes model training. For instance, determining the optimal learning rate for a neural network is pivotal for achieving peak model performance. In practice, iterative fine-tuning of hyperparameters involves training multiple models with diverse parameter combinations, evaluating performance, and selecting the optimal model. This procedure can be executed by utilising the default values for the algorithm, manually configuring them based on past examples or experiments, or seeking assistance from experts to define these parameters [82]. Random searching is one of the strategies for hyperparameter tuning that will consider a pool of values and choose different combinations randomly to find the best set of hyperparameter values [83]. On the other hand, in the grid search technique, all the combinations of the specified set of values will be tried to determine the best combination which gives the highest model performance. A study compared the grid search and random search techniques to predict chronic kidney failure using XGB [84]. The best results came from the grid search technique. There are other hyperparameter optimisation techniques such as Bayesian optimisation [85, 86], genetic algorithm, etc. Experimental outcomes were comparable; nevertheless, the genetic algorithm outperformed the grid search technique and the Bayesian algorithm to predict customer transactions using NN [87]. However, except for a few works, the studies we reviewed lacked this important step. The ultimate goal of constructing these ML models for risk prediction is to gain the highest performance for the given problem. Therefore, conducting hyperparameter optimisation is crucial in developing ML models as it can improve the models' performance. Researchers must delve into suitable techniques for tuning hyperparameters in the machine learning models they construct.

In addition to the limitations of the included studies, several limitations of the review process should be considered. Although a comprehensive search strategy was applied across multiple databases, it is possible that some relevant studies indexed in other databases or published outside the selected sources were not captured. Furthermore, the review was limited to English-language publications, which may have resulted in the exclusion of potentially relevant studies published in other languages. The considerable heterogeneity among the included studies, particularly in terms of data sources, outcome definitions, machine learning

algorithms, prediction windows, and performance measures, also made direct comparisons across studies challenging and prevented the conduct of a quantitative synthesis.

2.6 Summary of the Chapter

This chapter presents an SLR conducted following the well-established PRISMA guidelines. With the growing use of ML in healthcare, several studies have employed advanced computer-based techniques to predict impending asthma attacks using heterogeneous data sources. Firstly, this review provided a conceptual overview of the research landscape (see Figure 2.4). Secondly, this review has generated a list of data sources available for researchers, which is available in Table 2.2. Thirdly, for each branch of the review, we presented lists of studies predicting the risk of an asthma attack as a classification with and without a prediction window (Tables A.2 and A.3) and prediction as a probability (Table A.4).

The findings of this review showed that research in this scope has utilised various data sources. The analysis showed the applicability of ML algorithms and their performance on asthma patient data. For classifying future attacks, ensemble methods based on decision trees frequently demonstrated a strong predictive performance, while extended neural networks have achieved acceptable results on probability prediction. However, direct comparisons between algorithms should be interpreted with caution due to heterogeneity in datasets and the absence of common benchmark datasets among the included studies. Further, the review found that a shorter prediction window improves the prediction outcome. For future research on ML applications for asthma prediction, hyperparameter tuning and data imbalance handling need to be carefully addressed to integrate data from multiple sources. Additionally, optimal lookback and prediction windows for asthma risk prediction require further investigation. Moreover, deriving an interpretable asthma risk outcome using ML techniques needs to be explored to assist health practitioners.

The new insights generated from the SLR motivated the conducting of the research presented in the following Chapters 3 and 4 in the thesis.

Chapter 3

Predicting the Risk of Asthma Attacks in New Zealand Using Machine Learning (Publication #3)

Citation: W. K. D. Jayamini, F. Mirza, A. Naeem, A. Tomlin, H. Tibble, K. Beyene, and A. H. Y. Chan, "Predicting the Risk of Asthma Attacks in New Zealand Using Machine Learning," in *Proc. 58th Hawaii Int. Conf. on System Sciences (HICSS 2025)*, Jan. 2025, pp. 3732–3741, doi: https://doi.org/10.24251/hicss.2025.448.

3.1 Publication #3 Prelude

As identified in the SLR, various factors can affect asthma control and, in turn, trigger asthma attacks. One important source of this data is hospital data, which provides direct documentation of asthma attacks. Identifying the key factors that increase the risk of asthma attacks is crucial for timely patient management. Accordingly, this chapter identifies risk factors for asthma attacks in NZ and compares the performance of ML algorithms in predicting these risks using hospital data. Further, there is a high data imbalance in asthma hospital data, in which there are very few instances with asthma attacks, the minority class, and a huge number of instances without attacks. Therefore, to address the research gap RG1 and the research question RQ1, in this chapter, we apply various DIHTs to hospital data, combined with different ML algorithms, and compare their performance to determine which method is more accurate within the study scope.

3.2 Introduction

Asthma is a non-communicable disease and one of the most common chronic respiratory diseases. The World Health Organisation states that, in 2019, approximately 262 million individuals were impacted by asthma, resulting in 455,000 fatalities [14]. The highest prevalence of asthma in adults is seen in the Western Pacific region [88]. NZ is one of the countries with the highest prevalence of asthma, where more than 597,000 people are on medications for asthma [89]. Often beginning in childhood, asthma affects all ages [3]. According to the NZ health surveys, one in every seven children (13%) and one in every eight adults (12%) are on pharmacological treatments for asthma [90]. Overall, the cost of asthma to the NZ society is around NZ 1 billion dollars per year [91]. This highlights the massive economic burden on the health system and government because of this medical condition. A combination of coughing, wheezing, breathlessness, and chest tightness is recognised as a symptom of asthma [14]. Multiple factors, including temperature, humidity, pollen and particle matter concentration, can trigger the worsening of asthma, resulting in asthma attacks. Asthma symptoms can worsen for different reasons and may lead to an asthma attack, which might be life-threatening. Consequently, this may result in hospital admission or even death [31]. Therefore, asthma control and management are critical for asthma patients as individuals and for the country as a whole. The aforementioned facts prove that there is still space to assist individuals with asthma, as well as health practitioners. Therefore, research in this area will be significant to personal and population healthcare management.

Asthma is a common reason children get admitted to hospitals in NZ [92]. Asthma-related hospitalisations result in high morbidity and mortality. Therefore, early detection of changes or exacerbations is essential to ensure rapid mitigation. Asthma prediction is challenging and complex because of this heterogeneity in presentation and diverse multi-factorial triggers that are unique to individuals [33, 93]. Therefore, early detection of changes or exacerbations is essential to ensure timely intervention. ML is a broader area of computer science that focuses on the development of algorithms that enable computers to learn and make predictions or decisions without being explicitly programmed. In the last few decades, AI and ML have taken up a greater space in the research field, with the application of these algorithms in various sectors. One of the major areas of these applications is the health industry. Many studies have developed ML models to diagnose, predict, control, and manage diseases. ML models have been recently constructed to predict the risk of asthma attacks in different regions of the world, for example, the USA [56, 59, 60, 61, 62, 70, 71], Canada [64, 66], and Europe [58, 69]. However, these techniques have been applied only to a limited extent in the NZ context, where there is a higher prevalence of asthma. It is important to explore

the applicability of ML algorithms for asthma risk prediction in this geographical region, as there can be different associations between the features because of existing diversities among different regions. The other important reason to investigate attacks in NZ is the rising risk of asthma attacks, and there are unique factors to consider for NZ, such as the worse asthma attack rates in Māori and Pasifika [28, 94]. The study focused on constructing prediction models using a few ML algorithms for asthma attack risk prediction.

The rest of the chapter is structured as follows. The *Literature Review* section will cover the literature review conducted within the work, and the data source and methods used in the study are explained in the *Methodology* section. Subsequently, the *Results* section will present the findings of the experiments, followed by the *Discussion* section, which critically evaluate the findings.

3.3 Literature Review

Many studies have been conducted to build prediction models using ML algorithms using data from various sources. Most of the input features of these models were clinical [55, 56, 59, 62, 65], hospital [61, 64, 66, 72], socio-demographic [58, 59, 60, 64, 67] and weather factors [33, 65, 67, 70, 72]. Some of the studies aimed to diagnose asthma as a disease using ML techniques. Biological, medical, and clinical data were used for the classification of asthma and non-asthma. A study classified asthma patients into three groups: mild, moderate, and severe based on severity using spectral integrated features extracted from tidal breathing observations in the trachea and lower lung using a wireless digital stethoscope [95]. They applied K-Nearest Neighbour (KNN), SVM, and ensemble classifiers. Another study [96] identified five clusters of asthma patients using agglomerative clustering and seven rules to classify patients using the C5.0 Decision Tree algorithm.

Apart from diagnosing asthma disease itself, several other studies [54, 65, 66, 67, 68, 69, 71] focused on predicting the risk of asthma attacks using ML algorithms. All these studies developed models to predict whether or not an asthma attack can occur in the future, hence treating asthma attacks as a binary outcome. While many of the studies targeted the asthma attack directly as the target variable, others considered hospitalisation or visiting the emergency department as the occurrence of asthma attacks. A study developed ML models for early prediction of asthma attacks based on bio-signals and environmental data from 21 volunteers, integrated based on the patient's location and date [59]. They excluded wind speed during the feature selection process. However, wind speed can increase pollen transmission in the air, which can cause breathing difficulties [67]. Therefore, eliminating such important factors may impact the model's predictions. The authors applied the SMOTE using SVM

over the full data set to handle the data imbalance in the target class [65]. However, the use of SMOTE should only be done on the training data; otherwise, it can lead to data leakage. Thereby, this might have impacted the model performance. They developed five prediction models to forecast asthma attacks using the ML classifiers, DT, GBM, LR, RF and SVM. Almost all the models with the selected feature set performed better, while the GBM model had the highest performance with 97% accuracy, 98% recall and 0.97 AUROC on the test data. Comparatively, the poorest performance has been shown by the DT model. The study authors claimed the integration of bio-signal data with environmental data as their research contribution.

A group of researchers developed asthma prediction models with ML in the first phase and compared the results with the Pediatric Respiratory Assessment Measure (PRAM) score and predictions made by ED physicians in the second phase [66]. PRAM score is generated based on 5 clinical attributes, and it maps the values of those attributes into a 4-point scale. In phase 1, Naive Bayes (NB) had 68% accuracy and 0.74 AUROC, while the NB model was less accurate (70.7%) than the PRAM score and physicians in phase 2. However, the sensitivity, specificity, PPV and NPV of the NB model were 0.70, 0.73, 0.85 and 0.53, respectively. Therefore, the model has space for further improvement. The authors expect that the model's performance could be improved by adding attributes related to the tacit knowledge physicians use in patient assessments. Another study [67] built a classification model to predict the chances of asthma attacks, as high or low, based on weather triggers. Two NN models were developed, one for personalisation and the other for generalisation. In addition to weather attributes, the ACT scores were inputs for the models. Gender, age group, location, outdoor occupation, and outdoor activities were additionally introduced to the generalisation model. Although the models performed well, this needs to be further validated, considering the imbalance that prevailed in the dataset. Both models gave an accuracy of 94% in asthma risk prediction. However, despite the high accuracy, the authors do not mention how they produced the training and test data, and no information is available on data processing and model optimisation. In-depth details of the model architecture are lacking, which limits the replication of the work. Additionally, the model performance has not been validated, especially with imbalanced data sets. Employing only weather data will not be enough for a personalisation prediction model, as there are other personal factors from medical, clinical and hospital data that warrant consideration. An ML approach to predict the need for hospitalisation due to asthma exacerbation at the time of ED triage for paediatric asthma was explored [69]. The study incorporated clinical, weather, neighbourhood characteristics (home environment and socioeconomic status) and community viral load data. The authors included classical ML algorithms: DT, LASSO Regression, RF and XGB. The XGB model

yielded the highest with a 0.84 AUROC and showed that patient vital signs, acuity, age, and weight, followed by socio-economic status and weather features, were the most important for hospitalisation prediction. The authors found that adding weight, socio-economic status, and weather data improved the performance of the models. However, the study did not consider the high data imbalance in the analyses.

Using the RF classifier, the research work selected the most important SNPs collected from the Genome-Wide Association Study (GWAS) and developed a paediatric asthma exacerbation prediction model [70]. Important SNPs were selected using RF feature importance. The prediction models gave an AUROC of about 0.50. The performance is low and needs to be further improved. Exploring other ensemble algorithms, such as XGB, may improve the results. The study developed ML models using large-scale real-world local data to predict asthma exacerbations [71]. The researchers utilised one dataset for model development and a separate dataset for external validation, which is significant. Mutually exclusive models were developed to predict asthma exacerbation considering three actions: ED visits, hospitalisation and non-severe exacerbation. Like many other studies, LR, RF and Light Gradient Boosting Machine (LightGBM) algorithms were used to develop the models. As in the study [69], LightGBM performed better with the highest AUROC (None-severe exacerbation 0.71, ED visits 0.88, hospitalisation 0.85) on both test and external data sets. Further, to add interpretation to the models, they used Shapley Additive (SHAP) Values and represented them graphically, improving the prediction models' transparency. As the study used data from day-to-day patient care Electronic Health records (EHR), the data reliability is high. However, they might have missed other admission/treatment data by being limited to a single healthcare provider. Apart from these binary classification studies, one study predicted the risk of asthma in three categories: low, medium, and high [68]. They propose a fog-based healthcare system using the NN model to predict asthma risk. The prediction model takes health and environmental data, such as obesity, smoking, age, carbon monoxide (CO), temperature, humidity, and nitrogen dioxide. In addition, they derived the predictor, "asthma status", using PEFR, exhaled air temperature, and FEV1 values. The results showed that the Back-propagation Neural Network (BPNN) model had a better performance with an accuracy of 86% compared to KNN and SVM models. Other performance metric values, such as sensitivity and specificity, were unclear as they did not mention the corresponding target class across the values.

3.4 Methodology

3.4.1 Data Collection

Hospital and pharmaceutical data from the national datasets of the NZ Ministry of Health were used for this study (detailed in Appendix C). The National Minimum Dataset is a national collection of publicly funded NZ hospital admissions, including acute admissions resulting from ED visits. Primary and secondary diagnosis codes (ICD-10-AM) are recorded for every hospital event. The pharmaceutical collection has medication dispensing data for subsidised medications in NZ. The ethics approval was obtained from the Auckland Health Research Ethics Committee (Appendix D). The data used for analysis were from 1 Jan 2008 to 31 Dec 2017. The data set consisted of a feature named "patient quarter", representing periods of 3 months of patients during the follow-up. For this study, data records from the fifth patient quarter (the fifth 3-month period) were employed so that some yearly data would also be included in the feature set. The dataset consisted of 1,566,349 asthma patient records with 49 features. The dataset was refined, following discussions with respiratory specialists and clinicians, by removing patient records of those diagnosed with asthma before the age of 6 to avoid confusion between asthma, pre-school wheeze, and bronchiolitis. This generated 355,113 patient records with 49 features. The target variable was asthma attack with binary values: 1 for the presence of an asthma attack and 0 for the absence of an asthma attack within a 3-month period (quarter of a year). Asthma hospital admission and/or oral corticosteroid prescription was considered as an asthma attack [97, 98]. Figure 3.1 shows the framework we followed in developing the ML models for asthma risk prediction.

3.4.2 Data Preprocessing

3.4.2.1 Feature Elimination - Observation

Initially, 10 features were removed by observation and discussion with the multidisciplinary research group. Some features were excluded as they were date fields, and some had duplicate data. The datasets consisted of corresponding binary variables and numeric variables for representing previous asthma attacks. We eliminated the corresponding binary variables as numeric features add more information to the predictive models.

3.4.2.2 Feature Elimination - Wrapper Method

RFE is a wrapper method used to remove less important features from a dataset. This method, combined with an ML model, eliminates the least predictive variables from the

feature set one by one until the optimal performance of the model is reached. We used the RF model as the ML model and the F1-score as the evaluation metric. Accordingly, a set of features was identified as least important and was removed from the feature set.

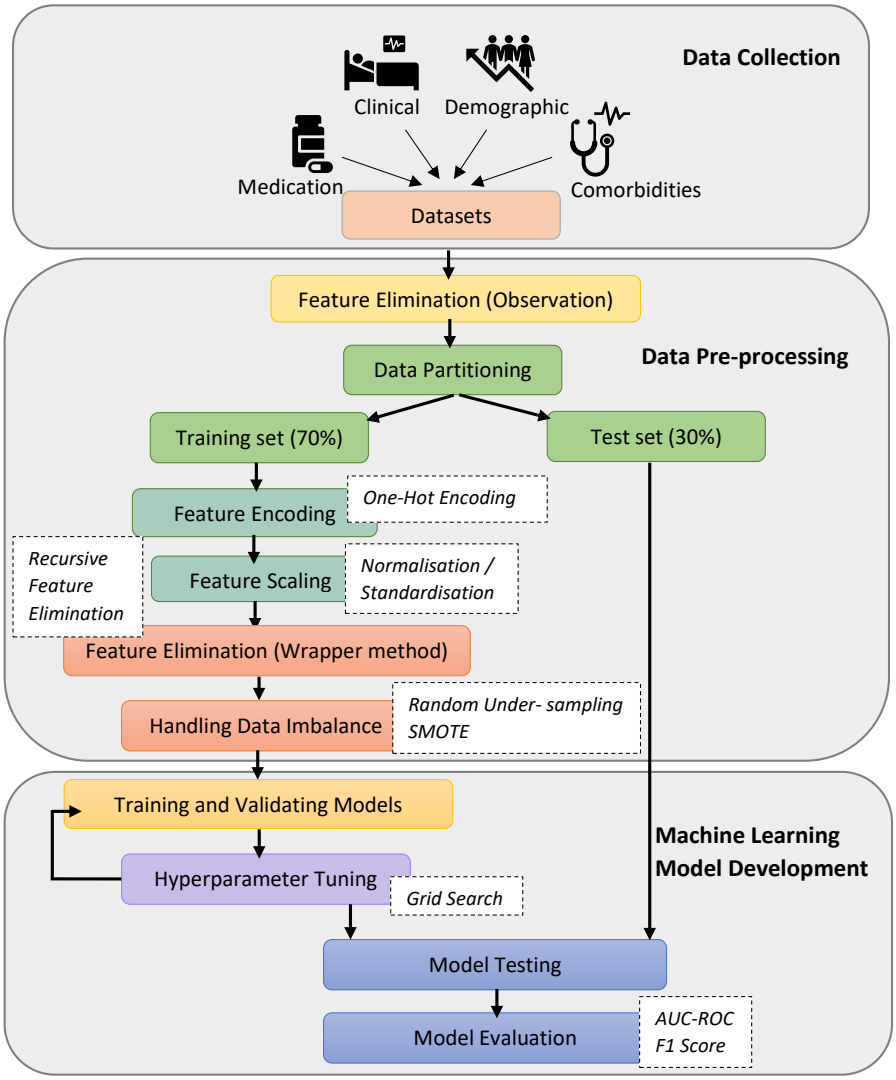


Fig. 3.1 The framework of ML-based asthma attack risk prediction model development.

3.4.2.3 Categorical Feature Encoding

The datasets comprise features with different data types, mainly binary, numerical and categorical. Some of the ML models can only deal with numbers; therefore, categorical values should be meaningfully converted to numbers. Different encoding techniques are available for this purpose. One-Hot-Encoding (OHE) is one of the commonly used encoding

techniques. It introduces new binary variables to represent each value of the categorical variables.

3.4.2.4 Numerical Feature Scaling

There were several numerical attributes in the dataset. These numerical features had different value ranges. Therefore, they were standardised to make them comparable. Standardisation converts the values between -1 and +1, which will preserve the significance of the initial values. The formula for standardising is,

$$Y = \frac{X - M}{SD}$$

where Y = standardized value, X = original value, M = mean, SD = standard deviation.

As another scaling method, normalisation was applied to the numeric variables, which brings all values in between 0 and 1. The formula for normalisation is,

$$Y = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}}$$

where Y = normalized value, X = original value, X_{min} = minimum value of X, X_{max} = maximum value of X.

3.4.2.5 Data Imbalance Handling

The raw data have an imbalanced response variable distribution in many fields, including fraud detection, medical diagnosis, and network intrusion detection; that is, at least one response variable value has a significantly greater proportion of instances. The current dataset had a target class imbalance in the ratio of 1:12, with the presence of attacks over the absence of attacks. Figure 3.2 graphically represents this data imbalance. When the ratio between the target classes is very high, it is considered a highly imbalanced dataset. When applying ML techniques in such scenarios, models would be biased towards the majority, and therefore, the models might not result in an accurate performance. Different techniques are available to overcome this issue. In such cases, either the minority class can be oversampled, or the majority class can be undersampled. Alternatively, both oversampling and undersampling can be performed at the same time. Among many techniques, RUS and SMOTE are more commonly used. In the RUS technique, examples of the majority class are eliminated randomly to balance the dataset [99]. The SMOTE technique will oversample the minority class by generating synthetic samples [100]. These two techniques were used in this study to handle data imbalance. They were applied only to the training data to produce a ratio of 1:1 among the two classes.

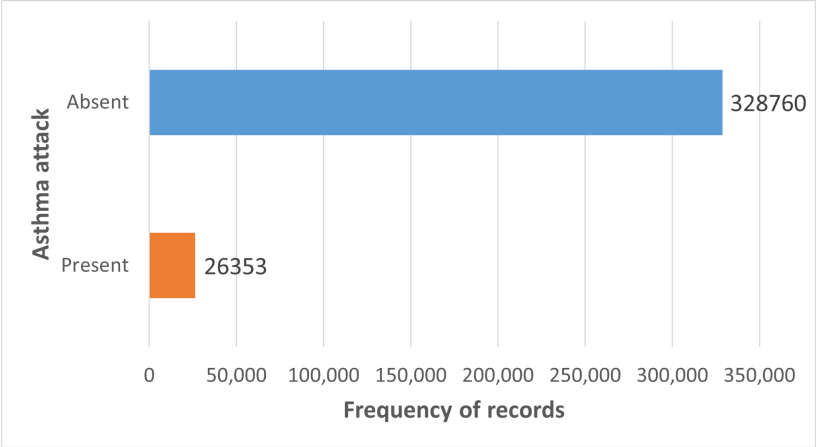


Fig. 3.2 Distribution of data among the two predictive classes.

3.4.3 ML Model Development

Asthma risk prediction models were developed using the ML algorithms RF and XGB. In addition to that, LR models were also developed. Training and test data were prepared in the proportions of 70% and 30%, respectively. The training data was balanced using the two data balancing methods. All models were trained and tested on the same training and test sets. We used stratified 10-fold CV to validate the models during the training process. Then, models were executed on the test data for evaluation purposes. Finally, the performance of the models was evaluated using AUROC and F1-score (for the positive class). Further, we determined the best threshold value for separating the model prediction into two classes based on the best F1-scores. These models were developed using Python programming language. Models were constructed on a computer with a processor of 3.50 GHz and RAM of 128 GB.

The performance of the algorithms also depends on the values chosen for the hyperparameters [101]. Therefore, as the next step, the Grid Search technique was employed to tune the model parameters. For the RF models, the hyperparameters tuned included the number of trees, criteria for splitting, maximum number of features at every split, minimum number of samples required for splitting a node, minimum number of samples required at each leaf node, and the maximum number of levels in a tree. The hyperparameters tuned in XGB models were the maximum depth of a tree, the minimum number of samples for a child node, the fraction of columns per tree, the fraction of samples per tree, the learning rate, and the number of trees. Along with the identification of the best combination of parameter values, models were retrained to obtain the results. The python codes used to develop these models are available at https://github.com/DarshaWK/Asthma-Attack-Prediction/

tree/main. Further, the feature importance was assessed using permutation importance (*sklearn.inspection.permutation_importance*). This method estimates the importance of a predictor by randomly permuting its values and measuring the resulting decrease in the model's F1-score. Predictors associated with larger decreases in F1-score would be considered more influential in the model's predictions.

3.5 Results

Most of the models improved their performance as a result of hyperparameter tuning. Firstly, a set of models was constructed by applying normalisation on numeric variables. Both RF-RUS and XGB-RUS gave a similar performance with an AUROC of 0.75 and an F1-score of 0.27 after hyperparameter tuning. The F1-score rose by nearly 18% (0.32) for a probability threshold value of 0.716. However, the LR model had a lower AUROC of 0.71, comparatively. The next set of models was developed by applying standardisation to numeric variables. The XGB-RUS achieved the best performance with an AUROC of 0.76 and an F1-score of 0.27. For a probability threshold value of 0.728, the F1-score improved by 22% (0.33). Also, RF and LR with RUS gained similar AUROC of 0.75, but F1-scores were 0.27 and 0.28, respectively.

The performance of the prediction models developed using the features selected from the RFE technique is listed in Tables 3.1, 3.2 and 3.3. Table 3.1 shows the results of the models fed with data that went through the normalisation process. Both the models, RF and XGB with RUS, showed a similar performance with an AUROC of 0.75 and an F1-score of 0.27. Also, the best threshold values (0.684 and 0.693, respectively) for both models were almost the same. However, XGB-RUS achieved a higher F1-score of 0.32, while it is 0.31 for RF-RUS. However, LR models showed comparatively lower performance. Among the three LR models, the model with RUS performed better with an AUROC of 0.72 and an F1-score of 0.25.

Table 3.2 presents the outcomes of the models developed to predict the risk of asthma attacks by feeding the predictors selected through the RFE technique. Instead of normalising, we applied standardisation to the numeric features of the inputs for these models. The results show that applying standardisation further improved the XGB-RUS. It gained an AUROC of 0.76 with an F1-score of 0.27. The best probability threshold of 0.720 enhanced the F1-score to 0.33. The second-best model is LR-RUS, which gained 0.75 AUROC and 0.28 F1-score. RF with RUS also obtained a similar 0.75 AUROC but with a 0.27 F1-score. Table 3.3 shows some other evaluation metrics of the models developed with RFE and standardisation after hyperparameter tuning. Overall, specificity and NPV values lie above 0.9. which implies the

models can predict the negative classes more accurately. However, the lower accuracy of the models in predicting the positive class is observed by the low sensitivity and PPV values.

The most important features were extracted from the XGB-RUS model developed with standardisation as shown in Figure 3.3.

Table 3.1 Performance of the models with RFE and normalisation.

Model	**DIHT**[1]	**AUROC**[2]		**BPT-OM**[3]	**F1-Score**		
		BM[4]	**OM**[5]		**BM**	**OM**	**For BPT**
RF	No-sample	0.72	0.75	0.165	0.08	0.01	0.31
	RUS	0.74	**0.75**	0.684	0.25	**0.27**	0.31
	SMOTE	0.73	0.74	0.357	0.23	0.26	0.30
XGB	No-sample	0.75	0.75	0.152	0.15	0.11	0.32
	RUS	0.74	**0.75**	0.693	0.25	**0.27**	**0.32**
	SMOTE	0.73	0.73	0.214	0.19	0.18	0.30
LR	No-sample	0.72	0.72	0.122	0.01	0.01	0.27
	RUS	0.72	0.72	0.627	0.25	0.25	0.27
	SMOTE	0.71	0.71	0.617	0.24	0.24	0.27

[1] Data imbalance handling technique
[2] Area under the receiver operating characteristic curve
[3] Best probability threshold for the optimised model
[4] Base model
[5] Optimised model

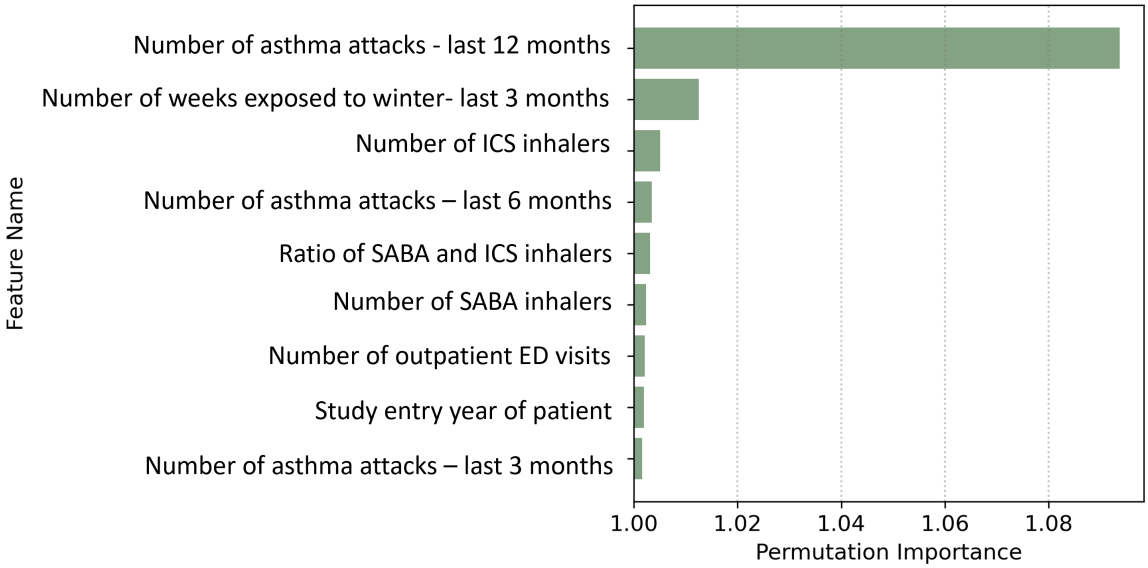


Fig. 3.3 Importance of features extracted from XGB-RUS model with standardisation.

Table 3.2 Performance of the models with RFE and standardisation.

Model	DIHT[1]	AUROC[2]		BPT-OM[3]	F1-Score		
		BM[4]	OM[5]		BM	OM	For BPT
RF	No-sample	0.73	0.75	0.139	0.13	0.09	0.32
	RUS	0.74	0.75	0.684	0.26	0.27	0.32
	SMOTE	0.73	0.74	0.357	0.16	0.18	0.32
XGB	No-sample	0.75	0.76	0.163	0.16	0.15	0.33
	RUS	0.75	**0.76**	0.720	0.26	**0.27**	**0.33**
	SMOTE	0.74	0.74	0.170	0.15	0.16	0.32
LR	No-sample	0.75	0.75	0.132	0.15	0.15	0.32
	RUS	0.75	0.75	0.641	0.28	0.28	0.32
	SMOTE	0.73	0.74	0.665	0.27	0.27	0.32

[1] Data imbalance handling technique
[2] Area under the receiver operating characteristic curve
[3] Best probability threshold for the optimised model
[4] Base model
[5] Optimised model

3.6 Discussion

In this study, we aimed to assess the performance of ML algorithms and traditional statistical models for predicting the risk of asthma attacks based on hospital and pharmaceutical data. The following subsections will discuss the findings of the study.

3.6.1 Important Asthma Attack Risk Predictors

The two features which represent the exposure to winter during the last 3 months and the ratio between Short-Acting Beta-Agonists (SABA) and Inhaled Corticosteroids (ICS) inhalers used in the past year were in the set of most important features as represented in Figure 3.3. Asthma is affected by cold weather [102]. Exposure to cold weather can increase the chance of getting severe asthma, which can lead to an asthma attack. The study [69] found that weather-related features are also important in predicting asthma attacks. Further, SABA and ICS inhalers are used by patients as medications for asthma. The study [103] concludes that the use of high doses of ICS reduces ED admissions due to asthma exacerbations, and the study [104] found that the frequency of SABA usage over three months correlated with adverse asthma results. Moreover, the study [59] shows that the mean number of daily SABA inhalations in the past has a greater importance in forecasting asthma attacks. Previous years' asthma attacks, including acute oral corticosteroid courses, emergency visits and the use of relievers, frequently are associated with future attacks [105]. This supports our finding that

Table 3.3 Performance of the models with RFE and standardisation based on other evaluation metrics.

Model	DIHT[1]	Precision (PPV[2])	NPV[3]	Sensitivity (Recall)	Specificity
RF	No-sample	0.278	0.949	0.384	0.920
	RUS	0.288	0.948	0.365	0.928
	SMOTE	0.299	0.946	0.335	0.937
XGB	No-sample	0.300	0.948	0.358	0.933
	RUS	0.306	0.947	0.351	0.936
	SMOTE	0.293	0.946	0.343	0.934
LR	No-sample	0.293	0.919	0.357	0.931
	RUS	0.266	0.950	0.402	0.911
	SMOTE	0.282	0.948	0.372	0.924

[1] Data imbalance handling technique
[2] Positive predictive value
[3] Negative predictive value

previous asthma attacks and SABA and ICS inhaler usage are important predictors of future asthma attacks.

3.6.2 Standardisation vs Normalisation

The dataset contained multiple numerical data which are very important in predicting the risk of asthma attacks. However, these variables have values between various ranges. For instance, the previous number of asthma attacks in the last 12 months had values between 0 – 20, while the number of outpatient and ED visits lie between 0-250. Different algorithms perform differently with and without scaling the numerical data. The study [106], which was conducted to assess the effect of different scaling methods on a dataset of heart failure patients, found that RF with standardisation had better performance, while the Decision Tree remained unchanged with scaling. Another study, which developed ML models for predicting future asthma exacerbations using daily home monitoring data, mentioned that standardisation performed well compared to normalisation [54]. With reference to the results (mainly F1-scores for the best probability thresholds) we obtained in this study (see Tables 3.1 and 3.2) confirm that standardisation achieves marginally better performance compared to normalisation. The numerical features in the dataset had varying scale values and quite large ranges, as we considered data from many patients. Therefore, some data points may act as outliers, and therefore, standardisation might have performed better as this method is based on the standard deviation.

3.6.3 Prediction Models

Among all the models developed, the best performance was achieved by the XGB-RUS and standardisation, which obtained an AUROC of 0.76 and an F1-score of 0.33. AUROC evaluates the discriminatory power of a binary classification model, measuring how well the model can distinguish between positive and negative instances. The F1-score, on the other hand, provides a measure of a model's accuracy in binary classification by considering the harmonic mean of precision and recall. This metric is particularly important when there is an imbalance between the target classes. Similarly, in the past studies [59, 60, 64, 69], which built prediction models using LR, RF and XGB algorithms, found that the XGB model performed better comparatively. The XGB model achieved 0.67 [64], 0.83 [59] and 0.84 [69] AUROC values for predicting the risk of asthma attacks. According to the findings, the model which obtained a higher AUROC of 0.85 on the training data has a lower performance (0.67) due to overfitting. Incorporating near-term data of the asthma patients like inhaler use, Peak Inspiratory Flow (PIF), inhalation volume, inhalation duration, and time to PIF to predict risk in the following 4 days might have helped the prediction model to gain a higher performance (AUROC=0.83) in [59] while including more weather-related factors such as wind, air pressure, humidity might have improved the predictions in [69]. Similar to our findings, the XGB model obtained around 0.76 AUROC [60]. In contrast, LR (AUROC = 0.88) did not beat ML models in [55] predicting an occurrence of asthma attack within 2 days based on home monitoring data. The adaptability offered by ML models might not always be necessary to achieve the most effective predictive model for medical data, and the interpretability of a prediction algorithm doesn't necessarily have to compromise model performance [55]. In our work as well, even though XGB was marginally performing better, some of the LR models also had the highest level of performance. Overall, from the results, it is clear that hyperparameter tuning has enhanced the model performance, and we also observed that a similar performance can be obtained with a lesser number of features we selected using the RFE method. The smaller the number of input features, the lower the model complexity. Feature reduction improves the model performance and avoids model overfitting [107, 108].

3.7 Summary of the Chapter

This chapter employed hospital and pharmaceutical data on asthma patients in the NZ context and applied various ML and traditional models to predict the risk of asthma attacks. The results showed that prior asthma attacks, length of winter exposure and the number of ICS

and SABA inhalers are the key risk predictors of impending asthma attacks. Health datasets contain variables of different types, including numerical and categorical. When dealing with numerical data, different features may have different ranges of values. Before feeding into a model, it is important to scale numeric features to a comparable range so that the model is not biased by their varying value ranges. Among these two techniques, standardisation marginally outperformed normalisation. Real-world datasets are often imbalanced with respect to the target variable. Similarly, we observed a significant imbalance in the data between the two target classes. Hence, we applied two DIHTs, SMOTE and RUS, and combined them with the ML models to make predictions. Numerical feature standardisation combined with the XGB model and RUS performed slightly better than the traditional LR model. The findings demonstrate that ML models can be used to predict future asthma attack risk. However, the current results are insufficient to ensure applicability in the real world. With the insights from this chapter, the asthma attack risk prediction model development was further expanded to determine the minimum number of features required for accurate prediction. This would allow the models to be simple and generalisable, hence applicable in clinical settings with limited data access. Also, this enhances the interpretability. The next chapter, Chapter 4, provides in-depth details of the work carried out to achieve this.

Chapter 4

Data-Efficient Machine Learning Approach for Predicting Asthma Attack Risk (Publication #4)

Citation: W. K. D. Jayamini, F. Mirza, M. A. Naeem, A. Tomlin, H. Tibble, K. A. Beyene, and A. H. Y. Chan, "Data-Efficient Machine Learning Approach for Predicting Asthma Attack Risk," *Scientific Reports*, Jul. 2026, doi:https://doi.org/10.1038/s41598-026-62302-y (**IF=3.9, Q1**)

4.1 Publication #4 Prelude

Asthma prediction can be generated through ML techniques with the involvement of variables from different data sources. However, the availability and accessibility of all these variables could be challenging in certain settings. Asthma is a prevalent medical condition worldwide. Therefore, it is essential to develop prediction models that require only limited data. This chapter presents the different strategies of selecting features and the performance of the prediction models developed with the new insights to showcase generalising asthma attack risk prediction models with limited data requirements.

4.2 Introduction

Asthma is a non-communicable, chronic respiratory disease that affects people worldwide. Globally, 37.86 million incident cases and 260.48 million prevalent cases of asthma were reported in 2021 [4]. Asthma attack risk prediction has become a significant area of research

in recent years as researchers seek to leverage ML and advanced data integration strategies to improve the expected outcome. These studies explored the use of heterogeneous datasets, including biological, hospital, weather, socio-demographic and radiology image data, to gain a higher accuracy in the asthma attack risk prediction [52, 109, 110]. This presents a significant demand for data collection, and it may not be feasible to find cases where it would be realistic to obtain such a diverse range of data to perform predictive analysis for escalating asthma risk. A previous study [111] found that there were a few important predictors among the vast array of hospital and pharmaceutical data features. We believe that predicting asthma attacks is a life-saving endeavour, and it is crucial for both patients and healthcare professionals. However, setting up a prediction model that is both accurate and validated is a significant challenge, requiring vast amounts of data.

Hence, we pose our research question: Can we simplify the method for predicting asthma attack risk? This simplification is targeted from two different angles. One approach is to experiment with simplifying the feature set used to develop the risk prediction model. The other approach is to use standard ML algorithms rather than complex deep learning models, which also simplifies the development of the model. Constructing deep learning models often demands large datasets, and the "black-box" nature limits the transparency and the interpretability, which is critical in clinical settings [112].

There is an increasing availability of high-dimensional data across various domains, including bioinformatics, text mining and finance. This has created a growing need for effective feature selection or feature reduction techniques to identify the most important features, thereby achieving the best accuracy for a given task. Feature selection reduces the dimensionality and helps to identify the most informative attributes, thereby enhancing predictive performance while reducing the training and prediction time [113, 114]. Also, high dimensionality not only increases computational cost, but it also contributes to issues such as model overfitting, data sparsity and the curse of dimensionality [115, 116, 117]. Feature selection is the strategy of identifying the most informative subset of original features, while feature reduction creates new features through mathematical transformations and projections [114, 117].

Feature selection methods are traditionally divided into filter, wrapper and embedded approaches. Filter methods assess individual variables using statistical measures such as correlation or mutual information, without the involvement of any predictive model, making them computationally efficient, scalable and model-agnostic, despite the limited ability to detect feature interactions [118]. However, filter methods traditionally overlook interactions between features [118]. In contrast, wrapper methods identify optimal feature subsets through iterative model training and evaluation, whereas embedded methods integrate feature

selection directly within the model construction process. Consequently, embedded methods are generally more computationally efficient, while wrapper methods often provide greater flexibility in evaluating feature subsets independently of the learning algorithm. A study on Electroencephalogram-based emotion recognition across two datasets found that applying the wrapper method, the genetic algorithm, generally improved the model performance [119]. However, they can become computationally intensive and are prone to overfitting [120]. Embedded methods integrate feature selection into the model training process, such as LASSO regression (L1 regularisation) or tree-based feature importance, while balancing efficiency and accuracy by integrating feature pruning into the learning algorithm itself [119, 121].

To address the limits of past research, the present study offers the following contributions to the field.

- Implementing various feature selection strategies to identify the most important predictors while evaluating the model performance under each strategy.
- Employing a range of ML techniques to develop prediction models under each feature selection strategy to determine the model with the best performance.
- Handling data imbalance in the data sets using several DIHTs to compare which technique performs best.

The remainder of this chapter is structured as follows. First, the *Related Work* section presents the literature published on asthma attack risk prediction. Next, the *Methodology* section outlines the data source, study design and the methods used in this study, including data collection, feature optimisation, data preprocessing and model development steps. The *Results* section presents the models' performance in terms of accuracy and efficiency. The *Discussion* section then interprets the findings in relation to the existing literature and highlights their implications.

4.3 Related Work

4.3.1 Asthma Attack Risk Predictors

In predicting asthma attack risk, past studies have utilised a wide range of features from various domains [109]. The most frequently used data sources are hospital, clinical and pharmaceutical data [58, 111, 122]. Asthma-related hospitalisations or emergency visits have been considered as the occurrence of asthma attacks [58, 59]. Further, asthma attacks can

be identified based on the prescription or usage of medications [58, 59, 64]. Comorbidities have been accounted for using the CCI. The history of previous asthma attacks has also been considered, as it is a key predictor of future attacks [58, 60, 61, 111]. In addition, smoking status and BMI have been included [58, 60]. Demographic data (e.g., age, gender and ethnicity) were included in the feature sets, as variations in these factors have been shown to influence the impact of asthma among patients [59, 61, 63, 64, 70, 111].

The risk prediction can be made at either the short-term individual level or the long-term population level. In addition to the above, clinical factors, such as asthma symptoms, peak expiratory flow rate and inhaler use, are included in previous studies, which are generally collected on a daily basis through telemonitoring systems and smart devices [55, 56, 57, 58] for short-term individual-level risk prediction. On the other hand, environmental and meteorological data have been integrated into models for predicting asthma attack risk, such as temperature, wind speed, air quality index, air pressure, pollen count, particle matter ($PM_{2.5}$) and different types of gases (CO, NO_2, SO_2) [60, 65, 67, 68, 69, 72] for long-term population-level risk prediction.

Past studies have identified the most important predictors. Several studies found that previous asthma attacks, asthma symptoms, medication usage and daily peak flow readings are important asthma attack risk predictors [56, 58, 59, 60, 64, 65, 122]. Additionally, the literature indicates that comorbidities and hospital stays are also significant predictors [58, 59, 60]. Moreover, it has been found that age, gender, race and smoking status are important in predicting impending asthma attacks [63, 65, 69, 72, 122]. Furthermore, prior studies have demonstrated the importance of weather factors including temperature, wind speed, relative humidity and air quality in predicting future asthma attacks [65, 69].

4.3.2 Identifying Important Predictors

Previous studies have employed various techniques to conduct feature selection in predicting asthma attack risk. A study has applied LASSO regularisation, which imposes a penalty on various features, thereby avoiding overfitting [65]. LASSO has the limitation that it tends to select only one variable from a group of highly correlated variables, while ignoring the others, which may lead to instability in feature selection [123]. A study identified the most important features using the feature importance scores generated from the RF algorithm [70]. This method provides a ranking of the features that contribute most to the prediction [124]. The ensemble nature of the algorithm reduces its sensitivity to noise, thereby generating robust outcomes. However, this method can be biased based on certain properties of the features, like feature correlation. If two features are highly correlated, their importance could be shared, or one feature may dominate, making it difficult to determine which feature is

truly important [125]. Another study employed the RFE method [54]. After the model is trained on the full feature set, it ranks the features by their importance and subsequently removes the least important features. It iteratively removes less informative features while validating the performance to identify the optimal subset of features [126]. Another study employed feature elimination through manual observation, followed by RFE with RF, and then generated feature importance using the XGB algorithm [111]. As the RFE technique utilises the actual model intended for prediction, the selected features are more likely to perform well with that model. This process continues until the key features are identified and does not affect predictive performance. However, it can be computationally costly due to iterative training. But this is only until the most important features are identified, and will not affect the prediction performance.

4.3.3 Handling Data Imbalance

Datasets for predicting asthma attack risk are often highly imbalanced because attacks occur relatively rarely compared to the total number of instances - i.e. there are more days without attacks when patients are well, rather than days with attacks. Most studies did not address the target class imbalance, although a few accounted for it. A study that detected asthma exacerbations using daily home monitoring data employed three different techniques: RUS, ROS and SMOTE to handle the data imbalance with a ratio of about 1:1265 (exacerbation:no exacerbation) [54]. The RUS technique performed best. Another study, which had a data imbalance of 1:2631 (exacerbation:no exacerbation), has shown that sequential RUS and weighting achieved the best performance [58]. A study reported a similar finding, identifying paediatric asthma patients at risk of hospital revisits, with a one-time to multiple-time visits ratio of about 1:4 [63]. SMOTE has been used in a study for asthma risk prediction, having a proportion of nearly 1:8 (present:absent) in the target class [65]. A key limitation of this study is the application of SMOTE on the full dataset, which leads to data leakage. Therefore, any of these sampling techniques should only be applied to the training data.

4.3.4 ML Algorithms Used in Asthma Attack Risk Prediction

Recent studies have widely employed ML approaches to predict the risk of asthma attacks. Commonly applied algorithms include LR [59, 60, 64, 65, 69, 71, 111], RF [59, 60, 64, 65, 69, 71, 111] and XGB [59, 60, 64, 65, 69, 111]. Across many studies, XGB achieved the best performance. XGB is an ensemble learning algorithm that sequentially constructs decision trees, where each new tree is developed to correct the residual errors in the preceding trees, thereby building a strong predictive model [127]. In some studies, LR performed

slightly better. LR showcases the importance of individual predictors through the estimated coefficients, providing interpretability. Although RF was not the top performer, many studies included it in their analysis. It is an ensemble algorithm that combines multiple decision trees to enhance predictive accuracy [128]. It has several advantages, including robustness to outliers and missing values, as well as the ability to handle a larger number of features [129]. Each tree is trained on a random bootstrap sample of the training data, and at each split, a random subset of features is considered. This dual randomness reduces variance, mitigates overfitting and enhances generalisation compared to a single decision tree. A few studies applied Decision Trees [65, 69], SVM [65, 68], LightGBM [71], Naïve Bayes [66] and NN [67, 68].

The literature shows that several types of predictors can be used for predicting asthma attack risk. However, the addition of more features complicates model development. Despite several studies employing feature selection methods, they have not compared multiple methods to determine the optimal feature set, which would simplify model development, making it generalisable and user-friendly. Hence, we propose systematically eliminating features while determining the most important ones using different strategies. This enhances the simplicity and generalisability of the asthma attack risk prediction models.

4.4 Methodology

The CRISP-DM methodology was employed in this study, as illustrated in Figure 4.1 and discussed in the following subsections.

4.4.1 Data Collection

Hospital and pharmaceutical datasets were obtained from national datasets of the NZ Ministry of Health (Appendix C). These datasets comprise publicly funded hospital admissions in NZ, including acute admissions through emergency departments, as well as data on the national dispensing of subsidised medications. The ethics approval for data collection was obtained from the Auckland Health Research Ethics Committee (Appendix D).

The data used for analysis spanned from January 1, 2008, to December 31, 2017. A feature named patient-quarter was derived to predict impending asthma attacks in the next 3 months. This represents a 3-month period for each patient. As some yearly-based features were included in the feature set, we extracted data for the fifth patient-quarter so that it could include data for the previous year. There were 355,113 patient records with 39 features.

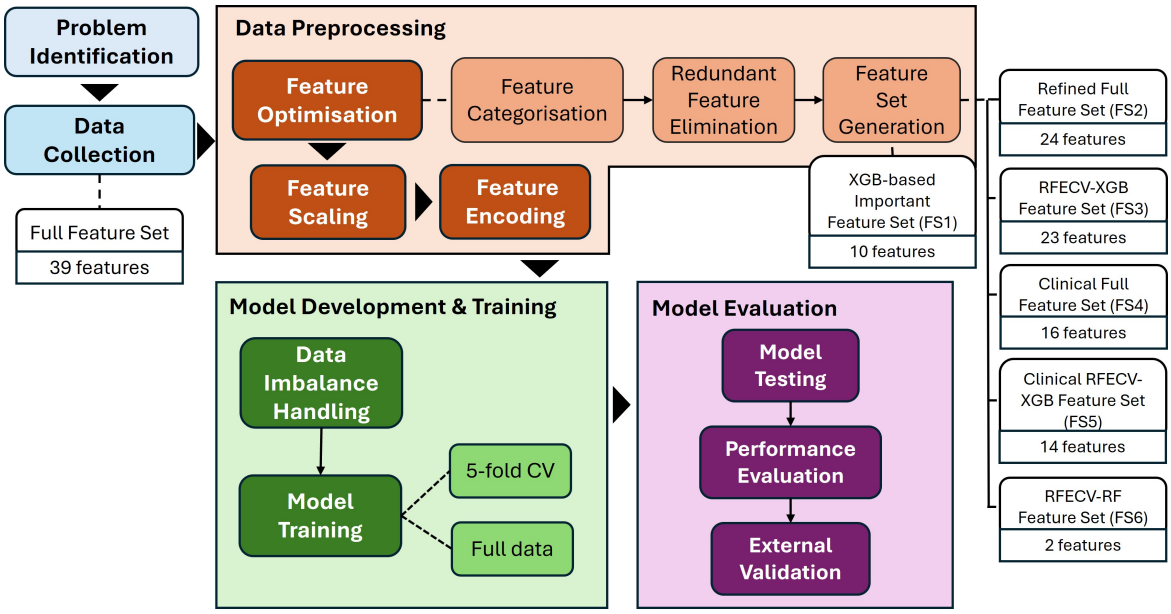


Fig. 4.1 Overview of the CRISP-DM methodology followed in the study.

The target variable represents the presence or absence of an asthma attack within a 3-month period. Asthma hospital admission and/or oral corticosteroid prescription were used to define an asthma attack [97, 98]. For external validation, similar data were obtained for another patient-quarter (the ninth patient-quarter), which included 1,093,645 instances. It is worth noting that during the model training for external validation, the dataset was expanded to a total of 1,142,322 instances. Appendix F provides the algorithm followed in this research.

4.4.2 Data Preprocessing

Before developing the prediction models, several preprocessing steps were followed to refine the feature set. The following subsections will discuss each of those steps in detail.

4.4.2.1 Feature Optimisation

Feature Categorisation and Redundant Feature Elimination

Table 4.1 presents the newly derived features *HistoryOfAllergies* and *CardiovascularDiseases*, by categorising the same group of medical conditions together. This approach was used to reduce the high number of categories. As all the original features are binary, the derived features were assigned 1 if one of the corresponding original features is 1; otherwise, 0. Furthermore, there were some features that appeared to exhibit higher collinearity among

them. Features were grouped by category, and a single representative from each was retained to avoid redundancy (see Table 4.2).

Previous asthma attacks were represented by the number of asthma attacks in the past 12 months, as indicated by the feature *P12MNoOfAsthAttacks*. Similarly, to represent the asthma medications, we chose *SABA_ICS_Ratio*. It is the ratio between the ICS and SABA inhaler usage, as a parameter to reflect both SABA and ICS inhaler use. In addition, *CharlsonComorbidityScore* was selected by excluding *DementiaAlzheimers* and *RheumatologicalDisease,* as these are already included in the score calculation.

Table 4.1 New features derived by grouping existing features.

Original features	Derived Feature
Rhinitis Nasal Polyps Atopic Dermatitis Anaphylaxis Pulmonary Eosinophilia	HistoryOfAllergies
Cardiovascular Cerebrovascular Disease Ischaemic Heart Disease Heart Failure	CardiovascularDiseases

Table 4.2 Features selected from different feature categories to exclude redundant features.

Category	Features	Selected Feature
Previous asthma attacks	P12MNoOfAsthAttacks P6MNoOfAsthAttacks P3MNoOfAsthAttacks	P12MNoOfAsthAttacks
Asthma medication	NoOfICSInhalers SABA-ICS ratio NoOfSABAInhalers NoOfAsthmaControllerMeds Nebulised SABA Asthma severity step	SABA-ICS ratio
Charlson Comorbidity Index	CharlsonComorbidityScore DementiaAlzheimers RheumatologicalDisease	CharlsonComorbidity Score

Feature Set Generation

The study included six feature sets. The workflow of feature set generation is presented in Figure 4.2. The first feature set (FS1) was extracted from the results of a previous study [111], which highlights the most important features for predicting asthma attack risk based on the optimal prediction model, XGB-RUS. In this study, the top 10 features from that

feature importance list were considered as the first feature set. Secondly, the refined full feature set (FS2) was considered to compare with the other strategies. Next, the Recursive Feature Elimination with Cross-Validation (RFECV) method was used to determine the most important features based on the two classifiers, RF and XGB, separately on FS2. This generated the feature sets FS3 and FS6, respectively. Additionally, with the clinical expertise knowledge, a full clinical feature set (FS4) was defined, which considers the most important features in the original feature set for predicting impending asthma attacks. Similarly, RFECV was applied with XGB and RF on FS4 to extract the feature sets FS5 and FS6, respectively.

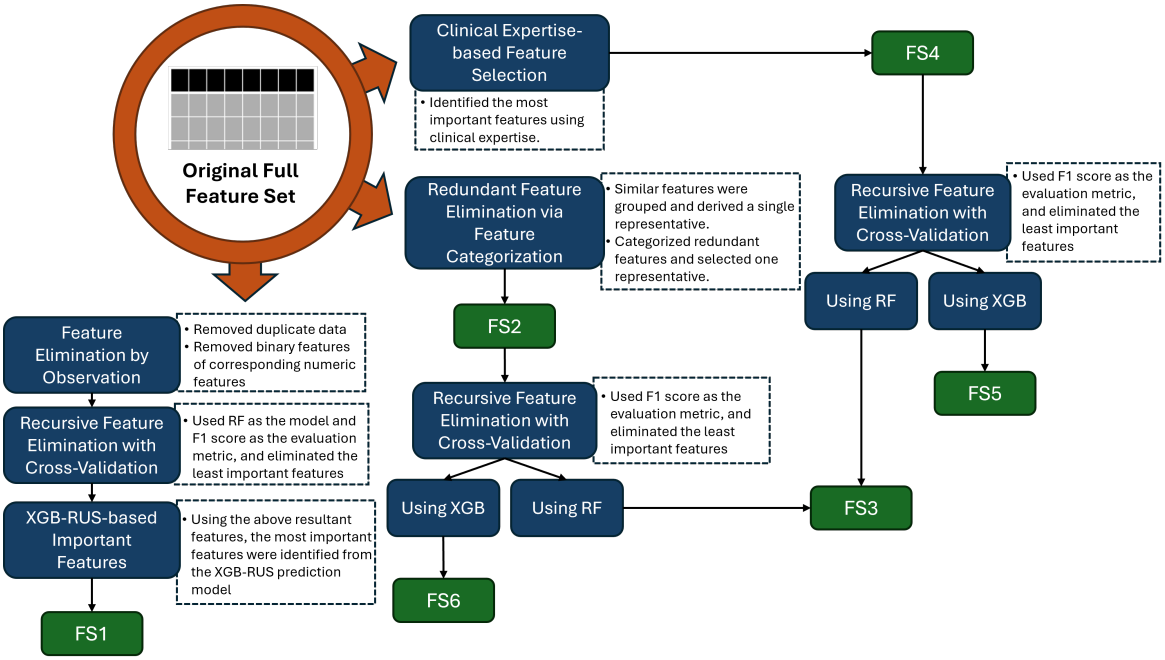


Fig. 4.2 Feature sets generation workflow.

Table 4.3 displays the division of the feature sets according to different strategies. The first column shows the top 10 features extracted from the previously developed XGB model (FS1). It consists of information on past asthma attacks, exposure to winter, SABA and ICS medication usage and some demographic factors, such as gender, ethnic group and region (District Health Board (DHB)). The FS2 is the full feature set generated after refining the original feature set (24 features). The FS3 has 23 features, excluding Psoriasis. Figure E.1 shows the selection of the optimal number of features using this strategy. Notably, the plot indicates that the optimal number of features is 48, as it has counted the one-hot encoded features separately. After analysing, the distinct feature count was 23. The FS4 has 16 features, excluding *Psoriasis*, *DHB*, *Obesity*, *Lower Respiratory Tract Infection (LRTI)*, *Anxiety Depression*, *Diabetes_NoMetformin*, *BetaBlockers*. FS5 excluded only the two features

NoOfHospitalisations and *NumberOfMetforminRx*, from FS4 (see Figure E.1b). Significantly, the FS6 comprises only two features: *SABA_ICS_Ratio* and *P12MNoOfAsthAttacks*. It is important to note that the RFECV-RF on FS2 and FS4 resulted in the same feature set (FS6) as the most important features, as shown in Figure E.1c and Figure E.1d.

4.4.2.2 Feature Scaling and Encoding

The feature set has several numeric features with varying value ranges. Therefore, to bring them onto a similar scale, standardisation was applied before feeding them into the prediction models, ensuring that no single feature would disproportionately influence the models. The previous study [111] found that standardisation performed better than normalisation. There were a few categorical features, including *EthnicGroup* and *DHB*. These features were encoded using the one-hot encoding technique, a method commonly employed for categorical feature encoding. In addition to that, the *Age* was grouped into separate bins as 12-14, 15-44, 45-64 and >=65 years, to best represent the age groups rather than individual ages. These categories were chosen in line with previous literature [130, 131, 132, 133].

4.4.3 Model Development and Training

4.4.3.1 Handling Data Imbalance

The dataset was highly imbalanced in terms of the target class, with a 1:12 ratio where the class of asthma attacks present was the minority. To address this data imbalance, four different DIHTs were employed. They included two down-sampling techniques and two oversampling techniques. The down-sampling techniques included RUS and ENN, while SMOTE and KMeans-SMOTE were used as oversampling techniques. For the feature sets having both categorical and numerical variables, an extension of SMOTE, SMOTE for Nominal and Continuous (SMOTE-NC) was used instead of standard SMOTE. It should be noted that wherever SMOTE-NC was applied, no OHE was applied on the categorical features. However, the term SMOTE is used commonly to refer to both of them in the text. For comparison, models were developed without applying any DIHTs, which is denoted as No-sample.

4.4.3.2 Model Training

Asthma attack risk prediction models were generated using three classifiers, LR, RF and XGB, as they performed better in similar studies. In each case, the dataset was split into a training set (70%) and a test set (30%). Initially, each model was built and tested with 5-fold CV using

the training set. Later, models were rebuilt on the full training dataset. Full specification of the hyperparameters of all the models is given in Table E.2. The models were implemented using the Python programming language and executed on a system equipped with a 3.50 GHz processor and 128 GB of RAM. The Python code used to develop these models are available at https://github.com/DarshaWK/asthma-prediction-feature-optimisation.

4.4.4 Model Evaluation

4.4.4.1 Model Testing and Performance Evaluation

Initially, each model was analysed using 5-fold CV, then trained on the full training set and tested on the test set. In each of these steps, model performance was evaluated using the AUROC and F1-score (for the positive class). AUROC measures how well a model distinguishes between the classes, while the F1-score produces a balanced value between precision and recall. In addition, we calculated Area Under the Precision-Recall Curve (AUPRC) and the calibration metric, Brier score. AUPRC measures how well a model identifies the positive class by balancing precision and recall, particularly when the outcome is rare or imbalanced. Brier score measures the accuracy of predicted probabilities by calculating the average squared difference between predicted probabilities and actual outcomes, where lower values indicate better performance. Additionally, identifying the top-performing models, prediction time (in seconds (s)) and model size (in megabytes (MB)) were compared as efficiency metrics. These efficiency factors are critical in clinical settings for making decisions faster without minor delays.

4.4.4.2 External Validation

To validate the performance of the top models, they were tested on an additional dataset, known as the validation data. Here, the model performance was evaluated based on the prediction accuracy. The best model combinations selected from the above were trained on the training data and then tested on the validation data. The validation data had around a 1:22 imbalance in the target class. To quantify potential dataset shift between the test and external validation cohorts, the Population Stability Index (PSI) was calculated for each predictor variable as presented in Table E.1.

Table 4.3 Details of the feature sets generated based on feature selection strategies.

	FS1-XGB-Based Important Feature Set	FS2-Refined Full Feature Set	FS3-RFECV-XGB Feature Set	FS4-Clinical Full Feature Set	FS5-RFECV-XGB Clinical Feature Set	FS6-RFECV-RF Feature Set
1	P12MNoAsthAttacks[a]	Q5_WeeksDuringWinter	Q5_WeeksDuringWinter	Q5_WeeksDuringWinter	Q5_WeeksDuringWinter	SABA_ICS_Ratio[b]
2	Q5_WeeksDuringWinter	Age	Age	Age	Age	P12MNoOfAsthAttacks[a]
3	NoOfICSInhalers	Gender	Gender	Gender	Gender	
4	SABA_ICS_Ratio[b]	EthnicGroup	EthnicGroup	EthnicGroup	EthnicGroup	
5	P6MNoAsthAttacks[c]	DeprivationQuintile	DeprivationQuintile	DeprivationQuintile	DeprivationQuintile	
6	P3MNoAsthAttacks[d]	DHB[e]	DHB[e]	SmokingStatus	SmokingStatus	
7	NoOfSABAInhalaers	Psoriasis	Obesity	CharlsonComorbidityScore	CharlsonComorbidityScore	
8	DHB[e]	Obesity	LRTI[f]	NumberOfMetforminRx	NoOfOPEDVisits	
9	Gender	LRTI[f]	AnxietyDepression	NoOfHospitalisations	Paracetamol	
10	EthnicGroup	AnxietyDepression	Diabetes	NoOfOPEDVisits	NSAIDs[g]	
11		Diabetes	SmokingStatus	Paracetamol	SABA_ICS_Ratio[b]	
12		SmokingStatus	CharlsonComorbidityScore	NSAIDs[g]	P12MNoOfAsthAttacks[a]	
13		CharlsonComorbidityScore	NumberOfMetforminRx	SABA_ICS_Ratio[b]	HistoryOfAllergies	
14		NumberOfMetforminRx	Diabetes_NoMetformin	P12MNoOfAsthAttacks[a]	CardiovascularDiseases	
15		Diabetes_NoMetformin	NoOfHospitalisations	HistoryOfAllergies		
16		NoOfHospitalisations	NoOfOPEDVisits	CardiovascularDiseases		
17		NoOfOPEDVisits	BetaBlockers			
18		BetaBlockers	Paracetamol			
19		Paracetamol	NSAIDs[g]			
20		NSAIDs[g]	SABA_ICS_Ratio[b]			
21		SABA_ICS_Ratio[b]	P12MNoOfAsthAttacks[a]			
22		P12MNoOfAsthAttacks[a]	HistoryOfAllergies			
23		HistoryOfAllergies	CardiovascularDiseases			
24		CardiovascularDiseases				

[a] Number of asthma attacks in the previous 12 months
[b] Ratio between short-acting beta2-agonist and inhaled corticosteroids inhaler count
[c] Number of asthma attacks in the previous 6 months
[d] Number of asthma attacks in the previous 3 months
[e] District health board
[f] Lower respiratory tract infection
[g] Non-steroidal anti-inflammatory drugs

4.5 Results

4.5.1 Comparison of Models Based on Accuracy

There are 90 different models generated by combining various ML algorithms, feature selection strategies and DIHTs. The performance of all these models is graphically represented in Figure 4.3. The columns of the facet grid plot show the specific feature sets used to develop the models, while each row shows the evaluation metric. The x-axis of each subplot represents the ML algorithm used to build the prediction model. Different coloured bars indicate the corresponding DIHT. The first row shows the mean AUROC values for each model in 5-fold CV, with the error bars representing the 95% confidence interval. Similarly, the second row shows the mean F1-scores of the models.

Table E.3 presents the models' performance on the 5-fold CV and the test set separately. Based on the results, it is challenging to select the best model, as some models that achieve higher AUROC scores have lower F1-scores. Overall, the best performance was achieved by a few of the LR and XGB classifiers built on the FS1.

Table 4.4 presents the top-performing models based on the values of the evaluation metrics. The models LR-RUS, LR-ENN and LR-SMOTE achieved the highest AUROC of 0.75 and F1-score of 0.28 on the FS1 test set. Similar results were shown by the XGB-ENN model on FS1 (see Figure 4.3). However, with a marginal drop in the AUROC values to 0.74, the LR-KMeans-SMOTE model on FS1 achieved the highest F1-score of 0.30. Despite the overall best performance of LR models, they have wider confidence intervals during the training phase, especially for F1-scores of the LR-KMeans-SMOTE models. Additionally, the F1 score uncertainty gap during the training phase is generally higher for models developed with ENN and KMeans-SMOTE using FS6. KMeans-SMOTE technique produces the highest gap in the confidence intervals for RF (0.12-0.29) and XGB (0.12-0.30) with FS6 (see Table E.3).

Among the data-balanced models, the lowest AUROCs of 0.55 and 0.52 were achieved on the test data by RF-ENN and XGB-ENN, respectively. For FS6, the lowest F1-score of 0.13 was observed for both models on the test set. However, in the 5-fold CV, the models show an AUROC of around 0.70 and an F1-score of 0.23. Overall, most models achieved an AUROC above 0.70. The best F1-score of 0.30 is achieved by the LR-KMeans-SMOTE models developed based on FS2 and FS3. However, the AUROCs for those two models are 0.73 each, slightly below the best.

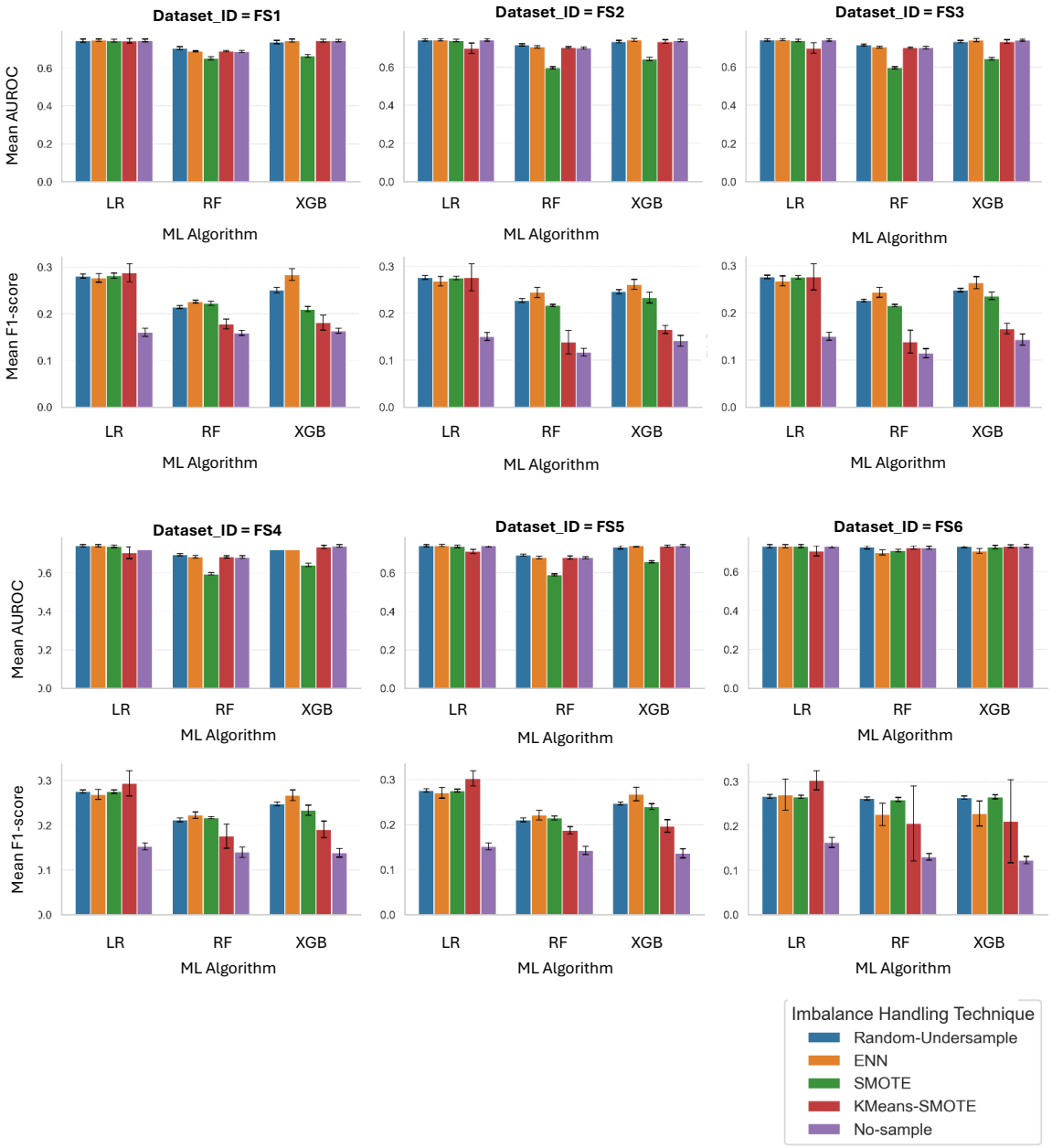


Fig. 4.3 Performance of the ML algorithms for different feature sets and DIHTs for 5-fold CV.

Overall, for most models built on the first five feature sets, RF models have slightly lower AUROC values. However, there are fluctuations in the F1-scores of these models for different ML algorithms and DIHTs. As shown in Figure 4.3, all the models without a data balancing technique (No-sample) have the lowest F1-scores, although the AUROC values remain higher. Comparing the performance of the models, especially the F1-scores, for different DIHTs, LR algorithms show more robust behaviour regardless of the DIHT applied.

Table 4.4 Top performing models for the FS1 feature set.

ML Algorithm	DIHT	AUROC		F1 Score		AUPRC	Brier Score
		CV Mean (95% CI)	Test set	CV Mean (95 % CI)	Test set	Test set	Test set
LR	RUS	0.75 (0.74-0.75)	0.75	0.28 (0.28-0.29)	0.28	0.29	0.06
	ENN	0.75 (0.74-0.76)	0.75	0.28 (0.27-0.29)	0.28	0.29	0.06
	SMOTE	0.75 (0.74-0.75)	0.75	0.28 (0.27-0.29)	0.28	0.28	0.19
	KMeansSMOTE	0.74 (0.73-0.76)	0.74	0.29 (0.27-0.31)	0.30	0.27	0.09
XGB	ENN	0.75 (0.74-0.75)	0.75	0.28 (0.27-0.30)	0.28	0.29	0.06

The findings clearly show that under-sampled data improves the performance of the RF and XGB algorithms regardless of the feature set, except for ENN on FS6. Among the two data under-sampling techniques, RUS and ENN, ENN performs slightly better with RF and XGB algorithms except for FS6. The oversampling technique, KMeans-SMOTE, shows marginally higher F1-scores for the LR models on all the feature sets. The SMOTE technique has significantly lowered the AUROC values in the RF models, except for FS6.

4.5.2 Comparison of Models Based on Model Efficiency

In addition to the evaluation metrics, we considered the efficiency of the models to determine the best-performing models. Prediction time and model size were the two metrics used for efficiency comparison. The efficiency of all the models developed is represented in Figure E.2.

In comparison to all models, RF was the lowest performer with the highest prediction time. Overall, the RF-SMOTE models ranked at the lowest efficiency, with a model size exceeding 540 MB and a prediction time exceeding 2.5 s. These values are not considered as any standards for a prediction model. These could change based on different settings. The sole idea was to identify a threshold to compare the other models developed. Based on the ML algorithms, FS2-LR-ENN (prediction time=0.13 s, model size=90.00 MB), RF-SMOTE with FS1, FS2 and FS3 with average measures of prediction time=2.67 s and model size=579.37 MB and XGB-ENN with FS3 and FS2 with average values of prediction time=0.14 s and model size=89.48 MB were the lowest performers.

Based on all of the values, models with a prediction time of less than 1 second and a model size of less than 1 MB were chosen as the top-performing models. It should be noted that, although a few models with the No-sample DIHT were among the top-ranked models based on efficiency, they were eliminated due to lower F1-scores. The primary objective was to identify the best-performing models based on all performance measures.

Figure 4.4 displays the comparison of the 12 top-performing models based on the efficiency metrics. Among these, a significant observation is that models with KMeans-SMOTE were larger in size (~0.90 MB), while the FS6-XGB-SMOTE took a similar amount of space (0.93 MB). But it was one of the models with the lowest prediction time (0.01 s). Notably, all LR-RUS models in this subset yield the lowest model size (0.25 MB), independent of the feature set. However, there are slight variations in the prediction time for the LR-RUS models using different feature sets, with a marginally lower time for FS6 and a higher time for FS2. The shortest prediction time ($\leq$ 0.01 s) is reported by the models using FS6, despite their varying model sizes. Oversampling techniques have increased the model size. Within this group of top models, approximately two-thirds are based on the RUS technique, whereas none incorporate ENN.

4.5.3 External Validation of the Prediction Models

Table 4.5 presents the performance of the models on the validation data. The results show that the top selected models yield AUROC values of 0.74, which is slightly lower than the testing results. However, reductions in the F1-scores were observed. Compared with the smallest feature set (FS6), the two models with higher efficiency and better accuracy were included in this validation. However, their overall performance dropped.

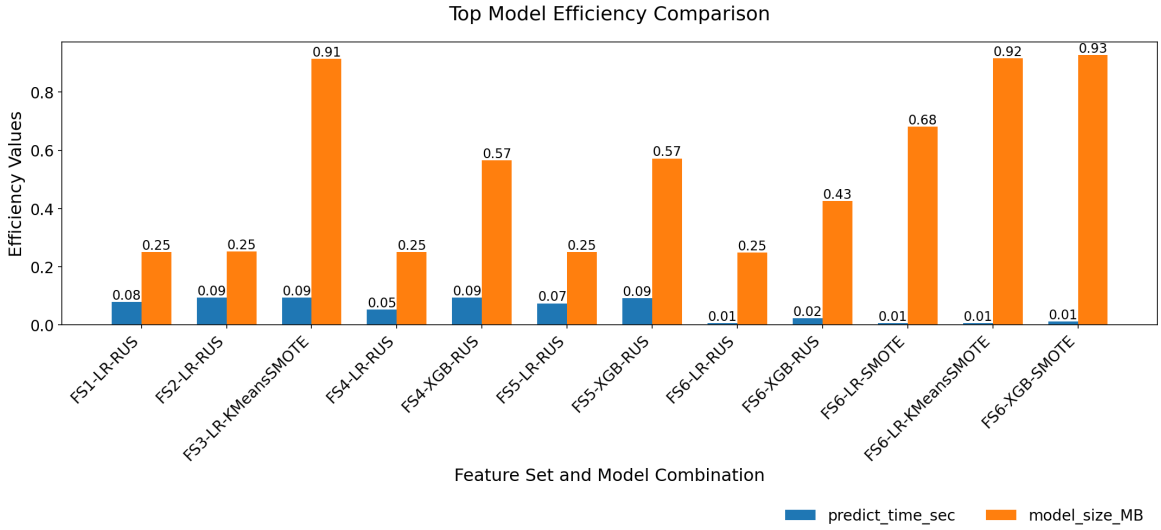


Fig. 4.4 Comparison of the top-performing models with respect to both efficiency metrics, prediction time (s) and model size (MB).

Table 4.5 Performance of the top models on the validation data.

Feature Set	ML Algorithm	DIHT	AUROC	F1 Score	AUPRC	Brier Score
FS1	LR	RUS	0.74	0.22	0.23	0.18
	LR	ENN	0.74	0.24	0.23	0.03
	LR	SMOTE	0.74	0.22	0.23	0.18
	XGB	ENN	0.74	0.24	0.24	0.03
	LR	KMeansSMOTE	0.71	0.27	0.22	0.06
FS6	LR	RUS	0.71	0.23	0.21	0.19
	XGB	RUS	0.71	0.19	0.20	0.19

4.6 Discussion

This is an important study that aimed to evaluate the performance of a range of ML algorithms and traditional statistical models in predicting impending asthma attacks using national health datasets with different feature optimisation strategies. These strategies generated several feature sets by excluding some of the less important features in making predictions of asthma attack risk. The FS1 was defined based on the previous study [111], which developed several prediction models for predicting asthma attack risk. Based on the best-performing XGB model, the study identified the most important features for predicting asthma attack risk.

Several common features are included in the feature sets generated in this study. One of them is the number of past asthma attacks. This has been identified as a key predictor of future asthma attacks in several previous studies [109, 130, 131, 134]. Demographic factors such as *Gender* and *EthnicGroup* were common among the first five feature sets. A study shows how asthma impacts different ethnic groups in NZ [44]. There is evidence that some ethnic groups, specifically Māori, the Indigenous population of NZ, have a higher prevalence of asthma [44]. Another feature was the *SABA_ICS_Ratio*, which shows the ratio between preventer and reliever medication. This has been considered a predictor of asthma attack risk in previous studies [94, 135]. However, the method we used to calculate this ratio differs from that employed in previous studies. In our study, we calculated the SABA to ICS ratio as the number of SABA inhalers divided by the number of ICS dispensed to a patient over the last year. In other studies [94, 135], it has been calculated as the average daily dose of SABA/ICS. However, this reflects key aspects of asthma control. The external validation results (Table 4.5) show a slight reduction in the models' performance. The PSI calculated (Table E.1) supports this performance reduction by indicating a significant and moderate shift in the key predictors *SABA_ICS_Ratio* and *P12MNoOfAsthAttacks*, respectively, across the test and validation cohorts.

RFECV-XGB on FS2 resulted in FS3, excluding only *Psoriasis*. Psoriasis is a long-term, immune-driven inflammatory condition that primarily affects the skin and/or joints [26]. The lower importance of this feature is evident in the values in Table E.4, which indicates that 99.98% of instances involve the absence of Psoriasis, regardless of whether asthma attacks occur or not. Therefore, Psoriasis has had no significant impact on differentiating the occurrence or non-occurrence of an asthma attack within the prediction models developed in the study. However, literature shows that individuals diagnosed with psoriasis were more likely to be susceptible to asthma [105, 136]. In this study, only 0.02% of patients had this health condition; therefore, the lack of data may have led the models to overlook it as an important predictor of asthma attacks.

Although previous asthma hospitalisations are a key predictor for determining future asthma attacks, in this cohort, we considered hospitalisations due to all causes; therefore, they might have become a less important factor in differentiating between the target class. This is evident from the descriptive statistics presented in Table E.5. Although the mean number of hospitalisations was slightly higher among those who experienced asthma attacks (mean=0.53) compared to those who did not (mean=0.32), the median and first quartile (Q1) values were both zero for both groups, indicating that most individuals had no prior hospitalisations. In fact, 80.6% of attack cases and 80.7% of non-attack cases resulted in zero hospitalisations in the previous 12 months. This limited variability might have reduced the predictive value of this feature. Furthermore, regarding *NumberOfMetforminRx*, although the mean number of metformin prescriptions was slightly higher among individuals with asthma attacks (mean=0.34) compared to those without (mean=0.32), the distribution is highly skewed, with the median and all quartiles equal to zero in both groups since metformin use is uncommon amongst asthma patients (Table E.6). Additionally, 95.4% of attack cases and 95.7% of non-attack cases had no metformin use. This suggests that the variation in this feature is limited and its contribution to distinguishing between the two groups is minimal.

Generally, the LR and XGB models performed marginally better than the RF models. The best AUROC was given by the LR classifier with both under-sampling and oversampling techniques on FS1, while the highest F1-scores were given by the KMeans-SMOTE with the LR classifier on FS1 although the AUROC was marginally lower than others. The FS1 feature set includes some redundant features, such as previous asthma attacks in different time windows and medication usage, such as SABA and ICS inhalers. Therefore, the multicollinearity between these features might have impacted the prediction models, especially on the LR classifier [137, 138]. However, all LR models were developed using ridge regularization, which constrains coefficient magnitude and reduces overfitting.

Although the AUROC values for the No-sample models were sufficient, all these models exhibited the lowest F1-scores (refer to Figure 4.3). This could be because AUROC summarises the model's ability to discriminate between the positive and negative classes at all possible thresholds. But not the number of correctly classified instances under each class. In highly imbalanced datasets, specificity (true negative rate) can appear very high because the negative class is much larger. This can make the Receiver Operating Curve (ROC) curve appear to indicate that the model performs well overall, even if it does not accurately identify the minority (positive) class. Therefore, using ROC curves alone in these scenarios may lead to an overestimation of the model's actual performance [139]. The F1-score is the harmonic mean of precision and recall, providing a more informative evaluation of the classifier's performance, especially for the minority class [139, 140]. Therefore, it is essential to consider the F1-score in model evaluation, particularly when there is a significant data imbalance. Accordingly, it can be concluded that when there is a huge data imbalance, using a DIHT improves the overall model performance.

On the other hand, the results showed that LR-RUS, LR-ENN, LR-SMOTE and XGB-ENN achieved the highest AUROCs and F1-scores. Further analysis on the performance using AUPRC and Brier score values confirmed the better performance of the LR-RUS, LR-ENN and XGB-ENN models on FS1. A similar finding that LR-RUS performs better has been reported in another study [57]. In the RUS technique, data from the majority class is randomly removed. In contrast, ENN eliminates most samples that differ significantly from most of their k-nearest neighbours, thereby improving class separability and reducing noise in the data. This sharpens the decision boundary, guiding classifiers to prioritise learning from more informative data. A study proposed a combined approach of SMOTE-ENN on diabetes prediction and found that RF with this approach achieved the best performance compared to SVM [141]. However, overall, our findings showed that the RF models had marginally lower performance compared to the other two classifiers. LR has outperformed the XGB in predicting asthma exacerbations in a study, although they have not considered the data imbalance [63], while it is the other way round in another study [60]. A similar observation was found in another work on developing asthma attack risk prediction models [111]. These findings suggest that LR and XGB classifiers generally perform better on highly imbalanced data. Despite the better performance of XGB, the XGB-ENN showed the lowest AUROC for FS6. This could be due to the lack of features in discriminating between the two predictive classes, as too few features might not provide enough information to the models [142]. These algorithms attempt to construct multiple decision trees with various combinations, and a lack of features may reduce their predictive power. Fewer features will result in fewer possible splits in the decision trees. This is further supported by the model

validation results, which show that a very limited number of features would significantly reduce the model's performance, especially when the datasets become larger and exhibit a huge data imbalance.

Comparatively, the findings showed that the LR-RUS models have shorter prediction times for FS6 and longer for FS2. A parallel observation was there for the RF and XGB algorithms. This could be mainly due to the difference in the number of features the algorithms need to consider during a prediction. Overall, RF-SMOTE was the least performer for the first five feature sets compared to the other prediction models developed in the study. Accordingly, Figure 4.5 proposes a generalised workflow for obtaining better results in asthma attack risk prediction. Considering both the factors, accuracy and efficiency, and feature selection strategies experimented within the study, using XGB-RUS to determine the most important features is identified as a better option. While it is critical to handle data imbalance, down-sampling data generates marginally better results. Overall, as discussed above, LR and XGB generated better results. Hence, the study findings helped in generating the proposed workflow which would be a hybrid model that combines two ML algorithms to make a better prediction with improved accuracy and efficiency.

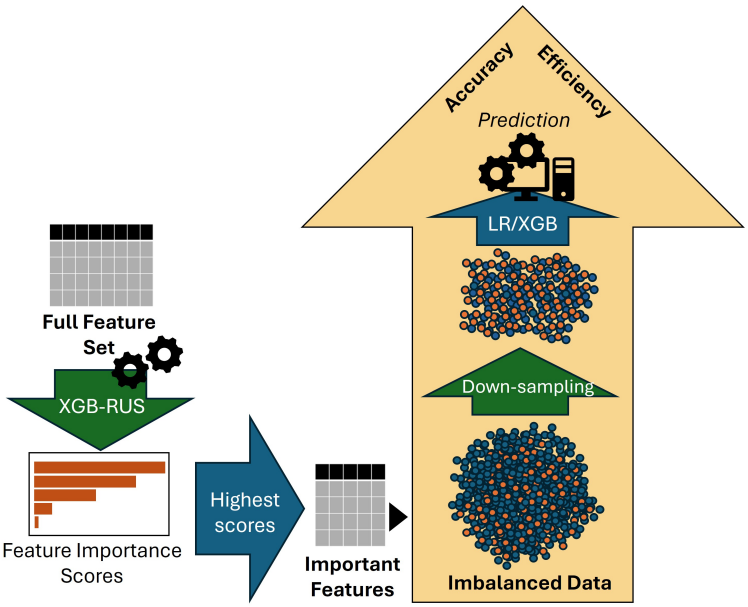


Fig. 4.5 Generalised workflow for asthma attack risk prediction.

Several factors strengthen this study. Most researchers will typically stick to one feature selection method in their individual work. However, in this study, several feature selection strategies were employed simultaneously. Additionally, the study focused on addressing data imbalance. Various DIHTs were applied to the dataset to handle the data imbalance before

the actual prediction models were constructed. Further, it utilised a few conventional ML algorithms to develop the prediction model. As a result, a total of 90 model combinations were to be compared. Moreover, the performance of the models was evaluated, not only based on accuracy but also on efficiency. Another key strength was the external validation of the models.

Apart from the significant strengths of the study, several limitations should be noted. The study only applied feature selection to a dataset that has hospital and pharmaceutical data. The findings might be different if the data parameters were broader to include weather, peak flow, inhalation data, etc. Including further data sources might help in identifying further key predictors for asthma attack risk prediction models. Also, there are several other feature selection strategies that could be applied to confirm the most important predictor identification. In the current study, it only applied RFECV and did not compare the other feature selection methods. Therefore, it is worth experimenting and comparing the other strategies/techniques for feature selection in future work.

4.7 Summary of the Chapter

This chapter explored a method to minimise the complexity of predicting asthma attack risk while determining the least amount of data required. This work contributes to asthma attack risk prediction by experimenting with variant feature sets, different data resampling techniques, and ML algorithms, resulting in 90 combinations for comparison. By considering both accuracy and efficiency, this chapter provided a more balanced perspective on model performance evaluation. The results showed that the number of asthma attacks in the previous year and medication use, particularly the SABA_ICS_Ratio, are strong predictors of upcoming attacks. Hybrid models that combine XGB-RUS-based feature selection (FS1) with down-sampling methods (RUS or ENN) for LR and XGB algorithms performed slightly better than the others. Interestingly, a very simple model using only two features (FS6) with LR-RUS also produced promising results, but with a higher variation, which needs further investigation. Overall, the findings indicated that accurate predictions can still be achieved with fewer features, improving both model simplicity and generalisability. Although asthma attacks are predictable, not all attacks are preventable. Patients admitted to the hospital due to severe asthma outcomes may need to stay for several days to receive appropriate treatment. During this time, hospitals are required to allocate adequate resources to support patient care. Therefore, anticipating the length of hospital stay for asthma patients in advance could help hospital staff plan and manage resources more efficiently and effectively. The next chapter, Chapter 5 provides the work carried out to predict asthma LoS.

Chapter 5

State of Asthma-Related Hospital Admissions in New Zealand and Predicting Length of Stay Using Machine Learning (Publication #5)

Citation: W. K. D. Jayamini, F. Mirza, M. A. Naeem, and A. H. Y. Chan, "State of Asthma-Related Hospital Admissions in New Zealand and Predicting Length of Stay Using Machine Learning," *Applied Sciences*, vol. 12, no. 19, p. 9890, Oct. 2022, doi: https://doi.org/10.3390/app12199890 (**IF=2.5, Q2**).

5.1 Publication #5 Prelude

Hospital admission is a potential consequence of severe asthma attacks. Patients who are hospitalised due to asthma exacerbations may require several days of inpatient care to receive appropriate treatment aimed at controlling symptoms and stabilising disease severity. Beyond the individual perspective, hospital LoS is a critical indicator for evaluating the efficiency and quality of healthcare services. It not only reflects the severity of an exacerbation but also represents the resource burden placed on healthcare systems. ML has enabled more accurate modelling and prediction of LoS, enabling early intervention and informed decision-making. Therefore, in this chapter, we address the research gap RG2 and the research question RQ2 by developing prediction models for asthma LoS using ML algorithms based on asthma hospital data. This enables us to investigate how these techniques perform in the context.

5.2 Introduction

Asthma is a common long-term condition due to abnormal airway functionality with airflow limitations and wide variations during a short time [143]. Those symptoms are generally intermittent and reversible with proper treatment [17]. Asthma affects children and adults, but often starts in childhood [3]. While many studies have been conducted to manage childhood asthma [46, 144, 145, 146] and adult asthma [147], limited research has explored how LoS may impact asthma treatment and outcomes. The study [34] found that hospitalisation due to COVID-19 has increased among children (5–17 years) with poorly controlled asthma compared to children with well-controlled or without asthma in Scotland. This highlights the risk that asthma patients face in their lives. It emphasises that this respiratory condition requires regular medical treatments and proper management because the disease's severity varies with other factors, including environmental and meteorological conditions [35, 36, 37], workplace conditions, and severe adverse life events [38]. If asthma is not adequately managed, it can lead to sudden exacerbations needing emergency health care and hospital admission for treatment and monitoring [35]. Therefore, asthma patients are at a high risk of needing medical care, which could be prolonged and costly to the health system.

The highest prevalence of asthma in adults is seen in the Western Pacific region [88]. NZ is one of the countries with the highest prevalence of asthma, where more than 597,000 people are on medical treatment for asthma [89]. Asthma is a common reason why children get admitted to hospitals in NZ [92]. According to NZ health surveys, one in every seven children (13%) and one in every eight adults (12%) take medical treatments for asthma [90]. By comparing the characteristics of patients admitted with critical asthma syndrome (CAS) to the ICU in the United States of America, Australia, and NZ (ANZ), ANZ patients made up a more significant percentage of ICU patients and had more prolonged ICU and hospital stays [148]. Therefore, asthma is a more prevalent health condition in NZ than in other countries, with a high impact on quality of life and a significant burden on health service providers.

Severe asthma symptoms could cause asthma exacerbations and result in hospital admission. When patients get admitted to the hospital, they may require an extended stay for treatment. LoS can be expressed as the number of days that patients stay in the hospital, i.e., the duration of a patient's stay, starting from the date of admission to the date of discharge. Predicting LoS is helpful for hospitals to determine the utilisation and management of available resources while providing required medical treatments to patients. LoS also reflects the complexity of patient status and hospital care efficiency [149]. Furthermore, hospital stays incur a considerable cost, impacting the economy to the extent of increased unpaid bills and bankruptcy rates [150]. According to a study, direct cost estimation of NZD 103 million

and NZD 62 million has been made for paediatric asthma-related hospital admissions and prescriptions in NZ, respectively, for the period 2010–2019 [151]. The estimated total cost for ED and outpatient visits in NZ in 2020 is nearly NZD 163 million [91]. Moreover, patient transportation during medical emergencies may cost healthcare providers and patients, which could be avoided if asthma hospitalisations and transfers were minimised [152]. Therefore, recognising the critical predictors of asthma LoS and the early prediction of LoS can lead to saving money for a country via the effective management of resources in hospitals. This study aims to explore regional-level NZ asthma admissions using statistical approaches to identify the importance of factors related to asthma LoS. Furthermore, we aim to use ML models to predict LoS. In summary, the contributions of this work are as outlined:

- **Preprocessing data:** We performed different preprocessing techniques on the raw data before data analysis. Missing values and outliers are commonly available in real datasets. Therefore, preprocessing eliminated the missing values manually identified and used the z-score to handle outliers. We grouped some feature values into major categories and excluded minor categories. Before feeding data into ML models, we encoded the categorical values using OHE technique or a manual mapping code. We scaled the numerical values via standardisation.

- **Data analysis:** It was essential to find the association between different factors with asthma LoS. First, we grouped LoS into stay and no stay. To identify the importance of factors associated with the asthma LoS group, we performed correlation analyses using chi-square and ANOVA tests on categorical and numerical features, respectively. Also, we performed bivariate and multivariate analyses to explore the association of features with the LoS group.

- **Feature extraction:** Initially, there were 13 variables, including LoS. We derived new features from existing date variables, adding more information to the dataset. Some variables with feature descriptions and dates were removed, which had no extra computational contribution. As a result, we ended up with a total of 9 variables.

- **Developing a methodology:** For predicting asthma LoS using ML algorithms, we extracted instances with LoS in the range of 1 to 14 days and then split them into training and test sets. Different ML models were developed and validated using 5-fold CV. Initially, we developed baseline models and, as the next step, applied the grid search technique to tune hyperparameters which could optimise model performance. After identifying the best-performing hyperparameter values, we redeveloped a model with the best set of hyperparameters and retrained the model using the whole training set. Finally, these models were tested on the test data.

- **Performance Evaluation:** To understand and compare the model performance, we evaluated the models on the test data using a few evaluation metrics, RMSE, and MAE, which are error metrics commonly used to evaluate regression models. We selected the model with the minimum error values as the best model.
- **Future Direction:** Following all the stages above, we highlight several key points as the future direction towards asthma LoS predictions.

The remaining part of the chapter has the following sections: The *Related Work* section presents the literature review conducted within this study. The *Methodology* section covers the data source and the methods used for analysing asthma-related admission data and predicting LoS using ML. The *Results* section presents the findings of the analysis and the ML models' performance. In the *Discussion* section, we discuss the results obtained.

5.3 Related Work

Research has been conducted internationally to identify the key determinants in predicting LoS for paediatric asthma hospitalisations using socio-demographic, temporal, and diagnostic factors [153, 154], all of which have statistically significant associations with LoS [153]. They have found age, gender, and day of the week as important predictors [153, 154]. Additionally, respiratory-related secondary diagnoses, year of admission [153], obstructive sleep apnoea, complex chronic conditions, and season (winter) [154] have been reported as the key predictors of asthma LoS. Another study examined the association of ambient air pollution on the LoS of children (aged 5–18 years old) with asthma in South Texas and found that ozone levels and $PM_{2.5}$ significantly correlate with LoS [155].

Data from previous paediatric asthma research in NZ show that Māori children are more likely to experience asthma hospitalisations compared to non-Māori children (7.2/1000 versus 3.5/1000, $p < 0.001$), and a higher percentage of Māori children are readmitted to hospitals within three months of their first admission (18% versus 14%, $p < 0.001$) [151]. This shows that there is an apparent ethnic disparity in terms of asthma hospitalisations. Thus, it is essential to investigate the association between ethnicity and asthma LoS. A model that accurately predicts LoS at the time of admission of an asthma patient could be beneficial for healthcare practitioners. This could enable the effective and efficient utilisation of human and other resources available. Previous research has considered this a regression problem and developed Multiple Linear Regression (MLR), Support Vector Regression (SVR), RF, and gradient boosting models to predict LoS for other long-term health conditions. One study [156] used an open-source Microsoft dataset for predicting LoS, and gradient boosting

outperformed other models, with a mean absolute error of 0.44, compared to MLR, SVR, and RF. However, it is unclear for which disease they have built LoS prediction models. A group of researchers has applied a regression tree (Cubist) model for predicting the LoS of patients diagnosed with congestive heart failure [157]. Another study [158] has developed artificial neural network and logistic regression models with more than 88% accuracy for Coronary Atherosclerosis (CAS) patients in the cardiovascular unit. Apart from regression models, LoS classification has been performed with three classes for a population in a general surgery department in Iran [159]. They used KNN, naïve Bayes, and DT ML algorithms to predict LoS and concluded that DT performs well with 85% accuracy. The study [160] followed binary classification in predicting LoS in two classes: short LoS and prolonged LoS following colorectal cancer resection. They used the median of LoS (9 days) to separate the records into two LoS groups. In predicting the LoS group, the study [160] obtained an AUROC of 0.82 for both SVR and logistic regression models. This kind of LoS classification could be helpful for clinicians in identifying patients at a high risk of prolonged stays. To predict whether a patient in the ICU can be discharged within 10 days, the study [161] developed a real-time learning one-class classification framework using Extreme Learning Machines (ELM). They chose the value of 10 because it was the dataset's median and geometric mean. However, when selecting the mean value of the dataset to separate the numeric LoS, it needs to be more specific because it solely depends on the dataset. Instead, researchers could ask for support from medical staff to decide on the appropriate value. Whilst there has been prior research to predict LoS for other medical conditions, to our knowledge, ML has not yet been applied to predict asthma LoS with a real dataset.

5.4 Methodology

5.4.1 Data Source

The data used in this study is an Auckland region-wide asthma-related admission dataset collected from the Auckland District Health Board in NZ from 1 January 2017 to 1 January 2021 (detailed in Appendix G). The algorithm followed in this work is provided in Appendix I. Patients admitted for asthma to any of four Auckland regional DHBs were included for analysis. Eligible patients admitted to the hospital with asthma were identified based on the International Classification of Diseases (ICD-10AM) diagnostic codes relating to asthma (Table 5.1).

The dataset comprised 11,414 anonymised records of children (<18 years) and adults with asthma. Admission data were recorded from the four DHBs in the Auckland region in

Table 5.1 Diagnostic codes and descriptions.

ICD-10AM Diagnostic Code	Description
J450	Predominantly allergic asthma
J451	Nonallergic asthma
J458	Mixed asthma
J459	Asthma, unspecified
J46	Status asthmaticus
R05	Cough
R060	Dyspnoea
R061	Stridor
R062	Wheezing

NZ (including Northland). Admission data included the socio-demographic and diagnostic variables of the patients.

5.4.2 Data Preprocessing

The dataset was cleaned and preprocessed before analysing and developing ML models. Only a few missing values were present in the dataset and were directly removed. A few columns with dates and descriptions were removed as they did not add any extra information to the dataset. The dependent variable had outliers; therefore, those were removed using z-score values. Records having z-scores greater than 3 were identified as outliers and removed from the dataset. After handling missing values and outliers, the dataset remained with 11,348 records. Figure 5.1 shows the activity diagram for the methodology followed in data analysis and ML model development.

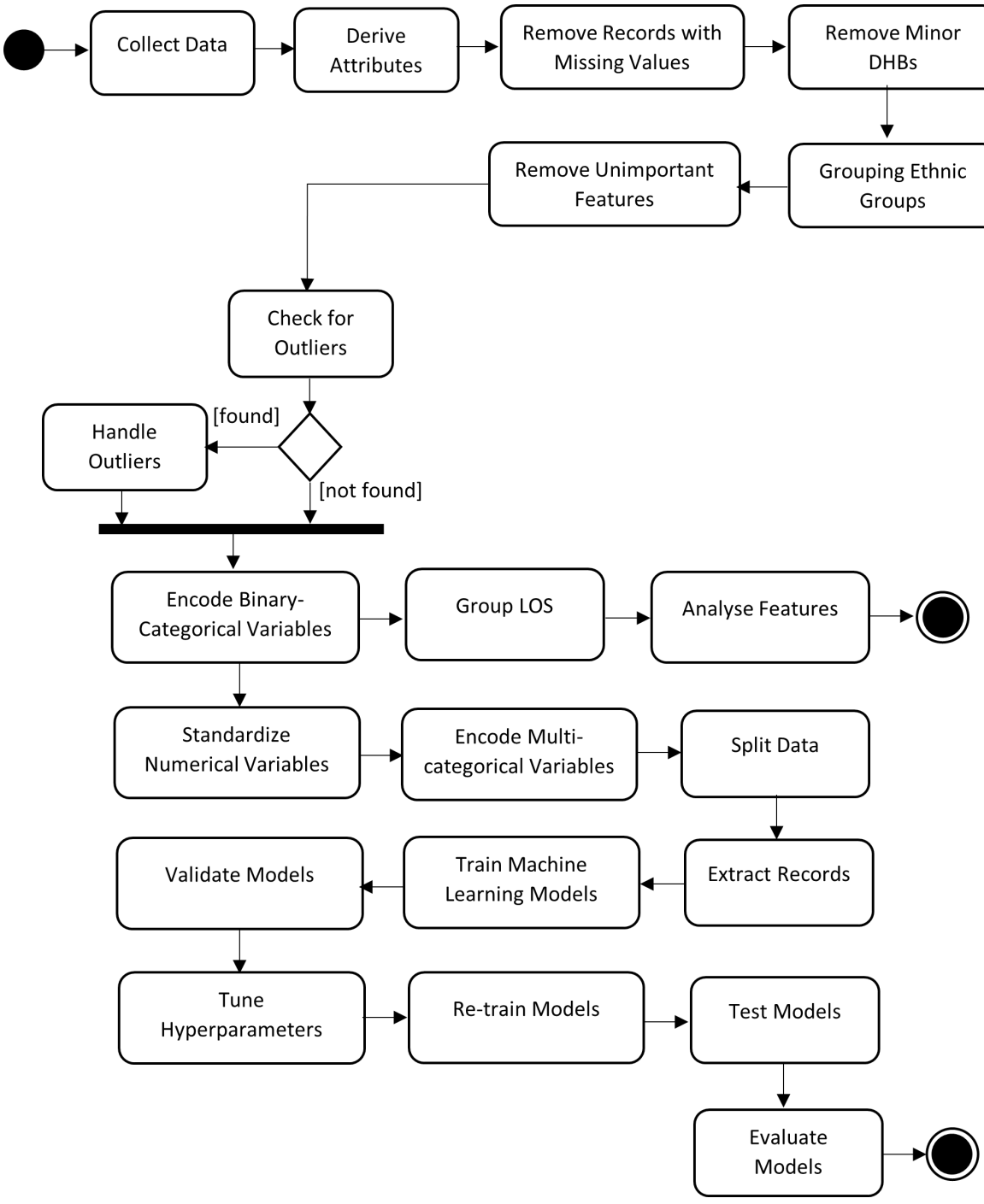


Fig. 5.1 Activity diagram for feature analysis and ML model development methodology.

5.4.3 Modifying and Deriving Features

The dataset had hospital admissions data from Auckland, Waitemata, Counties Manukau, and Northland DHBs. Additionally, some other features were derived from the existing

features. Each admission was comprised of the admission date and the discharge date. The admitted month was extracted from the admit date as a new temporal feature to explore the pattern of asthma admissions relating to the month of admission, as seasonality is known to affect asthma symptom control [162]. Ethnicity is self-identified and is registered at each health interaction rather than being a classification that remains unchanged over time [163]. Where an individual identifies as more than one ethnicity, prioritised ethnicity is used; that is, individuals are classified into one ethnic group in the following order: Māori, Pacific Peoples, Asian, MELAA (Middle Eastern/Latin American/African), Other, and European [164]. Due to the high variability in ethnicity, prioritised ethnicity values were grouped into the 6 major categories of ethnicities in NZ based on NZ Statistics level 2 categories. Accordingly, the major ethnicities used to analyse the dataset were European, Pacific peoples, Asian, Māori, MELAA, and Other ethnicities. Categorical features with binary values, gender and smoker status, were encoded to binary values (0 and 1) to represent them in a numeric form.

5.4.4 Grouping LoS

LoS is a continuous variable having an extensive range of values. Therefore, to analyse the data, LoS was labelled into two groups (binary classification): “Stay” (group 1) and “No stay” (group 0). When LoS > 0, they were labelled as “Stay” and denoted by 1. Admissions with LoS = 0 were labelled as “No stay” and represented by 0. Accordingly, groups “Stay” and “No stay” had 6,559 and 4,789 records, respectively. The LoS grouping was introduced solely for descriptive and exploratory analysis purposes. The intention was to distinguish encounters that did not result in an inpatient stay from those requiring hospital admission.

5.4.5 Exploring Feature Set and Development of ML Models

First, a descriptive analysis was performed to observe the characteristics of individual features related to asthma admissions. Then, bivariate and multivariate analyses were conducted to explore the relationship among different features. To understand the correlation between the features and the target variable, LoS group correlation analysis was performed to investigate the p-value for each feature. Since there are categorical and numerical features, the chi-square test was used to find the relationship between categorical features and the LoS group target variable. The ANOVA [38] test was performed to see the correlation between age and the LoS group. As a result, we found that all the features except the admit month were correlated with the LoS group. However, the admit month was not removed from the current work to retain the seasonality factor in the dataset.

After performing the analysis, different ML models were developed to predict asthma LoS. For this, we used only the LoS records between 1 and 14, including boundaries, as there was a minimal number of records with higher LoS values, which were not suitable for training the ML models. In addition to these regression models, classification models were developed using the LoS groups defined previously. Before the models were developed, categorical features were converted to numerical features. Accordingly, multi-categorical features, admit day of the week, diagnostic code, admit month, DHB group and ethnicity group, were encoded using the one-hot encoder approach since binary features had already been converted previously in the analysis stage. This encoding technique creates separate binary-valued variables to represent each category of a variable. For example, seven new binary variables were generated after one-hot encoding for the admit day of the week to replace the original variable. Since age had continuous values, this was standardised before applying to the ML models.

The dataset was split into training and test sets of 70% and 30%, respectively. The ML models developed were SVR, RF, and KNN. The boosting algorithm, XGB, also used to build the ML models. These ML models were developed using the training set, and hyperparameter tuning was conducted using the GridSearchCV technique with 5-fold CV. Commonly for tree-based models, hyperparameters such as number of estimators (trees) and number of features were tuned. Additionally, min_samples_leaf and min_samples_split were tuned for RF, while learning rate, sample size and maximum depth of trees were tuned for XGBoost. For SVR, gamma, kernel, degree, and C (penalty for misclassifying a data point) were tuned, while hyperparameters tuned for KNN were the algorithm to compute the nearest neighbour, number of neighbours, and power parameter, for distance calculation and weights. The models with the optimal parameters were then trained and evaluated using a test set. Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) were used to evaluate the model performance. Analysis and ML model development were performed using Python 3.8 programming language in Jupyter notebook using the library packages, including matplotlib and scikit-learn. All executions were conducted on a personal computer with 16 GB RAM, an Intel Core i5 processor, and a CPU of 1.60 GHz. The code to implement data preprocessing and develop ML models is available at https://github.com/DarshaWK/Predict-asthma-LOS-using-ML.

5.5 Results

5.5.1 Characteristics of Asthma Admission Data

The age of the patients ranged from 0 to 101 years, with a mean age of around 30 years. A considerable number of admissions were from children younger than 5 years old. Gender representation of females (49.2%) and males (50.8%) was roughly equivalent. Less than 10% were smokers. Most asthma admissions were European (35.68%) and Pacific (23.0%). Asians and Māori made up 21.8% and 15.8% of asthma-related admissions, respectively, whilst MELAA and other ethnicities made up the remaining 3.7%.

One third of patients (33.1% and 29.3%) were identified with unspecified asthma and wheezing diagnostic codes, respectively. May to August (winter season) had a higher number of asthma admissions during the year, while January and December (summer) showed the least number of admissions, with monthly proportions >9% and <7%, respectively. Figure 5.2 shows the yearly trend in asthma admissions. There were 2771 admissions in 2017, which increased to 2962 in 2018. In 2020, the number of admissions had reduced by 263 compared to 2019. One possible reason for this drop could be lockdowns and public health practices, such as mask-wearing, imposed by health practitioners to avoid COVID-19. This may have safeguarded asthma patients from getting exposed to asthma triggers.

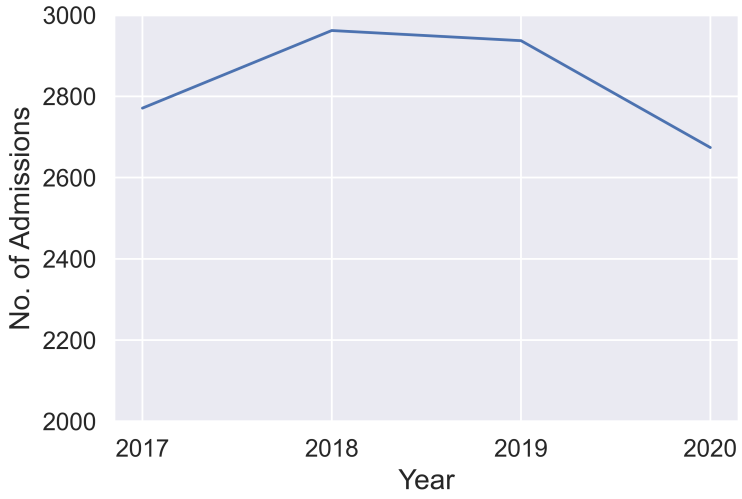


Fig. 5.2 The yearly trend in the number of asthma admissions.

Concerning the distribution of LoS, most LoS were of short duration (see Figure G.1). After handling the outliers using the z-score values, LoS had a mean of 2.28 days and a standard deviation of 4.97 days. Minimum and maximum LoS were 0 and 45 days. The total

number of instances in the dataset was reduced to 10,833. Asthma admissions were higher on Mondays (16.5%) and Tuesdays (15.9%) and lowest on Saturdays (11.2%).

The distribution of different features under each of the LoS groups is demonstrated in Figure 5.3. The y-axes of the graphs represent the number of admissions/count, while the x-axes have different values for each feature. A dual colour code (orange and blue) has been used to differentiate the LoS groups (stay (1) and no stay (0)). Mondays and Tuesdays have the highest admissions for LoS group 1, while Mondays and Fridays have the highest for LoS group 0. Each day had a greater proportion of admissions for patients whose LoS was more than 1 day compared to no stays. Over 60% of females stay more than 1 day when admitted, compared to around 55% of males. A greater proportion of smokers remain for more than 1 day (66.3%) compared to non-smokers (56.9%). More than half of the patients admitted with unspecified asthma, dyspnoea, cough, and stridor have a LoS > 1 day, with the proportions of 64.2%, 58.5%, 59.9%, and 81%, respectively. But for patients with admitted diagnoses of wheezing and allergic asthma, the majority have a LoS < 1 day. During the year, compared to no-stay groups, each month has more asthma admissions with a LoS > 1 day rather than 0 days. Among the major ethnic groups, 60% of Europeans, 58.9% of Pacific peoples, 53.5% of Asians, and 59.6% of Māori stay more than 1 day after admission. Nearly half of the MELAA and Other ethnic groups stay for more than 1 day.

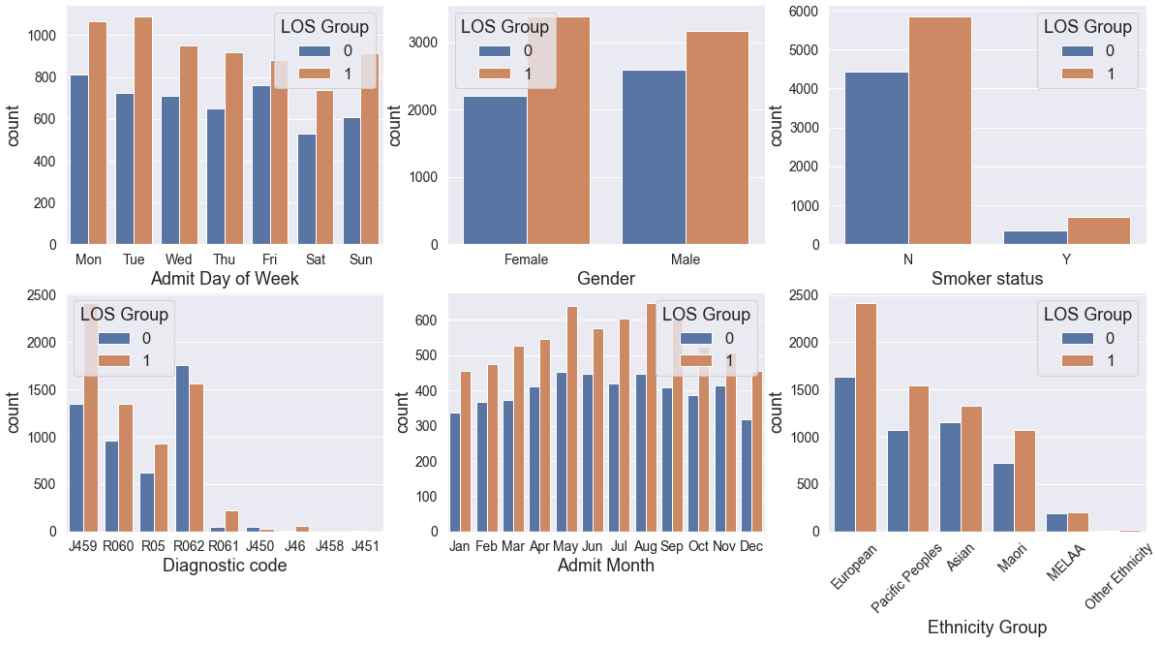


Fig. 5.3 Distribution of feature values among the LoS groups.

As seen in Figure 5.4, most children under 10 years of age do not have a prolonged LoS in the hospital. However, as age increases, the number of patients who stayed more than 1

day exceeds the no-stay group. Most patients older than 68 years have a longer LoS, with a small proportion being discharged on the same day of admission. Asthma admissions in the Auckland DHB are significantly higher than in other DHBs.

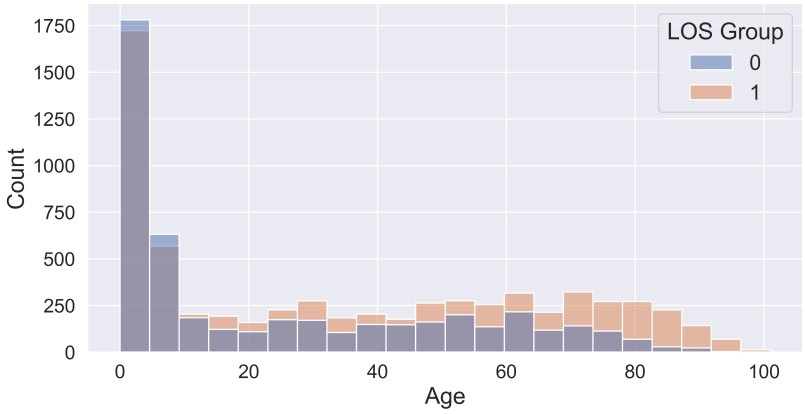


Fig. 5.4 Age distribution among the LoS groups.

5.5.2 Association of Features with LoS

This study observed several key features concerning asthma admission in NZ. Admission rates for children aged 0–5 years were significantly higher than for other ages. Figure 5.5 illustrates the LoS concerning age for males and females. One important observation is that very few males aged between 20 and 50 years stayed more than 20 days in hospitals compared to females. But like females, many males over 50 years had a LoS greater than 20 days. The average value of LoS is 2.3 days for the whole population, equal to the average LoS for asthma patients in 2018, as reported in a previous study [17].

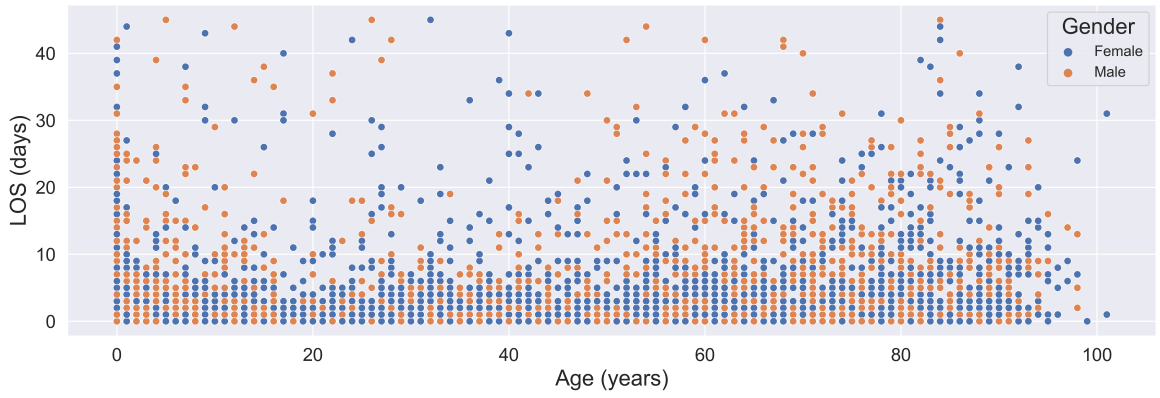


Fig. 5.5 Correlation between age, LoS, and gender.

The central tendency of LoS against ethnic groups is illustrated in Figure 5.6 based on gender. The average LoS for females was higher than for males in most ethnic groups, including Europeans, Pacific peoples, Māori, and Other ethnicities. For all the ethnic groups, the mean LoS for females was 1.5 days or more. The Other ethnic group had the lowest mean LoS values for female and male patients. Among all the ethnic groups, Māori had the highest mean LoS at nearly 2.8 days (95% CI) for females and European at 2.6 days (95% CI) for males. Table J.1 shows the p-values of correlation between the feature set and the LoS group, which were calculated from the correlation finding tests.

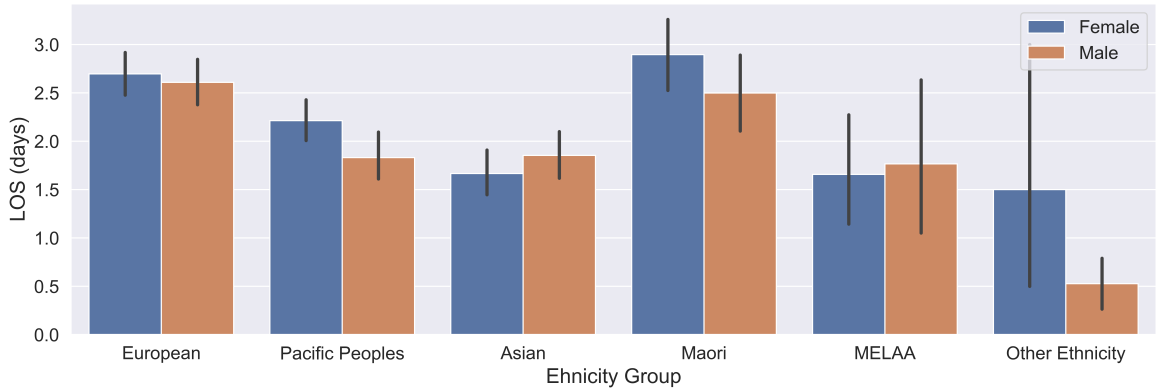


Fig. 5.6 Correlation between ethnicity, average LoS, and gender.

5.5.3 Development of ML Models

Several ML models were developed to predict the LoS using the asthma admission dataset with 5862 instances and with the LoS ranging from 1 to 14 days. As LoS is a continuous variable, ML regressors were developed to predict the LoS of asthma patients. Table 5.2 shows the performance of the models based on the evaluation metrics RMSE and MAE. The performance of additionally developed classification models are given in Table J.2.As all these are error values, the models with lower error values perform better. The RF model predictions made for the first 500 instances of the test data are plotted in Figure 5.7. Predicted LoS values ranged from 1 to 6 days, with a mean of 2.6 days and a standard deviation of 1.0 days. Figure 5.8 demonstrates the mean predicted LoS among the different values of the features in the test data, with the error bars showing the standard deviation. Almost all feature values had a mean predicted LoS between 2 and 3 days, except the diagnostic codes R062 and R05, which had a mean predicted LoS of around 1.5 and 4 days, respectively. According to our test dataset, there was nearly a 0.5-day gap in the predicted LoS among gender and smoker status differences. However, there were fluctuations among admitted weekday, diagnostic code, admit month and ethnicity. Wednesday to Friday had comparatively higher

mean predicted LoS values. February and November had the least mean predicted LoS, while June, August, and September had relatively higher values. The mean predicted LoS for Europeans was nearly 3 days, and for Māori, it was above 2.5 days. Significantly, the Other ethnic group had a larger standard deviation, which needs to be further analysed, only considering minority groups.

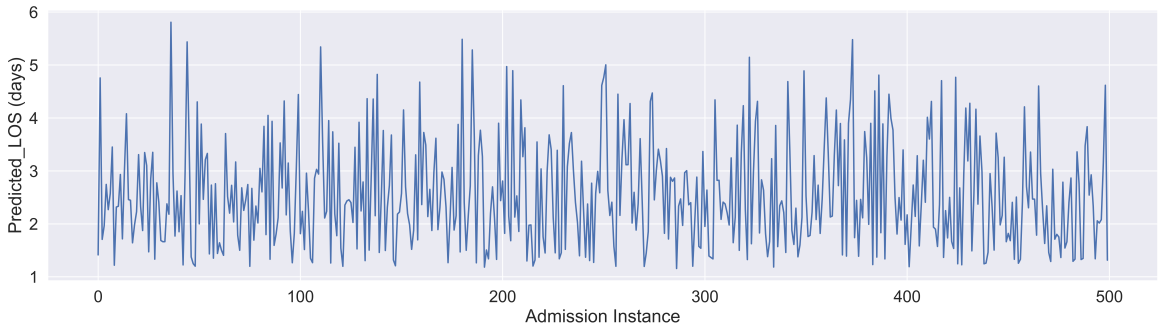


Fig. 5.7 Predicted LoS for the first 500 admission instances in the test data.

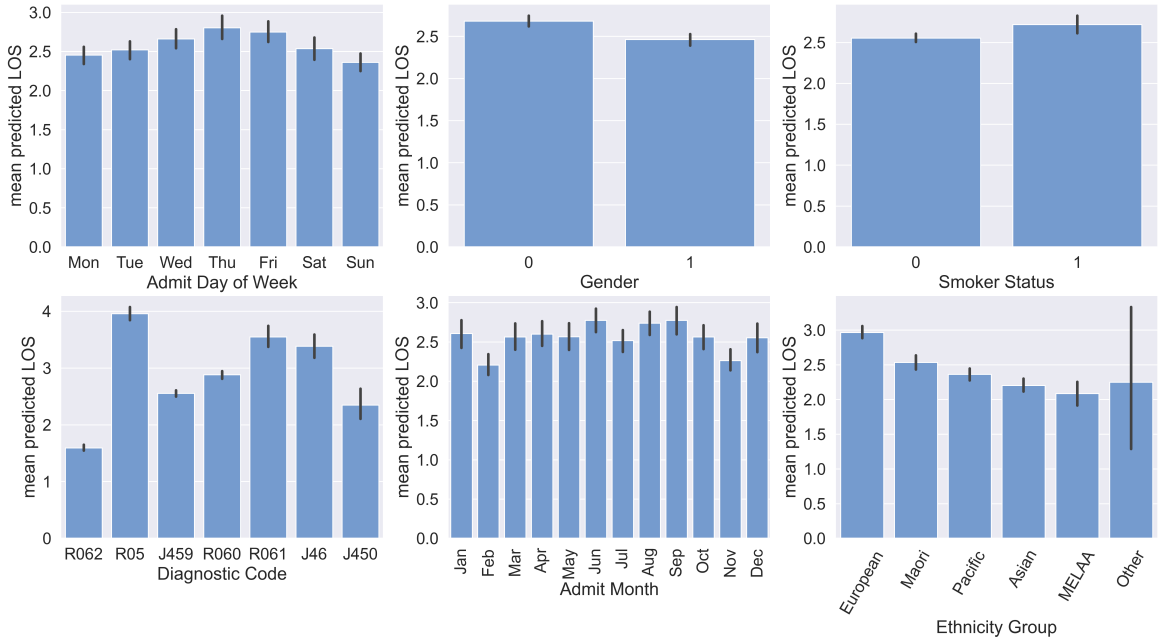


Fig. 5.8 Mean predicted LoS for different feature values.

Table 5.2 Performance of the regression models in predicting LoS.

Model	RMSE[1]	MAE[2]
SVR	2.65	1.50
Random Forest	2.48	1.67
KNN	3.08	1.58
XGBoost	2.50	1.67

[1] Root Mean Squared Error
[2] Mean Absolute Error

5.6 Discussion

This study explored predictors of LoS for asthma-related hospital admissions for children and adults in NZ. Our study found several predictors that increase LoS: gender, ethnicity, smoking status, day of the week of admission, month of admission, and presenting diagnosis.

Most of the admissions were recorded from children rather than adults, which is expected since asthma is predominantly a childhood medical problem. However, another study observed that most older adults have a higher LoS than younger adults [153]. The research work [147] suggested that extra attention should be paid to adult asthma patients. Asthma LoS had a clear association with the day of the week. Several studies [165, 166] have found that Mondays are associated with the highest asthma-related admission rates compared to weekends, which is consistent with our findings. One of the main reasons for this could be that air pollution peaks during weekdays due to more vehicle movements, factory operations, and other factors, while the opposite may be true on weekends [167]. Even indoor air pollution in urban schools due to $PM_{2.5}$ and PM_{10} will cause respiratory disorders among children, leading to more hospital admissions on weekdays [168]. Another reason could be that asthma patients are busy or out during weekends and wait until Monday to meet with their general physician or go to the hospital, as they try to manage themselves over the weekends.

We observed *unspecified asthma* (J459) as the most common diagnostic code in asthma-related admissions, similar to the findings of a past study [151]. However, the previous study [151] was on children (0–14 years) in NZ. It did not include wheezing (R062) as an inclusion criterion for asthma-related admissions, which was our study's second most common diagnosis. Our study included a broad range of diagnostic codes to identify people with asthma; however, due to the broad inclusion criteria, it is possible that people admitted for non-asthma-related diagnoses were included. We decided to use broad criteria as we wanted our LoS prediction model to be more sensitive than just specifically for asthma, allowing triage at the point of asthma-related presentations to the hospital.

Many studies have reported that asthma is affected by environmental and meteorological factors [35, 36, 169]. These factors change with the seasons of the year. We found that seasonality greatly affected asthma-related admissions, with higher admissions in the winter (June to August) and fewer admissions during the Summer. These findings are essential to help health workforce resourcing and planning to meet predicted hospital demand. A previous study [151] showed that Māori children were hospitalised with asthma at higher rates than non-Māori. Our study found that Māori females have the highest average LoS, potentially reflecting a greater difficulty with asthma management. This could be due to delays in seeking healthcare or treatment due to financial or logistical barriers, a reluctance to proceed with multiple treatment adjustments, and/or a lack of rapport with healthcare providers [170]. Unfortunately, we did not have access to asthma severity or symptom data that may explain the variations seen in the LoS. The higher admission rate for Pacific peoples reported in previous studies is consistent with the results of our study [89]. Also, the average LoS (2.3 days) we found for the asthma-related admissions is comparable with the result (1.9 days) of a previous study [154]. More research on minority ethnic groups and access to the severity and symptom data would be helpful for further comparisons.

Here, predicting the LoS was considered a regression problem. This could be extended for multiclass classification to make the prediction more informative for health care providers with more variables and examples. Predicting the LoS of an asthma patient enables healthcare teams to take immediate, appropriate clinical actions with patients. Also, before analysing and developing the ML models, a few age groups could be defined beforehand without considering this as a continuous value. Furthermore, in addition to these features, future research could be carried out to find other factors, such as deprivation quintile, domicile, and their association with the current factors, which would be relevant in predicting LoS for asthma patients, providing further information for hospital staff to better manage and allocate available health resources.

In predicting LoS using ML models, the RF algorithms performed better with an RMSE of 2.48. The predictions of the RF model had a mean LoS of 2.6 days with a standard deviation of 1.0, which is acceptable. XGB also performed well, giving an RMSE of 2.50. Both algorithms showed identical values for the performance metric MAE of 1.67. However, SVR had the lowest MAE of 1.50. Although KNN was the least performing algorithm, it gave an RMSE of 3.08 and MAE of 1.58, which are acceptable. This implies that ML could predict asthma LoS more accurately using demographic, socioeconomic, and temporal factors. These factors could be applied to clinical practice in real-time to help triage patients, for example, awareness that LoS could be higher for specific demographic groups at certain times of the year. Whether this translates to cost savings will need further investigation.

However, the models need to be improved to make them more applicable in clinical practice. For example, this could be considered a multiclass classification problem by grouping LoS into meaningful groups with knowledge from asthma experts. Furthermore, more robust ML models such as NN and deep learning algorithms could be applied to predict LoS, which has shown better performance in medical prediction and recommendation systems [35, 36, 58, 146, 171, 172, 173, 174].

5.7 Summary of the Chapter

Asthma hospitalisation represents a severe outcome of an asthma attack and may result in inpatient stays that contribute to LoS. In this chapter, we analysed a real-world asthma-related hospital admission dataset in NZ to explore the characteristics and existing associations between distinct variables. Subsequently, several ML models were developed to predict asthma LoS. The results showed that demographic, diagnosis and temporal factors are highly associated with asthma LoS. The highest average LoS was found among Māori females. This emphasises the need for particular attention to the Māori community, the Indigenous population of NZ, to explore how asthma attack prediction models can be informed and adapted to better support these communities in asthma management. As a result, it might help to reduce or mitigate the severe asthma outcomes, thereby reducing asthma hospitalisation. Although the findings showed the positive applicability of ML models in asthma LoS prediction, they require further validation before deploying these models in clinical settings. One of the key findings of this study confirmed the fact that asthma has a higher impact on Māori and Pasifika ethnic groups. Addressing the gap in the literature (RG3) and the findings of the current chapter lead to the Chapters 6 and 7.

Chapter 6

Perceptions Towards Using Artificial Intelligence and Technology for Asthma Attack Risk Prediction: A Qualitative Exploration of Māori Views (Publication #6)

Citation: W. K. D. Jayamini, F. Mirza, M. C. Bidois-Putt, M. A. Naeem, and A. H. Y. Chan, "Perceptions Toward Using Artificial Intelligence and Technology for Asthma Attack Risk Prediction: Qualitative Exploration of Māori Views," *JMIR Formative Research*, vol. 8, e59811, 2024, doi: https://doi.org/10.2196/59811. (**IF=2.1, Q2**)

6.1 Publication #6 Prelude

Asthma affects adults and children regardless of ethnicity. NZ is a high asthma prevalence region, and it consists of multiple ethnicities. Literature showed that Māori and Pasifika are high-risk asthma communities who are different ethnic groups in NZ. In our prior work, we found that asthma greatly affects the Māori and Pasifika groups. Therefore, when investigating the application of ML techniques to predict asthma attack risk at the national level, it is critical to examine the impact of asthma across vulnerable communities within the population. A primary issue in these ML-based prediction model development is that these models can under-represent the minority ethnic groups and introduce bias, potentially

exacerbating health disparities. Also, the best way to understand the technology drivers within these communities is through direct communication with them. As an initial step, in this chapter, we aimed to explore Māori views and perceptions of using AI and digital technologies for asthma self-management, identify key considerations for developing asthma attack risk prediction models, and ensure Māori are represented in ML models without worsening existing health inequities.

6.2 Introduction

Asthma is a long-term respiratory condition characterised by restricted airflow and rapid fluctuations in airway functioning over a short period [143]. Asthma affects around 262 million people worldwide and caused 455,000 deaths in 2019 [14]. Hence, optimal asthma management is critical, yet in NZ, asthma control continues to be poor, with the rate of asthma exacerbations increasing by a third in the last decade [28]. One of the key issues with asthma management is the lack of awareness of worsening asthma, with exacerbations often being detected too late [42]. Digital technologies and the use of AI can support earlier detection of worsening asthma and improved asthma self-management [175]. There exists an increasing number of digital solutions to support asthma self-management. For instance, the smart peak flow meter used to measure the airflow out of the lungs, and the smart inhaler used for taking medicines, integrated with mobile applications, can help the user automatically record peak flow readings and improve the use of inhaler medications with the audiovisual reminders, respectively [176]. There are also digital health management systems and decision support systems that can help control and manage asthma, which have been shown in a recent Cochrane review to significantly improve adherence and asthma control [28, 177]. These systems provide opportunities for monitoring treatment behaviours, supporting communication between patients and healthcare providers, as well as generating data that improves patient-provider interactions [132]. Together, these technologies can assist patients in preventing or mitigating severe outcomes resulting from worsening asthma symptoms. However, despite the vast growth in digital healthcare solutions to support asthma management, implementation of these technologies into routine practice remains limited [175], and even fewer studies exist to explore how health technologies affect indigenous populations. One possible reason for poor implementation could be the risk of both consumers and healthcare providers feeling overwhelmed by the large number of choices and functionalities provided by these technologies [178]. An observational study [179] conducted in Portugal reported that out of 336 participants, only 3% were using an asthma app to monitor asthma and improve inhaler adherence. Another reason is the concerns about data

privacy and security [175]. Concerns about data accuracy, training the users, and integration with current patient management systems can limit these implementations [180]. Further, poor e-health literacy, cost-effectiveness, limitations with technical infrastructure, and gaps in policy and regulatory systems can be causes for these limitations [180]. Understanding the factors that can improve engagement with digital health solutions is key if we are to design and develop solutions to facilitate asthma management as part of routine care.

Recently, there has been an increase in the use of AI, including ML techniques, for developing disease prediction models embedded in digital health solutions. AI is an advanced computer-based technology used to generate intelligent machines that can perform tasks requiring human intelligence. It allows the development of algorithms capable of reasoning, learning, and making data-based decisions. ML is a subset of AI that enables machines to learn and improve from experience without being explicitly programmed. It provides algorithms and techniques that allow computers to analyse and interpret complex patterns and relationships in data automatically. ML algorithms can learn from historical data, identify patterns, and make predictions or decisions based on the learning. Previous works show the potential of ML to predict asthma attacks and improve asthma management [31, 57].

However, ML models have raised ethical concerns, as these tools can potentially exacerbate existing health disparities. A study in the United States of America found that ML algorithms developed based on EHR tend to be biased towards poor and minor ethnic communities and those who lack the financial means for regular and ongoing medical treatment [181]. If the information utilised to train AI algorithms contains biases, the algorithms will uphold these biases, resulting in inaccurate predictions [182]. This is important to address as there is a risk that the use of these AI algorithms could worsen inequities. A narrative review reports that AI algorithms trained in limited demographic groups can lead to a lack of generalisability and unintended social bias [183]. Recent work has demonstrated that state-of-the-art clinical prediction models underperform on women, ethnic, and racial minorities due to the higher absolute number of non-Māori versus Māori represented in the population in NZ [184]. NZ has one of the highest rates of asthma worldwide, with one in seven children (13%) aged 2–14 years (110,000 children) and one in eight adults (12%; 452,000 adults) currently taking asthma medication [89]. Statistics show that asthma has a higher impact on Māori [89], with asthma admission rates of Māori children being twice the rate of Pākehā (non-Māori) children [151]. Our previous research work [94] found that Māori females have the highest mean asthma length of hospital stay compared to other ethnic groups in NZ. All these findings highlight the significant impact of asthma on the Māori community. Capacity in technology handling, level of education, perception of communities, and cultural sensitivity are other factors that need to be addressed as technology advances,

and to ensure that technologies do not worsen inequities [185]. However, there has been limited research in this area exploring the perceptions of Māori, the Indigenous population of NZ, towards the use of AI and digital technology to manage asthma.

The research aimed to understand the perceptions of Māori on using AI and digital technology to predict the risk of asthma attacks. We explored to identify factors that may need to be considered when developing ML models and systems to predict the risk of asthma attacks in Māori to ensure the models do not exacerbate existing health inequities. The findings can be used to inform the models and systems designed for asthma risk prediction for Māori and identify specific user requirements to inform customisation of future interventions.

6.3 Methodology

6.3.1 Overview

A qualitative study design was followed with inductive thematic analysis.

6.3.2 Participant Recruitment

Māori diagnosed with asthma were recruited through the team's Māori researcher's (MP) networks and via the research team's networks following a snowball sampling approach. The Māori researcher engaged with the Māori community through advertisements (Flyer provided in Appendix K), social media and word of mouth, in line with the Te Ao Māori (Māori worldview) principle of partnerships and relationships where people's relationships are central to Māori [186]. The potential participants identified through the above approaches were initially provided with an information sheet (Appendix L) which explained the study to inform them what kind of research they are going to be involved in. Contact details were recorded from those who were happy to engage in the study and satisfied the inclusion criteria. Inclusion criteria required participants to identify as being of Māori ethnicity, have a diagnosis of asthma, provide informed consent, and agree to interview recording or note-taking. The aim was to have a minimum sample of 20 participants to ensure data richness and thematic saturation [187].

6.3.3 Interview Process

The interviews were conducted by the Māori researcher using a semi-structured guide (Appendix M). It was culturally appropriate for the Māori researcher to conduct the interviews.

This guide was developed with reference to previous research, which explored Māori experiences and perceptions toward antibiotics for upper respiratory tract symptoms [188], and attitudes toward brain health [189]. The interview guide was reviewed by the team, including the Māori researcher who is affiliated with iwi (tribes) Ngāti Awa and Te Arawa. The interview guide consisted of six key sections with multiple open-ended questions. In the first section, the interviewer (MP) provided an introduction to inform the participant about the purpose and importance of the study. The participant was then asked some basic questions to facilitate their entry into the discussion. In the second section, the interviewer asked the participants open-ended questions about their experience with asthma to elicit information without steering the conversation in a potentially biased direction. The open-ended questions in the third section targeted identifying factors that trigger their asthma. This section of the interview helped the researchers identify any factors specific to Māori asthma patients that may need to be considered in asthma risk prediction models to ensure Māori needs are represented. The discussion was then directed towards applying technology in asthma prediction, and Māori views and perceptions towards the use of AI and digital technology. Each interview took 30 - 60 minutes to complete, and no repeat interviews were required.

Participants were given the choice to respond in either English or Te Reo Māori, according to their preference. Participants could invite their carer or other support person (for example, the primary caregiver involved in the patient's daily care) to partake in the interview if the carer was deemed more appropriate. This was solely based on the participant's choice, and there were no mandatory requirements to do so. The interviews were conducted at a location mutually convenient to the participant and the interviewer. Due to the time constraints for participants for travelling, severe weather conditions, sickness, and availability, online interviews via videoconferencing were an alternative offered to participants.

6.3.4 Ethical Considerations

Ethics approval for the study was received from the Auckland Health Research Ethics Committee (Appendix N). Participants provided written consent before interviews, and verbal consent was obtained beforehand for online interviews (consent form provided in Appendix O). Most interviews were audio-recorded with the participant's consent, except for two cases where participants opted for the researcher to take notes instead of recording. Participants were assigned two-digit unique identifiers, and their data were stored on a password-protected server managed by one of the authors' institutions. As approved by the ethics committee, all participants received a NZD $30 supermarket voucher to acknowledge their time and contributions.

6.3.5 Interview Data Analysis

The primary researcher employed the transcription tool built within Microsoft Word software to transcribe the recorded interviews. These preliminary transcripts were subsequently refined through manual review, involving repeated listening to the recordings for accuracy. The Māori researcher's input was sought to enhance the transcripts' credibility, particularly in identifying Māori terms employed by the participants. Hence, the transcripts were not returned to the participants for comments or correction. The transcripts were then stored anonymously, utilising a distinct participant identifier.

The final transcript data underwent a careful procedure for analysis and were imported into NVivo 12 software to facilitate organised examination and analysis. The analysis process encompassed manual coding, where key themes and patterns were identified through attentive content evaluation. The interview data were analysed using the Braun and Clarke method of inductive thematic analysis [41]. This study was underpinned by an interpretivist epistemological position and a constructivist theoretical perspective. This assumes that meaning is constructed through individuals' experiences, perceptions, and cultural context. This approach was appropriate as the study aimed to explore participants' perceptions of AI and technology for asthma attack risk prediction. Thematic analysis was therefore used to identify and interpret patterns of meaning within the data. An initial coding framework was developed by the primary researcher (WJ), which was then discussed with a senior researcher (AC) and the team's Māori researcher (MP). The coding tree had three main branches: concerns over AI use, expectations over AI-based systems and experience with the healthcare systems. Any disagreements were resolved by consensus discussion within the research team. The coding framework was reviewed and refined, then used to code the interview data, with common themes being extracted iteratively.

Further, we executed a sentiment analysis to understand the opinion of the interview participants regarding the use of AI and digital technology for asthma attack risk prediction. We employed a natural language processing model called BERT, which stands for Bidirectional Encoder Representations from Transformers [190]. BERT is designed to learn by looking at the words in a sentence from both directions (left to right and right to left) at the same time. This bidirectional approach enables BERT to capture the intricacies of context more effectively. Consequently, the pre-trained BERT representations can be fine-tuned with the addition of just a single output layer, facilitating the development of state-of-the-art models for various tasks such as question answering and language inference, without requiring substantial modifications to task-specific architectures. We utilised an existing model in this work, as there were not enough data samples to fine-tune the model for our own scenario. We utilised the BERT model downloaded from the Hugging Face website to conduct the analysis

[191]. This model is fine-tuned for sentiment analysis on product reviews in six languages: English, Dutch, German, French, Spanish, and Italian. Based on those learnings, it predicts the sentiment of a given input as a number of stars (between 1 and 5). The number of stars shows the positivity or the satisfaction level of the input. We utilised the existing model in this work, as there were not enough data samples to fine-tune the model for our scenario. The answers to the interview question, "What do you think about using technology to help manage your asthma?" were extracted for the sentiment analysis. This data was then fed into the model, and the sentiment scores were obtained from 1 to 5, which were labelled as very negative, negative, neutral, positive, and very positive for analysis.

The methodology followed in the study is graphically represented in Figure 6.1.

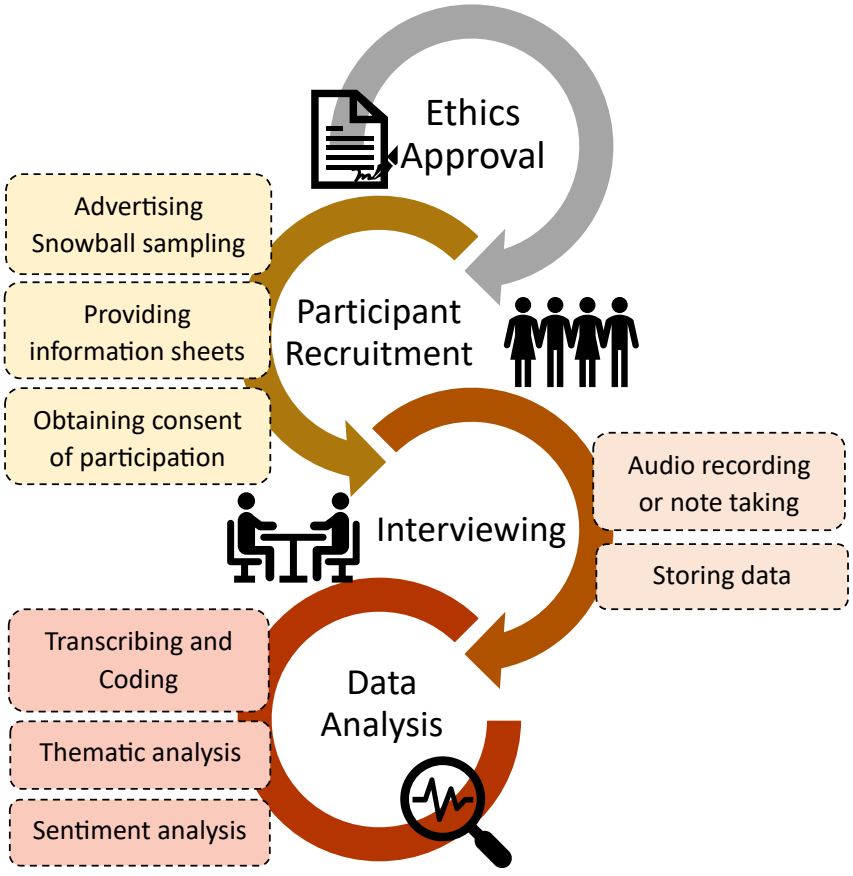


Fig. 6.1 Methodology of exploring Māori perceptions on using AI and technology for asthma risk prediction.

6.4 Results

6.4.1 Participant Characteristics

The final sample consisted of 20 participants (male: n=3, 15%; female: n=17, 85%) aged 18 to 76 (mean=42, SD=16.16) years. None of the participants refused to participate in the interviews after recruitment. The participants came from different regions of the North and South Islands of NZ. In total, 2 participants were interviewed together, while the remaining participants were interviewed one-on-one. Out of all participants, 35% had been diagnosed with asthma at less than 5 years of age, while 45% were diagnosed at ages 5-10 years. The remaining participants were diagnosed after the age of 50 years. Most managed their asthma by themselves, while 25% received support from family, friends, or health professionals. Half of the interviews (n=10) were conducted digitally using the Zoom videoconferencing platform, and the rest were conducted kanohi ki te kanohi, according to the participant's preference.

6.4.2 Key Themes Generated from Interview Data

In total, four key themes were identified relating to the perceptions of Māori toward the use of AI and digital technology for the prediction of asthma attacks: (1) concerns about AI use, (2) interest in using technology to support asthma, (3) characteristics of AI-based systems, and (4) experience in asthma management and opportunities for technology to improve care. Themes 1 and 4 were further divided into subthemes, as represented in Figure 6.2. For Theme 1, concerns about AI use were subdivided into three subthemes: (1) trust about technology, (2) face-to-face interaction with Māori, and (3) inadequate knowledge about AI and digital technology. Theme 4, experience in asthma management, also consists of three subthemes: (1) previous experience with medical staff, (2) accessibility to health care resources, and (3) asthma-triggering factors. Example quotes for each theme are provided in Appendix P. The following sections present the different themes identified, along with illustrative quotes extracted from the interview data.

6.4.2.1 Theme 1: Concerns About AI Use

6.4.2.1.1 Trust in Technology

The participants raised several concerns about using AI and digital technology to predict the risk of asthma attacks. A primary concern was the level of trust that they could place in AI and digital technology usage. The degree of confidence an individual places in something

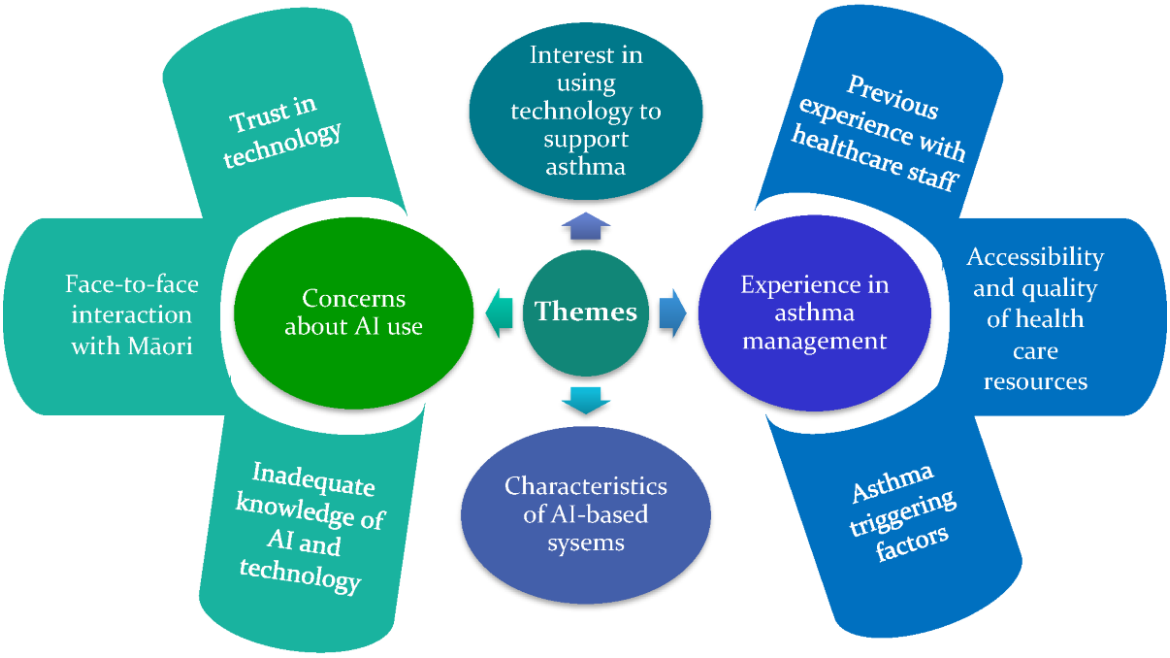


Fig. 6.2 Key qualitative themes relating to the perceptions of Māori towards using AI and digital technology for asthma attack prediction.

significantly influences their willingness to embrace and endorse that particular entity or concept [192]. Some participants felt that they could not trust these computer-based systems, particularly as they had received limited information. For Māori, trust is a vitally important aspect of all their relationships. Some participants emphasised that more trust is built up in a relationship with another person of Māori ethnicity.

> "*Like I would be like, yeah, you're going to save me, that would not be something I can trust 100%, ever. ...I also believe it can be [obscenity] dangerous when we're relying on that to protect things when everyone's bodies are different because of my health and my journey; I'd be dead if I relied on [obscenity] technology.*" [E2, F (female), 37 years old]

> "*think trust [emphasises], there's much trust building up relationships with Māori and stuff like that*" [E5, F, 38 years old]

The researchers note that one way that the general public gets to know AI is through watching movies, where AI is usually portrayed as being used to harm the world. The informal and limited information that many people receive makes them believe that AI is bad, and they should not interact with it, especially to manage their health.

> "*... that's what they're thinking of ... aww AI is dangerous.*" [M (male), 76 years old]

> "*You know things can harm; you got people talking about AI taking over the world. That's quite scary. AI's not doing what they've been told to do, and I've read so many things about that. And I and I freak out with that.*" [S7, F, 37 years old]

Some participants were hesitant to use AI support to manage their health due to a lack of information about its real benefits. Although they had read about AI, they seemed to have encountered predominantly negative perspectives on its use, which may have contributed to their trust issues.

6.4.2.1.2 Face-to-Face (Kanohi Ki Te Kanohi) Interaction With Māori

In the context of introducing new AI and digital technologies or concepts, some participants preferred engaging in a face-to-face, physical setting with someone from their own communities, for example, a Māori representative. They felt that an individual coming from their own cultural community and possessing cultural insights would facilitate seamless communication and understanding. Moreover, a few participants felt more comfortable and had greater trust when they received information about these technological applications through the involvement of senior members of the Māori community (kaumatua). Also, a few participants expressed their preference for having a face-to-face option within these technological solutions, "*face-to-face is better*" [M, 23 years old]. This approach would instil a greater sense of ease and comfort in their communication.

> "*My personal thing for Māori would be that kanohi ki te kanohi, you know, face-to-face and the fact that they're comfortable with you – A caregiver*

This shows that relationships and culture are important, as having face-to-face interactions with a person from the same culture explaining the information would likely be more readily accepted by the recipient.

6.4.2.1.3 Inadequate Knowledge of AI and digital technology

Interview discussions highlighted that many of the participants wanted to know more about AI and its uses. Several participants candidly expressed that they had limited knowledge of AI and recognised that they wanted to acquaint themselves with these technologies better. In addition, they recognised the importance of broadening their understanding, realising the positive impact AI could have on their lives and community.

"*...and when it comes to AI, I don't know much about it. I know little bits.*". [E2, F, 37 years old]

"*Yeah, I think I need to know more about them because I don't know much about technologies and the word AI*" [S7, F, 37 years old]

While the term AI was familiar to them, its practical uses remained unfamiliar. Some participants were apprehensive when the interviewer introduced the concept of using AI to predict asthma attack risks.

"*Just hearing it to say, you know, to help me with, dealing with my asthma, I just feel nervous about hearing that.*" [E5, F, 38 years old]

"*...not everybody's confident with technology, and there's been, you know, a lot of rumours around technology and stuff too.*" [S7, F, 37 years old]

The fear of technology stemmed from their limited understanding of AI and related concepts, as stated above in reference to AI in movies. Consequently, they tended to accept information given to them at face value due to their lack of information in this area.

6.4.2.2 Theme 2: Interest in Using Technology to Support Asthma

Many participants showed a strong interest in adopting and understanding AI and digital technology. Some participants already used smart devices such as smartphones and tablets with AI-related features. A few participants were intrigued by the potential of AI to assist in managing their asthma. Most participants responded positively when the interviewer explained how AI and digital technology can be used to predict asthma attack risk.

"*Ohh, now you pull up [sic] like that. I think that you know, it's not as scary as I imagined.*" [E5, F, 38 years old]

"*That sounds awesome, absolutely, yeah, absolutely... It's just interesting to know. Because, I mean, it's so much different to manually doing something, like doing it yourself and trying to write a diary on, like, how many times have you done that yet? This and that, and yeah, that'd be interesting.*" [S6, F, 44 years old]

"*I am a great believer in technology ... I think it will benefit humanity immensely. It must be handled the right way.*" [P15, M, 76 years old]

Likewise, many participants generally demonstrated an improved shift in their perceptions of AI, recognising its potential to make tasks more convenient and expressing confidence in its benefits for humanity when managed correctly.

6.4.2.3 Theme 3: Characteristics of AI-Based Systems

With regard to the attributes of a computer-based system for predicting asthma attack risks, many participants emphasised the importance of utmost simplicity in usage. Notably, they emphasised that the system should be easy to use for older individuals who may encounter challenges when engaging with technology.

> "*When I see my Papa, I see him really struggle to use the tech, so I think they would need to be relatively simple to use. User-friendly. Simple design, simple to use for our elderly.*" [C4, F, 32 years old]

Many participants highlighted that many Māori already need to engage in activities that pose significant complexity, particularly when it comes to interpreting medical terms and navigating the health system. As a result, their preference was for applications characterised by simplicity. While younger participants preferred technologically advanced solutions, they also emphasised the value of straightforwardness, enabling users to navigate the application with minimal effort.

> "*I think if it wasn't too bulky, I guess. We don't have to remember too much.*" [B9, M, 62 years old]

> "*High-tech, Straightforward. Easy to use.*" [K3, F, 18 years old]

> "*When it comes to medical stuff, it needs to be in English, and it needs to be simple, not using scientific words... not using abbreviations, not complicated, no scientific words.*" [P11, F, 51 years old]

Some participants preferred the use of simple English without any scientific words in a proposed system, whereas a few preferred using Te reo Māori, "*An option for Te reo Māori would be great.*" (C4, F, 32 years old). They suggested incorporating Te Reo Māori as an optional selection in technological systems, allowing users to choose based on personal preferences.

6.4.2.4 Theme 4: Experience in Asthma Management and Opportunities for Technology to Improve Care

Some participants discussed their experiences with asthma management under this theme and how these experiences may affect their use of digital technologies to manage asthma. These fell into 3 subthemes relating to accessibility and quality of health care resources, previous experience with health care staff, and triggers for asthma exacerbations.

6.4.2.4.1 Accessibility and Quality of Health Care Resources

A few participants highlighted the limitations and barriers to accessing health care resources, in particular for those living in rural areas, where access to health resources was particularly difficult. Māori living outside urban areas must often travel a long distance to visit a medical clinic to obtain appropriate medical treatments. Some participants described the experience as costly and even life-threatening in an emergency, as arriving at the right place would take longer. One participant mentioned this as follows:

> "*How is it going to help if you are having an attack if you live in the middle of nowhere (farms, the country?).*" [S1, F, 23 years old].

They also highlighted the difference between the quality of the medical services in their region and those in urban areas.

> "*The doctors up here are just so different to Auckland; would I say the DHB (District Health Board) up here is not as good as it is in Auckland, and I noticed that by moving up here. I understand what all my Nana and they were all talking about now. They're not good here at all.*" [S6, F, 44 years old]

This reflects how some participants are disappointed with health care services in different regions, suggesting that these services do not meet the standards they have experienced or heard about in urban areas.

6.4.2.4.2 Previous Experience With the Health Care Staff

A few participants were disappointed with the current health care system due to the experiences they had during hospitalisations and treatments. They described general dissatisfaction with the health care system, "*I've been stuck in our medical system until then, so...*" [E2, F, 37 years old]. One participant directly answered the question "*Who helps you manage your asthma?*" with "*It's me, not our system*" [E2, F, 37 years old]. A few participants expressed feeling like they had not received proper attention from health care staff, and they expected doctors to listen to them properly and take their concerns seriously.

> "*Yeah, the doctors are great at giving out inhalers, but I could do with a review and see if there's something better on the market. Because you don't know what's available. They just kind of put you on something, you think, hey, well, they know, and that's the best. But it's not always*" [C16, F, 59 years old]

> "*I just know that a lot.. and I'm gonna sound racist here, and I don't mean to be, but a lot of non-Māori therapists don't work well for Māori.*" [S17, F, 66 years old]

Even though they appreciated the doctor's service, they discussed how they independently researched different medications themselves, because they felt that the doctor's prescription was not always correct. Furthermore, they shared that they felt their negative experiences with the health care providers were a result of their ethnicity and a product of a lack of culturally appropriate services.

6.4.2.4.3 Triggers for Asthma Exacerbations

Almost all the participants did not highlight any specific factors relating to asthma that might be unique to Māori communities or need to be considered in ML prediction models.

> "*Normally it's the shift of like cold to warm, warm to cold, it's like from one season to the other, like when we're just going in from the end of summer rush into and then entering back at[sic].*" [E2, F, 37 years old]

> "*...when I start getting sick or anything like that... It like, triggers it. So, if I come down with a cold or anything like that, I get really bad asthma.*"[S6, F, 44 years old]

> "*Dust is another big thing for me. I can't dust my house because I get sick. So someone else has to do it for me, yeah.*" [S7, F, 37 years old]

It was clear that triggers for exacerbations, such as seasonal changes, environmental factors such as pollen and dust, intense workouts and exercises (tiredness), stress, smoking, and other sicknesses, were important considerations for them when considering a prediction model.

6.4.3 Sentiment Analysis

In total, 9 out of 20 (45%) participants had a "neutral" opinion on using AI and digital technology to predict the risk of asthma attacks, indicating neither positive nor negative thoughts, and 4 out of 20 (20%) participants expressed a "very positive" opinion, while 2 out of 20 (10%) participants were "positive". Altogether, 6 out of 20 (30%) participants showed positivity toward the use of AI and digital technology for this purpose. However, a quarter of the participants (20%) held a negative viewpoint. Figure 6.3 presents the distribution of sentiments.

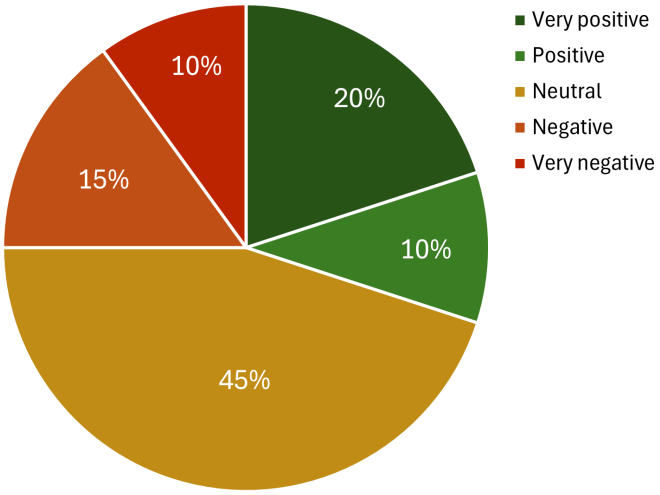


Fig. 6.3 Sentiment analysis of participants' opinions on using AI and digital technology for predicting the risk of asthma attacks.

6.5 Discussion

This is the first study to explore the views, perceptions, and experiences of Māori toward using AI and digital technology for asthma attack risk prediction, and to explore whether there are any particular factors that should be considered in ML models to predict the risk of asthma attacks. In total, 4 key themes were identified, 3 relating to factors that may affect the engagement and uptake of AI and digital technology in Māori, and the fourth describing the experience of Māori with the health care system in NZ. Participants raised concerns about using AI and digital technology, with many expressing that they needed more information about AI and digital technology. This could have been a reason for the majority proportion of the participants being neutral rather than positive or negative about the use of AI and digital technology to predict the risk of asthma attacks. Historically, Indigenous perspectives, for example, the views and perceptions of Māori people, have not had a significant influence in guiding the development of emerging technologies such as AI [193]. However, this seems to be changing, with a recent study proposing a method for developing and implementing AI that incorporates mātauranga Māori (Māori knowledge) [194] and government AI policies highlighting the importance of considering impacts of Māori and their perspectives. This will contribute to improving the cultural sensitivity and inclusiveness of AI systems. Providing sufficient information to all users is a key consideration when developing such applications, especially for different ethnic groups who experience knowledge gaps in technology and limited access to information [195]. Regardless of user age, having sufficient knowledge

about the usage of these digital technologies is necessary for potential users to gain the benefits from these technologies to control and manage their disease.

Some questioned the reliance on computer-based predictions and how this may affect the robustness of health care. This highlighted the importance of trustworthiness to participants to engage with computer-based systems. Trust is a measure of the ability to rely on technology, and low trust can lead to disuse of it [192, 196], which is a considerable disadvantage, especially as technology could bring about significant benefits for the users. One study [192] identifies the factors affecting trust in technology as characteristics based on the users, the environment, and the technology. This is aligned with this study's findings that limited knowledge about technology and access to it has created trust issues. Therefore, in the future, computer health app designers and developers should focus on enhancing the reliability for Māori to encourage them use those systems to obtain maximum benefits.

Furthermore, the participants emphasised the importance of face-to-face interactions as part of using technology and devices related to health, in addition to the technology's introductory sessions. Another study [197] illustrates the significance of incorporating Māori health values such as kanohi ki te kanohi (face-to-face) in perinatal telehealth. This is consistent with the results of our study, where participants also highlighted the importance of having a face-to-face element or introduction to technology in order to engage with it. Similarly, another group of researchers investigated the perceptions of using AI in diabetic eye screening [198] and found that the involvement of clinicians in the screening process is a must. Therefore, this face-to-face component is essential in future AI-based systems to make them culturally relevant and applicable.

Although some concerns were expressed about AI and digital technology usage, participants were open to the idea of using technology to manage their asthma if more information could be provided about the technologies. The percentages of very positive and positive sentiment classes confirm this. A group of researchers explored Māori experiences of telehealth consultations during the COVID-19 lockdown period, and they found how Māori appreciated the support they gained through technology in consulting health professionals [199]. In another study, participants expressed the potential benefits of using mobile health to support their aspirations of hauora (health and well-being) [200].

Participants also suggested some characteristics that technologies and software system developers should consider to encourage uptake and engagement by Māori. The participants' preference for having Te reo Māori as an option is supported by the findings of other research [201], which reports users having a preference for their own language and highlighting the importance of language for technology uptake. For example, a health professional suggested including language options in digital health systems [191], and that for culturally

and linguistically diverse individuals, language can be a barrier to technology use [199, 202]. In the NZ context, Te reo Māori is culturally important; therefore, it should be considered as an option available in software systems. One study identified the use of Te reo Māori as a symbol of recognition and therefore showing respect to the community within mental health practice [203]. Yet, even though one participant in this study mentioned that they were not fluent in Te Reo Māori, statistical data show that the use of Te Reo Māori among Māori has grown from 6.1% in 2018 to 7.9% in 2021 [204]. Hence, following the recommendations from several participants, offering the choice to switch to their native language appears to be a favourable approach for enhancing the usability of these systems among the users. Importantly, this would also be in line with NZ's Te Tiriti o Waitangi (Treaty of Waitangi) obligations and encourage wider use of Te Reo within NZ society. Te Tiriti o Waitangi is a historical agreement between the British Crown and Māori rangatira (chiefs), and it confirmed, among other things, that Te reo is a taonga (treasure) of Māori culture.

The other feature discussed by the participants, simplicity, was recognised to be important for everyone to minimise the burden of using the system to manage a chronic medical condition such as asthma. During the discussions, participants shared their experiences with the existing health system. One of the concerns they raised was the limited accessibility to health care resources and their quality. They questioned the feasibility of travelling from rural to urban areas to receive medical treatment in an emergency. A review paper exploring the experiences of Māori in NZ's public health system has identified practical barriers that have added to the negative experiences of Māori with the health system, such as cost, transport, and time [205]. The paper outlines the reasons for those barriers, such as financial difficulties, transportation issues, and time allocation difficulties, as they need to arrange leave from work or require childcare to attend clinics. Travel time to access care has been cited as a significant barrier to managing health for Māori in other studies, particularly those living away from main urban centres [206]. According to Environmental Health Intelligence NZ (ehinz), most of the Māori population resides in the rural northern and central regions of the North Island [207] where medical services are more limited, so access to care in a timely manner is challenging. Technology could address this barrier by supporting web-based consultations, telehealth options and remote monitoring without the need to travel to access in-person consultations.

A few of the participants mentioned the negative experiences they have had in the past when dealing with health care staff during consultations. In addition to apparent and hidden racism and bias, Māori patients and their families believed that the broader aspects of their spiritual and cultural customs were not adequately recognised within the conventional health care system [205]. Technological solutions could improve the experience and interaction

issues between Māori and Pākehā by offering more personalised options. These systems could incorporate images or other symbolic elements of Māori culture to enhance user responsiveness, which may not always be feasible in face-to-face interactions. Ensuring that the technology that is developed is culturally responsive and facilitates a more positive experience of the health system is a key consideration going forward.

Furthermore, this study did not identify specific asthma-triggering factors for Māori patients with asthma. This presents an opportunity for AI and digital technology to analyse data of patients with asthma and determine if there are any specific triggers for Māori individuals experiencing asthma. AI can identify these factors and make predictions based on personalised predictors, leading to more effective and individualised care [52, 57, 75].

While this study has highlighted important findings, there are several limitations that need to be mentioned. One of the limitations is the small number of participants who participated in the interviews. While we did reach data saturation with 20 participants and the sample of 20 is similar to other qualitative work in this area and may be considered sufficient [203], the current sample of participants may have a gender bias, as there were many more females than males. This is important to consider as asthma affects females and males differently, and their experiences may therefore be different [208]. Future studies building on this work could explore the influence of sex and gender on asthma and technology experiences and aim to recruit a more balanced sample. The study also conducted a sentiment analysis to present the sentiments of the participants regarding the use of AI and digital technology to predict the risk of asthma attacks. However, we did not have enough data to fine-tune the model, and therefore, we used an existing model. In future research, a model trained on the same type of data and then used for further analysis would ideally be used. Additionally, the study did not identify any specific element relating to Māori that could be used as an additional predictor in ML models, which is aligned with current literature. Future research could be conducted to further explore these factors. Another limitation of this study is that it did not explore the impact of Māori culture and knowledge on their experiences of technology. Studies involving Māori communities emphasise that "Māori culture at the core of Māori perceptions and experiences" [209], highlighting the need to consider cultural factors [210]. Future research should prioritise examining the role of Māori culture and knowledge to provide a more comprehensive understanding of these experiences. We did not specifically explore access to technology or the Digital Divide, asthma assistive technologies, or telehealth, which is another limitation of this study. However, it would be worth differentiating these technologies and exploring them, as there can be different opinions about each of them according to personal understandings. Despite this limitation, our findings are still crucial, as this is the first study to investigate this area.

6.6 Summary of the Chapter

Digital technologies are increasingly being used to manage health, including asthma. Given the greater asthma burden among Māori, there is a need to understand their perspectives on digital technologies for managing asthma. This chapter is the first to investigate Māori perspectives on the use of AI and digital technology for asthma, and to uncover any specific considerations to be integrated into ML-based asthma attack risk prediction models to mitigate health disparities. This work was based in interpretivist epistemological position, which assumes that meaning is constructed through individuals' experiences and perspectives. Four predominant themes were identified: (1) concerns about AI use, (2) interest in using technology and AI-based systems, (3) characteristics of AI-based systems, and (4) experience in asthma management and opportunities for technology to improve care. Notably, discussions highlighted a need for more information about AI and digital technology, with concerns about the trustworthiness of technology. Furthermore, participants' expectations for a simple, easy-to-use computer-based application, particularly for health care purposes, were identified. These findings hold implications for future research on AI and digital technology to manage asthma in Māori. It would be worthwhile to conduct a comparative analysis of other health conditions in Māori. Future work could involve implementing and co-designing technological solutions that incorporate Māori-based factors, with the potential to facilitate equitable health care and support positive self-management of asthma. Ensuring equitable uptake of technology by considering Māori perspectives across all aspects of technology use should be a priority. In addition to Māori, evidence has previously been provided confirming the higher prevalence of asthma within the Pasifika community in NZ. Having explored Māori perceptions, we extended the work to examine Pasifika perceptions of the use of AI and digital technology for predicting asthma attack risk. Chapter 7 provides this work in detail.

Chapter 7

Leveraging AI and Digital Technology for Holistic Asthma Management in the Pasifika Community (Publication #7)

Citation: F. Mirza, W. K. D. Jayamini, R. Lutui, K. Pole, A. M. Ikihele, A. F. Mu'aulama, and A. H. Y. Chan, "Leveraging AI and Technology for Holistic Asthma Management in the Pasifika Community," *Telematics and Informatics Reports*, p.100302, Feb. 2026, doi: https://doi.org/10.1016/j.teler.2026.100302. (**IF=4.7, Q1**)

7.1 Article #7 Prelude

Pasifika is among the ethnic groups with the highest prevalence of asthma in NZ. However, as Pasifika communities represent a relatively small proportion of the total NZ population, ML prediction models developed using data from the whole population might under-represent them. Therefore, this chapter addresses RQ3 by exploring the perceptions of the Pasifika ethnic group on the use of AI and digital technology for asthma management. A qualitative approach was employed as similar to Chapter 5. This chapter contributes to the sparse literature on Pasifika perspectives on AI and digital technology-driven asthma management.

7.2 Introduction

Asthma is a long-term condition affecting the respiratory system, due to persistent inflammation in the airways and episodes of airway narrowing [211]. In 2019, asthma affected an

estimated 262 million people globally and contributed to approximately 455,000 deaths [14]. Consequently, achieving effective asthma management is essential. However, in NZ, asthma control remains inadequate, with the rate of exacerbations rising by one-third over the past decade [28]. Asthma represents a major public health issue in NZ, with around 615,000 individuals estimated to be affected nationwide [8]. In this paper, we also refer to the Pacific Islander (PI) Community as *Pasifika*, another term used to identify this community. Asthma has a greater impact on Māori[1] and Pasifika, who face higher rates of severe symptoms and less effective disease management [28, 91, 212]. Notably, in NZ, Pasifika face significantly higher asthma mortality rates than the non-Pasifika population [91]. Among adults aged 15 to 49, Pasifika also experience asthma hospital admissions at nearly three times the rate of other ethnic groups [91].

A significant challenge in asthma management is the delayed recognition of worsening symptoms, often resulting in late detection of potential exacerbations [42]. The adoption of digital technologies and AI offers the potential to facilitate early identification of worsening asthma and enhance self-management practices [175, 213]. Digital interventions increasingly support asthma self-management, such as smart peak flow meters and inhalers integrated with apps to automate monitoring and encourage correct use via audiovisual reminders [176]. These technologies monitor treatment impact, improve patient-provider communication, and generate data to support clinical interactions [81, 214]. Such digital tools can help patients prevent or lessen the impact of asthma exacerbations.

Despite the rapid expansion of digital health solutions for asthma management, their integration into routine clinical practice remains limited [175], with even fewer investigations focusing on the impact of these technologies in different ethnic groups. Limited uptake of technologies in routine care may result from difficulties consumers and providers face in processing and evaluating numerous technologically complex options [178]. An observational study conducted in Portugal [179] found that, among 336 participants, only 3% actively used an asthma app to monitor their condition and support inhaler adherence. Issues such as data accuracy, the need for user training, and challenges with integrating new technologies into existing patient management systems can hinder implementation [180]. Furthermore, concerns regarding data privacy and security also present significant barriers to adoption [175]. Additional obstacles include low levels of eHealth literacy, concerns about cost-effectiveness, technical infrastructure limitations, and gaps in policy and regulatory frameworks [180]. Identifying factors that enhance engagement with digital health solutions is critical to developing tools supporting asthma management in standard care [214].

[1] Māori are the Indigenous Polynesian people of NZ.

Despite advancements in asthma treatments and intensified public health initiatives in NZ, asthma exacerbations have continued to rise over the past decade, highlighting ongoing challenges in effective disease control [28, 91]. Proper management remains essential to reducing asthma-related deaths, hospitalisations, and flare-ups.

Emerging digital health solutions, particularly those powered by AI, offer promising tools to enhance asthma outcomes. These technologies can support earlier detection of acute events and facilitate improved self-management [214]. AI-based predictive models analyse large health datasets to forecast exacerbation risks, enabling timely interventions that may prevent unnecessary hospital admissions [111, 183, 215]. While AI offers potential in healthcare, it raises equity concerns due to the underrepresentation of groups like Pasifika in NZ, potentially leading to biased predictions and worsening health disparities [216, 217]. A prior study on Māori views of AI for asthma prediction emphasised the need for culturally, socially, and linguistically responsive systems to ensure trust and engagement [218]. Whether similar perspectives are seen in PI populations remains to be explored.

With the growing integration of AI in healthcare and future advancements in asthma treatments, it is essential to examine end users' perspectives directly using these advancements. Therefore, a qualitative study is suitable to discover the challenges and opportunities that Pasifika foresee AI and digital technology could solve towards their asthma management. This research aims to explore the attitudes of asthmatic Pacific adults living in NZ towards using digital technologies and AI.

The rest of the chapter is organised as follows. The *Methods* section covers participant recruitment, interviews, ethical considerations, and data analysis. The *Results* section presents participants' demographics and key themes identified. Subsequently, a design framework is postulated. The *Discussion* interprets findings about existing literature, highlighting implications for culturally responsive interventions and reflecting on the study's methodological strengths and limitations.

7.3 Methods

7.3.1 Participants Recruitment

Recruitment for this study was conducted through the research team's professional, research and personal networks at Auckland University of Technology, The University of Auckland,

Moana Connect[2] and Asthma NZ[3]. Flyers (provided in Appendix Q) were distributed within these networks, including email lists and social media platforms like Facebook and Instagram. Pasifika adults diagnosed with asthma, residing in Auckland, NZ, were recruited following the purposeful sampling strategy. We used Pasifika researchers AMI (PhD), a registered nurse and KP, from Moana Connect, along with AFM (Master's in Professional Development; Registered Nurse; Certified Samoan Language Interpreter), a clinical nurse and researcher, to connect with Pasifika participants. The interviewing team has the skills and ability to speak a range of Pacific languages, such as Vagahau Niue, Gagana Samoa, and Lea Faka-Tonga, which supported the effective conduct of the interviews.

Participants were invited to contact the research team via email or phone. Participation was voluntary, with interviews scheduled online or in person at convenient times and locations. Potential participants received an information sheet (provided in Appendix R) to ensure understanding of the study. Contact details were obtained from consenting participants, and language preferences were considered to align them with appropriate interviewers. Eligibility required participants to identify as PI, be 18 or older, have been diagnosed with asthma, provide informed consent, and agree to recording or note-taking, with optional family involvement. This purposive and culturally grounded sampling strategy was employed to recruit participants with experience of asthma. This approach was adopted to maximise the depth, relevance, and ensure credibility of the data, rather than to achieve statistical representativeness, which was not the objective of this work. In this study, representation is understood analytically, meaning that the results reflect a variety of perceptions and experiences within the participant group, rather than offering population-level generalisation. It should be noted that for this study, we did not consider the level of digital literacy of the participants, as the goal is to explore their perceptions in using AI and digital technology for asthma care.

7.3.2 Interview Process

The Pasifika researchers (KP, AMI and AFM - all female) with formal training in qualitative interviewing conducted the interviews using a semi-structured interview guide (Appendix S). It was developed by referring to earlier studies that explored Māori perceptions towards the use of AI and digital technology for asthma attack risk prediction [218], Māori experiences and perspectives on antibiotic use for upper respiratory tract symptoms [188], as well as their

[2]Moana Connect, a Pasifika-led organization based in Auckland, dedicated to elevating experiences of Pasifika families through research, learning, and community.

[3]Asthma NZ, a non-profit health organization supporting individuals with asthma and Chronic Obstructive Pulmonary Disease (COPD) through education, clinical support, and advocacy.

views on brain health [189]. It should be highlighted that referring to the previous studies in preparation for this interview guide was solely for the purpose of comparison of the findings. No prior relationship existed between the interviewers and participants before recruitment. Before the interviews, participants were informed of the interviewers' professional roles, the aims of the study, and the purpose of the research.

The interview guide included six structured sections with open-ended questions. The first introduced the study's purpose, followed by basic engagement questions. The second explored participants' asthma experiences to gather detailed narratives. The third identified asthma triggers relevant to Pasifika, to support culturally appropriate risk models. The final sections examined Pasifika perspectives on applying AI and digital technology for asthma prediction. The use of a semi-structured interview guide strengthened the credibility by ensuring consistency across interviews, allowing flexibility to explore deeply into participants' experiences. Interviews lasted between 60 and 90 minutes, and no repeat sessions were necessary. The study aimed to recruit at least 20 participants to achieve sufficient data depth and thematic saturation, consistent with qualitative research standards for interview-based studies [187]. During data collection and analysis, saturation was assessed iteratively, and recruitment was ceased after 18 interviews, as thematic saturation was reached and no significant new themes emerged.

Participants could respond in English or their preferred Pacific language; some interviews, though in English, often included Pacific terms. They could invite a caregiver or primary carer if desired. Interviews were held at convenient locations or online via video conference when illness, time constraints, travel, or weather prevented in-person meetings.

7.3.3 Ethical Considerations

Participants provided written consent before the interviews, while verbal consent was obtained in advance for those interviewed online. The consent form is provided in Appendix U. All interviews were audio-recorded with participants' agreement. The data were stored securely on a password-protected server managed by one of the researchers' institutions. The study received ethical approval from the Auckland Health Research Ethics Committee in NZ (Appendix T).

7.3.4 Interview Data Analysis

All interviews were transcribed, with those conducted in a Pacific language transcribed in their original language. To mitigate any translational fidelity, the transcripts in the Pacific language were transcribed and validated by two researchers (AMI, AFM) who are fluent in

English and the Pacific language. To ensure accuracy, transcripts were manually reviewed through repeated listening to the audio recordings by three researchers (AMI, KP, AFM). Anonymity was preserved by assigning each transcript a unique participant identifier. Data analysis followed a general inductive approach [219]. This approach provides a systematic procedure for analysing qualitative data through iterative coding, categorisation, and the development of summary findings derived directly from the data. The research team engaged in multiple discussions to review the coded data, explore interpretations, and reach agreement on the final themes. Manual coding was employed to identify recurrent themes and meaningful patterns through a detailed examination of the content [220]. Accordingly, initial coding was conducted independently by three researchers. Through structured team discussions, the initial codes were then iteratively compared, refined and grouped into categories, which formed the basis for the development of the themes. Any interpretive discrepancies were resolved through collaborative discussions and mutual consensus within the team, ensuring that the final themes reflected both the data and the cultural context. The investigator-triangulation approach, in which multiple researchers independently performed different tasks and then discussed them with the team, enhanced the intercoder reliability and the trustworthiness of the interpretations and findings. To maintain reflexivity, ongoing discussions were held that integrated cultural and professional perspectives. These discussions helped the team to reflect on how their assumptions and positions could influence the interpretation of data and encouraged them to consider alternative ways of understanding the data. The methodology followed in the study is graphically represented in Figure 7.1.

Throughout data analysis, the research team engaged in reflexive discussions to consider how their professional and cultural backgrounds might influence data interpretation.

The reporting of this qualitative study was guided by the Consolidated Criteria for Reporting Qualitative Research (COREQ) checklist. The completed COREQ checklist is provided in Appendix V.

7.4 Results

7.4.1 Participant Demographic Overview

A total of 18 participants were involved in the study. These included 16 individual interviews and one conducted with a pair of participants, each lasting between 60 and 90 minutes. Participant demographics are outlined in Table 7.1. The majority were female (61%) and aged between 25 and 44 years (56%). Pasifika is a vast community involving many regions, languages and communities. We tried to appreciate this diversity by sampling individuals

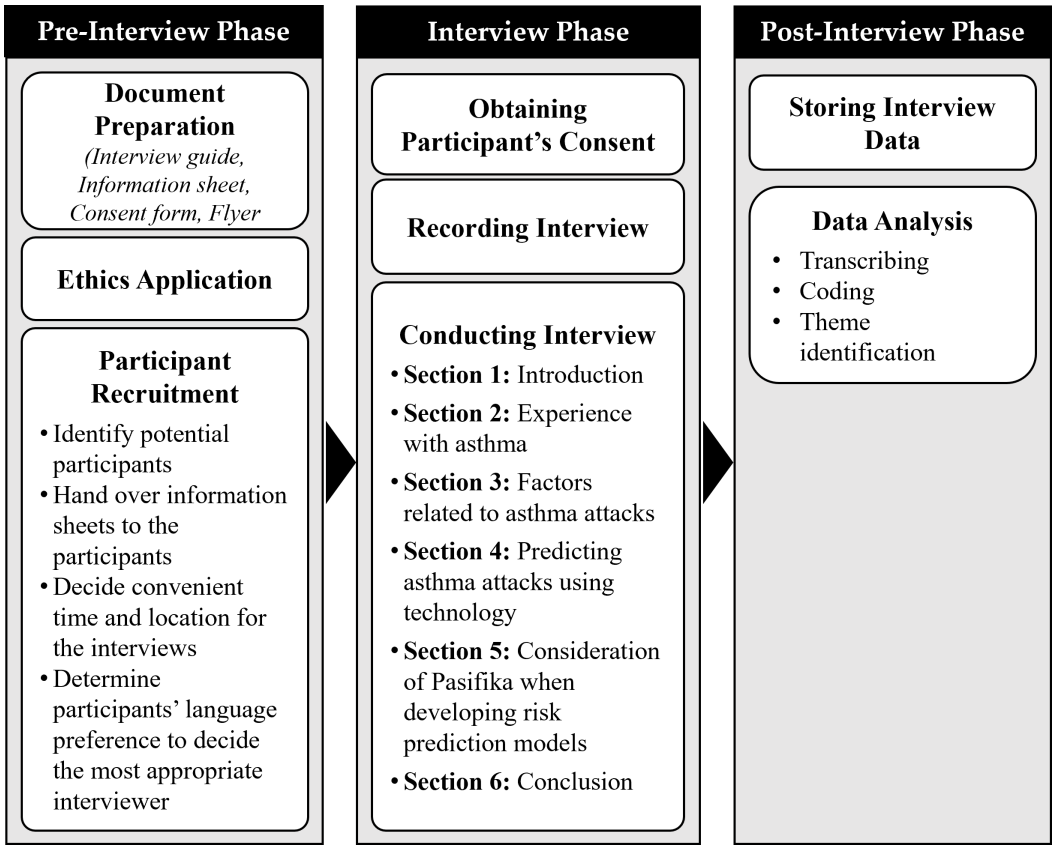


Fig. 7.1 Methodology of exploring Pasifika perceptions on using AI and digital technology for asthma attack risk prediction.

from a variety of Pacific ethnic backgrounds, with some identifying with more than one ethnicity. It should be noted that the presence of a higher proportion (67%) of Samoan among the participants is due to the Pasifika community in Auckland having a skewed population with a majority of Samoan (49%) [220]. While most participants were born in NZ (61%), 39% were born and raised in the Pacific Islands. Most participants (72%) had been diagnosed with asthma during childhood, while the remaining 28% were diagnosed in adulthood.

7.4.2 Key Themes Generated from Interview Data

Participants shared a range of views on the use of AI and technology in asthma care, with some highlighting its benefits, while others expressed hesitation, ambivalence, or resistance. Figure 7.2 visually represents the inductive analytical process, illustrating how codes were grouped into the six key themes as: (1) Awareness of AI; (2) Potential of AI; (3) Concerns of AI; (4) Digital divide in age groups; (5) General asthma management; and (6) Future

design considerations. The subsequent sections outline the key themes identified, supported by representative quotations from the interview transcripts.

Table 7.1 Summary of the demographic profiles of the interview participants

Patient Characteristic	**Frequency: N=18 (100%)**
Gender	
Female	11 (61%)
Male	7 (39%)
Age	
18-24	2 (11%)
25-34	5 (28%)
35-44	5 (28%)
45-54	2 (11%)
55-64	3 (17%)
65-74	1 (6%)
Place of birth	
New-Zealand born	11 (61%)
Pacific Island-born	7 (39%)
Asthma diagnosis	
Childhood-onset	13 (72%)
Adult onset	5 (28%)
Ethnicity[1]	
Samoan	12 (67%)
Niuean	6 (33%)
Fijian	2 (11%)
Māori	1 (6%)
Cook Islands Māori	1 (6%)
Tongan	1 (6%)

[1] Participants were able to select more than one response.

7.4.2.1 Theme 1: Awareness of AI

Individual participants reported differing levels of awareness and understanding of AI. For many, their familiarity was shaped indirectly through media exposure and public discourse. In several cases, participants' knowledge was influenced by observing how others discussed or interacted with AI. Overall, their awareness tended to be generic rather than specific, with limited understanding of how AI might be applied specifically in healthcare settings. This

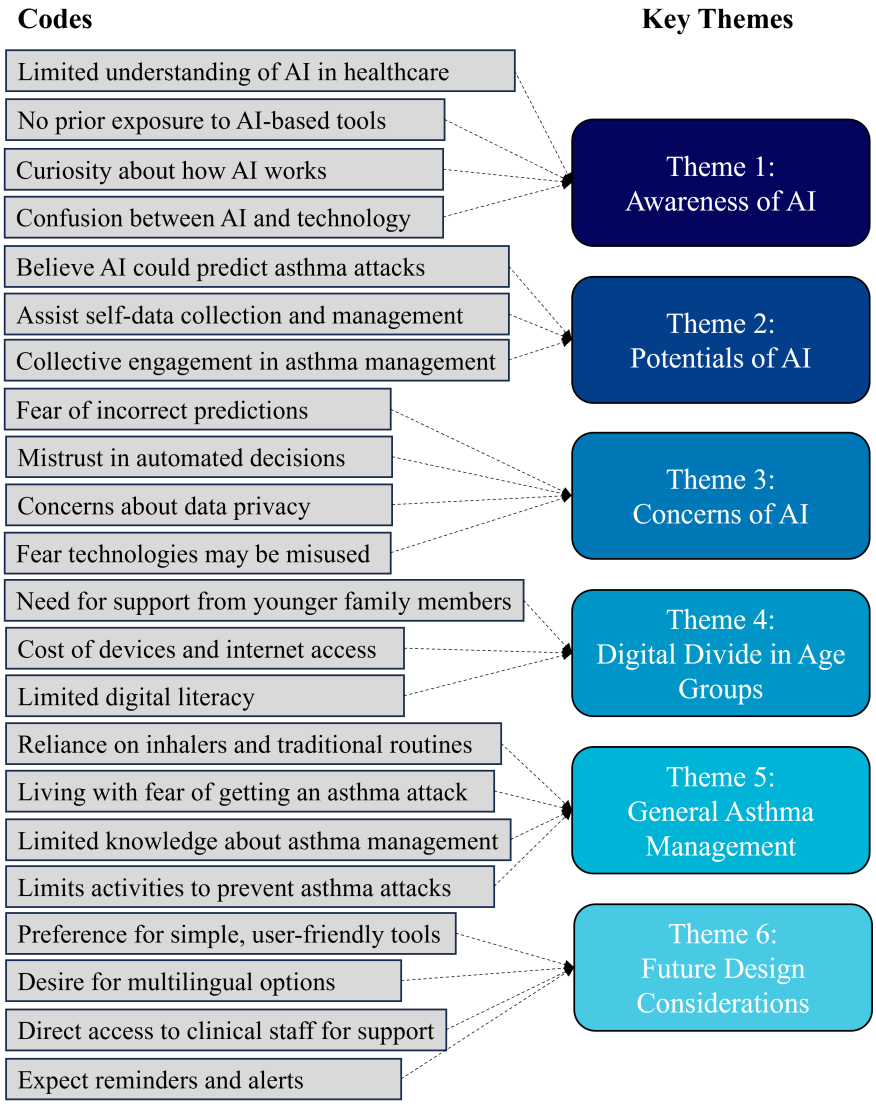


Fig. 7.2 Key qualitative themes generated from codes relating to the perceptions of PIs towards using AI and digital technology for asthma attack prediction.

ambivalence reflected a limited understanding of how AI and technology might be applied specifically within healthcare settings.

> "*Just stuff that you hear in the media about AI. It does seem like it has a lot of information to the point where it can be quite scary if used in the wrong way."* [F, 35 years old, Samoan/Niuean]

> *"I know that there is like a lot of potential to use it. I know people use it for like assignments or ChatGPT and like I use it sometimes for photo editing. So, I think it is a really smart, underused technology."* [F, 26 years old, Samoan]

Most participants viewed the use of AI for health management as a relatively new and unfamiliar concept within health care. While initial responses varied, some participants expressed openness to incorporating AI into asthma care. They indicated a willingness to consider such technologies, notably if they demonstrated clear benefits to their health and well-being. Several participants described themselves as open but not confident, reflecting limited knowledge of AI in healthcare.

> *"AI, I am aware of. I work in a technology company. So, nothing new. But AI and health, I don't know much about it, and where it's being used at the moment or how it's being used. My thoughts on it, to be fair, I'm all for it, because I have asthma."* [F, 35 years old, Niuena]
>
> *"I'm aware of AI... I wouldn't say my knowledge is that good in with like health tech, but like, I'm open to it. It's not something I'm against."* [M, 27 years old, Samoan/Niuean]

These insights illustrate a clear theme: while general awareness of AI exists among most of the participants, detailed understanding, particularly about health care applications, remains limited. The theme underscores that perceptions of AI are often shaped more by indirect exposure than by direct experience or formal knowledge. Nonetheless, not all participants showed an openness to its use in asthma care, especially when potential benefits were apparent. This suggests an important opportunity for targeted education and culturally relevant engagement to build trust and understanding around AI in Pasifika health contexts.

7.4.2.2 Theme 2: Potential of AI

Many of the participants highlighted the potential of AI to aid asthma management, especially through timely alerts that signal early signs of an exacerbation. These notifications were seen as helpful for taking prompt action, making informed care decisions, and possibly reducing the severity of upcoming asthma attacks.

> *"I guess it will help with the prevention as well, like you can identify when an attack is coming and pretty much shut it down then and there."* [M, 30 years old, Samoan]
>
> *"I think that'd be so helpful, so you can like, especially if you have anxiety around your asthma, it would be helpful to prepare yourself so that you know when you're going to have it."* [M, 19 years old, Niuean]

Another advantage identified by some participants was integrating technology into asthma management, which could streamline the process of organising and managing asthma-related

information. This integration was seen as beneficial for supporting self-monitoring and reducing reliance on manual symptoms and medication tracking.

> *"It'll make data collection easier. I feel, yeah, it'll make data collection easier."* [M, 35 years old, Niuean)
>
> *"I think for myself, if I could see that information, and have access to it, I think that'd be a lot easier for me. I mean, I don't have to record it."* [F, 32 years old, Niuean]

One participant emphasised that this approach could strengthen support networks within Pasifika families by promoting a collective response to asthma management. The notion that all family members could be informed and ready to respond to asthma-related risks, especially for elderly relatives, underscores the value of shared responsibility within Pasifika families.

> *"It would serve as a support mechanism not only for me as a patient, grandchildren and caregiver. In our aiga (family), if grandmother is an asthmatic person, then everyone will learn how to use this device so that grandma is warned. So that the family is warned that this could happen for their grandmother, and it could be a family tool... everyone is supporting each other because that's how our Pasifika families are, supporting everyone."* [F, 60 years old, Samoan]

7.4.2.3 Theme 3: Concerns of AI

While many participants expressed some interest in using AI, some of them raised several concerns, reflecting ambivalent attitudes towards AI and technology. A key worry was that excessive reliance on digital tools might diminish individuals' self-awareness and reduce their motivation to actively monitor and manage their asthma symptoms independently.

> *"The disadvantage of this is dependence; instead of trying to manage your asthma yourself, you rely on AI. Too lazy because you know the device has the control."* [F, 44 years old, Tongan]
>
> *"If I've got a tool that I can use to tell me when something's going to happen, does that then mean I'm not even going to care about all the other risk factors or try and manage my own conditions because I'm so reliant on this tool?"* [F, 34 years old, Samoan]

Several participants raised concerns about the accuracy of these tools, especially given the differences in asthma history, ethnicity, and overall health among individuals. Many

highlighted the need for risk prediction models to incorporate diverse health backgrounds and personalised medical information.

> *"If you're using someone's symptoms, that of a mild asthmatic person, would it really help someone who was severely asthmatic? I think for my own asthma, I prefer it to be personalised to me, just because it's much different."* [F, 19 years old, Niuean]
>
> *"You would need to look at my ethnicity because mine would not be the same as the Saiga (Chinese) person... the personal information that you would need from each person would need to be thorough and customised, so that we don't fall in a generic pool."* [F, 35 years old, Samoan]

While many participants recognised the importance of a personalised approach using individual health data, some participants highlighted the necessity for strong security and confidentiality. They stressed the need for robust privacy protections and increased control over access to their health information.

> *"One of the fears is sometimes when information is collected, it can be used elsewhere. So, for example, if our information has been collected by, say, MSD (Ministry of Social Development) or work and income, and then suddenly ping, someone else knows about what you're going through... where's that information going to and who has access to it? I think is more the concern."* [F, 34 years old, Samoan]
>
> *"My biggest concern is who else has access to this information that I do not know of? Like, my privacy... how am I reassured that this information is stayed or kept within or shared with people that I know or am I going to be asked for permission."* [F, 32 years old, Niuean]

These concerns highlight the participants' need for a balance between the potential benefits of AI-based solutions and the protection of their privacy. Ensuring secure and controlled access to personal health data is essential for building trust and encouraging the use of AI-based technologies to manage diseases.

7.4.2.4 Theme 4: Digital Divide in Age Groups

Some participants highlighted that while younger individuals, primarily 'Gen Z' and younger 'Millennials', were generally more tech-savvy and had a greater awareness of AI, older adults, including those from 'Gen X' and 'Baby Boomer' generations, often found it challenging

to navigate modern interfaces and adapt to new digital devices. This generational gap in familiarity with technology was seen as a significant barrier to the widespread adoption of AI tools.

> *"You'll have people who understand how to use technology, like the young people, but as you go up there, older elders aren't completely aware of how to use technology."* [F, 19 years old, Niuean]
>
> *"I feel for the older ones who aren't quite sure what AI is... I know 40-year-olds that don't even know how to use a phone... I feel like that's something that needs to be factored in as well... if it's something to do with AI, that's another thing that they have to try and learn on top of just learning the simple basics of technology."* [F, 35 years old, Samoan/Niuean]

Not all participants were confident in using AI and technology. Some participants expressed the challenge of engaging older generations with new technology, particularly those in the 'Baby Boomer' and older 'Gen X' cohorts. One participant noted that individuals in these age groups, often less interested in digital tools, would benefit from simple, intuitive solutions and require minimal technical knowledge. These older users frequently rely on younger family members, typically 'Gen Z' or 'Millennials', to help them set up basic devices, underlining the importance of user-friendly design.

> *"If I think from my grandparents or my in-law's perspective, they're not really interested in tech stuff and all that... trying to appeal to them will be really useful and making it easier because they don't really you know, they still get us to come set up the Chromecast for them."* [F, 26 years old, Samoan]

Another concern raised was the cost and accessibility of digital tools. A participant emphasised that requiring Wi-Fi or paid subscriptions would create a significant barrier, especially for individuals with limited financial resources. This issue could affect users across all age groups, but is particularly challenging for those on fixed incomes, such as older adults.

> *"If we had to always be connected to Wi Fi or had to have a subscription... no way I'll be subscribing."* [F, 34 year old, Samoan]

These insights underscore the importance of designing digital health tools that are not only affordable and accessible but also tailored to the diverse capabilities and needs of different generations. Considering the financial and technological limitations faced by 'Gen X', 'Baby Boomers', and even some 'Millennials', ensuring equitable access and broader adoption is essential.

7.4.2.5 Theme 5: General Asthma Management

Several participants reflected on managing asthma from childhood into adulthood, emphasising key strategies that supported their asthma control. These included keeping multiple inhalers at home, ensuring timely prescription refills, and maintaining regular communication with healthcare providers.

> *"I'm managing it better because my GP is here. So now I'm on regular Symbicort which I just started this year whereas last year I was on salbutamol which wasn't really being managed well and then prednisone here and there."* [F, 26 years old, Samoan]
>
> *"I do have my Symbicort as a part of my 3-month prescriptions... I keep one in every room, I keep one in my car and in my purse as well. The doctors always happy to provide when I do ask that I need it, given my history with asthma, so they're quite helpful as well."* [F, 35 years old, Samoan/Niuean]

Some participants expressed frustration with the ongoing challenges of living with asthma, particularly the stress and restrictions it imposed on daily life. They highlighted the constant need to stay alert, such as being aware of their surroundings to avoid triggers and ensuring they had their inhaler nearby. For some, this heightened awareness impacted their ability to fully engage in activities, such as sports, and was a constant reminder of the limitations that asthma placed on their lives.

> *"I'm always watching out for myself, I'm always aware of and try and check my surroundings, making it is safe for me and that it is ok for me to be there, so that I get to enjoy whatever or wherever I am at making sure I get to go home and not having an attack when I get home,that I come back home and that I'm ok."* [F, 46 years old, Samoan]
>
> *"It has affected my life in terms of being unable to play sports, and I'm always constantly thinking about my inhaler. Thinking, I'll always need it with me. I use my inhaler quite often as I need it quite a bit."* [M, 21 years old, Samoan]
>
> *"You stop yourself from doing a lot of things, because you, I guess, you're scared that... you'll have an asthma attack, so you lay back on a lot of stuff like running, just like heavy or maybe like sports. For me, it was sports."* [F, 19 years old, Niuean]

While family members with asthma, including parents and siblings, were influential in shaping participants' awareness and understanding of the condition, several participants felt that further education on asthma management was necessary.

"Still a lot of PIs are still living with asthma, and like I said, the education around asthma is not high." [M, 30 years old, Samoan]

This underscores the importance of providing comprehensive asthma education to individuals and their families. By enhancing knowledge and awareness, it may be possible to improve asthma management and ensure that those affected are better equipped to handle the condition.

7.4.2.6 Theme 6: Future Design Considerations

Most of the participants highlighted the need for an AI tool that is simple, affordable, and accessible. A clear and straightforward interface with minimal navigation was essential to ensure usability across different age groups. Some participants also recommended including bilingual functionality to support individuals from diverse linguistic backgrounds.

"Making it accessible like making it easy to understand, not too many buttons, yes, just simple." [F, 19 years old, Niuean]

"Thinking of mum and dad it needs to be bilingual; it would be nice to be in their preferred language. If in English, it should be basic or there's room for them to information if they need it or unsure." [F, 35 years old, Samoan]

"For people who are not really IT savvy, it needs to be very simple at the touch of the button and as simple as possible or something I could speak to. Voice activating one with a rapid response." [F, 46 years old, Samoan]

A small number of participants emphasised the importance of ensuring direct access to a clinician for support, particularly for addressing queries and receiving timely advice on asthma management. This access would allow individuals to seek guidance from healthcare professionals as needed, promoting a more responsive and personalised approach to care.

"For other people, direct access to a clinician for advice and support with queries." [F, 35 years old, Samoan]

While AI offers valuable opportunities for enhancing asthma management, others emphasised the importance of maintaining in-person interactions as a fundamental way to build and sustain meaningful connections with healthcare providers.

"Naturally, we are, we've always been a face-to-face type of people where we need that connection." [F, 34 years old, Samoan]

AI offers substantial potential to enhance asthma management, but in-person interactions are crucial for building meaningful connections between patients and healthcare providers.

7.5 A Holistic Framework for Culturally Responsive, AI-Enabled Asthma Care in Pasifika Communities

Drawing on the findings of this study, we postulated a holistic framework based on the identified themes, aiming to support the effective and culturally appropriate integration of AI and digital technologies into asthma care within Pasifika communities. The framework, generated as a descriptive synthesis of perceptions, reflects a need to enhance community readiness and confidence in using these tools and ensure that their design and implementation align with the lived experiences, values, and health priorities of PIs. Informed by our qualitative analysis, the proposed framework, presented in Figure 7.3. It consists of two interrelated layers. The outer layer, represented by a triangle, emphasises the foundational role of public awareness, while the inner layer outlines key considerations for the design of AI and digital technology-based systems. The following sections provide a detailed discussion of each layer.

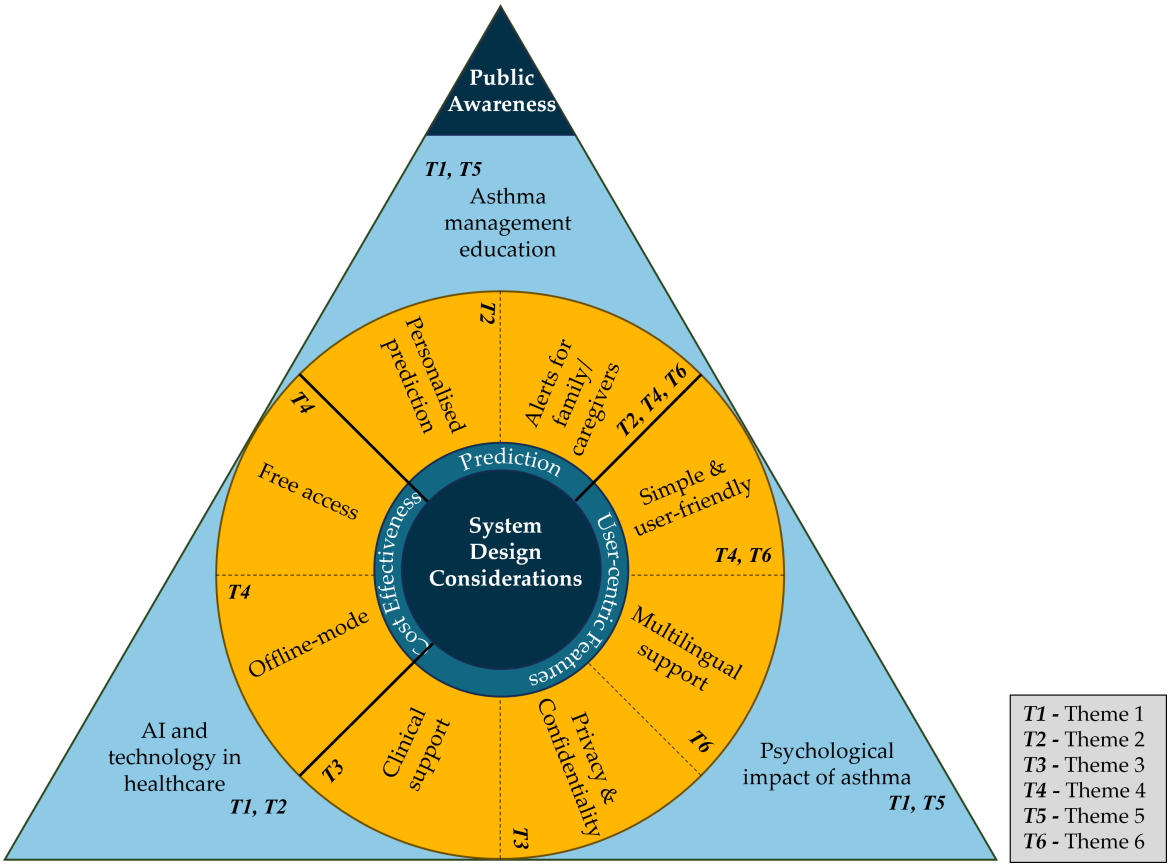


Fig. 7.3 Holistic framework for AI and digital technology for Pasifika asthma care.

7.5.1 Public Awareness

Asthma management among Pasifika relies not only on pharmacological treatment but also on culturally grounded trust, shared understanding, and systems that reflect communal values and lived experiences [221]. The outer ring of the proposed framework outlines three interconnected pathways to raise public awareness: asthma education, AI literacy in healthcare, and psychological support. Effective communication to the public by the government, including the healthcare and technical workforce, can stimulate their understanding and acceptance of AI solutions [222]. In the context of Pasifika communities, public awareness is a collective and relational process rather than an individual one. The awareness should be carried out through families, church groups and community networks, where trust is developed through interpersonal connections and shared life experiences. This approach would also minimise the digital divide in age groups. By prioritising community-informed awareness, this framework ensures that asthma care technologies can be made not only accessible but also deeply relevant to Pasifika families, fostering equitable adoption through culturally safe implementation. The following subsections elaborate on each of these pathways.

7.5.1.1 Asthma Education Management

Some participants raised a lack of knowledge about asthma and its management, which is supported by the findings in another study [223]. Effective asthma care for Pasifika communities must begin by meeting people where they usually gather, not just in clinics, but in homes, churches, and community spaces where health conversations naturally unfold. Instead of relying on medical jargon, education should flow through trusted channels. These ethnic groups place a high value on the participation of their elderly members in guiding, advising, and accepting things [218]. Hence, various strategies could be implemented to spread asthma education and its management effectively. For example, a Pasifika elder explaining inhaler use during family gatherings, youth leaders discussing asthma triggers at sports practices, or community nurses using visual aids that reflect Pasifika art and storytelling traditions. These approaches effectively transfer relevant health tips to individuals and families with a culturally grounded understanding, ultimately improving the quality of people's lives [224].

7.5.1.2 Psychological Impact of Asthma

For many asthmatic Pasifika individuals, the constant worry about potential attacks can overshadow daily life, which could lead to avoiding sports and physical activities like exercising. This fear limits activities unnecessarily, creating anxiety about asthma, which is

burdensome. Catastrophic thinking related to asthma influences the way patients perceive their symptoms; strangely, during periods *without exacerbations*, there was still increased symptom reporting [225]. Adequate support focuses on rebuilding confidence through culturally grounded approaches, like community-designed workshops that pair breathing techniques with traditional dance, or peer discussions that normalise managing flare-ups during activities. The goal is not to eliminate caution but to restore balance, where asthma is managed with awareness rather than fear, allowing for a satisfying, active life.

7.5.1.3 AI and Digital Technology in Healthcare

Introducing AI-driven healthcare requires more than technical explanations. It demands building trust in how algorithms serve collective well-being [188, 226]. Framing AI as a service or an assistant rather than a surveillance tool can ease adoption. For instance, showing how prediction models alert families to high-risk days based on local pollution data and other asthma-related factors could instil confidence. Community sessions could be held to demonstrate AI's role through tangible scenarios that apply to everyday situations. By communicating this information supported by clinical and technical expertise, the technology achieves transparency, enhancing trust in these technological solutions. Likewise, it can be aligned with the Pasifika worldviews. It highlights the importance of collective well-being, rather than focusing only on individuals. As a result, the awareness of AI and digital technologies could often be shaped through family, church and community networks while building trust [227].

7.5.2 Design Requirements

The inner layer in Figure 7.3 shows the design requirements that could be considered in developing AI and digital technological solutions to predict the risk of asthma attacks, which target Pasifika communities explicitly. The layer has three main directions: prediction, user-centric features and cost-effectiveness, which need to be focused on by the designers and developers of these solutions. The following sections will discuss each of these directions.

7.5.2.1 Prediction

A major concern raised by participants was the importance of using their own data for predictions, rather than relying on data from others. Individuals were concerned about making personal predictions based on external data, emphasising the need for personalised predictions tailored to their health information, achievable with current AI algorithms [228]. There are several AI algorithms that can be used to develop these prediction models, such as

logistic regression, random forest, gradient boosting machines, support vector machines, and neural networks, which can be developed mainly based on tabular data, including hospital, climate, and socio-demographic data [48, 109, 183]. Additionally, deep learning algorithms, such as convolutional neural networks, long short-term memory algorithms, can be utilised to construct models on other formats of data, voice (patients' speeches, breathing sounds) [229, 230] and images (CT scans, X-rays, thermal images) [48, 231, 232]. However, these models are developed especially for asthma detection rather than risk prediction.

Further, participants emphasised the importance of delivering risk alerts to the patient's device and the devices of family members or caregivers. This again highlights a key concept in the Pasifika worldview, where health and wellbeing should be a collective responsibility. As a community, they described that critical health events should be managed with the involvement and support of family networks. It reflects the Pasifika values of mutual support, relational accountability and collective protection [227]. Hence, it is important to allow risk alerts on multiple devices in future solutions with the necessary consents, privacy and security controls.

7.5.2.2 User-Centric Features

The framework incorporates several key user-centric features to ensure the tool meets the real-world needs of its intended users. Simplicity and user-friendliness are prioritised, recognising that individuals may have different levels of comfort with technology. Multilingual support is also included to make the system accessible to speakers of diverse languages within the community. Including culturally relevant visuals and narrative components will enhance the usability. Privacy and confidentiality are given careful attention, with measures in place to ensure users' personal health information remains secure and protected. Finally, the framework proposes including clinical support features, offering users the opportunity to contact a healthcare professional to seek advice or assistance whenever deemed necessary. These elements aim to create a supportive, trustworthy, and accessible user experience.

7.5.2.3 Cost-Effectiveness

Cost-effectiveness is a critical dimension incorporated into the proposed framework. Participants expressed concerns regarding the potential financial burden of using such systems, particularly the implications of subscription fees and the requirement for continuous Internet connectivity. In response, two specific features were proposed: offline-mode functionality and free access. Offline functionality allows users to operate the system without persistent internet access, reducing associated data costs. It allows continued access to core self-management

functions in situations where the internet connectivity is limited or unavailable. So that predictive model could be deployed locally on the device, allowing risk predictions to be generated without an internet connection. The online version would additionally support data synchronisation, model updates and integration with external healthcare systems. Ensuring free access was also deemed essential to eliminate financial barriers and promote equitable use. Together, these measures aim to enhance the system's affordability and long-term sustainability for users regardless of socioeconomic background.

7.6 Discussion

This study represents one of the earliest efforts in NZ to explore Pasifika communities' perspectives on the use of digital technologies and AI in managing asthma and predicting asthma exacerbations. One key factor identified in the study was that some participants might not have enough knowledge about asthma and its management. Though some individuals have had asthma for many years, they might still have inadequate knowledge about asthma as a condition and its management options [233]. Public awareness improves asthma management [224]. To overcome this limitation, conducting public awareness programs and workshops can help.

While AI in healthcare is advancing rapidly globally [234, 235, 236] and in NZ [237], its recognition and understanding among Pasifika communities remain limited.

Participants' awareness of AI and its potential uses in healthcare varied. Although some had limited knowledge, there was overall openness to adopting AI solutions that could support Pasifika communities and improve asthma management. A study conducted in China with older adults in a local Chinese community centre to explore the attitudes of AI-driven health technologies showed that they had limited understanding about these AI solutions [238], which echoes our findings with PIs. Another study conducted among older adults in Toronto, Canada, with asthma and chronic obstructive pulmonary disease, found similar findings to the current study: mobile health applications do support their disease care, while there is a tension between perceived benefits and discomfort with using technology [239]. Because asthma disproportionately affects PIs in NZ, targeted interventions are needed to address the unique challenges they face in managing the condition [28, 240]. The proposed descriptive framework covers all these key concerns, which can be used in future strategic approaches to enhance asthma management in different communities. As the framework is informed by the descriptive synthesis of perspectives on AI technologies, it is primarily designed to guide culturally responsive digital engagement and implementation strategies rather than technical AI system development or high-level algorithmic management.

Some of the participants in this study expressed concerns about data privacy and the use of personal health information in AI development. A recent NZ study found that patients generally accepted their data being used for AI in healthcare, provided key conditions were met, such as secure storage, effective de-identification, and protections against harm. Strong governance, rigorous testing, and clinician oversight were regarded as essential to ensure safety [241]. Similarly, Māori communities have previously highlighted the need for more information about AI and digital technology, and a need to address trust issues related to the use of technology [218]. Ensuring that all users are adequately informed is crucial in the development of such applications, particularly for minority ethnic groups as well as asthma vulnerable communities, who often encounter barriers in technological literacy and limited access to relevant information [195]. Multiple international studies emphasise the importance of addressing algorithmic biases and the careful integration of AI tools within clinical settings [217, 242, 243]. Respecting the cultural values and data sovereignty of diverse ethnic communities is essential, alongside ensuring strong protections in the collection, storage, and use of personal health data as AI technologies continue to evolve [244, 245]. The importance of data privacy and security is also reported in the study, which employed a patient and public involvement approach in the United Kingdom and engaged adults with asthma in shaping the design of future AI-supported self-management solutions [246].

It is crucial to understand that AI's benefits in healthcare may not be equally accessible to all communities, which could increase existing health disparities among underserved groups [242]. The study concludes that health equity should be a concern in ensuring these solutions do not disadvantage certain groups [246]. Several factors contribute to digital exclusion, including cultural differences like ethnicity, language, and religion; disabilities such as vision or hearing impairments; low levels of education; older age (over 65 years); living in rural or disadvantaged areas; and socioeconomic challenges like low income and unemployment [247]. A study conducted in the United Kingdom found that older age, lower socioeconomic status, disability, limited education, and unstable housing are significant factors linked to internet non-use [248]. Another study found that a widely used healthcare algorithm in the United States of America underestimated the needs of Black patients by using healthcare spending as a proxy for health status. Because Black patients often had lower healthcare costs due to unequal access, the algorithm introduced racial bias, underscoring the need for fairness and transparency in clinical AI tools [181]. Although access to the internet and digital technologies has improved in some places, many Pasifika communities still face obstacles such as limited access to devices, poor internet connection, and low digital skills [249, 250]. To ensure fair access and prevent widening health inequalities, AI-based health solutions need to address these barriers [245, 250, 251].

Many participants in this study said that asthma affects more than just their physical health; it also disrupts daily activities like sports, leisure, and spending time with family. These social and environmental challenges add to the emotional burden of asthma, including worries about attacks and fear of triggers. Such concerns often cause people to avoid certain activities, limiting their participation in everyday life [252, 253]. Managing asthma can be especially difficult for those with severe symptoms; even with medication, the condition still impacts their quality of life [252].

A key consideration is to leverage AI and digital technology to empower individuals and families to confidently monitor symptoms, enabling asthma management while staying actively engaged in their daily lives. Our proposed framework incorporates the learnings from this research to design impactful solutions. AI has the potential to enhance healthcare services equitably; however, a carefully considered strategy is essential to prevent the exclusion of any population group from the advancements of digital health [254].

Recommendations made by our Pasifika participants included: simplicity; cultural integration of Pacific languages; personalisation; and direct access to clinicians. Rather than relying only on data collected from a broader patient population, asthma attack prediction could be made more personalised by incorporating information specific to each individual, such as their medical history, current medication use, previous asthma attacks, and other relevant clinical and lifestyle factors. This would allow the model to generate predictions that better reflect an individual's unique circumstances. In addition, asthma management support could be tailored to each patient's specific needs and risk profile, making the recommendations more relevant and practical. For example, asthma triggers differ from one person to another. Instead of considering all possible triggers when generating management advice, the system could focus on the triggers that are known to affect a particular individual. By providing patient-specific predictions and recommendations, personalised asthma prediction and management support has the potential to improve both the relevance and effectiveness of these interventions. Notably, maintaining in-person interactions with health professionals to complement a digital approach was also emphasised. A health professional highlighted the importance of incorporating multiple language options into digital health systems [199], emphasising that for individuals from culturally and linguistically diverse backgrounds, language barriers can limit effective technology use [199, 202]. Also, in a previous study conducted within the Māori community, the authors found that the potential users preferred multiple languages in these technological solutions [218].

Overall, the findings of the study align with sociotechnical analyses of digital health that use AI and technology. It emphasises that adoption, non-adoption, and abandonment are influenced by how technology-based solutions interact with users, culture, and wider

social contexts [255]. This research contributes to a better understanding of how AI and digital technologies can support asthma care in Pasifika communities. By highlighting the importance of cultural relevance and equity, based on the descriptive synthesis of perceptions, it offers a foundation for developing future health interventions that are both innovative and appropriate for the needs of PIs.

This study offers several notable strengths. First, to the best of our knowledge, it is one of the first qualitative studies to explore attitudes toward AI and digital technology for asthma management within Pasifika communities. By employing a qualitative design, the study captured rich, nuanced insights into culturally embedded experiences, perceptions, and concerns that quantitative methods would likely miss. Including participants across a wide age range, from tech-savvy youth to older adults with limited digital literacy, provided a broad spectrum of perspectives on the challenges and opportunities associated with AI in healthcare. Notably, the study was conducted in a multilingual setting, allowing participants to express their views in their preferred language, contributing to a more authentic and culturally grounded understanding of the issues. Drawing from the thematic analysis, the study proposes a holistic framework, a descriptive synthesis of perceptions, to inform the development of culturally appropriate AI and digital technology-based interventions. This framework offers a strategic foundation for designing tools that support asthma self-management and foster digital inclusion and health empowerment among ethnically diverse populations. Ultimately, these contributions may help enhance the overall quality of life for asthmatic Pasifika individuals.

However, despite the strengths of this study, several limitations must be acknowledged. First, the study focused on the perceptions of Pasifika communities regarding AI and digital health tools but did not involve the testing or implementation of actual technological interventions. As a result, the practical usability, cultural appropriateness, and effectiveness of AI-driven solutions for asthma management remain unassessed. Additionally, while grounded in qualitative insights, the proposed framework was informed by hypothetical scenarios and not co-designed with participants or validated through participatory processes, limiting its immediate applicability in practice. The study did not undertake cross-comparative analysis with international Pasifika populations, which could provide a broader contextual understanding and reveal unique or shared challenges in digital health adoption. Also, given the qualitative, exploratory nature of this study, findings are interpreted as descriptive and illustrative. Further, the study did not compare competing explanatory theories or establish causal relationships. Rather, it takes an interpretive approach and offers a descriptive synthesis of perceptions that are grounded in empirical evidence. Finally, the research was limited to Pasifika adults residing in Auckland. Although thematic saturation was reached, the relatively

small sample size restricts the generalizability of the findings to broader Pasifika communities across NZ and beyond. Not all Pasifika sub-ethnic groups were adequately represented in the study, and while there were one or two participants from specific sub-groups, their limited numbers were insufficient to fully consider the diverse cultural distinctions that could influence perceptions of AI and digital technology.

7.7 Summary of the Chapter

This chapter expanded on limited research exploring perceptions of AI and digital technology use for asthma management and exacerbation prediction within Pasifika communities. This study was based in interpretivist epistemological position, which assumes that meaning is constructed through individuals' experiences and perspectives. The findings indicated that, although Pasifika communities are generally open to AI-based asthma tools, their widespread acceptance will depend on addressing concerns related to accuracy, data privacy, and accessibility, while ensuring the solution is personalised. The holistic framework postulated in this chapter identifies the factors that need to be addressed to enhance asthma management within the Pasifika community. A key focus is improving public awareness of asthma and culturally appropriate ways to manage it. Before introducing technological solutions, the potential users must be made aware of the usefulness of those tools for their disease management. By adhering to these steps, it is possible to develop solutions that equip individuals and families with the confidence to monitor symptoms and effectively manage asthma, thereby enabling them to maintain an active lifestyle. A practical approach to asthma management across diverse communities and age groups involves balancing technological solutions with personalised care and respecting cultural values and contexts. Additional studies are required to explore user experiences with AI tools, incorporating interviews conducted before and after use to evaluate changes in attitudes over time among Pasifika communities and other ethnic groups.

Chapter 8

Conclusion, Limitations and Future Directions

This chapter summarises the contributions of the thesis by synthesising the findings and insights generated through addressing the research questions. It highlights the key technical and scientific implications of the work and reflects on how asthma care can be improved through *data-driven* and *community-based* approaches. Further, we present the limitations of this thesis while outlining potential directions for future research to build on these findings and further explore opportunities to enhance support for asthma management.

8.1 Conclusion

This section is organised with reference to Figure 8.1, which helps to present the concluding remarks. It first summarises the *SLR* conducted to investigate the state of the art in ML applications for predicting asthma attack risk. Subsequently, it presents the conclusions on *ML-based asthma prediction* and *perception exploration of asthma prediction*. Each part revisits the research questions formulated within the thesis, allowing the discussion to reflect on how the findings generated research contributions.

8.1.1 The SLR on ML Applications for Asthma Attack Risk Prediction

An SLR was conducted and reported using the well-established PRISMA standards to investigate previous studies on the use of ML techniques to predict the risk of asthma attacks (Chapter 2). The review protocol was published in the international systematic review registry named PROSPERO. The review was important given the emerging use of ML in healthcare, particularly for asthma, which continues to pose a considerable health burden.

The review achieved the following significant outcomes. Firstly, a *conceptual overview* of the landscape of studies within this scope. Secondly, the review identified various *types of data* sources used in previous studies to develop ML-based prediction models. These included biological, clinical, environmental and meteorological, hospital, medical, and socio-demographic data sources. Among these, biological data were the least frequently used, whereas most studies relied on clinical, socio-demographic, and hospital data. Thirdly, we presented a *categorisation of studies* within this scope based on the nature of the prediction outcome.

Overall, the review findings showed that ML approaches can be effectively applied to asthma patient data, and DT-based ensemble models tend to achieve superior performance, while underscoring the need to handle data imbalance carefully. Also, many approaches prioritise predictive performance, with relatively little emphasis on interpreting model outputs or understanding the factors driving risk. Understanding the factors that trigger asthma can enable patients to make appropriate lifestyle adjustments to minimise risk. For instance, patients realising unusual patterns in vital signs or medication use may prompt them to see a doctor to avoid severe outcomes.

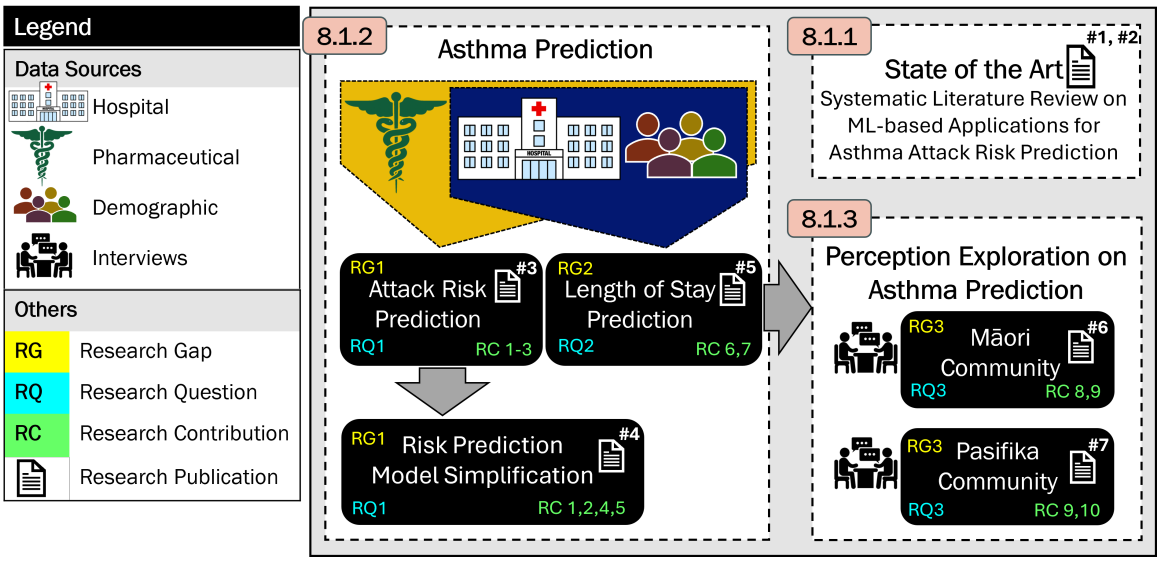


Fig. 8.1 Organisation of the conclusion, with subheadings reflecting the summarised content.

8.1.2 Asthma Prediction Using ML Techniques

8.1.2.1 Asthma Attack Risk Prediction Using ML

One of the research gaps identified in the literature was the lack of studies on developing robust ML models for imbalanced asthma attack data and on identifying key predictors for

inclusion in a generalisable prediction model (RG1). Based on this gap, the first research question was formulated as 'How do ML models perform in predicting asthma attacks with imbalanced data, and what are the key risk factors associated with future asthma attacks?' (RQ1). To address this, we implemented ML techniques to predict asthma attack risk via various combinations. This extensive experimentation was necessary to deal with the highly imbalanced data in NZ. As a result of this experimentation, we aim to determine the most important predictors for asthma attack risk prediction. This led to Publication #3, which is presented in Chapter 3.

During data preprocessing, applying feature scaling to numerical variables was necessary, and standardisation performed slightly better than normalisation. Several ML prediction models, including RF, XGB, and LR, were developed using hospital and pharmaceutical data. Asthma data was highly imbalanced; therefore, several imbalance handling techniques were utilised to reduce imbalance in the target class. Accordingly, SMOTE was used as an oversampling technique, and RUS was used as an undersampling technique. For comparison purposes, models were developed without applying any sampling technique. It should be noted that all these imbalance handling techniques were applied only on the training data to avoid data leakage during model development. The XGB model, combined with RUS and standardisation, showed marginally better performance compared to the other models, with an **AUROC of 0.76** and an F1 score of 0.33. However, LR with RUS also performed comparably, achieving an **AUROC of 0.75** and an F1 score of 0.32. This indicates that the LR and XGB models with RUS perform similarly and demonstrate better performance in this context. Hence, the thesis made technical contributions on *demonstrating the potential of ML techniques in asthma-related predictive modelling in the NZ context (RC1)* and *identifying effective strategies for data imbalance handling in asthma predictive modelling (RC2).*

In addition, the XGB-RUS model was used to identify important predictors as it achieved a better performance: *previous asthma attacks, duration of winter exposure, and the number of ICS and SABA inhalers* were identified as the primary predictors of impending asthma attacks. Therefore, in developing asthma attack risk prediction models, more focus can be placed on including these predictors. Thus, these insights contributed to both the scientific and technical domains in *identifying the most important asthma attack risk predictors (RC3).*

8.1.2.2 Asthma Attack Risk Prediction Model Simplification

Comprehensive data may be available in some clinical or hospital settings, whereas access to such information can be limited in others. In these situations, applying prediction models that rely on many predictors may be challenging. Accordingly, next, we focused on reducing the complexity of prediction models, with particular emphasis on minimising data require-

ments while maintaining prediction accuracy. It extends previous work, focusing mainly on minimising the feature set needed to predict asthma attack risk. Similar steps were followed to develop and evaluate the prediction models as in the previous modelling. Different feature sets were generated based on different criteria to determine the best combination of features that yields the best performance. In feature selection, two strategies were employed. One of these strategies was to use clinical expertise to generate feature sets, while the other was to use the RFECV technique. In addition to evaluating the models based on prediction accuracy, they were also assessed for efficiency. Accordingly, 90 model combinations were generated and evaluated. Importantly, the top-performing models were externally validated. This led to Publication #4 and is presented in Chapter 4.

Overall, the results showed that asthma attacks in the previous year and medication use, particularly the SABA_ICS_Ratio, play an important role in predicting upcoming attacks. Models that combined XGB–RUS-based feature selection with a downsampling technique such as RUS or ENN, applied to both LR and XGB, tended to perform slightly better than other approaches. In a setting where the feature set may differ, this hybrid approach could be used to filter for the most important features and thereby develop a simple yet interpretable model. Notably, the findings also indicate that even a very simple model using just two features (*SABA_ICS_Ratio* and *12MNoOfAsthAttacks*) with LR and RUS can deliver promising results. Overall, the results indicate that reliable predictions can be obtained with a reduced set of features, thereby improving model simplicity and supporting greater generalisability. While confirming the answers to RQ1 generated in the first phase, this added new insights into the important predictors to be considered in asthma attack predictive modelling. Therefore, while strengthening the contributions discussed before (RC1 and RC2), this also provided a *comprehensive performance analysis of ML models, highlighting their predictive accuracy and efficiency (RC4)*. Furthermore, it *provided insights into which feature combinations contribute to model generalisation through comparative performance analysis (RC5*).

8.1.2.3 Asthma LoS Prediction Using ML

Although predicting asthma attacks is possible, not all attacks are preventable. Patients experiencing an asthma attack could end up in a hospital to receive medical treatment, which could result in hospital stays. While asthma attack prediction is critical, it is also important to predict LoS, as this provides useful insights for an individual and informs resource utilisation and quality of care in hospitals. Hence, we formulated our second research question 'How accurately can we predict the asthma LoS using ML algorithms?' (RQ2). Several ML regression models, including SVR, RF, KNN, and XGB, were developed to predict the LoS

of an asthma patient utilising the hospital admission data in Auckland, NZ. This led to the Publication #5 and is presented in Chapter 5. The RF models performed better with an **RMSE of 2.48**. Therefore, it confirms that ML models can accurately predict asthma LoS. Hence, the thesis made a technical contribution as *evaluating the predictive performance of ML models for asthma LoS prediction (RC7)*.

Moreover, the thesis made another scientific contribution by *providing insights into asthma hospitalisation patterns in NZ through exploratory analysis of hospitalisation data (RC6)*. Hospital admission data were analysed to identify associations among the corresponding features. Several important insights emerged from this analysis. One of the key findings was that the highest average LoS was for Māori females. It was also found that the gender, ethnicity, smoking status, day of the week of admission, month of admission, and presenting diagnosis have an impact on the asthma LoS. This shows that demographics play an important role in asthma LoS. This is also supported by the literature, which reports the highest asthma prevalence in the Western Pacific region [6] and more importantly, Māori and Pacific children in NZ [10]. This motivated us to conduct empirical research to explore technological support for these communities at high risk of asthma.

8.1.3 Perceptions of Technological Support for Asthma Prediction

The high prevalence of asthma in the Māori and Pasifika communities motivated us to explore providing technological support for asthma management in these groups. The best way to understand the drivers of technology use among communities vulnerable to asthma is through direct interaction with them. Therefore, we formulated the third research question as: 'How do high-risk asthma communities in NZ perceive the use of AI and digital technology for predicting the risk of asthma attacks?' (RQ3). Semi-structured interviews were conducted separately with these two communities. Through this work, the thesis scientifically contributed *on investigating perceptions of different asthma-vulnerable communities regarding the use of AI and digital technology for asthma attack risk prediction (RC8)*. This work was intended as an initial step towards investigating technological support for these asthma-vulnerable communities, while minimising existing health disparities. To conduct these research, we engaged with the Māori and Pasifika community organisations and cultural experts. Comprehensive details about participant recruitment, conducting interviews, data collection and analysis are provided in the Chapters 6 and 7 of the thesis, which led to the Publications #6 and #7. As these communities speak different languages and are difficult to reach, participant engagement was best supported by involving health practitioners, nurses, and organisations that work closely with the respective communities. Also, we appreciated the participation of the interviewees by providing gift vouchers to recognise their time and

effort. Therefore, conducting these qualitative studies was costly, and we were awarded scholarships from the Health Research Council of NZ and the Faculty Contestable Research Funding to support our research.

From the Māori interview data analysis, four key themes were identified as: (1) concerns about AI use, (2) interest in using technology and AI-based systems, (3) characteristics of AI-based systems, and (4) experience in asthma management and opportunities for technology to improve care. The findings highlighted that the trustworthiness and simplicity of the digital solutions were key expectations of the participants. The six core themes identified in the Pasifika interview data analysis showed different aspects of their perception. The themes identified were (1) awareness of AI; (2) potential of AI; (3) concerns of AI; (4) digital divide in age groups; (5) general asthma management; and (6) future design considerations. The results suggest that, while Pasifika communities are broadly receptive to digital asthma management tools, their wider acceptance will depend on effectively addressing issues of data privacy and accessibility, and on ensuring that the solutions are appropriately personalised. Based on the interview data analysis, there was a general positive attitude towards the use of AI and digital technology for predicting asthma attack risk in the Pasifika community, compared to the Māori community. Therefore, addressing RQ3, the thesis made a contribution by *generating themes on using AI and digital technology to predict asthma attack risk for Māori and Pasifika communities (RC9)*.

Subsequently, the thesis makes a further contribution by *postulating a framework to strengthen asthma-related knowledge and inform future development of AI and digital technology–based solutions for Pasifika asthma care (RC10)*. This framework showed that, before implementing technological solutions, it is essential to ensure public awareness of asthma and its management. Subsequently, potential users should understand the value of these tools in managing their asthma. Following this, the developers of these solutions should focus on system design requirements, including factors such as prediction, user-centric features, and cost-effectiveness. Firstly, this holistic approach allows asthma-vulnerable communities to understand their own risk profiles and make lifestyle adjustments to minimise risk. For example, if an individual identifies that poor-quality insulation in their living space triggers asthma, they may seek support through relevant NZ schemes, such as government-funded insulation programmes, which can enable the installation of high-quality insulation at little or no cost. Secondly, it provides design requirements that can lead to technology-based solutions that motivate individuals and families to confidently manage their asthma effectively, helping them lead active, healthy lifestyles.

In summary, the thesis demonstrates that ML-based approaches can provide reasonably accurate and efficient predictions of asthma attacks, even with imbalanced data and reduced

feature sets. Also, it showed that ML-based models can predict asthma LoS with acceptable accuracy. At the same time, the qualitative findings from the asthma-vulnerable communities in NZ highlight that the successful adoption of such technological solutions depends not only on predictive performance, but also on trust, accessibility, and cultural relevance, particularly for Māori and Pasifika communities. Together, these findings illustrate that ML-based models can be utilised by doctors and nurses in healthcare settings as decision-support tools to generate predictions and identify high-risk patients, enabling medical practitioners to pay closer attention to those patients. Also, these models can be used to estimate how long an admitted asthma patient will remain in the hospital for medical treatment.

Overall, these early predictions will support health practitioners in managing their resources, minimising costs, enhancing their services, and saving lives, thereby improving patients' quality of life. In addition, following the holistic framework enables asthma-vulnerable communities to gain awareness of asthma and its management, thereby helping them understand their risk factors and make lifestyle adjustments accordingly. Furthermore, the framework provides factors to consider when developing future AI and digital technology-based solutions, thereby improving their acceptability among high-risk asthma populations.

8.2 Limitations and Future Directions

Respiratory health plays a vital role in overall human health. Therefore, this research area is rapidly emerging, with a growing volume of studies reflecting the critical importance of respiratory health, including asthma. Numerous studies have investigated this medical condition, with a particular focus on applications of ML techniques. Although there are many different technological solutions, there are still gaps for further research. Consequently, this thesis addressed the research questions RQ1, RQ2 and RQ3 generated from the existing gaps in the literature. Based on the findings and insights from the research addressing the research questions, this section identifies key limitations of the thesis and outlines potential future research directions to extend it. Thereby contributing to improved asthma management, enhanced quality of life, and reduced financial burden through accurate risk prediction and the incorporation of the perspectives of high-risk asthma communities.

Despite the strengths of the thesis, it is important to acknowledge the limitations that are inherent to the research process. One limitation is that the main literature review was limited to a particular time frame. The SLR covered studies published within a specific time period (from January 2010 to February 2023) at the time it was conducted, which is an inclusion criterion of the PRISMA guidelines. However, this is a rapidly developing

field of research, and as new forms of data are generated by advanced technology, ongoing monitoring of the latest models is required. There were always limitations on data, time and human resources. We were fortunate to collect data (refer to Appendices G and C) from the Auckland hospital and NZ Ministry of Health; however, it would be better to validate the prediction models on data from other regions to enable triangulation and make comparisons. Also, the thesis produced prediction outcomes as classifications rather than risk probabilities, which may provide additional clinical information for risk stratification. In addition, the thesis could not focus on individual-based data analysis due to limited data availability at the time of conducting the research. Despite time constraints, the thesis generated meaningful and interesting insights within the available timeframe. Subject-matter experts are essential for conducting research, and in health-related studies, engaging clinicians is particularly important, as we achieved in this project. Conducting interviews required cultural experts, and we were able to overcome this limitation with support from existing networks within the communities. However, the thesis did not explore in depth the sub-ethnic groups within the asthma-vulnerable communities, a limitation that may yield further insights.

Considering the thesis's strengths and limitations, this research can be further extended in future work, drawing on the findings and insights. The following outlines the recommendations for future work.

- **Integrating other data sources such as weather and peak flow data for asthma attack risk prediction**
  The quantitative methodology of this thesis used only hospital and pharmaceutical data to predict the risk of asthma attacks. In addition to these, other factors also impact asthma. For instance, in the literature, weather factors such as temperature, humidity, and wind have been associated with asthma attacks. Therefore, future research can integrate *multiple data sources* to develop a predictive model of asthma attack risk probability. Additionally, daily measured peak flow and inhaler data can also be combined to enhance the *personalised prediction*. As there are *smart digital devices* on the market used by asthma patients, corresponding prospective data can be collected daily and integrated with other data, thereby supporting more personalised risk predictions for the individuals. However, these data sources are generated more frequently than hospital data. Therefore, the *data integration* should be carefully examined in future research. Subsequently, the researchers could apply various feature selection methods to identify the most important predictors among the features used in model development, thereby enhancing model performance and simplicity. The thesis has led to another parallel project, integrating weather data with asthma hospital

data, conducted in collaboration with Earth Sciences NZ to understand the correlation between weather and asthma and develop prediction models.

- **Using advanced deep learning models for asthma attack risk prediction**
  Including multiple data sources to develop risk prediction models will increase the dataset's complexity. Therefore, ensemble techniques such as RF and XGB might not be able to learn the complex data patterns. Thus, more advanced and complex ML algorithms, specifically *deep learning algorithms* such as Long Short-Term Memory (LSTM) and other extensions of neural networks, could be explored in future research given the time series nature of some data sources. Also hybrid modelling could be tested when integrating stationary data with time series data. These might improve prediction accuracy; however, the models' efficiency should also be tested to determine the best model. The current study findings have led to a parallel research project named DIGIPREDICT which collects data from 300 asthma patients through various *smart devices* such as smart watches, peak flow meters, smart inhalers, in addition to the socio-demographics, asthma control, lung function and dietary intake, etc. to develop asthma attack risk prediction models. We have not included any information about this project in the thesis, but it has begun as an extension of research in this study's scope.

- **Conducting a pilot study for asthma attack risk prediction**
  This thesis examined different aspects of generating prediction models for impending asthma attacks, comparing feature selection, data imbalance handling, and ML techniques. Although the top-performing models were externally validated to some extent, to assess real-world applicability, future studies could conduct a pilot study to apply these models across different clinical settings, such as hospitals, where health professionals can use them as a decision-support tool while they personally make patient care decisions. Model predictions and human decisions can be compared to assess the performance and applicability of these models.

- **Co-designing and implementing AI and digital technology-based solutions**
  The findings from the qualitative approaches also have implications for future research on AI and digital technologies to manage asthma in Māori and Pasifika communities. The themes generated from the interviews showcased both positive and negative opinions. While acknowledging the positives, future research could involve implementing and co-designing technological solutions to address concerns such as trustworthiness and to embed design requirements with cultural impact.

- **Explore user experience on AI-based asthma management tools**
  Additional studies are required to explore user experiences with AI-based tools, incorporating interviews conducted before and after use to evaluate changes in attitudes over time among these distinct communities. Accordingly, these tools could be developed and tested on a similar group of participants through a pilot study. This type of future research will ensure the acceptance and usability of such customised tools and how they effectively contribute to asthma management within the asthma-vulnerable groups. Further, longitudinal studies will be needed to examine how these populations engage with these technology-based solutions.

References

[1] Mayo Clinic, *Asthma*, Accessed: 2025-11-20, Mar. 2025. [Online]. Available: https://www.mayoclinic.org/diseases-conditions/asthma/symptoms-causes/syc-20369653.

[2] Global Initiative for Asthma (GINA), *Global strategy for asthma management and prevention*, Accessed: 2025-11-20, 2024. [Online]. Available: https://ginasthma.org/2024-report/.

[3] National Health Service, *Asthma*, Accessed: 2025-09-18, National Health Service, 2025. [Online]. Available: https://www.nhs.uk/conditions/asthma/.

[4] Z. Mao et al., "Global, regional, and national burden of asthma from 1990 to 2021: A systematic analysis of the global burden of disease study 2021," *Chinese Medical Journal Pulmonary and Critical Care Medicine*, vol. 3, no. 1, pp. 50–59, 2025. DOI: 10.1016/j.pccm.2025.02.005.

[5] A. P. J. Rabe, W. J. Loke, K. Gurjar, A. Brackley, and D. E. Lucero-Prisno III, "Global burden of asthma, and its impact on specific subgroups: Nasal polyps, allergic rhinitis, severe asthma, eosinophilic asthma," *Journal of Asthma and Allergy*, vol. 16, pp. 1097–1113, 2023, Open access. DOI: 10.2147/JAA.S418145.

[6] P. Song et al., "Global, regional, and national prevalence of asthma in 2019: A systematic analysis and modelling study," *Journal of Global Health*, vol. 12, no. 04052, 2022. DOI: 10.7189/jogh.12.04052.

[7] N. U. Chowdhury, V. P. Guntur, D. C. Newcomb, and M. E. Wechsler, "Sex and gender in asthma," *European Respiratory Review*, vol. 30, no. 162, p. 210 067, 2021. DOI: 10.1183/16000617.0067-2021.

[8] Asthma and Respiratory Foundation New Zealand, *Asthma*, Accessed: 2025-12-10, 2025. [Online]. Available: https://www.asthmafoundation.org.nz/your-health/living-with-asthma.

[9] Asthma & Respiratory Foundation New Zealand, *The impact of respiratory disease in new zealand: 2023 update*, Report, Accessed: 2025-11-20, 2023. [Online]. Available: https://www.asthmafoundation.org.nz/assets/documents/Respiratory-Impact-Report-2023.pdf.

[10] Environmental Health Intelligence New Zealand. "Asthma prevalence and hospitalisations in children (0–14 years)," Accessed: Jun. 1, 2025. [Online]. Available: https://www.ehinz.ac.nz/assets/Surveillance-reports/Released_2025/Asthma_released062025.pdf.

[11] Medline Plus. "Asthma." Accessed: 2025-11-24. [Online]. Available: https://medlineplus.gov/ency/article/000141.htm.

[12] National Heart, Lung, and Blood Institute (NHLBI). "Asthma attack." Accessed: 2025-11-24. [Online]. Available: https://www.nhlbi.nih.gov/health/asthma/attacks.

[13] J. Martin, J. Townshend, and M. Brodlie, "Diagnosis and management of asthma in children," *BMJ Paediatrics Open*, vol. 6, no. 1, e001277, 2022. DOI: 10.1136/bmjpo-2021-001277.

[14] World Health Organization, *Asthma*, Accessed: 2024-02-15. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/asthma.

[15] Y. Ge, G. Wang, T. Liu, W. Ji, J. Sun, and Y. Zhang, "Predicting 30-day in-hospital mortality in icu asthma patients: A retrospective machine learning study with external validation," *BMC Pulmonary Medicine*, vol. 25, no. 1, p. 387, 2025. DOI: 10.1186/s12890-025-03881-w.

[16] M. Hasan, M. Arman, M. Bhuyain, F. Chowdhury, and M. Bathula, "Predictive analytics in healthcare: Strategies for cost reduction and improved outcomes in usa," *International Journal of Innovative Research and Scientific Studies*, vol. 8, no. 8, pp. 142–150, 2025.

[17] Organization for Economic Co-operation and Development (OECD), *Health care utilisation: Hospital average length of stay by diagnostic categories*, Accessed: 2022-06-05. [Online]. Available: https://stats.oecd.org/index.aspx?queryid=30165.

[18] J. A. Buendia, A. F. Zuluaga, and C. E. Martínez-Rodríguez, "Global and regional projections of the economic burden of asthma: A value of statistical life approach," *F1000Research*, vol. 14, p. 146, 2025. DOI: 10.12688/f1000research.157481.1.

[19] G. S. Mosnaim, J. Greiwe, S. P. Jariwala, R. Pleasants, and R. Merchant, "Digital inhalers and remote patient monitoring for asthma," *Journal of Allergy and Clinical Immunology: In Practice*, vol. 10, no. 10, pp. 2525–2533, 2022. DOI: 10.1016/j.jaip.2022.06.026.

[20] R. K. Merchant, R. Inamdar, and R. C. Quade, "Effectiveness of population health management using the propeller health asthma platform: A randomized clinical trial," *Journal of Allergy and Clinical Immunology: In Practice*, vol. 4, no. 3, pp. 455–463, 2016. DOI: 10.1016/j.jaip.2015.11.022.

[21] P. Tinschert, R. Jakob, F. Barata, J. N. Kramer, and T. Kowatsch, "The potential of mobile apps for improving asthma self-management: A review of publicly available and well-adopted asthma apps," *JMIR mHealth and uHealth*, vol. 5, no. 8, e113, 2017. DOI: 10.2196/mhealth.7177.

[22] M. Farzandipour, E. Nabovati, R. Sharif, M. Heidarzadeh Arani, and S. Anvari, "Patient self-management of asthma using mobile health applications: A systematic review of the functionalities and effects," *Applied Clinical Informatics*, vol. 8, no. 4, pp. 1068–1081, 2017. DOI: 10.4338/ACI-2017-07-R-0116.

[23] N. Tahmin et al., "Machine learning models incorporating genotype and ancestry improve severe asthma risk prediction," *Scientific Reports*, vol. 15, no. 1, p. 40 243, 2025. DOI: 10.1038/s41598-025-24080-x.

[24] M. Xie and C. Xu, "Predicting the risk of asthma development in youth using machine learning models," *PLOS ONE*, vol. 20, no. 11, e0336591, 2025. DOI: 10.1371/journal.pone.0336591.

[25] D. A. Aliyu, E. Akashah Patah Akhir, S. Yahaya, and H. Mamman, "Optimization techniques for asthma exacerbation prediction models: A systematic literature review," *IEEE Access*, vol. 12, pp. 110 862–110 890, 2024. DOI: 10.1109/ACCESS.2024.3440502.

[26] Z. Hussain, S. A. Shah, M. Mukherjee, and A. Sheikh, "Predicting the risk of asthma attacks in children, adolescents and adults: Protocol for a machine learning algorithm derived from a primary care-based retrospective cohort," *BMJ Open*, vol. 10, no. 7, e036099, 2020. DOI: 10.1136/bmjopen-2019-036099.

[27] H. Tibble et al., "Predicting asthma attacks in primary care: Protocol for developing a machine learning-based prediction model," *BMJ Open*, vol. 9, no. 7, e028375, 2019. DOI: 10.1136/bmjopen-2018-028375.

[28] A. H. Y. Chan, A. Tomlin, K. Beyene, and J. Harrison, "Asthma exacerbations in new zealand 2010-2019: A national population-based study," *Respiratory Medicine*, vol. 217, p. 107 365, 2023. DOI: 10.1016/j.rmed.2023.107365.

[29] A. Bourdin et al., "Severe asthma: Global and regional perspectives," *European Respiratory Journal*, vol. 54, no. 6, p. 1 900 900, 2019. DOI: 10.1183/13993003.00900-2019.

[30] Global Asthma Network, "The global asthma report 2018," Tech. Rep., 2018. [Online]. Available: https://www.globalasthmareport.org/2018/index.html.

[31] K. C. H. Tsang, H. Pinnock, A. M. Wilson, and S. A. Shah, "Application of machine learning to support self-management of asthma with mhealth," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, vol. 2020, pp. 5673–5677, 2020. DOI: 10.1109/EMBC44109.2020.9175679.

[32] R. Gauld, "The new zealand health care system," in *International Profiles of Health Care Systems. New York*. 2020, pp. 149–158.

[33] D. Patel, G. L. Hall, D. Broadhurst, A. Smith, A. Schultz, and R. E. Foong, "Does machine learning have a role in the prediction of asthma in children?" *Paediatric Respiratory Reviews*, 2021. DOI: 10.1016/j.prrv.2021.06.002.

[34] T. Shi et al., "Risk of covid-19 hospital admission among children aged 5–17 years with asthma in scotland: A national incident cohort study," *The Lancet Respiratory Medicine*, vol. 10, no. 2, pp. 191–198, 2022. DOI: 10.1016/S2213-2600(21)00491-4.

[35] K. L. Khatri and L. S. Tamil, "Early detection of peak demand days of chronic respiratory diseases emergency department visits using artificial neural networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 1, pp. 285–290, 2018. DOI: 10.1109/JBHI.2017.2698418.

[36] M.-S. Kim, J.-H. Lee, Y.-J. Jang, C.-H. Lee, J.-H. Choi, and T.-E. Sung, "Hybrid deep learning algorithm with open innovation perspective: A prediction model of asthmatic occurrence," *Sustainability*, vol. 12, no. 15, p. 6143, 2020. DOI: 10.3390/su12156143.

[37] T. Z. Maung, J. E. Bishop, E. Holt, A. M. Turner, and C. Pfrang, "Indoor air pollution and the health of vulnerable groups: A systematic review focused on particulate matter (pm), volatile organic compounds (vocs) and their effects on children and people with pre-existing lung disease," *International journal of environmental research and public health*, vol. 19, no. 14, 2022. DOI: 10.3390/ijerph19148752.

[38] S. Sandberg et al., "The role of acute and chronic stress in asthma attacks in children," *The Lancet*, vol. 356, no. 9234, pp. 982–987, 2000. DOI: 10.1016/S0140-6736(00)02715-X.

[39] S. R. Carroll, E. Herczog, M. Hudson, K. Russell, and S. Stall, "Operationalizing the CARE and FAIR principles for indigenous data futures," *Scientific Data*, vol. 8, p. 108, 2021. DOI: 10.1038/s41597-021-00892-0.

[40] R. Wirth and J. Hipp, "Crisp-dm: Towards a standard process model for data mining," in *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, vol. 1, 2000, pp. 29–39.

[41] V. Clarke and V. Braun, "Thematic analysis," *The Journal of Positive Psychology*, vol. 12, no. 3, pp. 297–298, 2017.

[42] M. L. Levy and R. Winter, "Asthma deaths: What now?" *Thorax*, vol. 70, no. 3, pp. 209–210, 2015. DOI: 10.1136/thoraxjnl-2015-206800.

[43] J. Gaudillo et al., "Machine learning approach to single nucleotide polymorphism-based asthma prediction," *PLOS One*, vol. 14, no. 12, pp. 1–12, 2019. DOI: 10.1371/journal.pone.0225574.

[44] R. J. Loymans et al., "Exacerbations in adults with asthma: A systematic review and external validation of prediction models," *The Journal of Allergy and Clinical Immunology: In Practice*, vol. 6, no. 6, pp. 1942–1952, 2018.

[45] A. I. Messinger, G. Luo, and R. R. Deterding, "The doctor will see you now: How machine learning and artificial intelligence can extend our understanding and treatment of asthma," *Journal of Allergy and Clinical Immunology*, vol. 145, no. 2, pp. 476–478, 2020. DOI: 10.1016/j.jaci.2019.12.898.

[46] M. Lovrić, I. Banić, E. Lacić, K. Pavlović, R. Kern, and M. Turkalj, "Predicting treatment outcomes using explainable machine learning in children with asthma," *Children*, vol. 8, no. 5, p. 376, 2021. DOI: 10.3390/children8050376.

[47] Y. Xiang et al., "Asthma exacerbation prediction and interpretation based on time-sensitive attentive neural network: A retrospective cohort study," *medRxiv*, p. 19 012 161, 2019. DOI: 10.1101/19012161.

[48] A. Kaplan et al., "Artificial intelligence/machine learning in respiratory medicine and potential role in asthma and copd diagnosis," *The Journal of Allergy and Clinical Immunology: In Practice*, vol. 9, no. 6, pp. 2255–2261, 2021. DOI: 10.1016/j.jaip.2021.02.014.

[49] M. J. Page et al., "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," *Systematic Reviews*, vol. 10, no. 1, p. 89, 2021. DOI: 10.1136/bmj.n71.

[50] M. Ouzzani, H. Hammady, Z. Fedorowicz, and A. Elmagarmid, "Rayyan – a web and mobile app for systematic reviews," *Systematic Reviews*, vol. 5, no. 1, p. 210, 2016, ISSN: 2046-4053. DOI: 10.1186/s13643-016-0384-4.

[51] S. Aggarwal et al., "Factors associated with future hospitalization among children with asthma: A systematic review," *Journal of Asthma*, vol. 60, no. 3, pp. 425–445, 2023. DOI: 10.1080/02770903.2022.2070762.

[52] E. T. Alharbi, F. Nadeem, and A. Cherif, "Predictive models for personalized asthma attacks based on patient's biosignals and environmental factors: A systematic review," *BMC Medical Informatics and Decision Making*, vol. 21, no. 1, pp. 1–13, 2021. DOI: 10.1186/s12911-021-01704-6.

[53] Critical Appraisal Skills Programme, *Casp (clinical prediction rule checklist)*, Critical Appraisal Skills Programme, 2022. [Online]. Available: https://casp-uk.net/.

[54] O. Zhang, L. L. Minku, and S. Gonem, "Detecting asthma exacerbations using daily home monitoring and machine learning," *Journal of Asthma*, vol. 58, no. 11, pp. 1518–1527, 2021. DOI: 10.1080/02770903.2020.1802746.

[55] A. A. H. De Hond, I. M. J. Kant, P. J. Honkoop, A. D. Smith, E. W. Steyerberg, and J. K. Sont, "Machine learning did not beat logistic regression in time series prediction for severe asthma exacerbations," *Scientific Reports*, vol. 12, no. 1, p. 20 363, 2022. DOI: 10.1038/s41598-022-24909-9.

[56] J. Finkelstein and I. C. Jeong, "Using cart for advanced prediction of asthma attacks based on telemonitoring data," in *2016 IEEE 7th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON)*, 2016, pp. 1–5. DOI: 10.1109/UEMCON.2016.7777890.

[57] J. Finkelstein and C. Jeong, "Machine learning approaches to personalize early prediction of asthma exacerbations," *Annals of the New York Academy of Sciences*, vol. 1387, no. 1, pp. 153–165, 2017. DOI: 10.1111/nyas.13218.

[58] K. Lisspers et al., "Developing a short-term prediction model for asthma exacerbations from swedish primary care patients' data using machine learning - based on the arctic study," *Respiratory Medicine*, vol. 185, p. 106 483, 2021. DOI: 10.1016/j.rmed.2021.106483.

[59] N. L. Lugogo et al., "A predictive machine learning tool for asthma exacerbations: Results from a 12-week, open-label study using an electronic multi-dose dry powder inhaler with integrated sensors," *Journal of Asthma Allergy*, vol. 15, pp. 1623–1637, 2022. DOI: 10.2147/jaa.S377631.

[60] J. H. Hurst, C. Zhao, H. P. Hostetler, M. Ghiasi Gorveh, J. E. Lang, and B. A. Goldstein, "Environmental and clinical data utility in pediatric asthma exacerbation risk prediction models," *BMC Medical Informatics and Decision Making*, vol. 22, no. 1, 2022. DOI: 10.1186/s12911-022-01847-0.

[61] J. W. Inselman et al., "A prediction model for asthma exacerbations after stopping asthma biologics," *Annals of Allergy Asthma Immunology*, vol. 130, no. 3, pp. 305–311, 2023. DOI: 10.1016/j.anai.2022.11.025.

[62] G. Luo, M. D. Johnson, F. L. Nkoy, S. He, and B. L. Stone, "Automatically explaining machine learning prediction results on asthma hospital visits in patients with asthma: Secondary analysis," *JMIR Medical Informatics*, vol. 8, no. 12, e21965, 2020. DOI: 10.2196/21965.

[63] E. K. Shin, R. Mahajan, O. Akbilgic, and A. Shaban-Nejad, "Sociomarkers and biomarkers: Predictive modeling in identifying pediatric asthma patients at risk of hospital revisits," *NPJ Digital Medicine*, vol. 1, p. 50, 2018. DOI: 10.1038/s41746-018-0056-y.

[64] T. Jiao, M. E. Schnitzer, A. Forget, and L. Blais, "Identifying asthma patients at high risk of exacerbation in a routine visit: A machine learning model," *Respiratory Medicine*, vol. 198, p. 106 866, 2022. DOI: 10.1016/j.rmed.2022.106866.

[65] E. Alharbi, A. Cherif, F. Nadeem, and T. Mirza, "Machine learning models for early prediction of asthma attacks based on bio-signals and environmental triggers," in *2022 IEEE/ACS 19th International Conference on Computer Systems and Applications (AICCSA)*, 2022, pp. 1–7. DOI: 10.1109/AICCSA56895.2022.10017305.

[66] K. J. Farion, S. Wilk, W. Michalowski, D. O'Sullivan, and J. Sayyad-Shirabad, "Comparing predictions made by a prediction model, clinical score, and physicians: Pediatric asthma exacerbations in the emergency department," *Appl. Clin. Informatics*, vol. 4, no. 3, pp. 376–391, 2013. DOI: 10.4338/ACI-2013-04-RA-0029.

[67] R. Haque et al., "Intelligent asthma self-management system for personalised weather-based healthcare using machine learning," in *Advances and Trends in Artificial Intelligence. Artificial Intelligence Practices*, Springer International Publishing, 2021, pp. 297–308. DOI: 10.1007/978-3-030-79457-6_26.

[68] C. K. Priya, M. Sudhakar, J. Lingampalli, and C. Z. Basha, "An advanced fog based health care system using ann for the prediction of asthma," in *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, 2021, pp. 1138–1145. DOI: 10.1109/ICCMC51019.2021.9418248.

[69] S. J. Patel, D. B. Chamberlain, and J. M. Chamberlain, "A machine learning approach to predicting need for hospitalization for pediatric asthma exacerbation at the time of emergency department triage," *Academic Emergency Medicine*, vol. 25, no. 12, pp. 1463–1470, 2018. DOI: 10.1111/acem.13655.

[70] M. Xu et al., "Genome wide association study to predict severe asthma exacerbations in children using random forests classifiers," *BMC Medical Genetics*, vol. 12, 2011. DOI: 10.1186/1471-2350-12-90.

[71] J. G. Zein, C. P. Wu, A. H. Attaway, P. Zhang, and A. Nazha, "Novel machine learning can predict acute asthma exacerbation," *Chest*, vol. 159, no. 5, pp. 1747–1757, 2021. DOI: 10.1016/j.chest.2020.12.051.

[72] Q. Do, S. Tran, and A. Doig, "Reinforcement learning framework to identify cause of diseases - predicting asthma attack case," in *2019 IEEE International Conference on Big Data (Big Data)*, 2019, pp. 4829–4838. DOI: 10.1109/BigData47090.2019.9006407.

[73] M. D. Eisner, P. P. Katz, E. H. Yelin, S. C. Shiboski, and P. D. Blanc, "Risk factors for hospitalization among adults with asthma: The influence of sociodemographic factors and asthma severity," *Respiratory research*, vol. 2, no. 1, pp. 1–8, 2000. DOI: 10.1186/rr37.

[74] I. N. Soyiri and D. D. Reidpath, "Evolving forecasting classifications and applications in health forecasting," *International journal of general medicine*, vol. 5, p. 381, 2012. DOI: 10.2147/IJGM.S31079.

[75] S. Bose, C. C. Kenyon, and A. J. Masino, "Personalized prediction of early childhood asthma persistence: A machine learning approach," *PLOS ONE*, vol. 16, no. 3, pp. 1–17, 2021. DOI: 10.1371/journal.pone.0247784.

[76] S. Mariani, M. M. Lahr, E. Metting, E. Vargiu, and F. Zambonelli, "Developing an ml pipeline for asthma and copd: The case of a dutch primary care service," *International Journal of Intelligent Systems*, vol. 36, no. 11, pp. 6763–6790, 2021. DOI: 10.1002/int.22568.

[77] J. M. B. Myers et al., "A pediatric asthma risk score to better predict asthma development in young children," *Journal of Allergy and Clinical Immunology*, vol. 143, no. 5, pp. 1803–1810, 2019.

[78] C. A. Jones et al., "Predicting asthma control using patient attitudes toward medical care: The react score," *Annals of Allergy, Asthma & Immunology*, vol. 102, no. 5, pp. 385–392, 2009.

[79] F. Xie, B. Chakraborty, M. E. H. Ong, B. A. Goldstein, N. Liu, et al., "Autoscore: A machine learning–based automatic clinical score generator and its application to mortality prediction using electronic health records," *JMIR medical informatics*, vol. 8, no. 10, e21798, 2020.

[80] S. Boer et al., "Development and validation of personalized prediction to estimate future risk of severe exacerbations and uncontrolled asthma in patients with asthma, using clinical parameters and early treatment response," *The Journal of Allergy and Clinical Immunology: In Practice*, vol. 7, no. 1, pp. 175–182, 2019.

[81] L. Chan et al., "Derivation and validation of a machine learning risk score using biomarker and electronic patient data to predict progression of diabetic kidney disease," *Diabetologia*, vol. 64, pp. 1504–1515, 2021.

[82] G. Luo, "A review of automatic selection methods for machine learning algorithms and hyper-parameter values," *Network Modeling Analysis in Health Informatics and Bioinformatics*, vol. 5, pp. 1–16, 2016.

[83] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization.," *Journal of Machine Learning Research*, vol. 13, no. 2, 2012.

[84] D. A. Anggoro and S. S. Mukti, "Performance comparison of grid search and random search methods for hyperparameter tuning in extreme gradient boosting algorithm to predict chronic kidney failure.," *International Journal of Intelligent Engineering & Systems*, vol. 14, no. 6, 2021.

[85] F. Hutter, H. H. Hoos, and K. Leyton-Brown, "Sequential model-based optimization for general algorithm configuration," in *Learning and Intelligent Optimization: 5th International Conference, LION 5, Rome, Italy, January 17-21, 2011. Selected Papers 5*, Springer, 2011, pp. 507–523.

[86] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," *Advances in Neural Information Processing Systems*, vol. 25, 2012.

[87] H. Alibrahim and S. A. Ludwig, "Hyperparameter optimization: Comparing genetic algorithm against grid search and bayesian optimization," in *2021 IEEE Congress on Evolutionary Computation (CEC)*, IEEE, 2021, pp. 1551–1559.

[88] T. To et al., "Global asthma prevalence in adults: Findings from the cross-sectional world health survey," *BMC Public Health*, vol. 12, p. 204, 2012, ISSN: 1471-2458. DOI: 10.1186/1471-2458-12-204.

[89] Health Quality and Safety Commission New Zealand, *Asthma*, Accessed: 2025-10-01, 2024. [Online]. Available: https://www.hqsc.govt.nz/our-data/atlas-of-healthcare-variation/asthma/.

[90] Ministry of Health, *Annual data explorer 2018/19: New zealand health survey. wellington: Ministry of health*, Accessed: 2024-04-15. [Online]. Available: https://minhealthnz.shinyapps.io/nz-health-survey-2018-19-annual-data-explorer.

[91] L. T. Barnard and J. Zhang, *The impact of respiratory disease in new zealand: 2020 update*, 2021. [Online]. Available: https://www.asthmafoundation.org.nz/assets/documents/Respiratory-Impact-report-final-2021Aug11.pdf.

[92] Asthma and Respiratory Foundation New Zealand, *Asthma*, 2022. [Online]. Available: https://www.asthmafoundation.org.nz/your-health/living-with-asthma.

[93] J. Gaudillo et al., "Machine learning approach to single nucleotide polymorphism-based asthma prediction," *PLoS ONE*, vol. 14, 12 2019. DOI: 10.1371/journal.pone.0225574.

[94] W. K. D. Jayamini, F. Mirza, M. A. Naeem, and A. H. Y. Chan, "State of asthma-related hospital admissions in new zealand and predicting length of stay using machine learning," *Applied Sciences*, vol. 12, 19 2022, ISSN: 2076-3417. DOI: 10.3390/app12199890.

[95] F. G. Nabi, K. Sundaraj, C. K. Lam, and R. Palaniappan, "Characterization and classification of asthmatic wheeze sounds according to severity level using spectral integrated features," *Computers in Biology and Medicine*, vol. 104, pp. 52–61, 2019, ISSN: 0010-4825. DOI: 10.1016/j.compbiomed.2018.10.035.

[96] D. Martinho et al., "A hybrid model to classify patients with chronic obstructive respiratory diseases," *Journal of Medical Systems*, vol. 45, pp. 1–11, 3 2021. DOI: 10.1007/s10916-020-01704-5.

[97] R. Y. Suruki, J. B. Daugherty, N. Boudiaf, and F. C. Albers, "The frequency of asthma exacerbations and healthcare utilisation in patients with asthma from the uk and usa," *BMC pulmonary medicine*, vol. 17, pp. 1–11, 2017. DOI: 10.1186/s12890-017-0409-3.

[98] H. K. Reddel et al., "An official american thoracic society/european respiratory society statement: Asthma control and exacerbations: Standardising endpoints for clinical asthma trials and clinical practice," *American journal of respiratory and critical care medicine*, vol. 180, no. 1, pp. 59–99, 2009. DOI: 10.1164/rccm.200801-060ST.

[99] S. B. Kotsiantis, D. Kanellopoulos, and P. E. Pintelas, "Data preprocessing for supervised leaning," *International journal of computer science*, vol. 1, pp. 111–117, 2 2006, ISSN: 1306-4428.

[100] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[101] J. N. V. Rijn and F. Hutter, "An empirical study of hyperparameter importance across datasets," 2017, pp. 91–98.

[102] H. Hyrkäs, T. M. Ikäheimo, J. J. K. Jaakkola, and M. S. Jaakkola, "Asthma control and cold weather-related respiratory symptoms," *Respiratory Medicine*, vol. 113, pp. 1–7, 2016, ISSN: 0954-6111. DOI: 10.1016/j.rmed.2016.02.005.

[103] N. Kearns, I. Maijers, J. Harper, R. Beasley, and M. Weatherall, "Inhaled corticosteroids in acute asthma: A systemic review and meta-analysis," *The Journal of Allergy and Clinical Immunology: In Practice*, vol. 8, 605–617.e6, 2 2020. DOI: 10.1016/j.jaip.2019.08.051.

[104] J. Paris et al., "Relationship between recent short-acting beta-agonist use and subsequent asthma exacerbations," *Annals of Allergy, Asthma and Immunology*, vol. 101, pp. 482–487, 5 2008. DOI: 10.1016/S1081-1206(10)60286-4.

[105] J. D. Blakey et al., "Identifying risk of future asthma attacks using uk medical record data: A respiratory effectiveness group initiative," *The Journal of Allergy and Clinical Immunology: In Practice*, vol. 5, 1015–1024.e8, 4 2017, ISSN: 2213-2198. DOI: 10.1016/j.jaip.2016.11.007.

[106] K. Balabaeva and S. Kovalchuk, "Comparison of temporal and non-temporal features effect on machine learning models quality and interpretability for chronic heart failure patients," *Procedia Computer Science*, vol. 156, pp. 87–96, 2019, ISSN: 1877-0509.

[107] P. Misra and A. S. Yadav, "Improving the classification accuracy using recursive feature elimination with cross-validation," *International Journal of Emerging Technologies*, vol. 11, pp. 659–665, 3 2020.

[108] K. Yan and D. Zhang, "Feature selection and analysis on correlated gas sensor data with recursive feature elimination," *Sensors and Actuators B: Chemical*, vol. 212, pp. 353–363, 2015. DOI: 10.1016/j.snb.2015.02.025.

[109] W. K. D. Jayamini, F. Mirza, M. A. Naeem, and A. H. Y. Chan, "Investigating machine learning techniques for predicting risk of asthma exacerbations: A systematic review," *Journal of Medical Systems*, vol. 48, no. 1, p. 49, 2024. DOI: 10.1007/s10916-024-02061-3.

[110] A. Budiarto, K. C. H. Tsang, A. M. Wilson, A. Sheikh, and S. A. Shah, "Machine learning–based asthma attack prediction models from routinely collected electronic health records: Systematic scoping review," *JMIR AI*, vol. 2, e46717, 2023. DOI: 10.2196/46717.

[111] W. K. D. Jayamini et al., "Predicting the risk of asthma attacks in new zealand using machine learning," in *Proceedings of the 58th Hawaii International Conference on System Sciences (HICSS 2025)*, 2025, pp. 3731–3741. DOI: 10.24251/HICSS.2025.448.

[112] H. Xi et al., "Forecasting hospitalization for adult asthma patients in emergency departments based on multiple environmental and clinical factors," *Journal of Asthma and Allergy*, vol. 18, pp. 861–876, 2025. DOI: 10.2147/JAA.S512405.

[113] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003. DOI: 10.5555/944919.944968.

[114] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Computers and Electrical Engineering*, vol. 40, no. 1, pp. 16–28, 2014. DOI: 10.1016/j.compeleceng.2013.11.024.

[115] V. Bolón-Canedo, N. Sánchez-Maroño, and A. Alonso-Betanzos, "Recent advances and emerging challenges of feature selection in the context of big data," *Knowledge-Based Systems*, vol. 86, pp. 33–45, 2015. DOI: 10.1016/j.knosys.2015.05.014.

[116] Z. Y. Taha, A. A. Abdullah, and T. A. Rashid, *Optimizing feature selection with genetic algorithms: A review of methods and applications*, 2024. DOI: 10.48550/arXiv.2409.14563. eprint: arXiv:2409.14563.

[117] E. Kraev, B. Koseoglu, L. Traverso, and M. Topiwalla, "Shap-select: Lightweight feature selection using shap values and regression," *arXiv preprint arXiv:2410.06815*, 2024.

[118] Y. Li, U. Mansmann, S. Du, and R. Hornung, "Benchmark study of feature selection strategies for multi-omics data," *BMC Bioinformatics*, vol. 23, no. 1, p. 412, 2022. DOI: 10.1186/s12859-022-04962-x.

[119] S. Abdumalikov, J. Kim, and Y. Yoon, "Performance analysis and improvement of machine learning with various feature selection methods for eeg-based emotion classification," *Applied Sciences*, vol. 14, no. 22, p. 10 511, 2024. DOI: 10.3390/app142210511.

[120] I. Ramos-Pérez, J. A. Barbero-Aparicio, A. Canepa-Oneto, Á. Arnaiz-González, and J. Maudes-Raedo, "An extensive performance comparison between feature reduction and feature selection preprocessing algorithms on imbalanced wide data," *Information*, vol. 15, no. 4, p. 223, 2024. DOI: 10.3390/info15040223.

[121] M. C. Barbieri, B. I. Grisci, and M. Dorn, "Analysis and comparison of feature selection methods towards performance and stability," *Expert Systems with Applications*, vol. 249, p. 123 667, 2024. DOI: 10.1016/j.eswa.2024.123667.

[122] Y. Xiang et al., "Asthma exacerbation prediction and risk factor analysis based on a time-sensitive, attentive neural network: Retrospective cohort study," *Journal of Medical Internet Research*, vol. 22, no. 7, e16981, 2020. DOI: 10.2196/16981.

[123] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 67, no. 2, pp. 301–320, 2005. DOI: 10.1111/j.1467-9868.2005.00503.x.

[124] Y. Rothacher and C. Strobl, "Identifying informative predictor variables with random forests," *Journal of Educational and Behavioural Statistics*, vol. 49, no. 4, pp. 595–629, 2024. DOI: 10.3102/10769986231193327.

[125] Y. Gerstorfer, M. Hahn-Klimroth, and L. Krieg, "A notion of feature importance by decorrelation and detection of trends by random forest regression," *Data Science Journal*, vol. 22, p. 42, 2023. DOI: 10.5334/dsj-2023-042.

[126] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, no. 1–3, pp. 389–422, 2002. DOI: 10.1023/A:1012487302797.

[127] J. Chen, F. Zhao, Y. Sun, and Y. Yin, "Improved xgboost model based on genetic algorithm," *International Journal of Computer Applications in Technology*, vol. 62, no. 3, pp. 240–245, 2020. DOI: 10.1504/IJCAT.2020.106571.

[128] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. DOI: 10.1023/A:1010933404324.

[129] S. J. Rigatti, "Random forest," *Journal of Insurance Medicine*, vol. 47, no. 1, pp. 31–39, 2017. DOI: 10.17849/insm-47-01-31-39.1.

[130] American Lung Association, *Asthma trends and burden*, 2022. [Online]. Available: https://www.lung.org/research/trends-in-lung-disease/asthma-trends-brief/trends-and-burden.

[131] C. I. Bloom, F. Nissen, I. J. Douglas, L. Smeeth, P. Cullinan, and J. K. Quint, "Exacerbation risk and characterisation of the uk's asthma population from infants to old age," *Thorax*, vol. 73, no. 4, pp. 313–320, 2018. DOI: 10.1136/thoraxjnl-2017-210650.

[132] A. H. Y. Chan, H. K. Reddel, A. Apter, M. Eakin, K. Riekert, and J. M. Foster, "Adherence monitoring and e-health: How clinicians and researchers can use technology to promote inhaler adherence for asthma," *The Journal of Allergy and Clinical Immunology: In Practice*, vol. 1, no. 5, pp. 446–454, 2013. DOI: 10.1016/j.jaip.2013.06.015.

[133] O. Selroos et al., "National and regional asthma programmes in europe," *European Respiratory Review*, vol. 24, no. 137, pp. 474–83, 2015. DOI: 10.1183/16000617.00008114.

[134] M. K. Miller, J. H. Lee, D. P. Miller, S. E. Wenzel, T. S. Group, et al., "Recent asthma exacerbations: A key predictor of future exacerbations," *Respiratory Medicine*, vol. 101, no. 3, pp. 481–489, 2007. DOI: 10.1016/j.rmed.2006.07.005.

[135] R. Lowden and S. Turner, "Past asthma exacerbation in children predicting future exacerbation: A systematic review," *ERJ Open Research*, vol. 8, no. 4, pp. 00 174–2022, 2022. DOI: 10.1183/23120541.00174-2022.

[136] W.-H. Boehncke, S. Boehncke, and M. P. Schön, "Managing comorbid disease in patients with psoriasis," *Bmj*, vol. 340, 2010. DOI: 10.1136/bmj.b5666.

[137] J. Wang et al., "Association between psoriasis and asthma risk: A meta-analysis," *Allergy and Asthma Proceedings*, vol. 39, no. 2, pp. 103–109, 2018. DOI: 10.2500/aap.2018.39.4109.

[138] H.-Y. Fang, W.-C. Liao, C.-L. Lin, C.-H. Chen, and C.-H. Kao, "Association between psoriasis and asthma: A population-based retrospective cohort analysis," *British Journal of Dermatology*, vol. 172, no. 4, pp. 1066–1071, 2015. DOI: 10.1111/bjd.13518.

[139] E. O. Bayman and F. Dexter, "Multicollinearity in logistic regression models," *Anesthesia & Analgesia*, vol. 133, no. 2, pp. 362–365, 2021. DOI: 10.1213/ANE.0000000000005593.

[140] G. Zeng and E. Zeng, "On the relationship between multicollinearity and separation in logistic regression," *Communications in Statistics - Simulation and Computation*, vol. 50, no. 7, pp. 1989–1997, 2021, ISSN: 0361-0918. DOI: 10.1080/03610918.2019.1589511.

[141] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets," *PLOS ONE*, vol. 10, no. 3, e0118432, 2015. DOI: 10.1371/journal.pone.0118432.

[142] L. A. Jeni, J. F. Cohn, and F. D. L. Torre, "Facing imbalanced data — recommendations for the use of performance metrics," in *Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII 2013)*, 2013, pp. 245–251. DOI: 10.1109/ACII.2013.47.

[143] F. Hargreave and P. Nair, "The definition and diagnosis of asthma," *Clinical and Experimental Allergy*, vol. 39, no. 11, pp. 1652–1658, 2009. DOI: 10.1111/j.1365-2222.2009.03321.x.

[144] N. Navanandan, J. Hatoun, J. C. Celedón, and A. H. Liu, "Predicting severe asthma exacerbations in children: Blueprint for today and tomorrow," *The Journal of Allergy and Clinical Immunology: In Practice*, vol. 9, no. 7, pp. 2619–2626, 2021. DOI: 10.1016/j.jaip.2021.03.039.

[145] H. Y. Seol et al., "Artificial intelligence-assisted clinical decision support for childhood asthma management: A randomized clinical trial," *PloS one*, vol. 16, no. 8, e0255261, 2021. DOI: 10.1371/journal.pone.0255261.

[146] D. M. Kothalawala et al., "Development of childhood asthma prediction models using machine learning approaches," *Clinical and Translational Allergy*, vol. 11, no. 9, e12076, 2021. DOI: 10.1002/clt2.12076.

[147] M. Perpiñá et al., "Expert consensus recommendations for the management of asthma in older adults," *Medicina Clínica (English Edition)*, vol. 159, no. 1, 53.e1–53.e14, 2022. DOI: 10.1016/j.medcle.2021.04.034.

[148] H. Abdelkarim et al., "A comparison of characteristics and outcomes of patients admitted to the icu with asthma in australia and new zealand and the united states," *Journal of Asthma*, vol. 57, no. 4, pp. 398–404, 2020, ISSN: 0277-0903. DOI: 10.1080/02770903.2019.1571082.

[149] L. Luo et al., "The effects of air pollution on length of hospital stay for adult patients with asthma," *The international journal of health planning and management*, vol. 33, no. 3, e751–e767, 2018. DOI: 10.1002/hpm.2532.

[150] C. Dobkin, A. Finkelstein, R. Kluender, and M. J. Notowidigdo, "The economic consequences of hospital admissions," *American Economic Review*, vol. 108, no. 2, pp. 308–52, 2018. DOI: 10.1257/aer.20161038.

[151] D. Schlichting, T. Fadason, C. C. Grant, and J. O'Sullivan, "Childhood asthma in new zealand: The impact of on-going socioeconomic disadvantage (2010–2019)," *New Zealand Medical Journal*, vol. 134, no. 1533, 2021.

[152] H.-D. Thai and J.-H. Huh, "Optimizing patient transportation by applying cloud computing and big data analysis," *The Journal of Supercomputing*, 2022. DOI: 10.1007/s11227-022-04576-3.

[153] I. N. Soyiri, D. D. Reidpath, and C. Sarran, "Asthma length of stay in hospitals in london 2001–2006: Demographic, diagnostic and temporal factors," *PLoS One*, vol. 6, no. 11, e27184, 2011. DOI: 10.1371/journal.pone.0027184.

[154] L. A. Shanley, H. Lin, and G. Flores, "Factors associated with length of stay for pediatric asthma hospitalizations," *Journal of Asthma*, vol. 52, no. 5, pp. 471–477, 2015. DOI: 10.3109/02770903.2014.984843.

[155] J. Baek, B. A. Kash, X. Xu, M. Benden, J. Roberts, and G. Carrillo, "Association between ambient air pollution and hospital length of stay among children with asthma in south texas," *International jJournal of Environmental Research and Public Health*, vol. 17, no. 11, p. 3812, 2020. DOI: 10.3390/ijerph17113812.

[156] M. Rachda Naila, P. Caulier, S. Chaabane, A. Chraibi, and S. Piechowiak, "A comparative study of machine learning models for predicting length of stay in hospitals," *Journal of Information Science and Engineering*, vol. 37, pp. 1025–1038, 2021.

[157] L. Turgeman, J. H. May, and R. Sciulli, "Insights from a machine learning model for predicting the hospital length of stay (los) at the time of admission," *Expert Systems with Applications*, vol. 78, pp. 376–385, 2017. DOI: 10.1016/j.eswa.2017.02.023.

[158] P.-F. Tsai et al., "Length of hospital stay prediction at the admission stage for cardiology patients using artificial neural network," *Journal of Healthcare Engineering*, vol. 2016, p. 7 035 463, 2016. DOI: 10.1155/2016/7035463.

[159] S. Aghajani and M. Kargari, "Determining factors influencing length of stay and predicting length of stay using data mining in the general surgery department," *Hospital Practices and Research*, vol. 1, no. 2, pp. 53–58, 2016. DOI: 10.20286/hpr-010251.

[160] O. J. Achilonu et al., "Use of machine learning and statistical algorithms to predict hospital length of stay following colorectal cancer resection: A south african pilot study," *Frontiers in Oncology*, vol. 11, 2021.

[161] X. Ma, Y. Si, Z. Wang, and Y. Wang, "Length of stay prediction for icu patients using individualized single classification algorithm," *Computer Methods and Programs in Biomedicine*, vol. 186, p. 105 224, 2020. DOI: 10.1016/j.cmpb.2019.105224.

[162] H. A. Cohen, H. Blau, M. Hoshen, E. Batat, and R. D. Balicer, "Seasonality of asthma: A retrospective population study," *Pediatrics*, vol. 133, no. 4, e923–e932, 2014, ISSN: 0031-4005.

[163] S. NZ, *Ethnicity standard classification: Findings from public consultation november 2019*, 2020. [Online]. Available: https://www.stats.govt.nz/consultations/ethnicity-standard-classification-findings-from-public-consultation-november-2019/.

[164] N. Boven, D. Exeter, A. Sporle, and N. Shackleton, "The implications of different ethnicity categorisation methods for understanding outcomes and developing policy in new zealand," *Kōtuitui: New Zealand Journal of Social Sciences Online*, vol. 15, no. 1, pp. 123–139, 2019, ISSN: 1177-083X. DOI: 10.1080/1177083x.2019.1657912.

[165] T. Mahony et al., "Weekend versus weekday asthma-related emergency department utilisation," *Academic Pediatrics*, vol. 22, no. 4, pp. 640–646, 2022. DOI: 10.1016/j.acap.2021.09.005.

[166] A. C. Krefis et al., "Temporal analysis of determinants for respiratory emergency department visits in a large german hospital," *BMJ Open Respiratory Research*, vol. 5, no. 1, e000338, 2018, ISSN: 2052-4439. DOI: 10.1136/bmjresp-2018-000338.

[167] M. Shakir and N. Rakesh, "Investigation on air pollutant data sets using data mining tool," in *2018 2nd International Conference on IoT in Social, Mobile, Analytics and Cloud (I-SMAC)*, pp. 480–485. DOI: 10.1109/I-SMAC.2018.8653757.

[168] J. Bennett et al., "Sources of indoor air pollution at a new zealand urban primary school; a case study," *Atmospheric Pollution Research*, vol. 10, no. 2, pp. 435–444, 2019, ISSN: 1309-1042. DOI: 10.1016/j.apr.2018.09.006.

[169] S. Ram, W. Zhang, M. Williams, and Y. Pengetnze, "Predicting asthma-related emergency department visits using big data," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 4, pp. 1216–1223, 2015. DOI: 10.1109/JBHI.2015.2404829.

[170] J. Busby et al., "Factors affecting adherence with treatment advice in a clinical trial of patients with severe asthma," *European Respiratory Journal*, vol. 59, no. 4, 2022, ISSN: 0903-1936.

[171] S. Padinjappurathu Gopalan, C. L. Chowdhary, C. Iwendi, M. A. Farid, and L. K. Ramasamy, "An efficient and privacy-preserving scheme for disease prediction in modern healthcare systems," *Sensors*, vol. 22, no. 15, p. 5574, 2022. DOI: 10.3390/s22155574.

[172] C. Iwendi, S. Khan, J. H. Anajemba, A. K. Bashir, and F. Noor, "Realising an efficient iomt-assisted patient diet recommendation system through machine learning model," *IEEE Access*, vol. 8, pp. 28 462–28 474, 2020, ISSN: 2169-3536.

[173] Y. Feng, Y. Wang, C. Zeng, and H. Mao, "Artificial intelligence and machine learning in chronic airway diseases: Focus on asthma and chronic obstructive pulmonary disease," *International Journal of Medical Sciences*, vol. 18, no. 13, pp. 2871–2889, 2021. DOI: 10.7150/ijms.58191.

[174] L. K. Ramasamy, F. Khan, M. Shah, B. V. V. S. Prasad, C. Iwendi, and C. Biamba, "Secure smart wearable computing through artificial intelligence-enabled internet of things and cyber-physical systems for health monitoring," *Sensors*, vol. 22, no. 3, p. 1076, 2022. DOI: 10.3390/s22031076.

[175] J. F. van Boven et al., "Ers "connect" clinical research collaboration–moving multiple digital innovations towards connected respiratory care: Addressing the over-arching challenges of whole systems implementation," *European Respiratory Journal*, vol. 62, no. 5, 2023. DOI: 10.1183/13993003.01680-2023.

[176] A. H. Chan, A. W. Stewart, J. Harrison, C. A. Camargo, P. N. Black, and E. A. Mitchell, "The effect of an electronic monitoring device with audiovisual reminder function on adherence to inhaled corticosteroids and school attendance in children with asthma: A randomised controlled trial," *The Lancet Respiratory Medicine*, vol. 3, no. 3, pp. 210–219, 2015. DOI: 10.1016/S2213-2600(15)00008-9.

[177] A. Chan et al., "Digital interventions to improve adherence to maintenance medication in asthma," *Cochrane Database of Systematic Reviews*, no. 6, 2022. DOI: 10.1002/14651858.CD013030.pub2.

[178] S. Bosnic-Anticevich, N. D. Bakerly, H. Chrystyn, M. Hew, and J. van der Palen, "Advancing digital solutions to overcome longstanding barriers in asthma and copd management," *Patient Prefer Adherence*, vol. 17, pp. 259–272, 2023. DOI: 10.2147/ppa.S385857.

[179] C. Jácome et al., "Asthma app use and interest among patients with asthma: A multicenter study," *Journal of Investigational Allergology and Clinical Immunology*, vol. 30, no. 2, pp. 137–140, 2020. DOI: 10.18176/jiaci.0456.

[180] H. Pinnock, C. Y. Hui, and J. F. van Boven, "Implementation of digital home monitoring and management of respiratory disease," *Current Opinion in Pulmonary Medicine*, vol. 29, no. 4, pp. 302–312, 2023. DOI: 10.1097/MCP.0000000000000965.

[181] Z. Obermeyer, B. Powers, C. Vogeli, and S. Mullainathan, "Dissecting racial bias in an algorithm used to manage the health of populations," *Science*, vol. 366, no. 6464, pp. 447–453, 2019. DOI: 10.1126/science.aax2342.

[182] C. O'neil, *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown, 2017, ISBN: 0553418831.

[183] N. A. Molfino, G. Turcatel, and D. Riskin, "Machine learning approaches to predict asthma exacerbations: A narrative review," *Advances in Therapy*, vol. 41, no. 2, pp. 534–552, 2024. DOI: 10.1007/s12325-023-02743-3.

[184] I. Y. Chen, E. Pierson, S. Rose, S. Joshi, K. Ferryman, and M. Ghassemi, "Ethical machine learning in healthcare," *Annual Review of Biomedical Data Science*, vol. 4, no. 1, pp. 123–144, 2021. DOI: 10.1146/annurev-biodatasci-092820-114757.

[185] A. K. Paul and M. Schaefer, "Safeguards for the use of artificial intelligence and machine learning in global health," *Bulletin of the World Health Organisation*, vol. 98, no. 4, p. 282, 2020. DOI: 10.2471/BLT.19.237099.

[186] E. Komene, B. Pene, D. Gerard, J. Parr, C. Aspinall, and D. Wilson, "Whakawhanaungatanga—building trust and connections: A qualitative study indigenous māori patients and whānau (extended family network) hospital experiences," *Journal of Advanced Nursing*, vol. 80, no. 4, pp. 1545–1558, 2024. DOI: 10.1111/jan.15912.

[187] J. M. Morse, "Determining sample size," *Qualitative Health Research*, vol. 10, no. 1, pp. 3–5, 2000. DOI: 10.1177/104973200129118183.

[188] K. Hika, M. Harwood, S. Ritchie, and A. H. Y. Chan, "Maori experiences and beliefs about antibiotics and antimicrobial resistance for acute upper respiratory tract symptoms: A qualitative study," *Antibiotics*, vol. 11, no. 6, p. 714, 2022. DOI: 10.3390/antibiotics11060714.

[189] R. L. Beard, D. J. Fetterman, B. Wu, and L. Bryant, "The two voices of alzheimer's: Attitudes toward brain health by diagnosed individuals and support persons," *The Gerontologist*, vol. 49, no. S1, S40–S49, 2009. DOI: 10.1093/geront/gnp083.

[190] K. Charmaz and L. Belgrave, "Qualitative interviewing and grounded theory analysis," in *The SAGE handbook of interview research: The complexity of the craft*. 2012, vol. 2, pp. 347–365. DOI: 10.4135/9781452218403.

[191] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv:1810.04805*, pp. 1–16, 2018.

[192] K. Siau and W. Wang, "Building trust in artificial intelligence, machine learning, and robotics," *Cutter Business Technology Journal*, vol. 31, no. 2, pp. 47–53, 2018.

[193] S. K. Cibangu, "Marginalization of indigenous voices in the information age: A case study of cell phones in the rural congo*," *Information Technology for Development*, vol. 26, no. 2, pp. 234–267, 2020. DOI: 10.1080/02681102.2019.1647403.

[194] J. Koh and G. Tuazon, "A proposal to include maori perspectives in aied," *ASCILITE Publications*, pp. 462–466, 2023. DOI: 10.14742/apubs.2023.498.

[195] N. P. Gordon and M. C. Hornbrook, "Differences in access to and preferences for using patient portals and other ehealth technologies based on race, ethnicity, and age: A database and survey study of seniors in a large health plan," *Journal of Medical Internet Research*, vol. 18, no. 3, e50, 2016. DOI: 10.2196/jmir.5105.

[196] E. Glikson and A. W. Woolley, "Human trust in artificial intelligence: Review of empirical research," *Academy of Management Annals*, vol. 14, no. 2, pp. 627–660, 2020. DOI: 10.5465/annals.2018.0057.

[197] N. Gasteiger, A. Anderson, and K. Day, "Rethinking engagement: Exploring women's technology use during the perinatal period through a kaupapa māori consistent approach," *New Zealand College of Midwives Journal*, no. 55, 2019. DOI: 10.12784/nzcomjnl55.2019.3.20-26.

[198] A. Yap et al., "Patients perceptions of artificial intelligence in diabetic eye screening," *The Asia-Pacific Journal of Ophthalmology*, vol. 11, no. 3, pp. 287–293, 2022. DOI: 10.1097/apo.0000000000000525.

[199] E. Wikaire et al., "Reducing healthcare inequities for māori using telehealth during covid-19," *New Zealand Medical Journal*, vol. 135, no. 1552, pp. 112–7, 2022.

[200] L. Te Morenga et al., "Co-designing an mhealth tool in the new zealand māori community with a "kaupapa māori" approach," *AlterNative: An International Journal of Indigenous Peoples*, vol. 14, no. 1, pp. 90–99, 2018. DOI: 10.1177/1177180117753169.

[201] L. Whitehead, J. Talevski, F. Fatehi, and A. Beauchamp, "Barriers to and facilitators of digital health among culturally and linguistically diverse populations: Qualitative systematic review," *Journal of Medical Internet Research*, vol. 25, e42719, 2023. DOI: 10.2196/42719.

[202] J.-a. P. Hughson, J. O. Daly, R. Woodward-Kron, J. Hajek, and D. Story, "The rise of pregnancy apps and the implications for culturally and linguistically diverse women: Narrative review," *JMIR mHealth and uHealth*, vol. 6, no. 11, e189, 2018. DOI: 10.2196/mhealth.9119.

[203] N. Moore, "Reducing seclusion use for tangata whai i te ora through integration of maori culture into practice," pp. 57–61, 2019. DOI: 10.34074/scop.3006009.

[204] Statistics New Zealand, *Te reo māori proficiency and support continues to grow*, Web page, Accessed 2025-08-18, 2022. [Online]. Available: https://www.stats.govt.nz/news/te-reo-maori-proficiency-and-support-continues-to-grow/.

[205] R. Graham and B. Masters-Awatere, "Experiences of māori of aotearoa new zealand's public health system: A systematic review of two decades of published qualitative research," *Australian and New Zealand Journal of Public Health*, vol. 44, no. 3, pp. 193–200, 2020. DOI: 10.1111/1753-6405.12971.

[206] J. Gilmour, A. Huntington, and B. Robson, "Oral health experiences of maori with dementia and whanau perspectives-oranga waha mo nga iwi katoa," *Nursing Praxis in New Zealand*, vol. 32, no. 1, pp. 20–27, 2016. DOI: 10.36951/ngpxnz.2016.003.

[207] Environmental Health in New Zealand, *Ethnic profile*, Accessed: 2025-11-10, 2024. [Online]. Available: https://www.ehinz.ac.nz/indicators/population-vulnerability/ethnic-profile/.

[208] C. R. Jenkins, L.-P. Boulet, K. L. Lavoie, C. Raherison-Semjen, and D. Singh, "Personalised treatment of asthma: The importance of sex and gender differences," *The Journal of Allergy and Clinical Immunology: In Practice*, vol. 10, no. 4, 963–971. e3, 2022. DOI: 10.1016/j.jaip.2022.02.002.

[209] A. Gillies, R. Tinirau, and N. Mako, "Whakawhanaungatanga—extending the networking concept," *He Pukenga Korero*, vol. 8, no. 2, pp. 29–37, 2007.

[210] M. P. Reilly, "Māori studies, past and present: A review," *The Contemporary Pacific*, vol. 23, no. 2, pp. 340–370, 2011. DOI: 10.1353/cp.2011.0039.

[211] J. W. Mims, "Asthma: Definitions and pathophysiology," in *International forum of allergy and rhinology*, vol. 5, Wiley Online Library, 2015, S2–S6.

[212] E. Davidson and A. Sheikh, "Tackling ethnic variations in asthma outcomes in new zealand's mãori and pacific children will need a long-term strategy," *Primary Care Respiratory Journal*, vol. 22, no. 3, pp. 269–270, 2013. DOI: 10.4104/pcrj.2013.00076.

[213] A. H. Y. Chan et al., "Digipredict: Physiological, behavioural and environmental predictors of asthma attacks—a prospective observational study using digital markers and artificial intelligence—study protocol," *BMJ Open Respiratory Research*, vol. 11, no. 1, 2024. DOI: 10.1136/bmjresp-2023-002275.

[214] A. H. Y. Chan, D. Drummond, S. Ramakrishnan, J. F. van Boven, P. G. Gibson, and D. Thomas, "Digitally mapping the asthma journey—from diagnosis to remission," *EClinicalMedicine*, 2025, ISSN: 2589-5370. DOI: 10.1016/j.eclinm.2025.103204.

[215] O. Grana-Castro, E. Izquierdo, A. Pinas-Mesa, E. Menasalvas, and T. Chivato-Perez, "Assessing the impact of new technologies on managing chronic respiratory diseases," *Journal of Clinical Medicine*, vol. 13, no. 22, 2024. DOI: 10.3390/jcm13226913.

[216] K. Witkowski, R. Okhai, and S. R. Neely, "Public perceptions of artificial intelligence in healthcare: Ethical concerns and opportunities for patient-centered care," *BMC Medical Ethics*, vol. 25, no. 1, p. 74, 2024. DOI: 10.1186/s12910-024-01066-4.

[217] C. A. Fazakarley, M. Breen, B. Thompson, P. Leeson, and V. Williamson, "Beliefs, experiences and concerns of using artificial intelligence in healthcare: A qualitative synthesis," *Digit Health*, vol. 10, p. 20 552 076 241 230 075, 2024. DOI: 10.1177/20552076241230075.

[218] W. K. D. Jayamini, F. Mirza, M. C. Bidois-Putt, M. A. Naeem, and A. H. Y. Chan, "Perceptions toward using artificial intelligence and technology for asthma attack risk prediction: Qualitative exploration of maori views," *JMIR Formative Research*, vol. 8, e59811, 2024. DOI: 10.2196/59811.

[219] D. R. Thomas, "A general inductive approach for analyzing qualitative evaluation data," *American Journal of Evaluation*, vol. 27, no. 2, pp. 237–246, 2006.

[220] B. E. Njee et al., "Perceptions and experiences of nondiabetic cameroonian immigrants in minnesota on access to affordable, culturally adapted healthcare services," *BMC Health Services Research*, vol. 25, no. 1, p. 640, 2025. DOI: 10.1186/s12913-025-12575-2.

[221] J. Fia'Ali'i, M. Law, C. O'Donovan, J. R. Skinner, and E. Broadbent, "Perspectives and experiences of māori and pasifika peoples living with cardiac inherited disease: A qualitative study," *Psychology and Health*, vol. 39, no. 6, pp. 728–748, 2024.

[222] J. Gerrard, R. Benson, E. Brown, and C. Varughese, "Capturing the benefits of ai in healthcare for aotearoa new zealand-full report," Office of the Prime Minister's Chief Science Advisor, Tech. Rep., 2023. [Online]. Available: https://www.dpmc.govt.nz/sites/default/files/2024-01/PMCSA-23-12-03-V3-PMCSA-AI-healthcare-LONG-REPORT-FINAL-%28pdf-version%29-v3.pdf.

[223] C. Baggott et al., "Patient preferences for asthma management: A qualitative study," *BMJ Open*, vol. 10, no. 8, e037491, 2020. DOI: 10.1136/bmjopen-2020-037491.

[224] B. Saini et al., "Asthma disease management-australian pharmacists' interventions improve patients' asthma knowledge and this is sustained," *Patient Education and Counceling*, vol. 83, no. 3, pp. 295–302, 2011. DOI: 10.1016/j.pec.2011.05.001.

[225] S. De Peuter, V. Lemaigre, I. Van Diest, and O. Van den Bergh, "Illness-specific catastrophic thinking and overperception in asthma," *Health Psychol*, vol. 27, no. 1, pp. 93–9, 2008. DOI: 10.1037/0278-6133.27.1.93.

[226] H. S. J. Chew and P. Achananuparp, "Perceptions and needs of artificial intelligence in health care to increase adoption: Scoping review," *Journal of Medical Internet Research*, vol. 24, no. 1, e32939, 2022.

[227] M. A. Teariki and E. Leau, "Understanding pacific worldviews: Principles and connections for research," *Kōtuitui: New Zealand Journal of Social Sciences Online*, vol. 19, no. 2, pp. 132–151, 2024. DOI: 10.1080/1177083X.2023.2292268.

[228] N. Rane, S. Choudhary, and J. Rane, "Towards autonomous healthcare: Integrating artificial intelligence (ai) for personalised medicine and disease prediction," *Available at SSRN 4637894*, 2023. DOI: 10.2139/ssrn.4637894.

[229] M.-H. Chen, G. Lee, and L.-P. Hung, "Ai-driven data analysis for asthma risk prediction," in *Healthcare*, vol. 13, MDPI, 2025, p. 774. DOI: 10.3390/healthcare13070774.

[230] M. Z. Alam et al., "Predicting pulmonary function from the analysis of voice: A machine learning approach," *medRxiv*, 2021. DOI: 10.3389/fdgth.2022.750226.

[231] A. Hamza, Y. Himeur, A. Amira, and A. Oulefki, "Predicting asthma attacks through ai-powered thermal imaging analysis of breathing patterns," in *2024 IEEE 10th International Conference on Big Data Computing Service and Machine Learning Applications (BigDataService)*, 2024, pp. 82–86. DOI: 10.1109/BigDataService62917.2024.00018.

[232] R. H. Abiyev and M. K. S. Ma'aitah, "Deep convolutional neural networks for chest diseases detection," *Journal of Healthcare Engineering*, vol. 2018, p. 4 168 538, 2018. DOI: 10.1155/2018/4168538.

[233] S. Alyas, R. Hussain, B. F. Ababneh, S. C. Ong, and Z.-U.-D. Babar, "Knowledge, perceptions, facilitators, and barriers towards asthma self-management among patients: A systematic review of the literature," *Exploratory Research in Clinical and Social Pharmacy*, p. 100 558, 2024. DOI: 10.1016/j.rcsop.2024.100558.

[234] N. Kalra, P. Verma, and S. Verma, "Advancements in ai based healthcare techniques with focus on diagnostic techniques," *Computers in Biology and Medicine*, vol. 179, p. 108 917, 2024. DOI: 10.1016/j.compbiomed.2024.108917.

[235] A. Bohr and K. Memarzadeh, "The rise of artificial intelligence in healthcare applications," in *Artificial Intelligence in healthcare*. Elsevier, 2020, pp. 25–60.

[236] Z. Zhang, Y. Genc, D. Wang, M. E. Ahsen, and X. Fan, "Effect of ai explanations on human perceptions of patient-facing ai-powered healthcare systems," *Journal of Medical Systems*, vol. 45, no. 6, p. 64, 2021. DOI: 0.1007/s10916-021-01743-6.

[237] R. Whittaker et al., "An example of governance for ai in health services from aotearoa new zealand," *NPJ Digital Medicine*, vol. 6, no. 1, p. 164, 2023. DOI: 10.1038/s41746-023-00882-z.

[238] A. K. C. Wong, J. H. T. Lee, Y. Zhao, Q. Lu, S. Yang, and V. C. C. Hui, "Exploring older adults' perspectives and acceptance of ai-driven health technologies: Qualitative study," *JMIR Aging*, vol. 8, e66778, 2025. DOI: 10.2196/66778.

[239] A. Kouri, S. Gupta, S. E. Straus, and J. E. M. Sale, "Exploring the perspectives and experiences of older adults with asthma and chronic obstructive pulmonary disease toward mobile health: Qualitative study," *Journal of Medical Internet Research*, vol. 25, e45955, 2023. DOI: 10.2196/45955.

[240] A. Matenga-Ikihele et al., "Respiratory research with maori and pacific children living in aotearoa, new zealand: A systematic review and narrative synthesis," *Journal of Primary Health Care*, vol. 17, no. 1, pp. 63–82, 2025. DOI: 10.1071/HC23066.

[241] R. Dobson and R. Whittaker, "What do health service users think about the use of their data for ai development?" *Student Health Technology Information*, vol. 310, pp. 1156–1160, 2024. DOI: 10.3233/SHTI231146.

[242] B. L. Green, A. Murphy, and E. Robinson, "Accelerating health disparities research with artificial intelligence," *Front Digit Health*, vol. 6, p. 1 330 160, 2024. DOI: 10.3389/fdgth.2024.1330160.

[243] M. Chustecki, "Benefits and risks of ai in health care: Narrative review," *Interactive Journal of Medical Research*, vol. 13, e53616, 2024. DOI: 10.2196/53616.

[244] L. Jones, K. Jacklin, and M. E. O'Connell, "Development and use of health-related technologies in indigenous communities: Critical review," *Journal of Medical Internet Research*, vol. 19, no. 7, e256, 2017. DOI: 10.2196/jmir.7520.

[245] Moana Research, "Pacific data sovereignty network- consultation document," Report, 2021. [Online]. Available: https://moanaconnect.co.nz/wp-content/uploads/2021/03/PDS-consultation-document.pdf.

[246] C. Y. Hui et al., "Patient and public involvement workshop to shape artificial intelligence-supported connected asthma self-management research," *PLOS Digital Health*, vol. 3, no. 5, e0000521, 2024. DOI: 10.1371/journal.pdig.0000521.

[247] S. Wilson et al., "Recommendations to advance digital health equity: A systematic review of qualitative studies," *NPJ Digital Medicine*, vol. 7, no. 1, p. 173, 2024. DOI: 10.1038/s41746-024-01177-7.

[248] A. Ueno, C. Dennis, and G. A. Dafoulas, "Digital exclusion and relative digital deprivation: Exploring factors and moderators of internet non-use in the uk," *Technological Forecasting and Social Change*, vol. 197, p. 122 935, 2023, ISSN: 0040-1625. DOI: 10.1016/j.techfore.2023.122935.

[249] A. A. A. Matenga-Ikihele Te Whanau et al., "Navigating digital inclusion and the digital va among niue mamatua through the provision of mobile phones during covid-19," *AlterNative (Nga Pae Maramatanga (Organ))*, vol. 19, no. 1, pp. 145–154, 2023. DOI: 10.1177/11771801221148343.

[250] Department of Internal Affairs, "The digital inclusion blueprint, te mahere mō te whakaurunga matihiko," Department of Internal Affairs, NZ, Report, 2019. [Online]. Available: https://www.digital.govt.nz/assets/Documents/113Digital-Inclusion-BlueprintTe-Mahere-mo-te-Whakaurunga-Matihiko.pdf.

[251] S. A. Saeed and R. M. Masters, "Disparities in health care and the digital divide," *Current Psychiatry Reports*, vol. 23, no. 9, p. 61, 2021. DOI: 10.1007/s11920-021-01274-4.

[252] V. M. McDonald, S. A. Hiles, K. A. Jones, V. L. Clark, and J. Yorke, "Health-related quality of life burden in severe asthma," *The Medical Journal of Australia*, vol. 209, no. S2, S28–S33, 2018. DOI: 10.5694/mja18.00207.

[253] S. Stanescu, S. E. Kirby, M. Thomas, L. Yardley, and B. Ainsworth, "A systematic review of psychological, physical health factors, and quality of life in adult asthma," *NPJ Primary Care Respiratory Medicine*, vol. 29, no. 1, p. 37, 2019. DOI: 10.1038/s41533-019-0149-3.

[254] Y. Xie, K. P. Fadahunsi, C. Kelleher, D. M. Tarn, A. Grace, and O. D. J, "Towards an inclusive digital health ecosystem," *Bull World Health Organ*, vol. 103, no. 2, pp. 170–173, 2025. DOI: 10.2471/BLT.24.292020.

[255] T. Greenhalgh et al., "Beyond adoption: A new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies," *Journal of Medical Internet Research*, vol. 19, no. 11, e8775, 2017. DOI: 10.2196/jmir.8775.

Appendix A

Risk of Bias Assessment and Asthma Attack Prediction Studies

Table A.1 Results of risk of bias assessment.

Study	Rule defined adequately[a]	Representativeness of cases[b]	Selection of controls[c]	Methods & validation explained[d]	Model reliable[e]
[65]	✓	✗	✗	✓	✓
[55]	✓	✓	✓	✓	✓
[72]	✓	✓	✗	✓	✓
[66]	✓	✓	✓	✓	✓
[56]	✓	!	✗	✗	!
[57]	✓	✓	✗	✓	✓
[67]	✓	✗	✗	✓	!
[60]	✓	✓	✓	✓	✓
[61]	✓	✓	✓	✓	✓
[64]	✓	✓	✓	✓	✓
[58]	✓	✓	✗	✓	✓
[59]	✓	✓	✓	✓	✓
[62]	✓	!	✗	✓	✓
[69]	✓	✓	✗	✓	✓
[68]	✓	✓	✗	✓	✓
[63]	✓	✓	✓	✓	!
[47]	!	✓	✓	!	!
[70]	✓	!	✓	!	!
[71]	✓	✓	✓	✓	✓
[54]	✓	✓	✗	✓	✓

✓= High quality; != Some concerns; ✗= Low quality

[a] Has the prediction rule been clearly defined?

[b] Was an appropriate spectrum of patients selected for model training?

[c] Was the model validated in a different group of patients?

[d] Were the methods and techniques to construct and validate the models clearly explained?

[e] Does the model perform well?

Table A.2 Summary of the studies on predicting risk of asthma attacks as a classification - with prediction window.

Study	Study aim	Important predictors identified	Feature selection	Data balancing	Target	ML algorithms	Prediction window	Lookback window	Outcomes/Findings
[55]	Develop and validate prediction models for short-term prediction of severe asthma exacerbations	Not given	None	None	Occurrence of a severe asthma exacerbation (0 or 1)	XGB, SVM, LR	2 days	5 days	ML models reached higher discriminative performance than a simple clinical rule
[56]	Predict asthma exacerbation one day in advance based on the previous 7-day window	Not given	Not given	Not given	Asthma exacerbation ("no–alert" or "high–alert")	CART	1 day	7 days	Developed a model that predicts asthma exacerbation on next day based on the previous 7-day window
[57]	Explore telemonitoring data for ML prediction of asthma exacerbations	Zone 7	MDL	Not clear	Asthma exacerbation ("no-alert" or "high-alert")	NB, ABN, SVM	1 day	7 days	Demonstrated the potential of ML techniques in personalised decision support, found the attributes of day seven alone had more predictive power
[[60]	Develop decision support tool for monthly prediction in children	obesity, atopy, medication, controller plan, service history	None	None	Asthma exacerbation (0 or 1)	LASSO, RF, XGB	30 days	30 days	Near-term prediction models performed better, and clinical factors alone had more predictive power
[61]	Predict asthma exacerbation after stopping biologics using ML	previous exacerbations, biologic treatment duration	None	None	Asthma exacerbation (0 or 1)	Elastic-net LR, RF, GBM	6 months	6 months	ML algorithms performed moderately in predicting asthma attacks after stopping asthma biologics
[62]	Explain asthma prediction model on imbalanced data	Not given	None	None	Asthma hospital visit (0 or 1)	XGB	1 year	Not given	Asthma risk prediction model with automatic explanation component
[58]	Predict risk of asthma exacerbations using ML models	comorbidity burden, past exacerbations	Collinearity	RUS, weighting	Asthma exacerbation (moderate or severe)	RF, XGB, RNN, LR (LASSO, Ridge, Elastic-net)	15 days	Multiple (10–365 days)	Prior clinical information is not sufficient for near-term asthma risk prediction
[59]	Use digital inhaler and clinical data to predict exacerbations	daily inhalations, baseline comparison, past exacerbations	None	None	Asthma exacerbation (0 or 1)	LR, RF, XGB	5 days	Not given	Digital inhaler data with demographic and clinical factors can be used for predicting impending asthma exacerbations
[63]	Predict hospital re-visit in pediatric asthma patients	age, stay duration, blight, neighborhood inequality	None	Under-sampling	Second hospital visit (0 or 1)	RF, SVM	1 year	1 year	Socio-markers can be reliable predictors of hospital re-visits and combined demographics, socio-markers and biomarkers can predict future hospital re-visits moderate accurately
[64]	Predict 1-year asthma exacerbations using electronic records	Asthma medications	None	None	Asthma exacerbation (0 or 1)	Elastic-net LR, RF, XGB	1 year	Not given	Inclusion of medication-related covariates increase the odds of having as asthma exacerbation
[54]	Predict future exacerbations using telemonitoring data	Not given	PCA, RFE	ROS, RUS, SMOTE	Asthma exacerbation (0 or 1)	LR, NB, DT, NN	4 days	None	ML models combined with telemonitoring data such as PEFR

Table A.3 Summary of the studies on predicting risk of asthma attacks as a classification - without prediction window

Study	Study aim	Important predictors identified	Feature selection technique	Data imbalance handling	Target	ML algorithms	Outcomes/Findings
[65]	Utilising bio-signals and environmental triggers for asthma risk prediction	smoking status, allergies, normal PEFR and reading, asthma symptoms (last 4 weeks), medication (last 4 weeks), no. of asthma attacks (last 4 weeks), temperature, wind speed, relative humidity, AQI	LASSO Regression	SMOTE using SVM (SMOTESVM)	Asthma attack (0 or 1)	DT, GBM, LR, RF, SVM	DT has the poorest performance while GBM has the best performance
[66]	To evaluate models for predicting the severity of asthma exacerbation in ED visits and compare with PRAM and physicians' predictions	Not identified	Not conducted	N/A (data not highly imbalanced)	Asthma exacerbation (Mild or Moderate/Severe)	NB, DT, E-DT, SVM, IB1, IB10	NB performed better among ML models, but both PRAM and NB were less accurate than physicians
[67]	Exploring the impact of weather on asthma exacerbations and tools for self-management	Not identified	Not conducted	None	Asthma attack (low or high)	NN	Personalisation and weather had positive effects on asthma self-management
[68]	Determine severity of asthma using real-time data from medical IoT devices	Not identified	Not conducted	None	Risk of asthma (low, medium, high)	BPNN, SVM, KNN	BPNN achieved higher accuracy than other classifiers
[69]	Compare ML models with community viral load for early prediction of hospital-level care in pediatric asthma	patient vital signs, acuity, age, weight, SES, temperature, humidity, wind speed, pressure	None	None	Hospitalisation (0 or 1)	DT, LASSO, RF, XGB	Except DT, all models perform well; adding SES and weather improved performance
[70]	Validate hypothesis that RF-selected SNPs improve asthma exacerbation prediction	set of SNPs	RF	None	Severe asthma exacerbation (0 or 1)	RF	RF modelling with selected SNPs was accurate; random SNPs were not informative
[71]	Develop ML models using real-world data to predict asthma exacerbations	age, long-acting agonist, inhaled or oral glucocorticoids, low FEV1 and FEV1/FVC	None	None	Non-severe exacerbation (0 or 1); ED visit (0 or 1); Hospitalisation (0 or 1)	LR, RF, LightGBM	Higher albumin linked to lower risk; FEV1/FVC affected individual risk prediction

Table A.4 Summary of the studies on predicting risk of asthma attacks as a probability

Study	Study aim	Important predictors	Feature selection technique/s	Data balancing	Target	ML algorithm	Outcomes/Findings
[72]	Develop an asthma risk prediction model by investigating the ability of Q-learning method	None	None	None	Relative risk (numeric)	LSTM	DQN framework for predicting asthma risk with the use of personal risk scores of triggers
[47]	Predict the risk of asthma exacerbations and explore potential risk factors	Gender, race, some diagnosis codes and medications	Attention weights of the model	None	Asthma exacerbation (probability)	TSANN	A TSANN model employing self-attention on both code-level and visit-level and addition of elapsed time improved model performance

Appendix B

PRISMA Checklist for Conducting Systematic Literature Review

Section and Topic	Item #	Checklist item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	Title
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	Abstract
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	Pages 1,2
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	Page 2
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	Page 3
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	Page 3
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	Page 3
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	Pages 3, 4
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	Page 4
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	Page 5
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	Page 5
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	Page 4
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	Page 4
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	Page 4
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	N/A
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	Pages 5-9
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	Page 4
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	N/A
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	N/A
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	Page 4
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	N/A

Section and Topic	Item #	Checklist item	Location where item is reported
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Page 3
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	Page 3
Study characteristics	17	Cite each included study and present its characteristics.	Page 5
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	Supplementary material
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	Tables 1-7
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	Pages 6-10
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	N/A
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	N/A
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	N/A
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	Not assessed
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	N/A
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Pages 10-13
	23b	Discuss any limitations of the evidence included in the review.	Pages 10-13
	23c	Discuss any limitations of the review processes used.	Page 13
	23d	Discuss implications of the results for practice, policy, and future research.	Page 10-13
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	Page 3
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	Page 3
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	N/A
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	Page 18
Competing interests	26	Declare any competing interests of review authors.	Page 18
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	No analytic code was generated for this review. Data extracted are available from the corresponding author upon reasonable request.

Page numbers refer to the published manuscript, not the thesis.

Appendix C

Asthma Attack Dataset Description

- Hospital and pharmaceutical datasets were collected from national datasets of the New Zealand Ministry of Health.
- These datasets comprise publicly funded hospital admissions in New Zealand, including acute admissions through emergency departments, as well as data on the national dispensing of subsidised medications.
- The ethics approval for data collection was obtained from the Auckland Health Research Ethics Committee (Appendix D).
- The data used for analysis spanned from January 1, 2008, to December 31, 2017 and consisted of 1,566,349 instances with 53 features.
- A variable named patient-quarter was identified to predict impending asthma attacks in the next 3 months. This represents a 3-month period for each patient.
- As some year-based features were included in the feature set, we extracted data for the fifth patient-quarter, which includes data from the previous year.
- There were 355,113 patient records with 53 features, including the target variable asthma attack, in the fifth patient-quarter dataset used to develop asthma prediction models.
- The Table C.1 shows a detailed description of the features of the dataset used in the thesis for developing asthma attack risk prediction models.

Table C.1 Description of asthma attack dataset.

Feature	Description	Data type	Values
ENHI	Encrypted NHI: used to identify different patients anonymously	String	
Q5_WeeksDuring Winter	Flag for number of weeks during winter season in each patient's Quarter 5; the three months could occur at any time of year	Numeric	1-13
Age	Age of patient in years	Numeric	Positive integer
Gender	Gender of patient	Categorical	F, M
EthnicGroup	Major/prioritised ethnic group of patient	Categorical	NZ European, Māori, Pacific, Asian, Other, Unknown
DeprivationQuintile	Neighbourhood deprivation level	Categorical	1 - 5
DHB	District Health Board	Categorical	List of DHB names (n=22)
Cardiovascular Cerebrovascular Disease	ICD 10AM: I00-I99; occurred within last year	Binary	1=Yes, 0=No
IschaemicHeartDisease	ICD 10AM: I20-I25; occurred within last year	Binary	1=Yes, 0=No
NasalPolyps	ICD 10AM: J33; occurred within last year	Binary	1=Yes, 0=No
Anaphylaxis	ICD 10AM: T78.2, T80.5; occurred within last year	Binary	1=Yes, 0=No
PulmonaryEosinophilia	ICD-10: J82; occurred within last year	Binary	1=Yes, 0=No
AtopicDermatitis	ICD-10AM: L20; occurred within last year	Binary	1=Yes, 0=No
Psoriasis	ICD10AM: L40; occurred within last year	Binary	1=Yes, 0=No
Obesity	ICD 10AM: E66; occurred within last year	Binary	1=Yes, 0=No
LRTI	Any of the following occurred in the previous year: ICD10-AM: J12-18 Pneumonia; J20-J22 Other acute lower respiratory infections; J40-J47 Chronic lower respiratory diseases	Binary	1=Yes, 0=No
RheumatologicalDisease	Diagnosis for rheumatoid arthritis (ICD-10: M05-M06, M35.3, M79.0) occurred within last year	Binary	1=Yes, 0=No
AnxietyDepression	Occurrence of depression (ICD-10-AM: F32, F33 or chemical IDs) or anxiety (ICD-10-AM: F40-F41 or meds) in the last year	Binary	1=Yes, 0=No
DementiaAlzheimers	Occurrence of ICD 10AM: F00-F03, G30, G31.0, G31.8 or medications for dementia (ChemID=3923, 3750) in the last year	Binary	1=Yes, 0=No

Feature	Description	Data type	Values
HeartFailure	Occurrence of ICD 10AM codes I110, I130, I132, I50; or dispensing of loop diuretics (ChemID=1544, 1171) or metolazone (ChemID=4006)	Binary	1=Yes, 0=No
Diabetes	Occurrence of ICD-10-AM: E10-E14 or dispensing of diabetes medications (oral hypoglycaemics, acarbose, insulin)	Binary	1=Yes, 0=No
Rhinitis	ICD 10AM: J30-J31 or spray formulations (fluticasone, beclometasone, budesonide) within last year	Binary	1=Yes, 0=No
SmokingStatus	Smoking status	Binary	1=Yes, 0=No
CharlsonComorbidityScore	Score generated from weighted values on comorbidities	Numeric	0 - 12
NumberOfMetforminRx	Use of Metformin in prior 12 months; evidence of association with asthma	Numeric	0 or Positive integer
Diabetes_NoMetformin	Diabetes patients not prescribed Metformin in prior 12 months	Binary	1=Yes, 0=No
NoOfHospitalisations	Number of hospitalisations during last 3 months	Numeric	0 or Positive integer
NoOfOPEDVisits	Number of outpatient or ED visits during last 3 months	Numeric	0 or Positive integer
BetaBlockers	$\geq$ 1 pharmacy claim for beta-blockers in previous 6 months	Binary	1=Yes, 0=No
Paracetamol	$\geq$ 1 pharmacy claim for paracetamol in previous 6 months	Binary	1=Yes, 0=No
NSAIDs	$\geq$ 1 pharmacy claim for NSAIDs or other analgesics in previous 6 months	Binary	1=Yes, 0=No
NebulisedSABA	Prescriptions for Salbutamol or Terbutaline nebulisers	Binary	1=Yes, 0=No
NoOfSABAInhalers	Number of SABA inhalers 6 months before index date	Numeric	0 or Positive integer
NoOfAsthmaControllerMeds	Any of inhaled corticosteroids or LABA medications in last 6 months	Numeric	0 or Positive integer
NoOfICSInhalers	Number of ICS inhalers dispensed in previous year	Numeric	0 or Positive integer
AsthmaSeverityStep	Generated based on 2+ controller meds per year: SABA, ICS, LABA, LTRA, Biologics	Numeric	0 - 3
SABA_ICS_Ratio	Ratio of SABA inhalers to ICS inhalers over previous year, adjusted for combination inhalers	Numeric (decimal)	0.0 to 1.0

Feature	Description	Data type	Values
P12MNoOfAsthAttacks	Number of asthma attacks in last 12 months	Numeric	0 or Positive integer
P6MNoOfAsthAttacks	Number of asthma attacks in last 6 months	Numeric	0 or Positive integer
P3MNoOfAsthAttacks	Number of asthma attacks in last 3 months	Numeric	0 or Positive integer
CohortYear	Year of index date; Date the patient entered the cohort	Numeric	2008 - 2017
IndexDate	Entry date of the patient into the data collection process	Date	
ObservationEndDate	Exit date of the patient from the data collection process	Date	
AsthmaHospitalisation	Hospitalisation due to asthma	Binary	1 = Yes, 0 = No
OralCorticosteroid Prescriptions	Prescription of oral corticosteroids to the patient	Binary	1 = Yes, 0 = No
P12MAsthAttack	Asthma attack in the last 12 months	Binary	1 = Yes, 0 = No
P6MAsthAttack	Asthma attack in the last 6 months	Binary	1 = Yes, 0 = No
P3MAsthAttack	Asthma attack in the last 3 months	Binary	1 = Yes, 0 = No
DateOfBirth	Date of birth of the patient	Date	
EthnicityCode	Unique code representing the patient's ethnicity	Categorical	
DomicileCode	Unique code representing the patient's domicile	Categorical	Unique domicile codes (n = 1500)
DHB_Code	Unique code representing the District Health Board	Categorical	Unique DHB codes (n = 22)
AsthmaAttack	Presence or absence of an asthma attack in the next patient-quarter (next 3 months) (Target variable)	Binary	1=Yes, 0=No

Appendix D

Asthma Attack Risk Prediction Using ML - Ethics Approval

AUCKLAND HEALTH RESEARCH ETHICS COMMITTEE (AHREC)

04/08/2021

Dr Amy Chan

Pharmacy

Re: Application for Ethics Approval (Our Ref. AH23069): Approved

The Committee considered your application for ethics approval for the study entitled "**Predicting acute asthma events using machine learning**".

We are pleased to inform you that ethics approval has been granted.

The expiry date for this approval is **04/08/2024**.

Amendments to the approved project: Should you need to make any changes to the approved project, please follow the steps below:

- Send a request to the AHREC Administrators to unlock the application form (using the Notification tab in the Ethics RM form).
- Make all changes to the relevant sections of the application form and attach revised documents (as appropriate).
- Change the Application Type to "Amendment request" in Section L.
- Add a summary of the changes requested in the text box.
- Submit the amendment request (PI/Supervisors only to submit the form).

If the project changes significantly, you are required to submit a new application.

Funded projects: If you received funding for this project, please provide this approval letter to your local Faculty Research Project Coordinator (RPC) or Research Project Manager (RPM) so that the approval can be notified via a Service Request to the Research Operations Centre (ROC) for activation of the grant.

The Chair and the members of AHREC would be happy to discuss general matters relating to ethics approvals. If you wish to do so, please contact the AHREC Ethics Administrators at ahrec@auckland.ac.nz in the first instance.

Additional information:

- Do not forget to fill in the 'approval wording' on the PISs, CFs and/or advertisements, using the date of this approval and the reference number, before you use the documents or send them out to your participants.

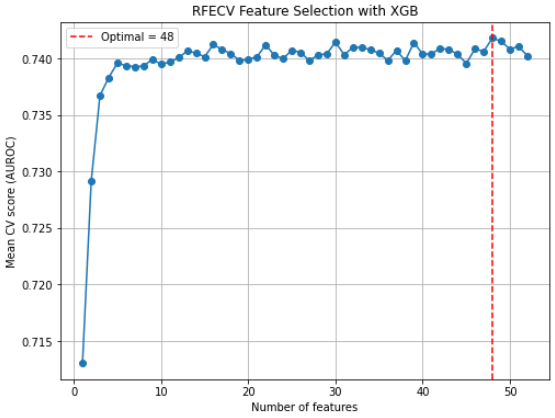
All communications with the AHREC regarding this application should indicate this reference number: **AH23069**.

AHREC Administrators

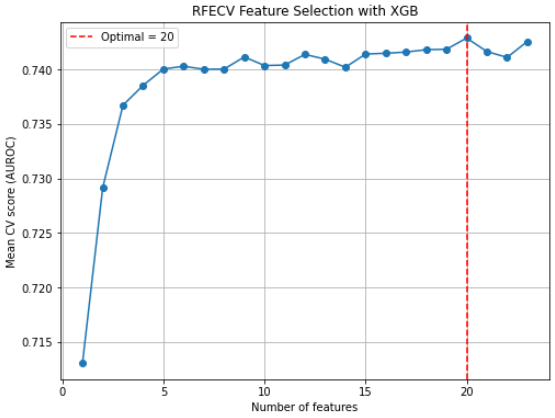
Auckland Health Research Ethics Committee

Appendix E

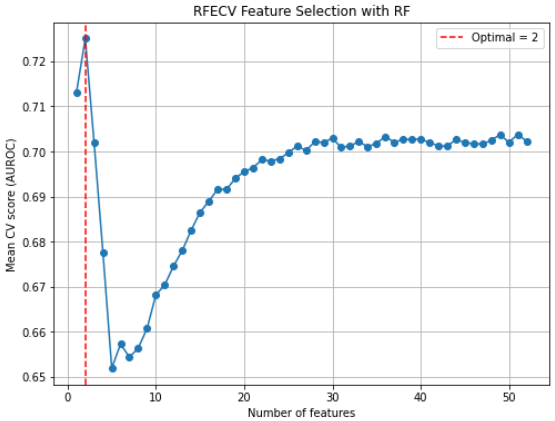
Feature Selection and Model Performance



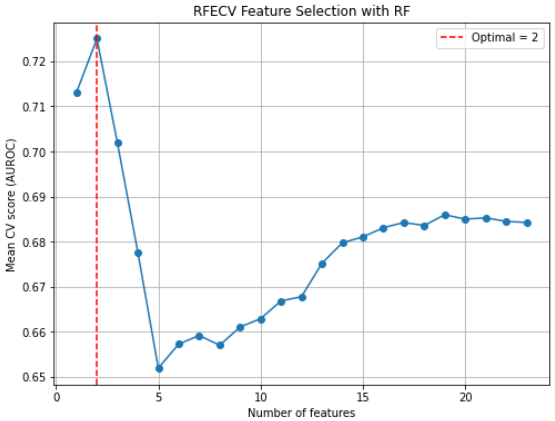
(a) RFECV using the XGB classifier on FS2. The optimal value is shown as 48 due to one-hot encoding, although only 23 distinct features are present.



(b) RFECV using the XGB classifier on FS4. The optimal value is shown as 20 due to one-hot encoding, although only 14 distinct features are present.



(c) RFECV using the RF classifier on refined full feature set (FS2).



(d) RFECV using the RF classifier on refined full feature set (FS4).

Fig. E.1 Mean AUROC values of feature sets during RFECV-based feature selection.

Table E.1 Population Stability Index (PSI) calculated to quantify the dataset shift between the test and validation cohorts.

Feature	PSI	Shift
Q5_WeeksDuringWinter	0.010698906279192382	No Shift
Age	0.026695889550533092	No Shift
Gender	4.850298672968513e-07	No Shift
EthnicGroup	0.007074	No Shift
DeprivationQuintile	0.000276	No Shift
DHB	0.010000207617798754	No Shift
Psoriasis	1.6810431545050762e-05	No Shift
Obesity	0.000706	No Shift
LRTI	0.008289	No Shift
AnxietyDepression	0.011774	No Shift
Diabetes	0.002716	No Shift
SmokingStatus	5.295783496472005e-05	No Shift
CharlsonComorbidityIndexScore	0	No Shift
NumberOfMetforminRx	1.3377286652305542e-11	No Shift
Diabetes_NoMetformin	0.004875	No Shift
NoOfHospitalisations	0.005555	No Shift
NoOfOPEDVisits	0.006818	No Shift
BetaBlockers	0.001506	No Shift
Paracetamol	0.005565	No Shift
NSAIDs	0.001062	No Shift
SABA_ICS_Ratio	1.8396067335412365	Significant Shift
P12MNoOfAsthAttacks	0.12315230139869743	Moderate Shift
HistoryOfAllergies	0.26665093912832427	Significant Shift
CardiovascularDiseases	0.008471	No Shift

Table E.2 Hyperparameter values used for developing the prediction models.

ML Model	Hyperparamenter	Value
LR	Penalty	L2
	Regularisation Strength (C)	1.0
	Solver	lbfgs
	Maximum Iterations	400
	Tolerance	0.0001
RF	Number of Trees	100
	Split Criterion	Gini
	Maximum Tree Depth	None (unlimited)
	Maximum Features per Split	sqrt
	Minimum Samples per Split	2
	Minimum Samples per Leaf	1
	Bootstrap Sampling	True
	Out-of-Bag Evaluation	False
XGB	Objective	binary:logistic
	Number of trees	100
	Learning rate	0.30
	Maximum tree depth	6
	Minimum child weight	1
	Gamma	0
	Subsample ratio	1.0
	Column sampling by tree	1.0
	L1 regularisation (α)	0
	L2 regularisation (λ)	1

Table E.3 Performance of the models in 5-fold CV and using test dataset.

Feature Set	Feature Set Description	ML Algorithm	Imbalance Handling Technique	Mean AUROC (95% CI) - CV	AUROC (Test set)	Mean F1-Score (95% CI) - CV	F1-score (Test set)	AUPRC (Test set)	Brier Score (Test set)
FS1	XGB-based importance 10 features	LR	ENN	0.75 (0.74-0.76)	0.75	0.28 (0.27-0.29)	0.28	0.29	0.06
FS1	XGB-based importance 10 features	LR	KMeansSMOTE	0.74 (0.73-0.76)	0.74	0.29 (0.27-0.31)	0.30	0.27	0.09
FS1	XGB-based importance 10 features	LR	No-sample	0.75 (0.74-0.75)	0.75	0.16 (0.15-0.17)	0.16	0.28	0.06
FS1	XGB-based importance 10 features	LR	RUS	0.75 (0.74-0.75)	0.75	0.28 (0.28-0.29)	0.28	0.28	0.19
FS1	XGB-based importance 10 features	LR	SMOTE	0.75 (0.74-0.75)	0.75	0.28 (0.27-0.29)	0.28	0.28	0.19
FS1	XGB-based importance 10 features	RF	ENN	0.69 (0.69-0.69)	0.69	0.23 (0.22-0.23)	0.23	0.16	0.10
FS1	XGB-based importance 10 features	RF	KMeansSMOTE	0.69 (0.69-0.69)	0.69	0.18 (0.17-0.19)	0.21	0.20	0.07
FS1	XGB-based importance 10 features	RF	No-sample	0.69 (0.68-0.69)	0.69	0.16 (0.15-0.16)	0.17	0.21	0.07
FS1	XGB-based importance 10 features	RF	RUS	0.71 (0.70-0.71)	0.71	0.21 (0.21-0.22)	0.22	0.21	0.22
FS1	XGB-based importance 10 features	RF	SMOTE	0.65 (0.65-0.66)	0.66	0.22 (0.22-0.23)	0.20	0.17	0.08
FS1	XGB-based importance 10 features	XGB	ENN	0.75 (0.74-0.75)	0.75	0.28 (0.27-0.30)	0.28	0.29	0.06
FS1	XGB-based importance 10 features	XGB	KMeansSMOTE	0.75 (0.74-0.75)	0.74	0.18 (0.16-0.20)	0.21	0.26	0.06
FS1	XGB-based importance 10 features	XGB	No-sample	0.75 (0.74-0.75)	0.75	0.16 (0.16-0.17)	0.16	0.29	0.06
FS1	XGB-based importance 10 features	XGB	RUS	0.74 (0.73-0.75)	0.74	0.25 (0.25-0.26)	0.25	0.27	0.20
FS1	XGB-based importance 10 features	XGB	SMOTE	0.67 (0.66-0.67)	0.67	0.21 (0.20-0.22)	0.19	0.24	0.07
FS2	Refined fullset 24 features	LR	ENN	0.74 (0.74-0.75)	0.74	0.27 (0.26-0.28)	0.27	0.28	0.06
FS2	Refined fullset 24 features	LR	KMeansSMOTE	0.70 (0.67-0.73)	0.73	0.28 (0.25-0.31)	0.30	0.27	0.08
FS2	Refined fullset 24 features	LR	No-sample	0.74 (0.74-0.75)	0.73	0.15 (0.14-0.16)	0.15	0.27	0.06
FS2	Refined fullset 24 features	LR	RUS	0.74 (0.74-0.75)	0.74	0.28 (0.27-0.28)	0.27	0.27	0.19

Feature Set	Feature Set Description	ML Algorithm	Imbalance Handling Technique	Mean AUROC (95% CI) - CV	AUROC (Test set)	Mean F1 Score (95% CI) - CV	F1-score (Test set)	AUPRC (Test set)	Brier Score (Test set)
FS2	Refined fullset 24 features	LR	SMOTE	0.74 (0.73-0.75)	0.73	0.28 (0.27-0.28)	0.27	0.27	0.19
FS2	Refined fullset24 features	RF	ENN	0.71 (0.70-0.71)	0.71	0.24 (0.23-0.26)	0.24	0.21	0.07
FS2	Refined fullset 24 features	RF	KMeansSMOTE	0.70 (0.70-0.71)	0.70	0.14 (0.11-0.16)	0.17	0.20	0.06
FS2	Refined fullset 24 features	RF	No-sample	0.70 (0.69-0.71)	0.70	0.12 (0.11-0.13)	0.12	0.20	0.06
FS2	Refined fullset 24 features	RF	RUS	0.72 (0.71-0.72)	0.72	0.23 (0.22-0.23)	0.23	0.20	0.21
FS2	Refined fullset 24 features	RF	SMOTE	0.60 (0.59-0.60)	0.70	0.22 (0.21-0.22)	0.13	0.17	0.07
FS2	Refined fullset 24 features	XGB	ENN	0.74 (0.73-0.75)	0.75	0.26 (0.25-0.27)	0.26	0.27	0.06
FS2	Refined fullset 24 features	XGB	KMeansSMOTE	0.73 (0.72-0.74)	0.74	0.17 (0.16-0.17)	0.19	0.26	0.06
FS2	Refined fullset 24 features	XGB	No-sample	0.74 (0.73-0.75)	0.74	0.14 (0.13-0.15)	0.14	0.27	0.06
FS2	Refined fullset 24 features	XGB	RUS	0.73 (0.73-0.74)	0.74	0.25 (0.24-0.25)	0.25	0.25	0.20
FS2	Refined fullset 24 features	XGB	SMOTE	0.64 (0.63-0.65)	0.71	0.23 (0.22-0.24)	0.15	0.23	0.06
FS3	RFECV-XGB 23 features	LR	ENN	0.74 (0.74-0.75)	0.74	0.27 (0.26-0.28)	0.26	0.28	0.06
FS3	RFECV-XGB 23 features	LR	KMeansSMOTE	0.70 (0.67-0.73)	0.73	0.28 (0.25-0.30)	0.30	0.27	0.08
FS3	RFECV-XGB 23 features	LR	No-sample	0.74 (0.74-0.75)	0.74	0.15 (0.14-0.16)	0.15	0.27	0.06
FS3	RFECV-XGB 23 features	LR	RUS	0.74 (0.74-0.75)	0.74	0.28 (0.27-0.28)	0.27	0.27	0.19
FS3	RFECV-XGB 23 features	LR	SMOTE	0.74 (0.73-0.75)	0.73	0.28 (0.27-0.28)	0.27	0.27	0.19
FS3	RFECV-XGB 23 features	RF	ENN	0.70 (0.70-0.71)	0.70	0.24 (0.23-0.25)	0.24	0.21	0.07
FS3	RFECV-XGB 23 features	RF	KMeansSMOTE	0.70 (0.70-0.71)	0.70	0.14 (0.11-0.16)	0.17	0.20	0.07
FS3	RFECV-XGB 23 features	RF	No-sample	0.70 (0.69-0.71)	0.70	0.11 (0.10-0.12)	0.11	0.20	0.21
FS3	RFECV-XGB 23 features	RF	RUS	0.72 (0.71-0.72)	0.72	0.23 (0.22-0.23)	0.23	0.20	0.21
FS3	RFECV-XGB 23 features	RF	SMOTE	0.60 (0.59-0.60)	0.70	0.22 (0.21-0.22)	0.13	0.17	0.07
FS3	RFECV-XGB 23 features	XGB	ENN	0.74 (0.73-0.75)	0.74	0.26 (0.25-0.28)	0.26	0.27	0.06
FS3	RFECV-XGB 23 features	XGB	KMeansSMOTE	0.73 (0.72-0.74)	0.74	0.17 (0.16-0.18)	0.19	0.25	0.06
FS3	RFECV-XGB 23 features	XGB	No-sample	0.74 (0.74-0.75)	0.74	0.14 (0.13-0.16)	0.14	0.27	0.06
FS3	RFECV-XGB 23 features	XGB	RUS	0.73 (0.73-0.74)	0.74	0.25 (0.24-0.25)	0.25	0.25	0.20
FS3	RFECV-XGB 23 features	XGB	SMOTE	0.64 (0.64-0.65)	0.71	0.24 (0.23-0.24)	0.15	0.23	0.06
FS4	Clinical fullset 16 features	LR	ENN	0.74 (0.74-0.75)	0.74	0.27 (0.26-0.28)	0.27	0.28	0.06
FS4	Clinical fullset 16 features	LR	KMeansSMOTE	0.71 (0.68-0.74)	0.72	0.29 (0.27-0.32)	0.29	0.27	0.19
FS4	Clinical fullset 16 features	LR	No-sample	0.74 (0.74-0.75)	0.74	0.15 (0.14-0.16)	0.15	0.27	0.19
FS4	Clinical fullset 16 features	LR	RUS	0.74 (0.74-0.75)	0.74	0.28 (0.27-0.28)	0.28	0.27	0.19
FS4	Clinical fullset 16 features	LR	SMOTE	0.74 (0.73-0.75)	0.74	0.28 (0.27-0.28)	0.28	0.27	0.19
FS4	Clinical fullset 16 features	RF	ENN	0.68 (0.68-0.69)	0.68	0.22 (0.22-0.23)	0.22	0.23	0.07
FS4	Clinical fullset 16 features	RF	KMeansSMOTE	0.68 (0.68-0.69)	0.68	0.18 (0.15-0.20)	0.17	0.17	0.07
FS4	Clinical fullset 16 features	RF	No-sample	0.68 (0.68-0.69)	0.68	0.14 (0.13-0.15)	0.14	0.20	0.06
FS4	Clinical fullset 16 features	RF	RUS	0.70 (0.69-0.70)	0.70	0.21 (0.21-0.22)	0.21	0.21	0.21
FS4	Clinical fullset 16 features	RF	SMOTE	0.60 (0.59-0.60)	0.66	0.22 (0.21-0.22)	0.17	0.16	0.07

Feature Set	Feature Set Description	ML Algorithm	Imbalance Handling Technique	Mean AUROC (95% CI) - CV	AUROC (Test set)	Mean F1 Score (95% CI) - CV	F1-score (Test set)	AUPRC (Test set)	Brier Score (Test set)
FS4	Clinical fullset 16 features	XGB	ENN	0.74 (0.73-0.75)	0.75	0.27 (0.26-0.28)	0.26	0.27	0.06
FS4	Clinical fullset 16 features	XGB	KMeansSMOTE	0.74 (0.73-0.74)	0.74	0.19 (0.17-0.21)	0.16	0.24	0.06
FS4	Clinical fullset 16 features	XGB	No-sample	0.74 (0.73-0.75)	0.74	0.14 (0.13-0.15)	0.14	0.27	0.06
FS4	Clinical fullset 16 features	XGB	RUS	0.73 (0.73-0.74)	0.74	0.25 (0.24-0.25)	0.25	0.26	0.20
FS4	Clinical fullset 16 features	XGB	SMOTE	0.634 (0.63-0.65)	0.67	0.23 (0.22-0.24)	0.18	0.24	0.06
FS5	Clinical RFECV-XGBF 14 features	LR	ENN	0.74 (0.74-0.75)	0.74	0.27 (0.26-0.28)	0.27	0.28	0.06
FS5	Clinical RFECV-XGBF 14 features	LR	KMeansSMOTE	0.71 (0.70-0.72)	0.70	0.30 (0.29-0.32)	0.24	0.27	0.09
FS5	Clinical RFECV-XGBF 14 features	LR	No-sample	0.74 (0.74-0.75)	0.74	0.15 (0.14-0.16)	0.15	0.27	0.06
FS5	Clinical RFECV-XGBF 14 features	LR	RUS	0.74 (0.74-0.75)	0.74	0.28 (0.27-0.28)	0.28	0.27	0.19
FS5	Clinical RFECV-XGBF 14 features	LR	SMOTE	0.74 (0.73-0.74)	0.74	0.28 (0.27-0.28)	0.28	0.27	0.19
FS5	Clinical RFECV-XGBF 14 features	RF	ENN	0.68 (0.67-0.69)	0.71	0.22 (0.21-0.23)	0.25	0.14	0.10
FS5	Clinical RFECV-XGBF 14 features	RF	KMeansSMOTE	0.68 (0.67-0.69)	0.70	0.19 (0.18-0.20)	0.16	0.18	0.17
FS5	Clinical RFECV-XGBF 14 features	RF	No-sample	0.68 (0.67-0.68)	0.70	0.14 (0.13-0.15)	0.14	0.18	0.07
FS5	Clinical RFECV-XGBF 14 features	RF	RUS	0.69 (0.69-0.70)	0.71	0.21 (0.21-0.21)	0.22	0.17	0.23
FS5	Clinical RFECV-XGBF 14 features	RF	SMOTE	0.59 (0.59-0.60)	0.69	0.21 (0.21-0.22)	0.16	0.13	0.09
FS5	Clinical RFECV-XGBF 14 features	XGB	ENN	0.74 (0.74-0.75)	0.74	0.27 (0.25-0.28)	0.27	0.27	0.06
FS5	Clinical RFECV-XGBF 14 features	XGB	KMeansSMOTE	0.74 (0.73-0.75)	0.72	0.20 (0.18-0.21)	0.15	0.24	0.06
FS5	Clinical RFECV-XGBF 14 features	XGB	No-sample	0.74 (0.74-0.75)	0.74	0.14 (0.13-0.15)	0.14	0.27	0.06
FS5	Clinical RFECV-XGBF 14 features	XGB	RUS	0.73 (0.72-0.74)	0.74	0.25 (0.24-0.25)	0.25	0.25	0.20
FS5	Clinical RFECV-XGBF 14 features	XGB	SMOTE	0.66 (0.65-0.66)	0.72	0.24 (0.23-0.25)	0.16	0.20	0.07
FS6	RFECV-RF 2 features	LR	ENN	0.73 (0.72-0.74)	0.70	0.27 (0.24-0.31)	0.29	0.26	0.06

Feature Set	Feature Set Description	ML Algorithm	Imbalance Handling Technique	Mean AUROC (95% CI) - CV	AUROC (Test set)	Mean F1 Score (95% CI) - CV	F1-score (Test set)	AUPRC (Test set)	Brier Score (Test set)
FS6	RFECV-RF 2 features	LR	KMeansSMOTE	0.71 (0.68-0.73)	0.70	0.30 (0.28-0.32)	0.29	0.26	0.06
FS6	RFECV-RF 2 features	LR	No-sample	0.73 (0.72-0.74)	0.73	0.16 (0.15-0.17)	0.16	0.20	0.07
FS6	RFECV-RF 2 features	LR	RUS	0.73 (0.72-0.74)	0.73	0.27 (0.26-0.27)	0.27	0.27	0.19
FS6	RFECV-RF 2 features	LR	SMOTE	0.73 (0.73-0.74)	0.73	0.27 (0.26-0.27)	0.27	0.27	0.19
FS6	RFECV-RF 2 features	RF	ENN	0.70 (0.69-0.71)	0.55	0.23 (0.20-0.25)	0.13	0.08	0.45
FS6	RFECV-RF 2 features	RF	KMeansSMOTE	0.73 (0.72-0.73)	0.72	0.21 (0.12-0.29)	0.22	0.25	0.06
FS6	RFECV-RF 2 features	RF	No-sample	0.72 (0.72-0.73)	0.72	0.13 (0.12-0.14)	0.13	0.25	0.06
FS6	RFECV-RF 2 features	RF	RUS	0.73 (0.72-0.74)	0.73	0.26 (0.26-0.27)	0.26	0.23	0.20
FS6	RFECV-RF 2 features	RF	SMOTE	0.71 (0.70-0.72)	0.71	0.26 (0.26-0.26)	0.26	0.24	0.19
FS6	RFECV-RF 2 features	XGB	ENN	0.71 (0.70-0.72)	0.52	0.23 (0.20-0.26)	0.13	0.08	0.43
FS6	RFECV-RF 2 features	XGB	KMeansSMOTE	0.73 (0.73-0.74)	0.73	0.21 (0.12-0.30)	0.23	0.26	0.06
FS6	RFECV-RF 2 features	XGB	No-sample	0.73 (0.73-0.74)	0.73	0.12 (0.11-0.13)	0.11	0.26	0.06
FS6	RFECV-RF 2 features	XGB	RUS	0.73 (0.72-0.74)	0.73	0.26 (0.26-0.27)	0.27	0.26	0.19
FS6	RFECV-RF 2 features	XGB	SMOTE	0.73 (0.72-0.74)	0.73	0.27 (0.26-0.27)	0.27	0.26	0.19

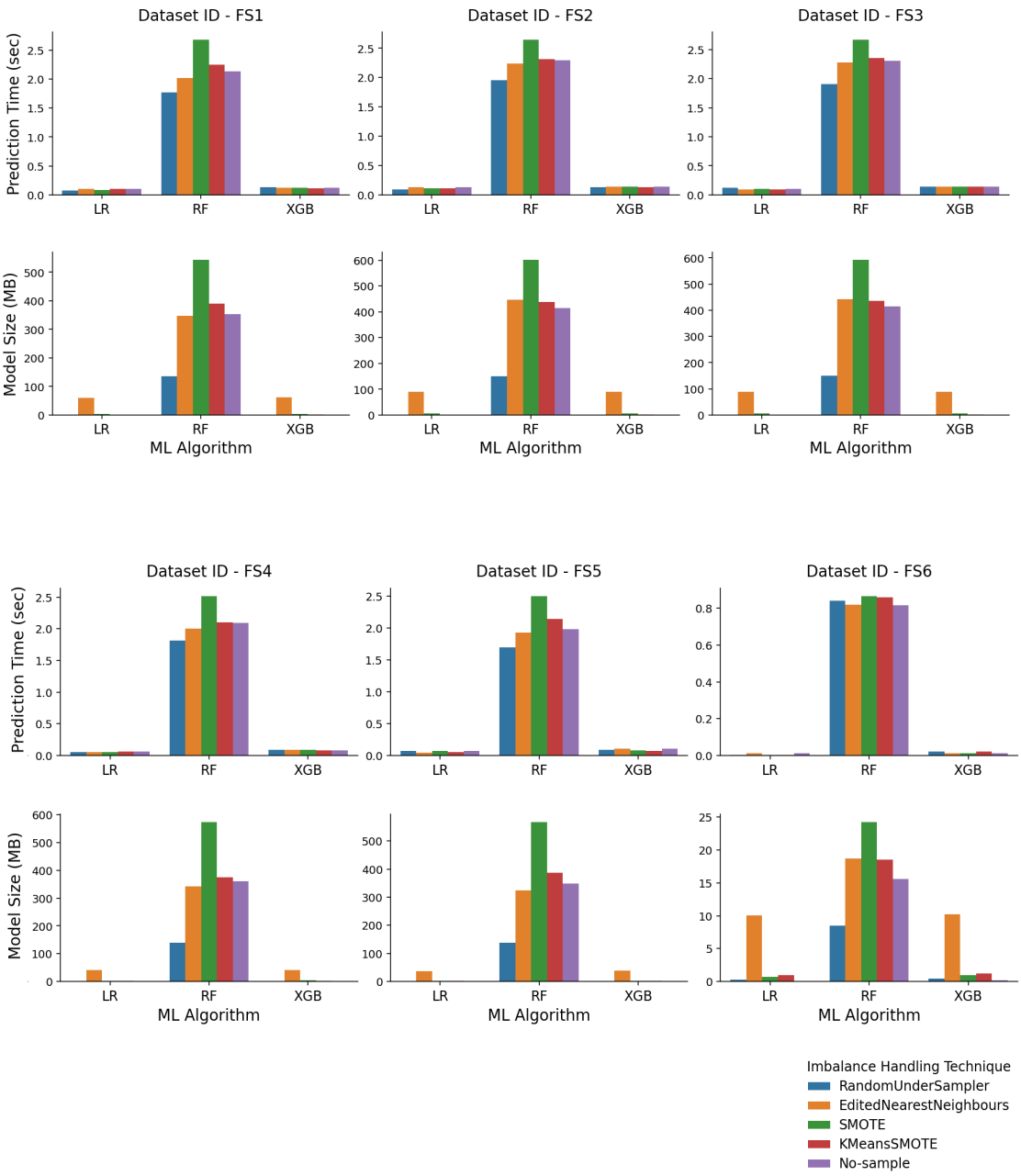


Fig. E.2 Comparison of the efficiency of the models based on different feature sets and data imbalance handling techniques.

Table E.4 Psoriasis vs asthma attacks in the training set.

Psoriasis	Asthma Attack=0	Asthma Attack=1	Total
0	207,016	16,107	223,123
1	30	5	35
Total	207,046	16,112	223,158

Table E.6 Number of Metformin Rx vs asthma attacks in the training set.

Metric / Category	Asthma Attack=0	Asthma Attack=1
No Metformin Use (0)	198,129	15,377
Used Metformin (1+)	8,917	735
Total	207,046	16,112
Minimum	0	0
Q1	0	0
Median	0	0
Q3	0	0
Maximum	352	101
Mean	0.32	0.34

Table E.5 Number of hospitalisations vs asthma attacks in the training set.

Metric / Category	Asthma Attack=0	Asthma Attack=1
No Hospitalisations (0)	167,127	11,784
Hospitalized (1+)	39,919	4,328
Total	207,046	16,112
Minimum	0	0
Q1	0	0
Median	0	0
Q3	0	1
Maximum	157	152
Mean	0.32	0.53

Appendix F

Asthma Attack Modelling Algorithms

Algorithm 1 Data Preparation

Load datasets:
    `Asthma_PatientQuarter5.txt` $\rightarrow \mathscr{D}_{\text{rawQ5}}$ Training and testing data
    `Asthma_PatientQuarter9.txt` $\rightarrow \mathscr{D}_{\text{rawQ9}}$ Validation data
Filter patients with `Age` ≥ 12:
    $\mathscr{D}_{\text{rawQ5}} \rightarrow \mathscr{D}_{\text{12YearsAboveQ5}}$
    $\mathscr{D}_{\text{rawQ9}} \rightarrow \mathscr{D}_{\text{12YearsAboveQ9}}$
Drop irrelevant, administrative or redundant features:
    `Excluding_Features = [ENHI, IndexDate, ObservationEndDate, CohortYear, AsthmaHospitalisation, OralCorticosteroidPrescriptions, P12MAsthAttack, P6MAsthAttack, P3MAsthAttack, DateOfBirth, EthnicityCode, DomicileCode, DHB_Code]`
Define new clinical variables:
    `HistoryOfAllergies` = `Rhinitis` OR `NasalPolyps` OR `AtopicDermatitis` OR `Anaphylaxis` OR `PulmonaryEosinophilia`
    `CardiovascularDiseases` = `CardiovascularCerebrovascularDisease` OR `IschaemicHeartDisease` OR `HeartFailure`
Replace `EthnicGroup` = `'Unknown'` with `'Other'`
Remove variables used to derive new clinical features
Bin `Age` into categories: 12-14, 15-44, 45-64, 65+
Remove redundant features and retain `P12MAsthAttack`, `SABA_ICS_Ratio`, `CharlsonComorbidityScore`
Retain outcome variable `AsthmaAttack`
Save modified datasets as:
    $\mathscr{D}_{\text{modifiedQ5}} \rightarrow$ "AsthmaPatients_12YearsAbove_Quarter5_FullData.csv"
    $\mathscr{D}_{\text{modifiedQ9}} \rightarrow$ "AsthmaPatients_12YearsAbove_Quarter9_FullData.csv"
Split the full datasets into training and testing sets using a 70:30 split ratio, resulting in:
    Training set $\rightarrow \mathscr{D}_{\text{modifiedQ5training}}$
    Testing set $\rightarrow \mathscr{D}_{\text{modifiedQ5testing}}$
Output: Preprocessed datasets for feature set definition

Algorithm 2 Defining Initial Main Feature Sets

Input:
Datasets from 2 $\mathscr{D}_{\text{modifiedQ5training}}$ and $\mathscr{D}_{\text{modifiedQ5testing}}$
Define XGB-based important feature set (FS1) from the previous findings as:

```
FS1 = [P12MNoAsthAttacks, Q5_WeeksDuringWinter, NoOfICSInhalers,
SABA_ICS_Ratio, P6MNoAsthAttacks, P3MNoAsthAttacks, NoOfSABAInhalers, DHB
 Gender, EthnicGroup]
```

Define full refined feature set (FS2) as:

```
FS2 = [Q5_WeeksDuringWinter, Age, Gender, EthnicGroup,
DeprivationQuintile, DHB, Psoriasis, Obesity, LRTI, AnxietyDepression
Diabetes, SmokingStatus, CharlsonComorbidityScore
NumberOfMetforminRx, Diabetes_NoMetformin, NoOfHospitalisations
NoOfOPEDVisits, BetaBlockers, Paracetamol, NSAIDs, SABA_ICS_Ratio
P12MNoOfAsthAttacks, HistoryOfAllergies, CardiovascularDiseases]
```

Define full refined feature set (FS4) as:

```
FS4 = [Q5_WeeksDuringWinter, Age, Gender, EthnicGroup, DeprivationQuintile
SmokingStatus, CharlsonComorbidityScore, NumberOfMetforminRx
NoOfHospitalisations, NoOfOPEDVisits, Paracetamol, NSAIDs, SABA_ICS_Ratio
P12MNoOfAsthAttacks, HistoryOfAllergies, CardiovascularDiseases]
```

Output: Main feature sets for further feature set generation

Algorithm 3 Iterative Recursive Feature Elimination with Cross-Validation (RFECV) on Main Feature Sets

Input: Training data $\mathscr{D}_{\text{modifiedQ5training}}$, feature sets FS2 and FS4, classifiers RF and XGB
For each feature set $FS \in \{FS2, FS4\}$:
    Extract numeric features and categorical features for FS
    Initialise numeric feature scaler:
        `ST_scalar` $\leftarrow$ `StandardScaler()`
    Initialise categorical feature encoder:
        `OH_Encoder` $\leftarrow$ `OneHotEncoder(handle_unknown='ignore', sparse_output=False)`
    Construct preprocessing pipeline `preprocessor` with `ColumnTransformer`:
        Apply `OH_Encoder` to categorical features
        Apply `ST_scalar` to numeric features
    Define dependent and independent variables:
        $(X_{\text{train}}, y_{\text{train}})$ and $(X_{\text{test}}, y_{\text{test}})$
    For each classifier $C \in \{\text{RF}, \text{XGB}\}$:
        Initialise base estimator C_{model} with fixed random state
        Apply preprocessing to training features:
            $X_{\text{transformed}} \leftarrow$ `preprocessor.fit_transform`(X_{train})
        Set target labels
            $y_{\text{resampled}} \leftarrow y_{\text{train}}$
        Define stratified cross-validation strategy:
            `StratifiedKFold(n_splits=5, shuffle=True, random_state=42)`
        Initialise RFECV:
            `RFECV(estimator=`$C_{\texttt{model}}$`, step=1, cv=CV, scoring='roc_auc', n_jobs=-1)`
        Fit RFECV on preprocessed features:
            `rfecv.fit(`$X_{\text{transformed}}$, $y_{\text{resampled}}$`)`
        Identify selected feature mask:
            `selected_mask` $\leftarrow$ `rfecv.support_`
        Extract selected and removed features:
            $\mathscr{F}_{\text{selected}} \leftarrow X_{\text{transformed.columns}}$[`selected_mask`]
            $\mathscr{F}_{\text{removed}} \leftarrow X_{\text{transformed.columns}}$[$\neg$`selected_mask`]
        Record optimal number of features:
            $|\mathscr{F}_{\text{selected}}| \leftarrow$ `rfecv.n_features_`
        Plot mean CV AUROC vs. number of retained features
Output: For each $FS \in \{FS3, FS5, FS6\}$ and classifier C, save $\mathscr{F}_{\text{selected}}$, $\mathscr{F}_{\text{removed}}$, and RFECV performance plots

Algorithm 4 Evaluation of Model Accuracy and Computational Efficiency Across Model and Sampling Combinations

Input training data $\mathscr{D}_{\text{modifiedQ5training}}$ and feature sets FS1, Fs2, FS3, FS4, FS5 and FS6
Construct pipeline:
    `preprocessor` → `sampler` → `classifier`
Define candidate classifiers:
    `LogisticRegression, RandomForestClassifier, XGBClassifier`
Define data imbalance handling techniques (sampler):
    `RandomUnderSampler, EditedNearestNeighbours, SMOTE, KMeansSMOTE, No-sample`
for each `model` in candidate classifiers **do**
  for each `sampler` in imbalance handling techniques **do**
    if `sampler == "No-sample"` **then**
      Construct pipeline: `preprocessor` → `classifier`
    else
      Construct pipeline: `preprocessor` → `sampler` → `classifier`
    end if
    Evaluate pipeline using 5-fold stratified cross-validation
    Compute predictive performance (AUROC, F1-Score) and efficiency (prediction time, model size) metrics
  end for
end for
Training and evaluating final models:
    Select a single model–sampler configuration
    Fit the corresponding `preprocessor` on the $\mathscr{D}_{\text{modifiedQ5train}}$
    Evaluate final model performance on the $\mathscr{D}_{\text{modifiedQ5test}}$
Analysing efficiency:
    Aggregate efficiency logs across iterations
    Compute mean training time, standard deviation, and standard error
    Estimate 95% confidence intervals using the Student's t distribution
Statistical analysis of accuracy metrics:
    Import fold-level cross-validation metrics
    Fit factorial ANOVA models for:
      F1-score $\sim$ Model $\times$ Imbalance Technique $\times$ Dataset
      AUROC $\sim$ Model $\times$ Imbalance Technique $\times$ Dataset
Confidence interval estimation for accuracy metrics:
    Compute 95% confidence intervals for AUROC and F1-score using 5-fold CV for model, dataset, and imbalance technique
Output:
    Comparative accuracy, efficiency, ANOVA results, and confidence intervals for all model and imbalance-handling combinations

Appendix G

Asthma LoS Dataset Description

- This dataset was collected from the Auckland Hospital, NZ, through the ethics approval (Appendix H).
- It consisted of 11,414 instances with 12 features, including asthma LoS.
- Table G.1 provides a detailed description of the features used in predicting asthma LoS.
- Table G.2 provides a sample of data.

Table G.1 Description of features in the LoS dataset.

Feature	**Description**	**Data type**	**Values**
Admit Date	Date the patient was admitted to the hospital	String	
Admit CBU Desc	Clinical business unit to which the patient was admitted	String	
Admit day of week	Day of the week the patient was admitted	String	Mon, Tue, Wed, Thu, Fri, Sat, Sun
Discharge date	Date the patient was discharged	String	
DHB Desc	District health board corresponding to the admission location	String	
Age	Age of the patient	Numeric	Positive integer
Gender Desc	Gender of the patient	Binary	Female, Male
Ethnicity	Ethnicity of the patient	String	
Smoker Flag	Indicates whether the patient is a smoker	Binary	Yes, No
Diag Desc	Description of the diagnosis at the time of admission	String	Predominantly allergic asthma, Nonallergic asthma, Mixed asthma, Asthma–unspecified, Status asthmaticus, Cough, Dyspnoea, Stridor, Wheezing
Diag Code	Diagnosis code at the time of admission	String	J450, J451, J458, J459, J46, R05, R060, R061, R062
LoS	Length of stay after admission	Numeric	Positive integer

Table G.2 Example records from the LoS dataset.

Admit Date	Admit CBU Desc	Admit Day of Week	Discharge Date	DHB Desc	Age	Gender	Ethnicity	Smoker Flag	Diag Desc	Diag Code	LoS
8/07/2019	Immunology	Mon	8/07/2019	Auckland	44	Female	New Zealand European	N	Asthma, unspecified	J459	0
5/07/2019	Emergency Medicine	Fri	8/07/2019	Auckland	60	Female	Cook Island Māori	N	Asthma, unspecified	J459	3
4/06/2019	Emergency Medicine	Tue	8/07/2019	Auckland	42	Male	Indian	N	Asthma, unspecified	J459	34
8/08/2019	Gen Medicine	Thu	9/8/2019	Auckland	56	Male	New Zealand European	N	Dyspnoea	R060	1
1/08/2019	Ophthalmology	Thu	1/08/2019	Waitemata	24	Male	Māori	N	Dyspnoea	R060	0
1/08/2019	Emergency Medicine	Thu	1/08/2019	Auckland	30	Female	New Zealand European	Y	Wheezing	R062	0

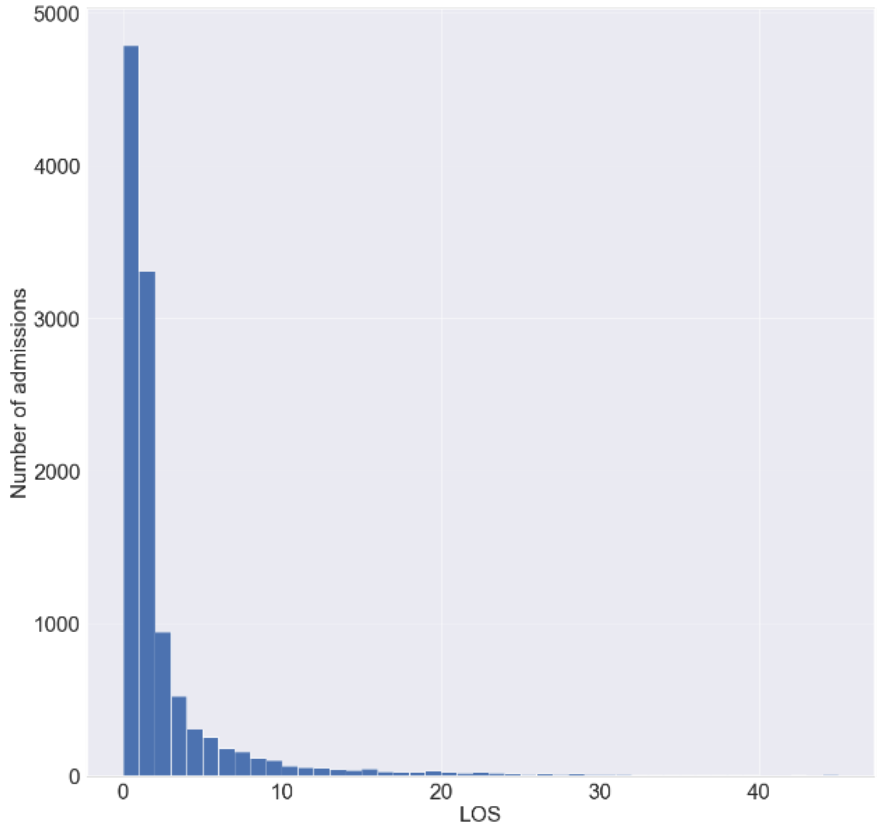


Fig. G.1 Distribution of LoS.

Appendix H

Asthma LoS Prediction - Ethics Approval

AUCKLAND HEALTH RESEARCH ETHICS COMMITTEE (AHREC)

11/05/2021

Dr Amy Chan

School of Pharmacy

Re: Application for Ethics Approval (Our Ref. AH22149): Approved

The Committee considered your application for ethics approval for the study entitled "**A Practical Tool to Predict Asthma-Related Emergency Department Visits and Admissions at Population Level**".

We are pleased to inform you that ethics approval has been granted.

The expiry date for this approval is **11/05/2024**.

Amendments to the approved project:

Should you need to make any changes to the approved project, please follow the steps below:

- Send a request to the AHREC Administrators to unlock the application form (using the Notification tab in the Ethics RM form).
- Make all changes to the relevant sections of the application form and attach revised documents (as appropriate).
- Change the Application Type to "Amendment request" in Section L.
- Add a summary of the changes requested in the text box.
- Submit the amendment request (PI/Supervisors only to submit the form).

If the project changes significantly, you are required to submit a new application.

Funded projects: If you received funding for this project, please provide this approval letter to your local Faculty Research Project Coordinator (RPC) or Research Project Manager (RPM) so that the approval can be notified via a Service Request to the Research Operations Centre (ROC) for activation of the grant.

The Chair and the members of AHREC would be happy to discuss general matters relating to ethics approvals. If you wish to do so, please contact the AHREC Ethics Administrators at ahrec@auckland.ac.nz in the first instance.

Additional information:

- Do not forget to fill in the 'approval wording' on the PISs, CFs and/or advertisements, using the date of this approval and the reference number, before you use the documents or send them out to your participants.

All communications with the AHREC regarding this application should indicate this reference number: **AH22149**.

AHREC Administrators

Auckland Health Research Ethics Committee

Appendix I

Asthma LoS Modelling Algorithm

Algorithm 5 Preprocessing Asthma Admission Data for LoS Prediction

Input: Raw asthma admission dataset $\mathscr{D}_{\text{raw}}$ with columns including `Admit Date`, `LOS`, `Ethnicity Prioritised Desc`, `DHB Desc`, `Gender Desc`, `Smoker Flag`, `Diag Desc`, etc.
Deriving a new feature `Admit Month`:
    Convert `Admit Date` to datetime object
    Derive new variable `Admit Month` from `Admit Date` (numeric month)
Handling missing or ambiguous values:
    Remove records where `Ethnicity Prioritised Desc` $\in$ { "Don't Know", "Not Stated" }
    Remove records where `Gender Desc` = "Unspecified"
Grouping ethnicity into major categories:
    European: {New Zealand European, Other European, European NFD}
    Maori: {Maori}
    Pacific Peoples: {Tongan, Samoan, Cook Island Maori, Niuean, Fijian, Other Pacific Peoples, Tokelauan, Pacific Island NFD}
    Asian: {Indian, Chinese, Other Asian, Southeast Asian, Asian NFD}
    Middle Eastern / Latin American / African: {Middle Eastern, African, Latin American}
    Other Ethnicity: {Other Ethnicity}
Grouping DHBs:
    Create new column `DHB Group` $\leftarrow$ `DHB Desc`
    Less frequent or unknown DHBs $\rightarrow$ "Other"
Removing unnecessary columns for modelling:
    Drop columns: `[Admit Date, Admit CBU Desc, Discharge Date, DHB Desc, Ethnicity Desc, Ethnicity Prioritised Desc, Diag Desc]`
Handling outliers in LoS:
    Compute Z-score for `LOS`: $Z = (LOS - \mu_{LOS})/\sigma_{LOS}$
    Remove records with $Z > 3$
Cleaning data:
    Remove records where `LOS` = 0
    Remove records where `LOS` > 14
Output: Processed dataset $\mathscr{D}_{\text{processed}}$ for modelling

Algorithm 6 Predicting Asthma Patient LoS Using Regression Models

Input: Processed dataset $\mathscr{D}_{\text{processed}}$
Defining preprocessing pipeline:
    Categorical features: `[Admit Day of Week, Diagnostic Code, Admit Month, Ethnicity Group, DHB Group]`
    Numerical features: `[Age]`
    Initialise encoders and scalers:
        `OneHotEncoder(sparse=False)` for categorical features
        `StandardScaler()` for numerical features
    Construct `ColumnTransformer` `preprocessor`:
        Apply OneHotEncoder to categorical features
        Apply StandardScaler to numerical features
        Pass through remaining features unchanged
Split data into training and test sets:
    Define input features $X = \mathscr{D}_{\text{processed}} \setminus \{\texttt{LOS}\}$
    Define target $y = \texttt{LoS}$
    Split X and y into training set (70%) and test set (30%)
Define regression models:
    RandomForestRegressor
    KNeighborsRegressor
    Support Vector Regressor (SVR)
    XGBRegressor
Define evaluation metrics for comparing model performance:
    Mean Squared Error (MSE)
    Root Mean Squared Error (RMSE)
    Mean Absolute Error (MAE)
Developing prediction models:
    For each regression model:
        Fit the model on preprocessed training data X_{train}, y_{train}
        Predict LOS on test set X_{test} to obtain $\hat{y}_{\text{test}}$
Evaluate model performance:
    Compute metrics on test set:
        $MSE = \text{mean\_squared\_error}(y_{\text{test}}, \hat{y}_{\text{test}})$
        $RMSE = \sqrt{MSE}$
        $MAE = \text{mean\_absolute\_error}(y_{\text{test}}, \hat{y}_{\text{test}})$
    Optionally, print cross-validated metrics means
Output: Trained regression models, predicted LOS on test set, evaluation metrics (MSE, RMSE, MAE) for comparison of model performance

Appendix J

Correlation of Features Used in Predicting Asthma Length of Stay

Table J.1 Correlation of features and LoS group.

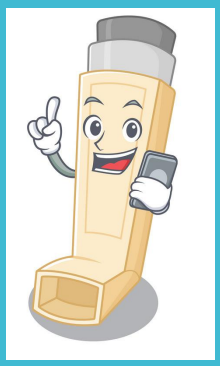
Feature	p-Value (95% CI)
Admit Day of Week	<0.05
Gender	<0.05
Smoker Status	<0.05
Diagnostic Code	<0.05
Admit Month	0.969
Ethnicity Group	<0.05
DHB Group	<0.05
Age	<0.05

Table J.2 Performance of the classification models in predicting LoS.

Model	Precision	Recall	F1 score
SVM	0.63	0.76	0.69
Random Forest	0.63	0.78	0.69
Logistic Regression	0.63	0.72	0.67
AdaBoost	0.63	0.73	0.68
Gradient Boosting	0.63	0.77	0.70

Appendix K

Māori Interviews - Recruitment Flyer



Interested in AI and technology in ASTHMA?

Come and discuss with us!

AUT UNIVERSITY
AUCKLAND, NEW ZEALAND

UNIVERSITY OF AUCKLAND
Waipapa Taumata Rau
NEW ZEALAND

We are looking for **Māori**:

- Aged 18 and above
- Have asthma
- Who wants to share their experience and views about the use of AI and technology in asthma in an interview or hui (focus group) (according to your preference)
- Be able to consent and speak in English

Interviews will be 30-60 mins and focus groups up to 2 hours
By taking part, you can help improve asthma care for Māori

To thank you, you will receive a $30 supermarket voucher

If interested, contact:
Darsha Widana darsha.jayamini@autuni.ac.nz or scan the below QR code to send an email

QR Code

Approved by the Auckland Health Research Ethics Committee on 23/3/2023 for three years. Reference number AH25484.

Appendix L

Māori Interviews - Participation Information Sheet

You are invited to take part in a study to help us find out how Māori think and feel about the use of Artificial Intelligence (AI) and technology to predict asthma attacks. Asthma attacks disproportionately affect many more Māori than non-Māori in New Zealand. One of the ways that we can help reduce the impact of asthma attacks is to use technology to help us predict when an attack might happen.

We are doing this study to find out what Māori think and feel about the use of technology such as AI for risk prediction and to identify key risk factors that may be affecting asthma patients in the Māori community so we can include these in any future computer-based prediction models such as machine learning models to ensure we avoid worsening health disparities.

This Participant Information Sheet will help you decide if you'd like to take part. It sets out why we are doing the study, what your participation would involve, what the benefits and risks to you might be, and what will happen after the study ends. We will go through this information with you and answer any questions you may have. You do not have to decide today whether or not you want to participate in this study. Before you decide you may want to talk about the study with other people, such as family, whānau, friends, or healthcare providers. Feel free to do this.

If you agree to take part in this study, you will be asked to sign the Consent Form on the last page of this document. You will be given a copy of both the Participant Information Sheet and the Consent Form to keep.

Why are we doing this study?

In one of our previous studies, we found out that Māori females had the highest average length of stay in hospital for asthma and that Māori have higher rates of asthma hospitalisations than

non-Māori. This finding showed how important it is to investigate key risk factors for asthma attacks in Māori and to find ways to predict attacks before they happen. Machine learning is an advanced computer technology and type of AI that has been used for predicting asthma attacks. However, these models need to capture Māori-specific risk factors to ensure that they work just as well in Māori and non-Māori. To do this, we need to find out how Māori can be best represented in machine learning models in a way that does not exacerbate existing health inequities. The aim of our study is to explore the views and attitudes of Māori towards the use of AI and technology in asthma, and identify the risk factors specific to Māori asthma patients, which are important to include in the asthma risk prediction models. We will use this information to help us develop a machine-learning-based risk score for Māori patients that can tell us a person's risk of asthma attacks.

What will my participation in the study involve?

You have been invited to participate as you have indicated that you are interested in taking part in this study, are aged 18 years or older, have asthma and identify as Māori ethnicity.

If you choose to participate, you will be asked to share information about your experience of asthma and opinions about AI and technology use in asthma and what risk factors there may be for asthma attacks that may be unique to the Māori community. This will take place as either a one-on-one or focus group interview, where a group of participants will be engaged in the discussion. You can choose which format you wish to be interviewed depending on what suits you best. For zoom video interviews, you will need to return the signed consent form by taking a photo or scan of the form and returning this by email.

Where will the study take place?

For one-on-one interviews, this can take place in-person at either The University of Auckland or Auckland University of Technology or over Zoom video conferencing, and will last about 30-60 minutes, at a time that suits you. You are welcome to have a support person attend if you wish. For focus group interviews, this will consist of up to 8 people per group and will be conducted in person either at The University of Auckland or Auckland University of Technology. The focus group will last up to a maximum of 2 hours.

How will my information be collected?

To make sure we can accurately remember the discussion points we will need to either audio or video record the conversation, whichever you are most comfortable with. Audio/video recordings will be retained for six years after the interview. For one-on-one interviews, we

can take notes instead of recording if you prefer. If you choose to be in a focus group, it is not possible to decline to be recorded or ask for recordings to be stopped as there will be others in the group who have consented to recording. However you can choose to remain silent or leave the room if you do not wish to answer a question, or withdraw participation (though it will not possible to withdraw your data from the focus group – see below section about your rights).

The interview recordings will be transcribed by the interviewer and /or a professional transcribing company to help us identify your views and identify common risk factors for asthma attacks. To ensure we keep your views as confidential as possible, we will assign a participant number to the recordings and the transcripts so we cannot identify you (i.e. the data is coded (de-identified)). The recordings and the transcriptions will be stored securely on a password-protected server on the University of Auckland research network. Only the researchers on the team will have access to the information from the interview; the recordings and transcript data will be deleted after six years. Any third-party transcribers used to transcribe the interviews will be asked to sign a confidentiality agreement to protect the privacy of the participants.

Note that for focus groups, it is not possible to guarantee confidentiality due to the group discussion nature but we do ask all participants in the focus group to not discuss what was said with others, and the group facilitator will actively encourage all participants to maintain confidentiality.

What will be discussed during the interview / focus group?

During the interview or focus group, we will ask you to describe what your experiences have been with asthma, what you think about AI and technology use, and what risk factors are related to asthma attacks in Māori. We will ask you for some background information about yourself including age, gender, support for daily activities, and confidence with using technology. If you have any questions about your health, we encourage you to discuss these with your GP or other health care team involved in your regular care to address these.

What will happen to my information after?

Your views will be combined with other participants in this study to work out how we can best ensure that the Māori community well represented in machine learning models developed for predicting the risk of asthma attacks. The information will be summarized in a written report or presentation. All contributions will remain deidentified (coded by a unique participant number) and none of the material in the results will personally identify you.

What are the rights of participants in the study?

Whether or not you take part is your choice. If you don't want to take part, you don't have to give a reason, and it won't affect the care you receive. If you do want to take part now, but change your mind later, you can pull out of the study at any time – before, during or following the interview. You do not need to give a reason for withdrawing. However, what happens to your information when you withdraw will be different depending on whether you choose to be part of a one-on-one interview or focus group:

- If you have been interviewed one-on-one, then you have one week to withdraw your responses, after which your data will have been combined with others and cannot be separated out.
- If you took part in a group interview (focus group), it will not be possible to withdraw your data as it has been collected along with other participants and it risks compromising the integrity of the data from other participants who do wish to continue to be part of the research.

If you choose to be interviewed one-on-one, you can also have a chance to review the transcripts of the interview recording and make edits to this if you wish – if you would like to do this, please provide your email address below. You will have 7 calendar days to provide comments on the transcripts.

For focus group interviews, it will not be possible to review and edit transcripts due to the group nature of the discussions as other participant's information will be in the transcript.

All patients who take part can be emailed a short summary of the results when they are completed – if you wish to receive this, please provide your email address at the end of the consent form.

All participants will be offered at $30 grocery (supermarket) voucher to thank them for their time and information provided. Please provide a postal address to send this at the end of the consent form if you are not doing the interview in-person.

Who do I contact for more information or if I have concerns?

If you have any questions, concerns or complaints about the study at any stage, you can contact the lead researcher or research team:

Amy Chan,
Senior Clinical Research Fellow
School of Pharmacy, University of Auckland
a.chan@auckland.ac.nz

Farhaan Mirza,
Associate Professor,
School of Computer and Mathematical Sciences,
Auckland University of Technology
farhaan.mirza@aut.ac.nz

Darsha Widana,
Student researcher,
School of Computer and Mathematical Sciences,
Auckland University of Technology
darsha.jayamini@autuni.ac.nz

If you wish to contact the Head of School of Pharmacy:

Professor Shane Scahill
Head of School of Pharmacy,
The University of Auckland
s.scahill@auckland.ac.nz

If you require Māori cultural support, talk to your whānau in the first instance. If you require further support, please contact us and we can put you in touch with our Māori advisor Marie-Claire Bidois-Putt for this project.

For concerns of an ethical nature, you can contact the Chair of the Auckland Health Research Ethics Committee at ahrec@auckland.ac.nz or at 373 7599 x 83711, or at Auckland Health Research Ethics Committee, The University of Auckland, Private Bag 92019, Auckland 1142.

Thank you for reading this and considering participating.

This study has been funded by the Health Research Council as part of their Ethics Studentship programme (HRC 22-925).

Approved by the Auckland Health Research Ethics Committee on 23/3/2023 for three years (Reference number AH25484).

Appendix M

Māori Interviews - Semi-Structured Interview Guide

The following open-ended questions were posed to the study participants under each section.

1. **Introduction**

   Introduce the interviewer, study purpose, and rights while going through the information sheet and the consent form.

   Asthma attacks affect many people in New Zealand. We are looking at ways to improve asthma care by being able to predict asthma attacks before they happen so that we can keep people well and out of hospital. One way to do this is the use of technology.

   The aim of today is to explore your thoughts and feelings about the use of technology in asthma and if there are any specific risk factors for attacks in Māori that we should think about.

   Firstly: Tell me about you (pepeha)

2. **Experience with Asthma**

   We will start off with talking about your experience with Asthma.

   - How do you feel about living with Asthma? How has it affected your life?

   For instance: How often do you take medications (such as inhalers)? How often do you go to meet GPs or go to the hospital for asthma?

3. **Factors related to Asthma attacks**

   Let's talk about the things related to attacks

   - What happens when you have an attack?

- What do you do when you have an attack?

- How do you feel?

- What triggers an asthma attack for you?

For instance: Pollen, Temperature, Wind, Pets, Stress

- Have you tried avoiding those triggers, and what have you experienced in doing that?

- Were you able to avoid them, or was there anything that prevents you from doing that?

4. **Predicting asthma attacks using technology**

One way of predicting asthma attacks is to use technology such as smart devices to monitor asthma (e.g. smart inhalers) and to use Artificial Intelligence (AI) to predict when an attack may happen.

- What do you think about using technology to help manage your asthma?

Are you aware of what AI is – what do you know about AI and its use in healthcare? What do you think of it?

- How would you feel if AI was used to help predict when your next attack might happen?

- What may be the pros/cons for you?

- What factors should be considered when using technology such as AI in Māori living with asthma?

- Why do you think those factors need to be considered?

- Why may it stop you from using technology? What may help?

5. **Considerations for Māori when developing risk prediction models**

- What factors should be considered when developing risk prediction models for asthma attacks?

e.g. kanohi-ki-te-kanohi, whanaungatanga. How do you think Māori culture and beliefs may be recognised and incorporated when considering risk prediction for asthma attacks?

6. **Conclusion**

- Is there anything else you would like to tell me that you think would help us in using AI to predict asthma attacks in Māori patients?

- Will I be able to contact you later in case I have additional questions?

Appendix N

Māori Interviews - Ethics Approval

AUCKLAND HEALTH RESEARCH ETHICS COMMITTEE (AHREC)

23/03/2023

Dr Amy Chan
Pharmacy

Re: Application for Ethics Approval (Our Ref. AH25484): Approved

The Committee considered your application for ethics approval for the study entitled "**Attitudes towards use of AI in asthma in Māori**".

We are pleased to inform you that ethics approval has been granted.

The expiry date for this approval is **23/03/2026**.

Locality approval: Before starting your research, ensure that all the required locality approvals have been obtained. If one or more DHBs will be a locality, please contact their Research Office(s) to determine the locality approval requirements of the DHB(s).

Final report: In order that up-to-date records are maintained, you must notify the Committee once your project is completed and submit a final report.

Amendments to the approved project: Should you need to make any changes to the approved project, please follow the steps below:

- Send a request to the AHREC Administrators to unlock the application form (using the Correspondence tab in Ethics RM).
- Make all changes to the relevant sections of the application form and attach revised documents (as appropriate).
- Change the Application Type to "Amendment request" in Section L.
- Add a summary of the changes requested in the text box.
- Submit the amendment request (PI/Supervisors only to submit the form).

If the project changes significantly, you are required to submit a new application.

Funded projects: If you received funding for this project, please provide this approval letter to your local Faculty Research Project Coordinator (RPC) or Research Project Manager (RPM) so that the approval can be notified via a Service Request to the Research Operations Centre (ROC) for activation of the grant.

The Chair and the members of AHREC would be happy to discuss general matters relating to ethics approvals. If you wish to do so, please contact the AHREC Ethics Administrators at ahrec@auckland.ac.nz in the first instance.

Additional information:

- Do not forget to fill in the 'approval wording' on the PISs, CFs and/or advertisements, using the date of this approval and the reference number, before you use the documents or send them out to your participants.

All communications with the AHREC regarding this application should indicate this reference number: **AH25484**.

AHREC Administrators

Auckland Health Research Ethics Committee

Appendix O

Māori Interviews - Consent Form

Attitudes towards use of Artificial Intelligence (AI) to predict asthma attacks in Māori and exploration of risk factors affecting asthma attacks: A qualitative study

- I have read and I understand the Participant Information Sheet.
- I have been given sufficient time to consider whether or not to participate in this study.
- I have had the opportunity to use a legal representative, I/ family support or a friend to help me ask questions and understand the study.
- I understand that taking part in this study is voluntary (my choice) and that I may withdraw (not continue to participate in the study) at any time (before, during or after the interview) without this affecting my care.
- If I have been interviewed one-on-one and I decide to withdraw from the study, I understand I have a week to ensure my information is not included in the study, and that after that time, the information collected about me will be collated with other participants' information and therefore cannot be separated and excluded.
- If I have been interviewed in a focus group, I understand that it will not be possible to withdraw my information as it has been collected as part of a larger group with other participants, and therefore cannot be separated and excluded
- I consent to the research staff collecting my views through an interview either one-on-one or in a focus group, including information about my health and asthma experiences.
- I agree to the interview being recorded either in audio or video form, or for one-on-one interviews, in written notes
- I understand that for focus groups, confidentiality cannot be guaranteed
- I understand that focus groups will be recorded and it is not possible to decline recording of focus groups

- I agree to the interviewer or a professional transcription company transcribing my interview recording
- For one-one-one interviews: I wish to review the transcript of the interview
- For one-one-one interviews, I understand that edits of the interview transcript only must be returned within 7 calendar days of receipt
- I understand that all interview data will be stored in a de-identified form
- I understand that my participation in this study is confidential and that no material, which could identify me personally, will be used in any reports on this study.
- I know who to contact if I have any questions about the study.
- I understand my responsibilities as a study participant.
- I wish to receive a summary of the results from the study.

Email I would like the transcript and results summary sent to (please write):

_____

Postal address I would like the supermarket voucher to be sent to (please write):

_____

Declaration by participant:

I consent to take part in this study.
Participant's name: _____
Signature: _____
Date: _____

Declaration by a member of the research team:

I have given a verbal explanation of the research project to the participant and have answered the participant's questions about it.
I believe that the participant understands the study and has given informed consent to participate.
Researcher's name: _____
Signature: _____
Date: _____

Approved by the Auckland Health Research Ethics Committee on 23/3/2023 for three years(Reference number AH25484).

Appendix P

Māori Interviews - Example Quotes

Table P.1 Example quotes for the themes and sub-themes identified in the study.

Sub-theme	Example quotes	Participant (Sex, Age)
Theme 1: Concerns about AI use		
Trust in technology	"… my Papa is dead against it. Papa is dad's dad, he's really against new technology. Tricknology, you know?"	F, 32 years
	"… I also believe it can be dangerous when we're relying on that to protect things when everyone's bodies are different because of my health and my journey; I'd be dead if I relied on technology."	F, 37 years
	"… Like I would be like, yeah, you're going to save me, that would not be something I can trust 100%, ever."	F, 37 years
	"I think trust [emphasises], there's a lot of trust building up relationships with Māori and stuff like that"	F, 38 years
	"… having something being absorbed by the computer and no physical contact, I think is a bit threatening for them, because where is it all going and what's happening to it?"	A caregiver
	"You know things can harm, you got people talking about AI taking over the world. That's quite scary. AI's not doing what they've been told to do, and I've read so many things about that. And I and I freak out with that."	F, 37 years
	"… Why are we getting controlled by technology? And who's taking all the data? Who's listening to the technology?"	F, 37 years
Preference for kanohi ki te kanohi (face-to-face) interaction with Māori	"face-to-face is better"	M, 23 years
	"… for Māori would be that kanohi ki te kanohi, you know, face-to-face and the fact that they're comfortable with you"	A caregiver
Inadequate knowledge of AI and technology	"… and when it comes to AI, I don't know much about it. I know little bits."	F, 37 years
	"Yeah, I think I need to know more about them because I don't actually know much about technologies and the word AI"	F, 37 years

Sub-theme	Example quotes	Participant (Sex, Age)
	"Just hearing it to say, you know, to help me with, dealing with my asthma, I just feel nervous about hearing that."	F, 38 years
	"… not everybody's confident with technology, and there's been, you know, there's a lot of rumours around technology and stuff too, so."	F, 37 years
Theme 2: Interest in using technology to support asthma		
	"Ohh, now you pull up like that. I think that, you know, it's not as scary as I imagined."	F, 38 years
	"That sounds awesome, absolutely, yeah, absolutely… It's just interesting to know. Because, I mean, it's so much different to manually doing something, like doing it yourself and trying to write a diary on…, that'd be interesting."	F, 44 years
	"I think it would be learning about it, making sure I'm very well aware of what it's doing, why it's doing what it's doing and making sure there are no side effects to it I guess. Mm. I think that's the main reason why I don't, and I'm not a technology kind of person. Like, I wasn't brought up with technology. So, I've had to adapt to all the technology, learn it all myself…"	F, 37 years
	"I am a great believer in technology" showed the trust that AI can be used to make people's lives better – "I think it will benefit mankind immensely. It must be handled the right way"	M, 76 years
Theme 3: Desired characteristics of AI-based systems		
	"When I see my Papa I see him really struggle to use the tech, so I think they would need to be relatively simple to use. User-friendly. Simple design, simple to use for our elderly."	F, 32 years
	"I think it's going to be accessible to all, like young and the old and easy to use basically"	F, 59 years
	"No more than four steps"	F, 66 years
	"I think if it wasn't too bulky, I guess. We don't have to remember too much."	M, 62 years
	"Make it simple… So, I mean, a lot of people shy away, especially, our people, they shy away from things that are a little bit hard to do so. Maybe if it was a bit easier."	F, 44 years
	"High-tech, Straightforward. Easy to use."	F, 18 years
	"When it comes to medical stuff, it needs to be in English, and it needs to be simple, not using scientific words… not using abbreviations, not complicated…"	F, 51 years
	"An option for Te reo Māori would be great."	F, 32 years
Theme 4: Experience in asthma management and opportunities for technology to improve care		
Accessibility and quality of healthcare resources	"How is it going to help if you are having an attack if you live in the middle of nowhere (farms, the country?)."	F, 23 years
	"The doctors up here are just so different to Auckland; would I say the DHB up here is not as good as it is in Auckland, and I noticed that by moving up here. I understand what all my Nana and them we're all talking about now. They're not good here at all."	F, 44 years
Prior experience with the healthcare staff	"It's me, not our system"	F, 37 years
	"I've been stuck in our medical system until then, so…"	F, 37 years
	"Doctors taking our concerns seriously and listening to us"	F, 23 years

Sub-theme	Example quotes	Participant (Sex, Age)
	"Yeah, the doctors are great giving out inhalers, but I could do with a review and see if there's something better out there on the market. Because you don't really know what's available. They just kind of put you on something, you think, hey, well, they know, and that's the best. But it's not always"	F, 59 years
	"I just know that a lot and I'm gonna sound really racist here and I don't mean to be, but a lot of non-Māori therapists don't work well for Māori"	F, 66 years
Triggers for asthma exacerbations	"Well, if he's been out in the cold and cold and the rain during the day. He will get quite chesty. Or is it the humidity like she said from your shower?"	A caregiver
	"Normally it's the shift of like cold to warm, warm to cold…"	F, 37 years
	"… because I know when I am very active, that really gets me Chesty…"	F, 38 years
	"I feel it only gets worse when I do intense workouts or exercises."	F, 23 years
	"… when I start getting sick or anything like that… It like triggers it. So, if I come down with a cold or anything like that, I get really bad asthma."	F, 44 years
	"Dust is another big thing for me."	F, 37 years

Appendix Q

Pasifika Interviews - Recruitment Flyer

Appendix R

Pasifika Interviews - Information Sheet

You are invited to take part in a study to help us find out how Māori and Pasifika think and feel about the use of artificial intelligence (AI) and technology to predict asthma attacks. Asthma attacks disproportionately affect many more ethnic groups including Māori and Pasifika in New Zealand. One of the ways that we can help reduce the impact of asthma attacks is to use technology to help us predict when an attack might happen.

We are doing this study to find out what Māori and Pasifika think and feel about the use of technology such as AI for risk prediction and to identify key risk factors that may be affecting asthma patients in the Māori and Pasifika communities so we can include these in any future computer-based prediction models such as machine learning models to ensure we avoid worsening health disparities.

This Participant Information Sheet will help you decide if you'd like to take part. It sets out why we are doing the study, what your participation would involve, what the benefits and risks to you might be, and what will happen after the study ends. We will go through this information with you and answer any questions you may have. You do not have to decide today whether or not you want to participate in this study. Before you decide you may want to talk about the study with other people, such as family, whānau, friends, or healthcare providers. Feel free to do this.

If you agree to take part in this study, you will be asked to sign the Consent Form on the last page of this document. You will be given a copy of both the Participant Information Sheet and the Consent Form to keep.

Why are we doing this study?

In one of our previous studies, we found out that Māori females had the highest average length of stay in hospital for asthma and that Māori have higher rates of asthma hospitalisations than non-Māori. Among the ethnic groups, next highest were for the

Europeans and Pacific people. These findings showed how important it is to investigate key risk factors for asthma attacks in Māori and Pasifika, and to find ways to predict attacks before they happen.

Machine learning is an advanced computer technology and type of AI that has been used for predicting asthma attacks. However, these models need to capture Māori-specific and Pasifika-specific risk factors to ensure that they work just as well in Māori, Pasifika and others. To do this, we need to find out how Māori and Pasifika can be best represented in machine learning models in a way that does not exacerbate existing health inequities.

The aim of our study is to explore the views and attitudes of Māori and Pasifika towards the use of AI and technology in asthma, and identify the risk factors specific to Māori and Pasifika asthma patients, which are important to include in the asthma risk prediction models. We will use this information to help us develop a machine-learning-based risk score for Māori and Pasifika patients that can tell us a person's risk of asthma attacks.

What will my participation in the study involve?

You have been invited to participate as you have indicated that you are interested in taking part in this study, are aged 18 years or older, have asthma and identify as Māori and/or Pasifika ethnicity.

If you choose to participate, you will be asked to share information about your experience of asthma and opinions about AI and technology use in asthma and what risk factors there may be for asthma attacks that may be unique to the Māori and Pasifika communities. This will take place as either a one-on-one or focus group interview, where a group of participants will be engaged in the discussion. You can choose which format you wish to be interviewed depending on what suits you best. For zoom video interviews, you will need to return the signed consent form by taking a photo or scan of the form and returning this by email.

Where will the study take place?

For one-on-one interviews, this can take place in-person at either The University of Auckland or Auckland University of Technology or over Zoom video conferencing, and will last about 30-60 minutes, at a time that suits you. You are welcome to have a support person attend if you wish. For focus group interviews, this will consist of up to 8 people per group and will be conducted in person either at The University of Auckland or Auckland University of Technology. The focus group will last up to a maximum of 2 hours.

How will my information be collected?

To make sure we can accurately remember the discussion points we will need to either audio or video record the conversation, whichever you are most comfortable with. Audio/video recordings will be retained for six years after the interview. For one-on-one interviews, we can take notes instead of recording if you prefer. If you choose to be in a focus group, it is not possible to decline to be recorded or ask for recordings to be stopped as there will be others in the group who have consented to recording. However you can choose to remain silent or leave the room if you do not wish to answer a question, or withdraw participation (though it will not possible to withdraw your data from the focus group – see below section about your rights).

The interview recordings will be transcribed by the interviewer and /or a professional transcribing company to help us identify your views and identify common risk factors for asthma attacks. To ensure we keep your views as confidential as possible, we will assign a participant number to the recordings and the transcripts so we cannot identify you (i.e. the data is coded (de-identified)). The recordings and the transcriptions will be stored securely on a password-protected server on the University of Auckland research network. Only the researchers on the team will have access to the information from the interview; the recordings and transcript data will be deleted after six years. Any third-party transcribers used to transcribe the interviews will be asked to sign a confidentiality agreement to protect the privacy of the participants. Note that for focus groups, it is not possible to guarantee confidentiality due to the group discussion nature but we do ask all participants in the focus group to not discuss what was said with others, and the group facilitator will actively encourage all participants to maintain confidentiality.

What will be discussed during the interview / focus group?

During the interview or focus group, we will ask you to describe what your experiences have been with asthma, what you think about AI and technology use, and what risk factors are related to asthma attacks in Māori and Pasifika. We will ask you for some background information about yourself including age, gender, support for daily activities, and confidence with using technology. If you have any questions about your health, we encourage you to discuss these with your GP or other health care team involved in your regular care to address these.

What will happen to my information after?

Your views will be combined with other participants in this study to work out how we can best ensure that the Māori and Pasifika communities are well represented in machine

learning models developed for predicting the risk of asthma attacks. The information will be summarized in a written report or presentation. All contributions will remain deidentified (coded by a unique participant number) and none of the material in the results will personally identify you.

What are the rights of participants in the study?

Whether or not you take part is your choice. If you don't want to take part, you don't have to give a reason, and it won't affect the care you receive. If you do want to take part now, but change your mind later, you can pull out of the study at any time – before, during or following the interview. You do not need to give a reason for withdrawing. However, what happens to your information when you withdraw will be different depending on whether you choose to be part of a one-on-one interview or focus group:

- If you have been interviewed one-on-one, then you have one week to withdraw your responses, after which your data will have been combined with others and cannot be separated out.
- If you took part in a group interview (focus group), it will not be possible to withdraw your data as it has been collected along with other participants and it risks compromising the integrity of the data from other participants who do wish to continue to be part of the research.

If you choose to be interviewed one-on-one, you can also have a chance to review the transcripts of the interview recording and make edits to this if you wish – if you would like to do this, please provide your email address below. You will have 7 calendar days to provide comments on the transcripts.

For focus group interviews, it will not be possible to review and edit transcripts due to the group nature of the discussions as other participant's information will be in the transcript.

All patients who take part can be emailed a short summary of the results when they are completed – if you wish to receive this, please provide your email address at the end of the consent form.

All participants will be offered at $30 grocery (supermarket) voucher to thank them for their time and information provided. Please provide a postal address to send this at the end of the consent form if you are not doing the interview in-person.

Who do I contact for more information or if I have concerns?

If you have any questions, concerns or complaints about the study at any stage, you can contact the lead researcher or research team:

Farhaan Mirza,
Associate Professor,
School of Computer and Mathematical Sciences,
Auckland University of Technology
farhaan.mirza@aut.ac.nz

Arieta Fa'apesolo
Pacifika Researcher
School of Computer and Mathematical Sciences,
Auckland University of Technology
Arieta.Fa'apesoloMu'aulama@middlemore.co.nz

Amy Chan,
Associate Professor
School of Pharmacy, University of Auckland
a.chan@auckland.ac.nz

Darsha Widana,
Student Researcher,
School of Computer and Mathematical Sciences,
Auckland University of Technology
darsha.jayamini@autuni.ac.nz

If you require cultural support, talk to your whānau in the first instance. If you require further support, please contact us and we can put you in touch with our Māori and Pasifika advisors Marie-Claire Bidois-Putt and Raymond Lutui for this project.

For concerns of an ethical nature, you can contact the Chair of the Auckland Health Research Ethics Committee at ahrec@auckland.ac.nz or at 373 7599 x 83711, or at Auckland Health Research Ethics Committee, The University of Auckland, Private Bag 92019, Auckland 1142

Thank you for reading this and considering participating.

This study has been part-funded by the Health Research Council as part of their Ethics Studentship programme (HRC 22-925).

Approved by the Auckland Health Research Ethics Committee on 23/3/2023 for three years. Reference number AH25484.

Appendix S

Pasifika Interviews - Semi-Structured Guide

The following open-ended questions will be posed to the participants under each section.

1. **Introduction**

   Introduce the interviewer, study purpose, and rights while going through the information sheet and the consent form.

   Asthma attacks affect many people in New Zealand. We are looking at ways to improve asthma care by being able to predict asthma attacks before they happen so that we can keep people well and out of the hospital. One way to do this is to use technology.

   The study aims to explore your thoughts and feelings about the use of technology in asthma, and if there are any specific risk factors for attacks in the Pasifika that we should think about. Firstly: Tell me about you (pepeha)

2. **Experience with Asthma**

   We will start by talking about your experience with Asthma. - How do you feel about living with Asthma? How has it affected your life?

   For instance, how often do you take medications (such as inhalers)? How often do you go to meet GPs or go to the hospital for asthma?

3. **Factors related to Asthma attacks**

   Let us talk about the things related to attacks

   - What happens when you have an attack?
   - What do you do when you have an attack?

- How do you feel?
- What triggers an asthma attack for you?
For instance: Pollen, Temperature, Wind, Pets, Stress

Have you tried avoiding those triggers, and what have you experienced in doing that?

Were you able to avoid them, or was there anything that prevented you from doing that?

4. Predicting asthma attacks using technology

One way of predicting asthma attacks is to use technology such as smart devices to monitor asthma (e.g., smart inhalers) and to use artificial intelligence to predict when an attack may happen.

- What do you think about using AI and technology to help manage your asthma? Are you aware of what AI is? What do you know about AI and its use in healthcare? What do you think of it?
- How would you feel if AI were used to help predict when your next attack might happen?
- What may be pros/cons for you?
- What factors should be considered when using technology such as AI in Pasifika asthma patients?

Why do you think those factors need to be considered?

What may stop you from using technology? What may help?

5. Considerations for Pasifika when developing risk prediction models

- What factors should be considered when developing risk prediction models for asthma attacks?

How do you think Pasifika culture and beliefs may be recognized and incorporated when considering risk prediction for asthma attacks?

6. Conclusion

- Is there anything else you would like to tell me that you think would help us in using AI to predict asthma attacks in Pasifika patients?
- Will I be able to contact you later in case I have additional questions?

Appendix T

Pasifika Interviews - Ethics Approval

AUCKLAND HEALTH RESEARCH ETHICS COMMITTEE (AHREC)

08/08/2024

Dr Amy Chan
Pharmacy

Re: Request for amendment of Ethics Approval (Our Ref. AH25484): Amendments approved

The Committee considered the amendment request for your study entitled "**Attitudes towards use of AI in asthma in Māori**".

Approval was granted for the following amendments:

1. To extend recruitment to include another 20-30 participants who identify as Māori-Pacific or Pacific ethnicity.
2. To extend recruitment to throughout 2025.
3. To update advertisement with above changes.

The ethics approval for this project expires on **23/03/2026**.

Further amendments to the approved project: Should you need to make any further changes to the approved project, please follow the steps below:

- Send a request to the AHREC Administrators to unlock the application form (using the Notification tab in the Ethics RM form).
- Make all changes to the relevant sections of the application form and attach revised documents (as appropriate).
- Change the Application Type to "Amendment request" in Section L.
- Add a summary of the changes requested in the text box.
- Submit the amendment request (PI/Supervisors only to submit the form).

If the project changes significantly, you are required to submit a new application.

The Chair and the members of AHREC would be happy to discuss general matters relating to ethics approvals. If you wish to do so, please contact the AHREC Ethics Administrators at ahrec@auckland.ac.nz in the first instance.

Additional information:

- Do not forget to fill in the 'approval wording' on the Participant Information Sheets, Consent Forms and/or advertisements, giving the dates of the initial approval and the reference number before you use them or send them out to your participants.

All communications with the AHREC regarding this application should indicate this reference number: **AH25484**.

AHREC Administrators

Auckland Health Research Ethics Committee

Appendix U

Pasifika Interviews - Consent Form

Attitudes towards use of artificial intelligence (AI) to predict asthma attacks in Māori and Pasifika and exploration of risk factors affecting asthma attacks: A qualitative study

- I have read and I understand the Participant Information Sheet.
- I have been given sufficient time to consider whether or not to participate in this study.
- I have had the opportunity to use a legal representative, I/ family support or a friend to help me ask questions and understand the study.
- I understand that taking part in this study is voluntary (my choice) and that I may withdraw (not continue to participate in the study) at any time (before, during or after the interview) without this affecting my care.
- If I have been interviewed one-on-one and I decide to withdraw from the study, I understand I have a week to ensure my information is not included in the study, and that after that time, the information collected about me will be collated with other participants' information and therefore cannot be separated and excluded.
- If I have been interviewed in a focus group, I understand that it will not be possible to withdraw my information as it has been collected as part of a larger group with other participants, and therefore cannot be separated and excluded
- I consent to the research staff collecting my views through an interview either one-on-one or in a focus group, including information about my health and asthma experiences.
- I agree to the interview being recorded either in audio or video form, or for one-on-one interviews, in written notes
- I understand that for focus groups, confidentiality cannot be guaranteed
- I understand that focus groups will be recorded and it is not possible to decline recording of focus groups

- I agree to the interviewer or a professional transcription company transcribing my interview recording
- For one-one-one interviews: I wish to review the transcript of the interview
- For one-one-one interviews, I understand that edits of the interview transcript only must be returned within 7 calendar days of receipt
- I understand that all interview data will be stored in a de-identified form
- I understand that my participation in this study is confidential and that no material, which could identify me personally, will be used in any reports on this study.
- I know who to contact if I have any questions about the study.
- I understand my responsibilities as a study participant.
- I wish to receive a summary of the results from the study.

Email I would like the transcript and results summary sent to (please write):

_____

Postal address I would like the supermarket voucher to be sent to (please write):

_____

Declaration by participant:
I consent to take part in this study.

Participant's name: _____
Signature: _____
Date: _____

Declaration by a member of the research team: I have given a verbal explanation of the research project to the participant and have answered the participant's questions about it. I believe that the participant understands the study and has given informed consent to participate.
Researcher's name: _____
Signature: _____
Date: _____

Approved by the Auckland Health Research Ethics Committee on 23/3/2023 for three years. Reference number AH25484.

Appendix V

COREQ Checklist for Reporting Pasifika Qualitative Study

COREQ (COnsolidated criteria for REporting Qualitative research) Checklist

A checklist of items that should be included in reports of qualitative research. You must report the page number in your manuscript where you consider each of the items listed in this checklist. If you have not included this information, either revise your manuscript accordingly before submitting or note N/A.

Topic	Item No.	Guide Questions/Description	Reported on Page No.
Domain 1: Research team and reflexivity			
Personal characteristics			
Interviewer/facilitator	1	Which author/s conducted the interview or focus group?	114
Credentials	2	What were the researcher's credentials? E.g. PhD, MD	114
Occupation	3	What was their occupation at the time of the study?	114
Gender	4	Was the researcher male or female?	114
Experience and training	5	What experience or training did the researcher have?	114
Relationship with participants			
Relationship established	6	Was a relationship established prior to study commencement?	115
Participant knowledge of the interviewer	7	What did the participants know about the researcher? e.g. personal goals, reasons for doing the research	115
Interviewer characteristics	8	What characteristics were reported about the inter viewer/facilitator? e.g. Bias, assumptions, reasons and interests in the research topic	115
Domain 2: Study design			
Theoretical framework			
Methodological orientation and Theory	9	What methodological orientation was stated to underpin the study? e.g. grounded theory, discourse analysis, ethnography, phenomenology, content analysis	116
Participant selection			
Sampling	10	How were participants selected? e.g. purposive, convenience, consecutive, snowball	114
Method of approach	11	How were participants approached? e.g. face-to-face, telephone, mail, email	115
Sample size	12	How many participants were in the study?	116
Non-participation	13	How many people refused to participate or dropped out? Reasons?	N/A
Setting			
Setting of data collection	14	Where was the data collected? e.g. home, clinic, workplace	116
Presence of non-participants	15	Was anyone else present besides the participants and researchers?	115
Description of sample	16	What are the important characteristics of the sample? e.g. demographic data, date	118
Data collection			
Interview guide	17	Were questions, prompts, guides provided by the authors? Was it pilot tested?	115
Repeat interviews	18	Were repeat inter views carried out? If yes, how many?	N/A
Audio/visual recording	19	Did the research use audio or visual recording to collect the data?	114
Field notes	20	Were field notes made during and/or after the inter view or focus group?	114
Duration	21	What was the duration of the inter views or focus group?	115
Data saturation	22	Was data saturation discussed?	115
Transcripts returned	23	Were transcripts returned to participants for comment and/or	N/A

Topic	Item No.	Guide Questions/Description	Reported on Page No.
		correction?	
Domain 3: analysis and findings			
Data analysis			
Number of data coders	24	How many data coders coded the data?	116
Description of the coding tree	25	Did authors provide a description of the coding tree?	117,119
Derivation of themes	26	Were themes identified in advance or derived from the data?	117
Software	27	What software, if applicable, was used to manage the data?	N/A
Participant checking	28	Did participants provide feedback on the findings?	N/A
Reporting			
Quotations presented	29	Were participant quotations presented to illustrate the themes/findings? Was each quotation identified? e.g. participant number	119-126
Data and findings consistent	30	Was there consistency between the data presented and the findings?	130-134
Clarity of major themes	31	Were major themes clearly presented in the findings?	117
Clarity of minor themes	32	Is there a description of diverse cases or discussion of minor themes?	119-126

Developed from: Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care*. 2007. Volume 19, Number 6: pp. 349 – 357

Once you have completed this checklist, please save a copy and upload it as part of your submission. DO NOT include this checklist as part of the main manuscript document. It must be uploaded as a separate file.

Page numbers refer to the thesis.