

## Article

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

Widana Kankanamge Darsha Jayamini <sup>1,2</sup> , Farhaan Mirza <sup>1</sup> , M. Asif Naeem <sup>3</sup> and Amy Hai Yan Chan <sup>4,\*</sup> <sup>1</sup> School of Engineering, Computer and Mathematical Sciences, Auckland University of Technology, Auckland 1010, New Zealand<sup>2</sup> Department of Software Engineering, Faculty of Computing and Technology, University of Kelaniya, Kelaniya 11600, Sri Lanka<sup>3</sup> School of Computing, National University of Computer and Emerging Sciences (NUCES), Islamabad 44000, Pakistan<sup>4</sup> School of Pharmacy, Faculty of Medical and Health Sciences, University of Auckland, Auckland 1010, New Zealand

\* Correspondence: a.chan@auckland.ac.nz

**Abstract:** Length of stay (LOS) is a key indicator of healthcare quality and reflects the burden on the healthcare system. However, limited studies have used machine learning to predict LOS in asthma. This study aimed to explore the characteristics and associations between asthma-related admission data variables with LOS and to use those factors to predict LOS. A dataset of asthma-related admissions in the Auckland region was analysed using different statistical techniques. Using those predictors, machine learning models were built to predict LOS. Demographic, diagnostic, and temporal factors were associated with LOS. Māori females had the highest average LOS among all the admissions at 2.8 days. The random forest algorithm performed well, with an RMSE of 2.48, MAE of 1.67, and MSE of 6.15. The mean predicted LOS by random forest was 2.6 days with a standard deviation of 1.0. The other three algorithms were also acceptable in predicting LOS. Implementing more robust machine learning classifiers, such as artificial neural networks, could outperform the models used in this study. Future work to further develop these models with other regions and to identify the reasons behind the shorter and longer stays for asthma patients is warranted.

**Keywords:** asthma; length of stay; machine-learning; prediction

**Citation:** Jayamini, W.K.D.; Mirza, F.; Naeem, M.A.; Chan, A.H.Y. State of Asthma-Related Hospital Admissions in New Zealand and Predicting Length of Stay Using Machine Learning. *Appl. Sci.* **2022**, *12*, 9890. <https://doi.org/10.3390/app12199890>

Academic Editors: Jose Antonio Cañas and Blanca Cárdena

Received: 22 August 2022

Accepted: 28 September 2022

Published: 1 October 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Asthma is a common long-term condition due to abnormal airway functionality with airflow limitations and wide variations during a short time [1]. Those symptoms are generally intermittent and reversible with proper treatment [2]. Globally, asthma affects around 339 million people and causes about 455,000 deaths [3,4]. Furthermore, 100 million people are expected to be affected by asthma by 2025 [4]. Asthma affects children and adults but often starts in childhood [5]. While many studies have been conducted to manage childhood asthma [6–9] and adult asthma [10], limited research has explored how LOS may impact asthma treatment and outcomes. The study [11] 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 have 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 [12–14], workplace conditions, and severe adverse life events [15]. If asthma is not adequately managed, it can lead to sudden exacerbations needing emergency health care and hospital admission for treatment and monitoring [12]. 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 [16]. New Zealand (NZ) is one of the countries with the highest prevalence of asthma, where more than 597,000 people are on medical treatment for asthma [17]. Asthma is a common reason why children get admitted to hospitals in NZ [18]. According to NZ health surveys, one in every seven children (13%) and one in every eight adults (12%) take medical treatments for asthma [19]. By comparing the characteristics of patients admitted with critical asthma syndrome (CAS) to intensive care units (ICUs) in the USA, Australia, and NZ (ANZ), ANZ patients made up a more significant percentage of ICU patients and had more prolonged ICU and hospital stays [20]. 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 a more extended stay in the hospital for treatment. Length of Stay (LOS) can be expressed as the number of days that patients stay in 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 [21]. Furthermore, hospital stays incur a considerable cost, impacting the economy to the extent of increased unpaid bills and bankruptcy rates [22]. According to a study, direct cost estimation of \$103 m and \$62 m has been made for paediatric asthma-related hospital admissions and prescriptions in NZ, respectively, for the period 2010–2019 [23]. The estimated total cost for ED and OP visits in NZ in 2020 is nearly \$163 m [24]. Moreover, patient transportation during medical emergencies may cost healthcare providers and patients, which could be avoided if asthma hospitalisations and transfers were minimised [25]. 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 machine learning 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 z-score to handle outliers. We grouped some feature values into major categories and excluded minor categories. Before feeding data into machine learning models, we encoded the categorical values using one-hot encoding 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 machine learning algorithms, we extracted instances with LOS in the range of 1 to 14 days and then split them into training and test sets. Different machine learning models were developed and validated using 5-fold cross-validation. 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, MSE and MAE, which are error terms commonly used to evaluate regression models. We selected the model having 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 paper has the following sections: Section 2 presents the related work to this study. Section 3 covers the methodology for analysing asthma-related admission data and predicting LOS using machine learning. Section 4 presents the analysis results and machine learning models' performance. In Section 5, we discuss the results obtained. Finally, in Section 6, we conclude and propose some future perspectives.

## 2. 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 [26,27], all of which have statistically significant associations with LOS [26]. They have found age, gender, and day of the week as important predictors [26,27]. Additionally, respiratory-related secondary diagnoses, year of admission [26], obstructive sleep apnoea, complex chronic conditions, and season (winter) [27] 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<sub>2.5</sub> significantly correlate with LOS [28].

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%,  $p < 0.001$ ) [23]. This shows that there is an apparent ethnicity 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), SVM, random forest, and gradient boosting Models to predict LOS for other long-term health conditions. One study [29] 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, SVM, and random forest. 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 [30]. Another study [31] 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 [32]. They used k-nearest neighbours (KNN), naïve Bayes, and Decision Tree (DT) machine learning algorithms to predict LOS and concluded that DT performs well with 85% accuracy. The study [33] 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 [33] obtained an area under the receiver operating curve (AUROC) of 0.82 for both SVM 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 intensive care unit (ICU)

can be discharged within 10 days, the study [34] 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, machine learning has not yet been applied to predict asthma LOS with a real dataset.

### 3. Materials and Methods

#### 3.1. 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. 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 Disease (ICD-10AM) diagnostic codes relating to asthma Table 1.

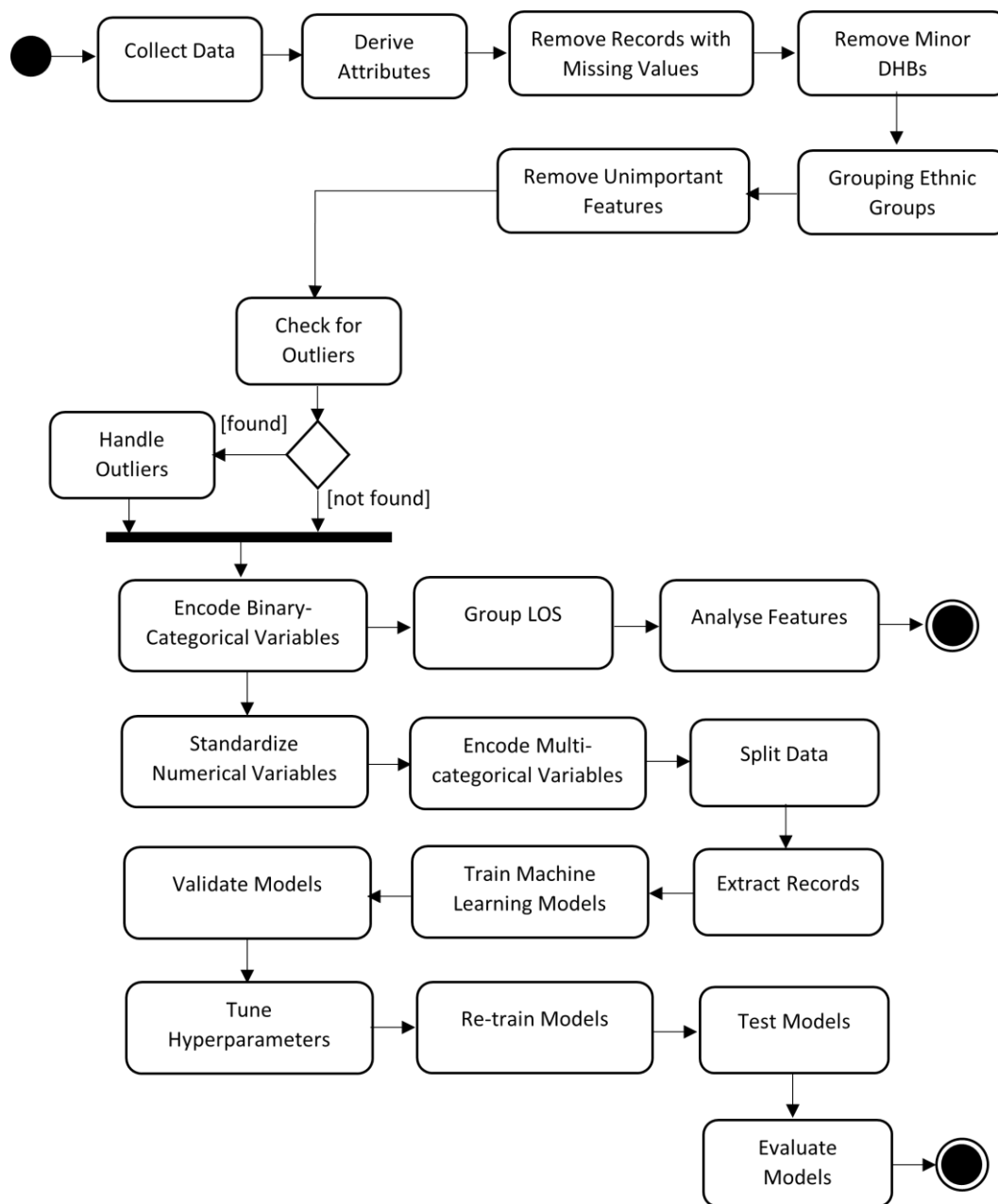
**Table 1.** Diagnostic codes and descriptions.

| 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                      |

The dataset comprised 11,414 anonymised records of children (<18 years) and adults with asthma. Admission data were recorded from the four District Health Boards (DHBs) in the Auckland region in NZ (including Northland). Admission data included the socio-demographic and diagnostic variables of the patients.

#### 3.2. Data Pre-Processing

The dataset was cleaned and preprocessed before analysing and developing machine learning 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 1 shows the activity diagram for the methodology followed in data analysis and machine learning model development.



**Figure 1.** Activity diagram for feature analysis and machine learning model development methodology.

### 3.3. 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 [35]. Ethnicity is self-identified and is registered at each health interaction rather than being a classification that remains unchanged over time [36]. 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, Other, and European [37]. 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, Middle Eastern/Latin

American/African (MELAA), and Other ethnicity. Categorical features with binary values, gender and smoker status, were encoded to binary values (0 and 1) to represent them in a numeric form.

### 3.4. 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 6559 and 4789 records, respectively.

### 3.5. Exploring Feature Set and Development of Machine Learning 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 admit month were correlated with the LOS group. However, admit month was not removed from the current work to retain the seasonality factor in the dataset.

After performing the analysis, different machine learning models were developed to predict asthma LOS. For this, we used only the LOS records between 1 and 14, including boundaries, as there were a minimal number of records with higher LOS values which was not suitable for training the machine learning models. 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 admit day of the week to replace the original variable. Since age had continuous values, this was standardised before applying to the machine learning models.

The dataset was split into training and test sets of 70% and 30%, respectively. The machine learning models developed were SVM, random forest (RF), and KNN. The boosting algorithm, Extreme Gradient Boosting (XGBoost), was used to build the machine learning models. These machine learning models were developed using the training set, and hyperparameter tuning was conducted using the GridSearchCV technique with 5-fold cross-validation. 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 SVM, gamma, kernel, degree, and C (penalty for misclassifying a data point) were tuned, while hyperparameters tuned for KNN were 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), mean absolute error (MAE) and mean squared error (MSE) were used to evaluate the model performance. Analysis and machine learning 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 machine learning models is available at

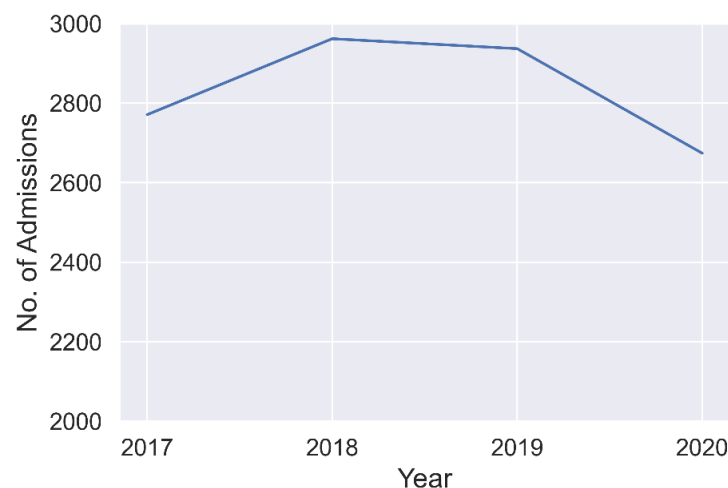
<https://github.com/DarshaWK/Predict-asthma-LOS-using-ML> (accessed on 17 September 2022).

## 4. Results

### 4.1. 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%) were 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%.

A 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 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.

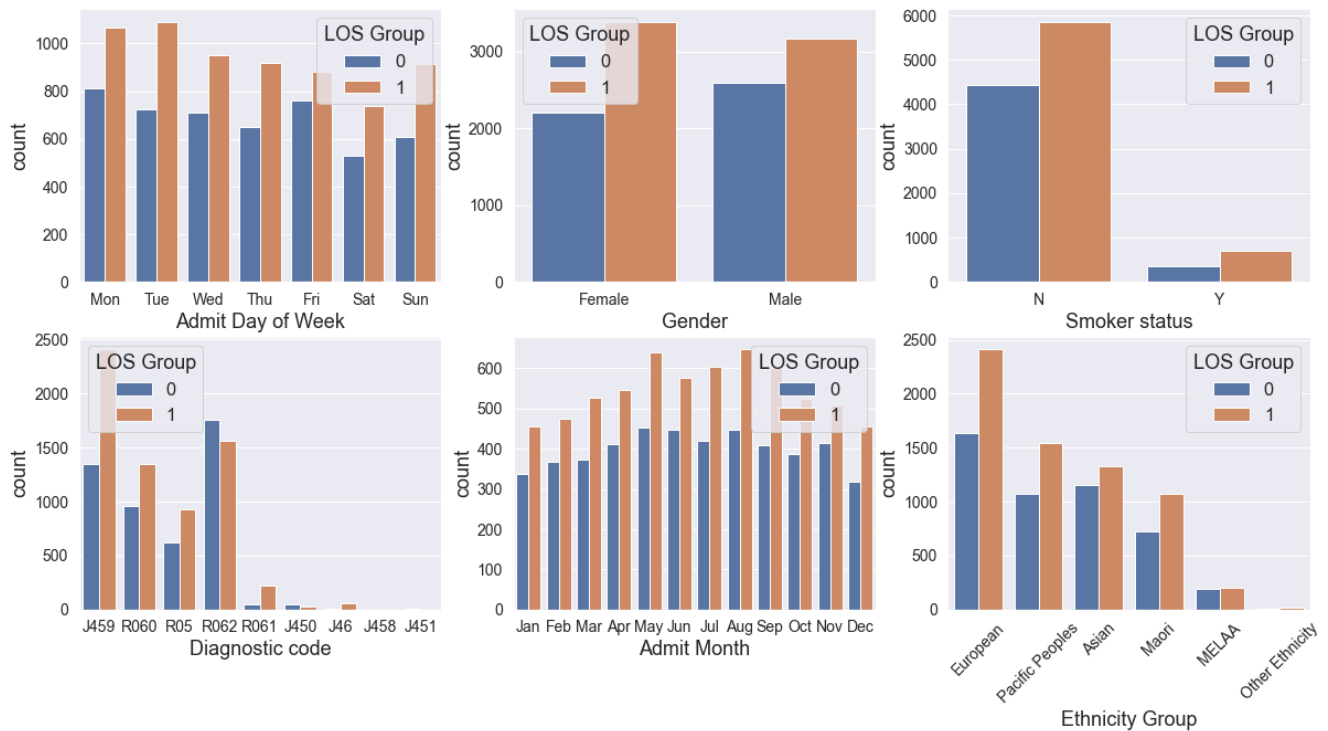


**Figure 2.** The yearly trend in the number of asthma admissions.

Concerning the distribution of LOS, most LOS were of short duration. 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%).

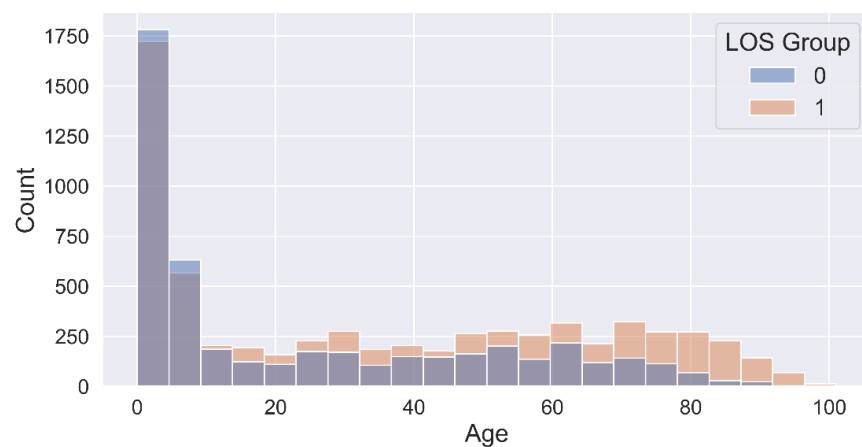
The distribution of different features under each of the LOS groups is demonstrated in Figure 3. The y-axes of the graphs represent the number of admissions/count, while the x-axes have different values for each feature. A dual colour code (orange and blue) has been used to differentiate the LOS groups (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 ethnicity 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 ethnicity groups stay for more than 1 day.



**Figure 3.** Distribution of feature values among the LOS groups.

As seen in Figure 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.

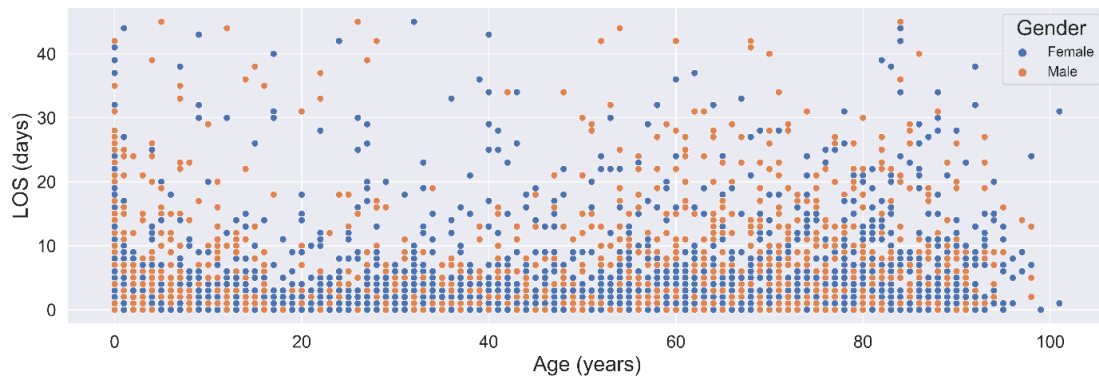


**Figure 4.** Age distribution among the LOS groups.

#### 4.2. Association of Features with LOS

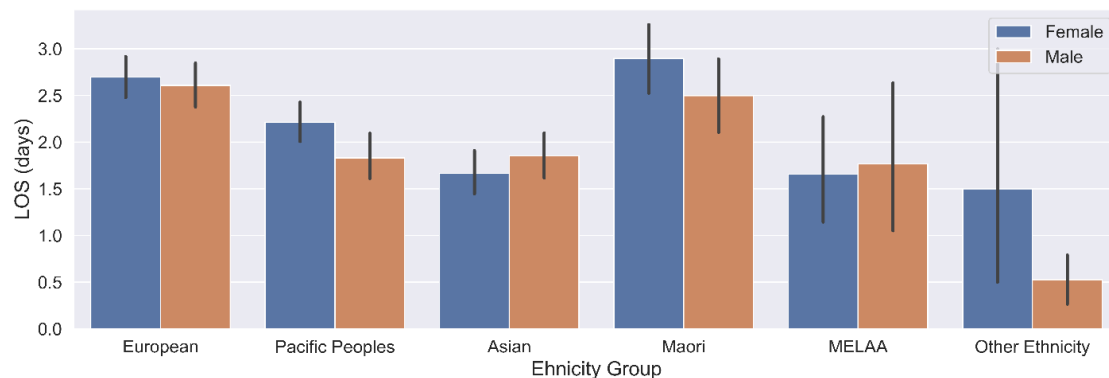
This study observed several key features concerning asthma admission in NZ. Admission rates for children aged 0–5 years were significantly higher than for other ages. Figure 5

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 [2].



**Figure 5.** Correlation between age, LOS, and gender.

The central tendency of LOS against ethnic groups is illustrated in Figure 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 ethnicity. For all the ethnic groups, the mean LOS for females was 1.5 days or more. The Other ethnicity group had the lowest mean LOS values for female and male patients. Among all the ethnicity 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 A1 shows the  $p$ -values of correlation between the feature set and the LOS group, which were calculated from the correlation finding tests.



**Figure 6.** Correlation between ethnicity, average LOS, and gender.

#### 4.3. Development of Machine Learning Models

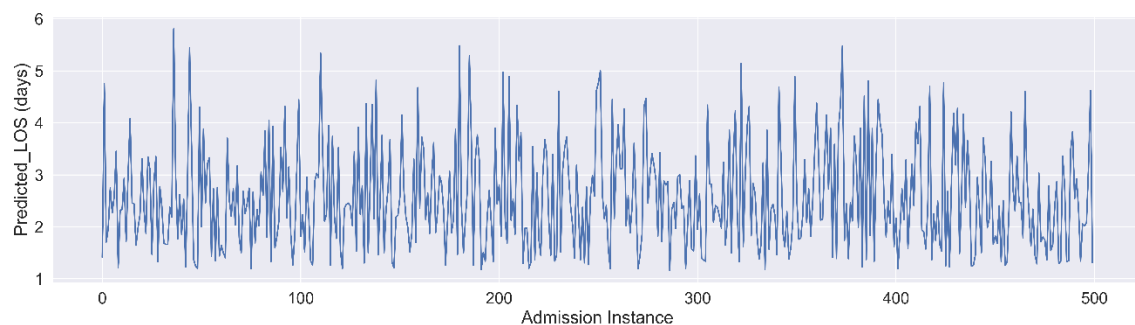
Several machine learning models were developed to predict the LOS using the asthma admission dataset with 5862 instances and with the LOS ranging from 1 to 14 days. As LOS is a continuous variable, machine learning regressors were developed to predict the LOS of asthma patients. Table 2 shows the performance of the models based on the evaluation metrics RMSE, MAE and MSE. As all these are error values, the models with lower error values perform better. The random forest model predictions made for the first 500 instances of the test data are plotted in Figure 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 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 ethnicity group had a larger standard deviation which needs to be further analysed, only considering minority groups.

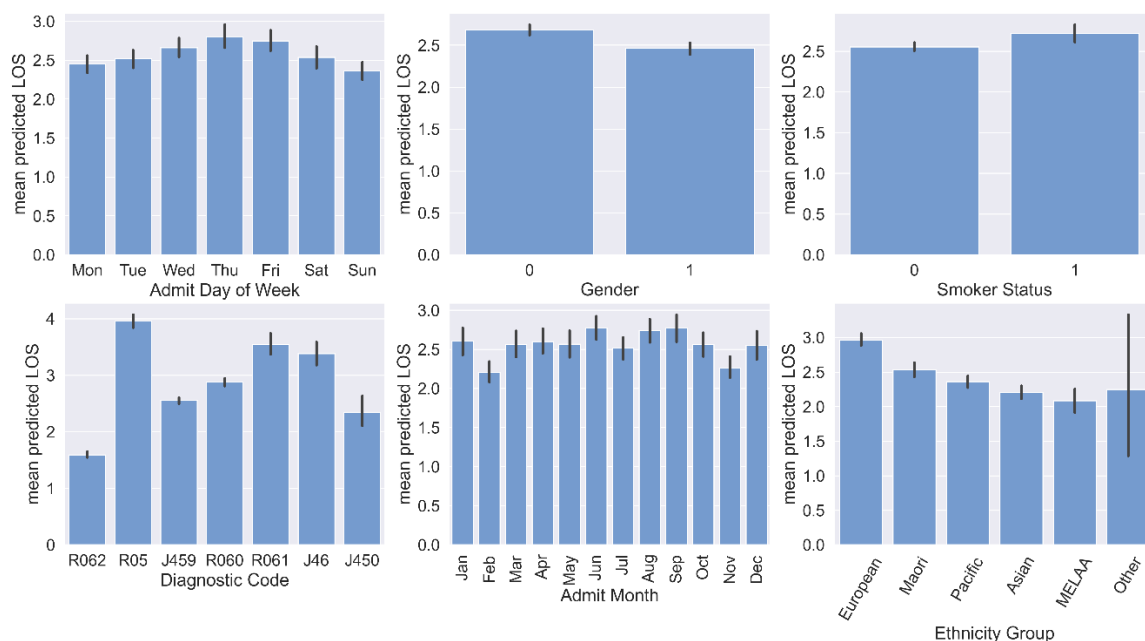
**Table 2.** Performance of the machine learning models in predicting LOS.

| Model         | RMSE <sup>1</sup> | MAE <sup>2</sup> | MSE <sup>3</sup> |
|---------------|-------------------|------------------|------------------|
| SVM           | 2.65              | 1.50             | 7.03             |
| Random Forest | 2.48              | 1.67             | 6.15             |
| KNN           | 3.08              | 1.58             | 9.46             |
| XGBoost       | 2.50              | 1.67             | 6.26             |

<sup>1</sup> RMSE—Root Mean Squared Error. <sup>2</sup> MAE—Mean Absolute Error. <sup>3</sup> MSE—Mean Squared Error.



**Figure 7.** Predicted LOS for the first 500 admission instances in the test data.



**Figure 8.** Mean predicted LOS for different feature values.

## 5. 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 [26]. Research from [10] 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 [39,40] 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 [41]. Even indoor air pollution in urban schools due to PM<sub>2.5</sub> and PM<sub>10</sub> will cause respiratory disorders among children, leading to more hospital admissions on weekdays [42]. 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 [23]. However, the previous study [23] 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 [12,13,43]. These factors change within 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 Summer. These findings are essential to help health workforce resourcing and planning to meet predicted hospital demand. A previous study [23] 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 [44]. 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 [17]. 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 [27]. 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 machine learning 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 machine learning models, the random forest algorithms performed better with an RMSE of 2.48 and MSE of 6.15. The predictions of the random forest model had a mean LOS of 2.6 days with a standard deviation of 1.0, which is acceptable. XGBoost also performed well, giving an RMSE of 2.50 and MSE of 6.26. Both algorithms showed identical values for the performance metric MAE as 1.67. However, SVM had the lowest MAE of 1.50. Although KNN was the least performing algorithm, it gave an RMSE of 3.08, MAE of 1.58 and MSE of 9.46, which are acceptable. This implies that machine learning 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 machine learning models such as artificial neural networks and deep learning algorithms could be applied to predict LOS, which has shown better performance in medical prediction and recommendation systems [9,12,13,45–49].

## 6. Conclusions and Future Direction

Asthma is a respiratory disease that is a significant public health concern. We analysed a real-world asthma-related dataset in NZ to explore the characteristics and build several machine learning models to predict LOS. Our results found that demographic, diagnosis and temporal factors are highly associated with asthma LOS. It was found that the highest average LOS was for Māori females. This emphasises that special attention could be separately given to the Māori community, the indigenous population in NZ, to explore any unique factors related to asthma LOS. Overall, asthma LOS prediction should be improved using more robust algorithms applied in real environments.

Accordingly, future research could investigate the admissions of other minority ethnic groups and other DHBs separately. Also, corresponding environmental, meteorological, and socioeconomic data could be combined with this real-world dataset to explore the impact of those new factors in predicting asthma LOS. In addition, past data on the previous hospital stays of asthma patients could be included as a factor, as well as data on asthma severity and control. Further, machine learning models could be built to classify multiple asthma-related LOS groups and identify the factors affecting LOS for asthma-related admissions. Moreover, future work could be carried out to explore the reasons behind the shorter and longer hospital stays.

**Author Contributions:** Conceptualization, F.M. and M.A.N.; methodology, W.K.D.J.; validation, W.K.D.J.; formal analysis, W.K.D.J.; investigation, W.K.D.J.; resources, A.H.Y.C.; data curation, A.H.Y.C.; writing—original draft preparation, W.K.D.J.; writing—review and editing, F.M., M.A.N. and A.H.Y.C.; supervision, F.M., M.A.N. and A.H.Y.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding. Data was kindly provided by the Auckland District Health Board with support from the Business Intelligence Unit.

**Institutional Review Board Statement:** This study was reviewed and approved by the Auckland Health Research Ethics Committee, reference number AH22149.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data available on request from authors, within ethics approvals.

**Conflicts of Interest:** A.H.Y.C has received research grant funding from The University of Auckland, Health Research Council and is the recipient of the Auckland Medical Research Foundation Senior Research Fellowship for asthma-related research. A.H.Y.C is also a fellow of the Asthma UK Centre of Applied Research (AUKCAR) and on the Board of Asthma New Zealand. All other authors declare no other relevant conflict of interest.

## Appendix A

**Table A1.** Correlation of features and LOS group.

| Feature           | <i>p</i> -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                    |

## References

- Hargreave, F.; Nair, P. The definition and diagnosis of asthma. *Clin. Exp. Allergy* **2009**, *39*, 1652–1658. [CrossRef] [PubMed]
- Organization for Economic Co-Operation and Development (OECD). Health Care Utilisation: Hospital Average Length of Stay by Diagnostic Categories. Available online: <https://stats.oecd.org/index.aspx?queryid=30165> (accessed on 5 June 2022).
- World Health Organization. Asthma. Available online: <https://www.who.int/news-room/fact-sheets/detail/asthma> (accessed on 7 February 2022).
- Global Asthma Network. *The Global Asthma Report 2018*; Global Asthma Network: Auckland, New Zealand, 2018.
- National Health Service. Asthma. Available online: <https://www.nhs.uk/conditions/asthma/> (accessed on 7 February 2022).
- Navanandan, N.; Hatoun, J.; Celedón, J.C.; Liu, A.H. Predicting severe asthma exacerbations in children: Blueprint for today and tomorrow. *J. Allergy Clin. Immunol. Pract.* **2021**, *9*, 2619–2626. [CrossRef] [PubMed]
- Seol, H.Y.; Shrestha, P.; Muth, J.F.; Wi, C.-I.; Sohn, S.; Ryu, E.; Park, M.; Ihrke, K.; Moon, S.; King, K. Artificial intelligence-assisted clinical decision support for childhood asthma management: A randomized clinical trial. *PLoS ONE* **2021**, *16*, e0255261. [CrossRef]
- Lovrić, M.; Banić, I.; Lacić, E.; Pavlović, K.; Kern, R.; Turkalj, M. Predicting treatment outcomes using explainable machine learning in children with asthma. *Children* **2021**, *8*, 376. [CrossRef]
- Kothalawala, D.M.; Murray, C.S.; Simpson, A.; Custovic, A.; Tapper, W.J.; Arshad, S.H.; Holloway, J.W.; Rezwan, F.I.; STE-LAR/UNICORN investigators. Development of childhood asthma prediction models using machine learning approaches. *Clin. Transl. Allergy* **2021**, *11*, e12076. [CrossRef] [PubMed]
- Perpiñá, M.; Gómez-Bastero, A.; Trisán, A.; Martínez-Moragón, E.; Álvarez-Gutiérrez, F.J.; Urrutia, I.; Blanco-Aparicio, M. Expert consensus recommendations for the management of asthma in older adults. *Med. Clínica* **2022**, *159*, 53.e1–53.e14. [CrossRef]
- Shi, T.; Pan, J.; Katikireddi, S.V.; McCowan, C.; Kerr, S.; Agrawal, U.; Shah, S.A.; Simpson, C.R.; Ritchie, L.D.; Robertson, C.; et al. Risk of COVID-19 hospital admission among children aged 5–17 years with asthma in Scotland: A national incident cohort study. *Lancet Respir. Med.* **2022**, *10*, 191–198. [CrossRef]
- Khatrri, K.L.; Tamil, L.S. Early detection of peak demand days of chronic respiratory diseases emergency department visits using artificial neural networks. *IEEE J. Biomed. Health Inform.* **2018**, *22*, 285–290. [CrossRef]
- Kim, M.S.; Lee, J.H.; Jang, Y.J.; Lee, C.H.; Choi, J.H.; Sung, T.E. Hybrid deep learning algorithm with open innovation perspective: A prediction model of asthmatic occurrence. *Sustainability* **2020**, *12*, 6143. [CrossRef]
- Maung, T.Z.; Bishop, J.E.; Holt, E.; Turner, A.M.; Pfrang, C. 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. *Int. J. Environ. Res. Public Health* **2022**, *19*, 8752. [CrossRef] [PubMed]
- Sandberg, S.; Paton, J.Y.; Ahola, S.; McCann, D.C.; McGuinness, D.; Hillary, C.R.; Oja, H. The role of acute and chronic stress in asthma attacks in children. *Lancet* **2000**, *356*, 982–987. [CrossRef]
- To, T.; Stanojevic, S.; Moores, G.; Gershon, A.S.; Bateman, E.D.; Cruz, A.A.; Boulet, L.-P. Global asthma prevalence in adults: Findings from the cross-sectional world health survey. *BMC Public Health* **2012**, *12*, 204. [CrossRef]
- Health Quality and Safety Commission New Zealand. Asthma. Available online: <https://www.hqsc.govt.nz/our-data/atlas-of-healthcare-variation/asthma/#:~:text=Internationally%2C%20New%20Zealand%20has%20a,reporting%20taking%20current%20asthma%20medication>. (accessed on 4 June 2022).
- Asthma and Respiratory Foundation NZ. Asthma. Available online: <https://www.asthmafoundation.org.nz/your-health/living-with-asthma> (accessed on 7 February 2022).
- Ministry of Health. Annual Data Explorer 2018/19: New Zealand Health Survey. Wellington: Ministry of Health. Available online: <https://minhealthnz.shinyapps.io/nz-health-survey-2018-19-annual-data-explorer> (accessed on 4 June 2022).
- Abdelkarim, H.; Durie, M.; Bellomo, R.; Bergmeir, C.; Badawi, O.; El-Khawas, K.; Pilcher, D. A comparison of characteristics and outcomes of patients admitted to the ICU with asthma in Australia and New Zealand and the United States. *J. Asthma* **2020**, *57*, 398–404. [CrossRef]

21. Luo, L.; Ren, J.; Zhang, F.; Zhang, W.; Li, C.; Qui, Z.; Huang, D. The effects of air pollution on length of hospital stay for adult patients with asthma. *Int. J. Health Plan. Manag.* **2018**, *33*, e751–e767. [[CrossRef](#)]
22. Dobkin, C.; Finkelstein, A.; Kluender, R.; Notowidigdo, M.J. The Economic Consequences of Hospital Admissions. *Am. Econ. Rev.* **2018**, *108*, 308–352. [[CrossRef](#)] [[PubMed](#)]
23. Schlichting, D.; Fadason, T.; Grant, C.C.; O'Sullivan, J. Childhood asthma in New Zealand: The impact of on-going socioeconomic disadvantage (2010–2019). *N. Z. Med. J.* **2021**, *134*, 80–95. [[PubMed](#)]
24. Barnard, L.T.; Zhang, J. *The Impact of Respiratory Disease in New Zealand: 2020 Update*; University of Otago: Dunedin, New Zealand, 2021.
25. Thai, H.-D.; Huh, J.-H. Optimizing patient transportation by applying cloud computing and big data analysis. *J. Supercomput.* **2022**, 1–30. [[CrossRef](#)]
26. Soyiri, I.N.; Reidpath, D.D.; Sarran, C. Asthma length of stay in hospitals in London 2001–2006: Demographic, diagnostic and temporal factors. *PLoS ONE* **2011**, *6*, e27184. [[CrossRef](#)] [[PubMed](#)]
27. Shanley, L.A.; Lin, H.; Flores, G. Factors associated with length of stay for pediatric asthma hospitalizations. *J. Asthma* **2015**, *52*, 471–477. [[CrossRef](#)]
28. Baek, J.; Kash, B.A.; Xu, X.; Benden, M.; Roberts, J.; Carrillo, G. Association between ambient air pollution and hospital length of stay among children with asthma in South Texas. *Int. J. Environ. Res. Public Health* **2020**, *17*, 3812. [[CrossRef](#)]
29. Rachda Naila, M.; Caulier, P.; Chaabane, S.; Chraïbi, A.; Piechowiak, S. A Comparative Study of Machine Learning Models for Predicting Length of Stay in Hospitals. *J. Inf. Sci. Eng.* **2021**, *37*, 1025–1038.
30. Turgeman, L.; May, J.H.; Sciulli, R. Insights from a machine learning model for predicting the hospital Length of Stay (LOS) at the time of admission. *Expert Syst. Appl.* **2017**, *78*, 376–385. [[CrossRef](#)]
31. Tsai, P.-F.; Chen, P.-C.; Chen, Y.-Y.; Song, H.-Y.; Lin, H.-M.; Lin, F.-M.; Huang, Q.-P. Length of Hospital Stay Prediction at the Admission Stage for Cardiology Patients Using Artificial Neural Network. *J. Healthc. Eng.* **2016**, *2016*, 7035463. [[CrossRef](#)]
32. Aghajani, S.; Kargari, M. Determining Factors Influencing Length of Stay and Predicting Length of Stay Using Data Mining in the General Surgery Department. *Hosp. Pract. Res.* **2016**, *1*, 53–58. [[CrossRef](#)]
33. Achilonu, O.J.; Fabian, J.; Bebington, B.; Singh, E.; Nimako, G.; Eijkemans, R.M.; Musenge, E. Use of Machine Learning and Statistical Algorithms to Predict Hospital Length of Stay Following Colorectal Cancer Resection: A South African Pilot Study. *Front. Oncol.* **2021**, *11*, 644045. [[CrossRef](#)] [[PubMed](#)]
34. Ma, X.; Si, Y.; Wang, Z.; Wang, Y. Length of stay prediction for ICU patients using individualized single classification algorithm. *Comput. Methods Programs Biomed.* **2020**, *186*, 105224. [[CrossRef](#)]
35. Cohen, H.A.; Blau, H.; Hoshen, M.; Batat, E.; Balicer, R.D. Seasonality of asthma: A retrospective population study. *Pediatrics* **2014**, *133*, e923–e932. [[CrossRef](#)]
36. Stats NZ. *Ethnicity Standard Classification: Findings from Public Consultation November 2019*; Stats NZ: Wellington, New Zealand, 2020.
37. Boven, N.; Exeter, D.; Sporle, A.; Shackleton, N. The implications of different ethnicity categorisation methods for understanding outcomes and developing policy in New Zealand. *Kōtuitui: New Zealand J. Soc. Sci. Online* **2020**, *15*, 123–139. [[CrossRef](#)]
38. Girden, E.R. *ANOVA: Repeated Measures*; Sage: Thousand Oaks, CA, USA, 1992.
39. Mahony, T.; Harder, V.S.; Ang, N.; McCulloch, C.E.; Shaw, J.S.; Thombly, R.; Cabana, M.D.; Kleinman, L.C.; Bardach, N.S. Weekend Versus Weekday Asthma-Related Emergency Department Utilization. *Acad. Pediatrics* **2022**, *22*, 640–646. [[CrossRef](#)]
40. Krefis, A.C.; Fischereit, J.; Hoffmann, P.; Pinnschmidt, H.; Sorbe, C.; Augustin, M.; Augustin, J. Temporal analysis of determinants for respiratory emergency department visits in a large German hospital. *BMJ Open Respir. Res.* **2018**, *5*, e000338. [[CrossRef](#)]
41. Shakir, M.; Rakesh, N. Investigation on Air Pollutant Data Sets using Data Mining Tool. In Proceedings of the 2018 2nd International Conference on IoT in Social, Mobile, Analytics and Cloud (I-SMAC), Palladam, India, 30–31 August 2018; pp. 480–485.
42. Bennett, J.; Davy, P.; Trompetter, B.; Wang, Y.; Pierse, N.; Boulic, M.; Phipps, R.; Howden-Chapman, P. Sources of indoor air pollution at a New Zealand urban primary school; a case study. *Atmos. Pollut. Res.* **2019**, *10*, 435–444. [[CrossRef](#)]
43. Ram, S.; Zhang, W.; Williams, M.; Pengetnze, Y. Predicting asthma-related emergency department visits using big data. *IEEE J. Biomed. Health Inform.* **2015**, *19*, 1216–1223. [[CrossRef](#)] [[PubMed](#)]
44. Busby, J.; Matthews, J.G.; Chaudhuri, R.; Pavord, I.D.; Hardman, T.C.; Arron, J.R.; Bradding, P.; Brightling, C.E.; Choy, D.F.; Cowan, D.C. Factors affecting adherence with treatment advice in a clinical trial of patients with severe asthma. *Eur. Respir. J.* **2022**, *59*, 2100768. [[CrossRef](#)]
45. Padinjappurathu Gopalan, S.; Chowdhary, C.L.; Iwendi, C.; Farid, M.A.; Ramasamy, L.K. An Efficient and Privacy-Preserving Scheme for Disease Prediction in Modern Healthcare Systems. *Sensors* **2022**, *22*, 5574. [[CrossRef](#)] [[PubMed](#)]
46. Iwendi, C.; Khan, S.; Anajemba, J.H.; Bashir, A.K.; Noor, F. Realizing an efficient IoMT-assisted patient diet recommendation system through machine learning model. *IEEE Access* **2020**, *8*, 28462–28474. [[CrossRef](#)]
47. Lisspers, K.; Stållberg, B.; Larsson, K.; Janson, C.; Müller, M.; Łuczko, M.; Bjerregaard, B.K.; Bacher, G.; Holzhauser, B.; Goyal, P. Developing a short-term prediction model for asthma exacerbations from Swedish primary care patients' data using machine learning-based on the ARCTIC study. *Respir. Med.* **2021**, *185*, 106483. [[CrossRef](#)]

- 
48. Feng, Y.; Wang, Y.; Zeng, C.; Mao, H. Artificial intelligence and machine learning in chronic airway diseases: Focus on asthma and chronic obstructive pulmonary disease. *Int. J. Med. Sci.* **2021**, *18*, 2871–2889. [[CrossRef](#)]
  49. Ramasamy, L.K.; Khan, F.; Shah, M.; Prasad, B.V.V.S.; Iwendi, C.; Biamba, C. Secure Smart Wearable Computing through Artificial Intelligence-Enabled Internet of Things and Cyber-Physical Systems for Health Monitoring. *Sensors* **2022**, *22*, 1076. [[CrossRef](#)]