Published: 14 January 2026
Online AM: 18 September 2025
Accepted: 21 August 2025
Revised: 04 June 2025
Received: 15 November 2024

RESEARCH-ARTICLE

Effective Predictive Modelling for Emergency Department Visits and Evaluating Exogenous Variables Impact: Using Explainable Meta-Learning Gradient Boosting

MEHDI NESHAT, University of Technology Sydney, Sydney, NSW, Australia

MICHAEL PHIPPS, Canberra Hospital, Canberra, ACT, Australia

NIKHIL JHA, School of Medicine and Psychology, Canberra, ACT, Australia

DANIAL KHOJASTEH, UNSW Sydney, Sydney, NSW, Australia

MICHAEL TONG, National Centre for Epidemiology & Population Health, Canberra, ACT, Australia

AMIR HOSSEIN GANDOMI, Khazar University, Baku, Azerbaijan

Open Access Support provided by:

Canberra Hospital

Khazar University

UNSW Sydney

National Centre for Epidemiology & Population Health

School of Medicine and Psychology

University of Technology Sydney

Effective Predictive Modelling for Emergency Department Visits and Evaluating Exogenous Variables Impact: Using Explainable Meta-Learning Gradient Boosting

MEHDI NESHAT, Canberra Health Services, Canberra, Australia and University of Technology Sydney, Sydney, Australia
MICHAEL PHIPPS, Canberra Health Services, Canberra Hospital, Canberra, Australia
NIKHIL JHA, Canberra Health Services, Canberra Hospital, Canberra, Australia and School of Medicine and Psychology, The Australian National University, Garran, Australia
DANIAL KHOJASTEH, School of Civil and Environmental Engineering, University of New South Wales, Sydney, Australia
MICHAEL TONG, National Centre for Epidemiology and Population Health, ANU College of Health and Medicine, The Australian National University, Canberra, Australia
AMIR H. GANDOMI, Faculty of Engineering and Information Technology, University of Technology Sydney, Australia University Research and Innovation Center (EKIK), Sydney, Australia, Obuda University, Budapest, Hungary, and Department of Computer Science, Khazar University, Baku, Azerbaijan

Accurately predicting Emergency Department (ED) visits is essential for optimising resource allocation, including staffing adjustments and Operating Room scheduling. Despite the proliferation of AI-driven models, effective ED visit prediction remains challenging due to limited generalisability, susceptibility to overfitting and underfitting, scalability and the complexity of fine-tuning hyper-parameters. To address these challenges, we propose a novel Meta-Learning Gradient Booster (Meta-ED) approach to forecast daily ED visits. Meta-ED leverages a comprehensive dataset spanning 23 years from Canberra Hospital, incorporating exogenous variables such as socio-demographic characteristics, healthcare usage, chronic diseases, diagnoses and climate parameters. Meta-ED combines four foundational learners—CatBoost, Random Forest, Extra Trees and LightGBM—with a Multi-Layer Perceptron (MLP) as the master-level learner, thereby enhancing predictive precision by integrating the strengths of diverse base models. Our comparative analysis, which involved testing 23 models against a set of predefined criteria, demonstrates the superior performance of Meta-ED, achieving an accuracy of 85.7% (95% CI [85.4%, 86.0%]) and outperforming prominent models like XGBoost, Random Forest, AdaBoost, LightGBM and Extra Trees by up to 106.3%. Furthermore, incorporating climate features resulted in a 3.25% improvement in prediction accuracy, effectively

Authors' Contact Information: Mehdi Neshat, Canberra Health Services, Canberra, Australia and University of Technology Sydney, Sydney, Australia; e-mail: mehdi.neshat@uts.edu.au; Michael Phipps, Canberra Health Services, Canberra Hospital, Canberra, Australia; e-mail: Michael.Phipps@act.gov.au; Nikhil Jha, Canberra Health Services, Canberra Hospital, Canberra, Australia and School of Medicine and Psychology, The Australian National University, Garran, Australia; e-mail: nikhil.jha@act.gov.au; Danial Khojasteh, School of Civil and Environmental Engineering, University of New South Wales, Sydney, Australia; e-mail: danial.khojasteh@unsw.edu.au; Michael Tong, National Centre for Epidemiology and Population Health, ANU College of Health and Medicine, The Australian National University, Canberra, Australia; e-mail: Michael.Tong@anu.edu.au; Amir H. Gandomi (corresponding author), Faculty of Engineering and Information Technology, University of Technology Sydney, Australia University Research and Innovation Center (EKIK), Sydney, Australia, Obuda University, Budapest, Hungary, and Department of Computer Science, Khazar University, Baku, Azerbaijan; e-mail: gandomi@uts.edu.au.

ACM 2637-8051/2026/1-ART3
https://doi.org/10.1145/3768317

capturing seasonal variations that influence patient volumes. These results underscore the potential of Meta-ED to advance predictive analytics in complex healthcare environments.

CCS Concepts: • **Computing methodologies → Machine learning**; • **Applied computing → Health informatics**;

Additional Key Words and Phrases: Emergency department, Retrospective cohort study, Prediction, Exogenous variable, Machine learning, Ensemble learning model, Meta learning

ACM Reference format:
Mehdi Neshat, Michael Phipps, Nikhil Jha, Danial Khojasteh, Michael Tong, and Amir H. Gandomi. 2026. Effective Predictive Modelling for Emergency Department Visits and Evaluating Exogenous Variables Impact: Using Explainable Meta-Learning Gradient Boosting. *ACM Trans. Comput. Healthcare* 7, 1, Article 3 (January 2026), 33 pages.
https://doi.org/10.1145/3768317

1 Introduction

The **Emergency Department (ED)**, distinct from other hospital units, faces challenges in efficiently managing resources due to its inability to regulate patient influx [30]. Serving as the hospital's primary point of entry, the ED confronts time constraints in responding to fluctuations in patient volume [15]. Swift and critical care within the emergency medical systems is expanding. Yet, the persistent issue of ED overcrowding poses a notable risk to patient welfare, particularly for those with chronic ailments and infectious conditions [40, 54]. ED congestion leads to delays in patient transfer and financial setbacks for hospitals as they divert cases to alternate facilities [37]. Precisely forecasting ED visitation rates is pivotal for optimising resource distribution and enabling proactive staffing adjustments, operational modifications and resource planning [31].

Forecasting the number of ED visitors presents a tough challenge primarily as no singular factor exhibits the precision essential for day-to-day or weekly planning [19]. Therefore, a holistic multivariate approach becomes imperative to identify the pivotal predictive elements, encompassing demographic variables, historical ED visit trends, meteorological data [11], air quality metrics [47], major events and other relevant factors. Currently, most healthcare institutions rely on basic algorithms for short-term forecasting of emergency admissions, utilising elementary rolling averages for daily estimations [63].

An uncomplicated approach involves enhancements through the integration of Bayesian methodologies or autoregressive inductive moving averages, incorporating diverse datasets [4, 28]. Nonetheless, Bayesian-driven methodologies may pose computational challenges, particularly with intricate models or extensive data sets, potentially constraining their real-time application feasibility.

Machine Learning (ML) models emerge as a more attractive method for predicting ED presentations [25] than Bayesian and classical prediction approaches (like linear regression [22], Naïve Bayes [24]), Support Vector Regression [76, 78] and Nonlinear Least Square Regression [74] thanks to their adaptability in managing intricate relationships and various data types. This adaptability allows a superior grasp of the complex patterns and interdependencies inherent in ED visitor data.

Some researchers [62, 78] have highlighted the success of time-series forecasting models, including **Long Short-Term Memory (LSTM)** [80], **Convolutional Neural Networks (CNNs)** [39] and stacked structures that integrate **Gated Recurrent Units (GRUs)** [39, 49] and Recurrent **Neural Networks (NNs)** [49]. These sophisticated architectures excel at modelling patient flow dynamics, particularly in cases involving nonlinear and high-dimensional datasets, underscoring their effectiveness in handling complex data.

Attention mechanisms [69] have remarkably propelled the field of time series forecasting forward, particularly with the advent of the Transformer model. In contrast to conventional sequential learning architectures, attention facilitates the elimination of recurrence and memory components by evaluating the entire input sequence simultaneously, pinpointing crucial elements for enhanced predictions. The self-attention feature in Transformers empowers them to adeptly grasp dependencies among distant elements in a sequence, rendering them exceptionally well-suited for managing extensive datasets. Recently, a few studies investigated the application of attention [6] and Transformers [8] to tackle the intricacies of forecasting ED attendance, highlighting their promise in unravelling temporal patterns and interconnections within health-related data.

The integration of external factors, often referred to as exogenous variables, has emerged as a prevalent and effective strategy in numerous recent research endeavours to considerably enhance the accuracy of predicting the number of visitors to ED [3, 6]. This innovative approach, where a variety of influences, including but not limited to climatic conditions and calendar-related information, have been extensively employed, as evidenced by the notable contributions made by researchers [6, 25, 49]. For example, regarding the application of exogenous variables, a recent study [6] developed an NN model with an Attention layer to predict ED admissions, achieving promising results with an R-squared value of 80%. The Attention model is also smaller and more computationally efficient than the LSTM-based model.

Another study [62] meticulously incorporated various weather-related parameters, which encompassed elements like atmospheric temperature, wind velocity, the directional flow of winds and the extent of cloud visibility while also integrating calendar-based data that featured significant time markers, such as public holidays and academic calendars. In addition to this, other scholarly investigations [5, 17] similarly adopted a range of external variables to inform their analyses, whereas certain studies. Beyond these factors, a broader spectrum of variables, such as air quality metrics (as examined by [27]), socio-economic determinants (as highlighted in the research by [33]) and the prevalence of influenza outbreak levels (discussed in the study by [70]), have also been employed to refine the predictive models. Collectively, these diverse and innovative studies underscore the significant value and importance of incorporating a wide array of exogenous data points in order to substantially enhance the overall accuracy of forecasting visitor numbers in EDs. However, combining a wide array of parameters results in heterogeneous data that poses significant challenges for conventional predictive models, as well as for current ML and other popular frameworks [73], in accurately forecasting short-term or long-term ED traffic due to the intricate nature and interactions of these variables.

Based on our literature review, hybrid ML models [7, 21, 31] have demonstrated superior performance over standalone models in predicting ED visitor numbers by combining the strengths of several algorithms. For instance, linear regression models are effective at capturing overall trends, whereas nonlinear models, such as NNs or DLs, are adept at identifying complex interactions and relationships. Integrating these methodologies produces a model capable of both generalising effectively and capturing intricate patterns in the data. One recent study [34] introduced a hybrid ML method for predicting ED attendance by combining linear regression, NNs and regression tree ensembles. The approach uses external variables like socio-economic, weather and temporal data to improve weekly forecasts. Trained on 11 years of data and tested over 1 year sequentially, the model achieved a validation error of around 5%, showing its potential for enhancing ED management. Furthermore, Lopez et al. [34] reported that external data such as climate and socio-economic information could not significantly enhance the accuracy of the prediction.

Although traditional regression models might predict patient visit volumes with acceptable accuracy, it is also as crucial to explain and justify the predictions, especially to healthcare decision-makers. Making the predictive models more interpretable enhances confidence among clinicians, administrators and the general public since it makes it more apparent how and why specific predictions are being produced [49]. This openness is essential for implementing predictive tools effectively in real-world healthcare settings [58]. On this front, **Explainable AI (XAI)** has also been in the spotlight due to its power to unmask the black-box nature of AI-driven models. XAI attempts to transparently reveal the impenetrable algorithms used by sophisticated algorithms by providing precise explanations of prediction construction. While few studies have addressed explainable models in forecasting for ED, earlier studies indicated strong connections between environmental factors—e.g., rain and temperature [1]—and patterns of patient visits. **Shapley Additive Explanations (SHAP)** has also been applied in some studies [51, 64] to analyse the impact of different predictors on patient volumes. Similar interpretability methods have also been applied to examples involving human behaviour, such as the usage of public bikes [48] and attendance at school [67].

Ensemble learning techniques are particularly suitable for predicting ED visitor numbers because they take advantage of the strengths of numerous base learners in order to improve predictive accuracy, stability and robustness [50]. ED visiting patterns are influenced by an inter-related multi-faceted set of factors [57] such as weather, public holidays, outbreaks and demographic characteristics that may not be well-captured by a single model. Ensemble methods—such as bagging [55], boosting [18, 81] and stacking [44]—reduce variance, bias or both by taking multiple algorithms or different training sets and combining them. Ensemble methods help identify nonlinear relationships, handle noisy or imbalanced data, and enhance accuracy, especially in high-variability environments like healthcare. Specific ensemble techniques (e.g., **Random Forest (RF)** [49, 50], **Gradient Boosting (GBR)** [43], XGBoost [12, 42, 77]) provide feature importance statistics as well, which facilitate interpretability, a significant factor in clinical decision-making and policy-making. Overall, ensemble learning is a flexible and strong methodology for modelling the unstable and multivariate character of ED demand.

This study presents a pioneering solution to the challenge of forecasting daily ED visitor numbers. The Meta-ED model, a novel meta-ensemble model, is introduced as an advanced hybrid ML technique. It is designed to optimise the aggregation of predictions from multiple sub-models and is trained on real, heterogeneous data from Canberra Hospital, ACT, Australia. The use of real, heterogeneous data from a healthcare setting enhances the model's credibility and applicability to real-world scenarios. This feature diversity was intentionally incorporated to allow the model to capture patterns that may be relevant in other geographical and institutional settings. The key contributions of this work are:

- This study establishes a comprehensive comparative framework for predicting ED attendance. It incorporates 23 widely recognised ML, sequential deep learning and ensemble methods, ensuring a meticulous and thorough evaluation of the Meta-ED model's capabilities.
- A new, effective predictive model is introduced, which combines an optimal selection of sub-learners. These sub-learners, chosen based on their superior comparative performance, include two boosting methods (**CatBoost (CatB)** [53] and **LightGBM (LGB)**) [29, 60] and two tree-based methods (RF and **Extra Trees (ExT)**), with a Master-learner (**Multi-Layer Perceptron (MLP)**) used to enhance prediction accuracy.
- The architecture and hyper-parameters of the proposed meta-ensemble model are optimised using an efficient **Differential Evolution (DE)** method paired with a fast local search algorithm (**Nelder–Mead (NM)**). This two-level adaptive strategy enhances the model's flexibility and reliability, enabling it to adapt more effectively to different datasets with varying feature distributions. The optimisation method is not hard-coded to Canberra-specific trends and can be retrained or fine-tuned for local data in other regions with minimal reconfiguration.
- Technical explainability experiments (XAI) are conducted to assess and compare the importance of exogenous features in ED visitor prediction, utilising both local and global interpretation techniques.
- The performance of the Meta-learner model is rigorously evaluated using several performance metrics. These metrics provide a comprehensive understanding of the model's performance and its advantages over existing models. It is demonstrated that the proposed Meta-learner outperforms the other 23 predictive models in both accuracy and robustness.

In Section 2, we outline our study design and provide an extensive analysis of dataset exploration. We then present an overview of the methodology, including the introduction of various ML and sequential deep learning techniques, as well as the technical specifics of the proposed meta-learning approach in Section 3. This section covers essential components, including feature selection, hyper-parameter optimisation, XAI and a detailed description of dataset characteristics and analysis. Subsequently, Section 4 presents a comprehensive analysis of the numerical results obtained from the applied methods, enabling a clear comparison of the performance and efficiency of the proposed framework. Finally, we conclude the article by summarising key findings and emphasising the advantages of our proposed method over existing solutions, as highlighted in Section 5.

2 Study Design and Datasets Exploration

2.1 ED Data

The ED of Canberra Hospital is crucial in providing healthcare services [9] to the ACT populace. This investigation harnesses the **Department of Health and Rehabilitation (DHR)** dataset sourced from the ED to delve into patient profiles, treatment trends and outcomes. Spanning an impressive 23 years, from January 1999 to December 2022, the dataset is a treasure trove of insights for in-depth analysis. The DHR dataset comprises around 1.6 million episodes that involved 535,474 distinct patients. It includes a demographic registry that details sex and birth dates. Each care event's primary diagnosis is categorised using the International Classification of Diseases, tenth revision (ICD-10-AM), while up to 10 most frequent diagnoses follow ICD-10 coding. Additional information encompasses admission and discharge dates, triage levels and patient outcomes. Canberra Hospital's ED processes a considerable influx of patients, with 76,505 reported presentations in ACT public hospitals [2]. In the first and second quarters of 2023–2024, 61.8% of the patients received timely ED treatment, with an average wait time of 25 minutes. Furthermore, 55.6% of patients were released from the ED within 4 hours of arrival, while 3.9% opted not to wait for assessment [2]. The details of these variables are listed in the Supplementary Tables 2 and 3.

2.2 Characteristics of Participants

The distribution of detailing various parameters linked to the number of ED visitors is illustrated by Figure 1 and Supplementary Figure 1. In particular, in 2019, the highest surge in ED visitor numbers, exceeding 90,000 episodes, was attributed to the COVID-19 pandemic. In addition, peak traffic occurs in August, on Mondays, between 10:00 AM and 1:00 PM, surpassing 1.35×10^5, 2.35×10^5 and 9×10^4 episodes, respectively. In terms of gender distribution, men contributed more than women, with a marginal difference of 2%. Statistical analyses also reveal that children ($age \leq 14$) constitute the predominant demographic, comprising 23% of all ED episodes. Interestingly, the data indicate a negative correlation between visitor age and ED attendance rates ($age < 50$). Furthermore, concerning triage levels, levels 4 and 3 (Urgent and Less Urgent) were the most frequently recorded situations in ED episodes. These insights gleaned from the histogram shed light on the complex dynamics of ED utilisation and demographics, aiding in the refinement of healthcare resource allocation and patient care strategies. Additionally, Table 1 presents descriptive statistics for the ED visitors' population based on 535,474 recorded episodes from 1999 to 2022. The statistics provide insights into various demographic and clinical characteristics of the population served by the DHR. The percentages shown in brackets represent values relative to the column base, offering a comparative view of each attribute within the dataset. This analysis aims to identify trends and patterns in ED usage over a 23-year period.

2.3 Exploratory Analysis of ICD-10 Features

The analysis of the 20 most common reasons for visiting the ED reveals that the group of patients who did not wait for treatment had the highest number of visitors. In the following, chest pain, pain in the abdomen and viral infections are second to fourth rank among all 20 viral ICD-10 recorded (see Supplementary Figure 2). Other studies [66] have confirmed outbreaks of seasonal viruses cause a surge in visits to EDs. This high prevalence of chest pain, abdominal pain and viral infections as leading reasons for ED visits highlights the need for effective triage and resource distribution, as these conditions are common and can require urgent attention, depending on severity.

The insights gleaned from the Supplementary Figure 3, which delineates the distribution percentages of the top 15 reasons for ED visits stratified by gender, reveal intriguing trends. The data indicates that males exhibited

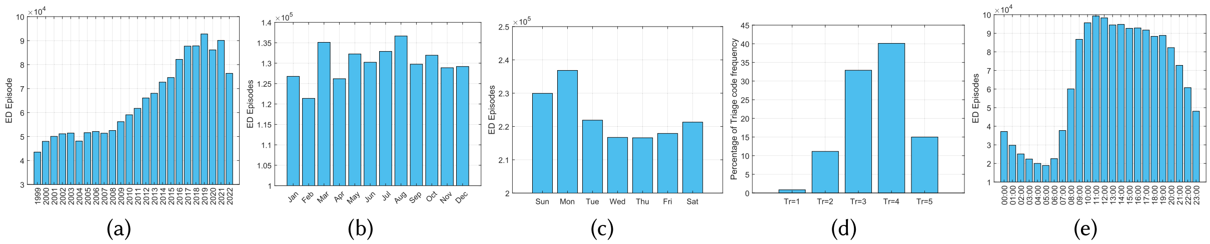


Fig. 1. Histogram detailing various parameters linked to the number of ED visitors at Canberra Hospital between 1999 and 2022.

Table 1. Descriptive Statistics for the Visitors' Population Based on ED Episodes (Population = 535,474) from 1999 to 2022

Variable		Total (%)
Total population		535,474 (100)
Total episode number		1,561,222 (100)
Female		764,998 (49)
Age	<=14	366,850 (**23.50**)
	14–24	238,753 (15.29)
	25–34	221,889 (14.21)
	35–44	180,278 (11.55)
	45–54	154,698 (9.91)
	55–64	129,775 (8.31)
	65–74	112,191 (7.19)
	>74	156,788 (10.04)
Triage	1 (Red)	13,237 (0.85)
	2 (Orange)	174,000 (11.15)
	3 (Yellow)	513,686 (32.90)
	4 (Green)	626,124 (**40.10**)
	5 (Blue)	234,175 (15.00)
State	ACT	1,315,365 (84.25)
	NSW	218,408 (13.99)
Disposition	Admit	503,784 (32.27)
	Home	911,854 (**58.41**)
	DNW	103,379 (6.62)
	LOR before treatment completed	7,181 (0.46)
	Discharged to DHR	21 (0.0013)
	Referred to other TCH service	25,938 (1.66)
	Transferred to other hospital	6,756 (0.43)
	Went to GP	1,236 (0.079)
	Died in ED	1,070 (0.068)
	Dead on arrival	3 (0.0001)
Most frequent ICD-10	'Z53.1' (Did not wait for treatment)	89,874 (5.76)
	'R07.4' (Chest pain)	62,962 (4.03)
	'R10.4' (Pain in abdomen)	62,511 (4.00)
	'B34.9' (Viral infection)	31,155 (2.00)
	'Z09.9' (For review)	30,513 (1.95)
	'J45.9' (Asthma, acute)	19,436 (1.24)
	'N39.0' (Urinary tract infection)	18,028 (1.15)
	'R11' (Nausea and vomiting except in pregnancy)	16,532 (1.06)
	'S09.9' (Injury, unspecified or suspected of head)	15,874 (1.02)
	'M54.5' (Low back pain)	14,060 (0.90)
	'R45.81' (Suicidal ideation)	13,425 (0.86)

Percentages in brackets are relative to the column base. The bold number shows the first rank in each group. DHR, Department of Health and Rehabilitation; DNW, Did Not Wait; LOR, Left at Own Risk.

a visit rate for open wounds that were approximately double that of females, implying a higher prevalence of this issue among male ED attendees. In contrast, females demonstrated a visit rate for urinary tract infections that was

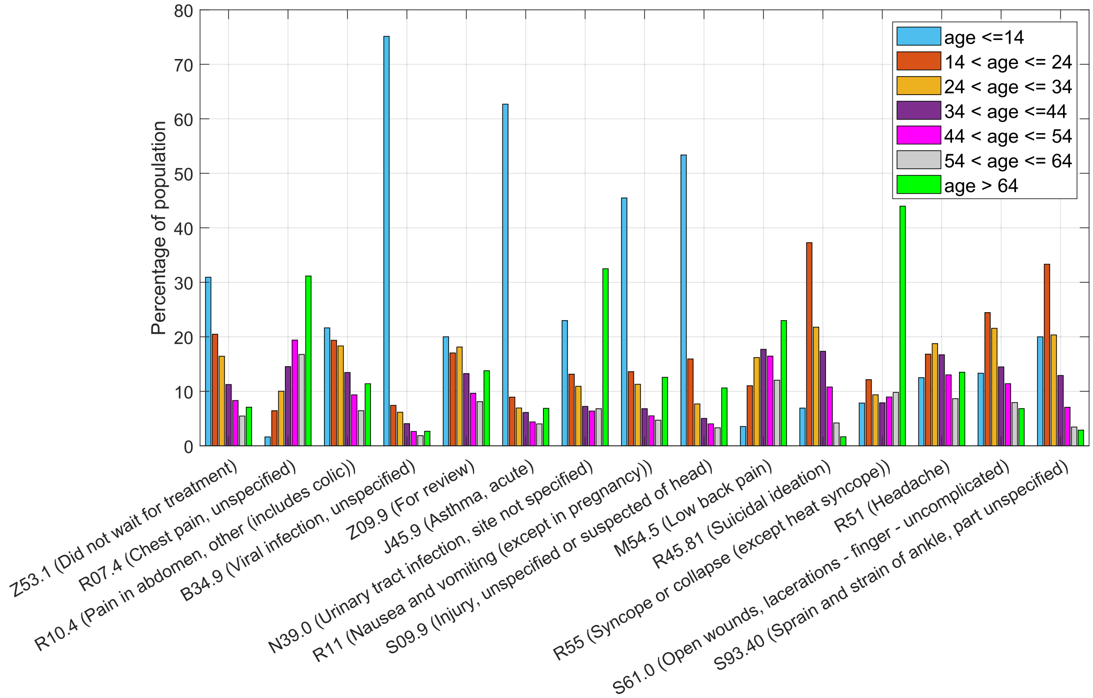


Fig. 2. The distribution percentage of the top 15 most viral reasons for ED visitors based on the age groups.

roughly twice that of males, signifying a higher incidence of this ailment among female ED visitors. In addition, in cases of abdominal pain, nausea, vomiting, lower back pain, suicidal ideation, syncope and headaches, women displayed a higher frequency of ED visits compared to men, underscoring a distinct gender bias in emergency care utilisation for these health concerns. Furthermore, in 7 out of the 15 leading reasons, females contributed more to these visits than males. This data underscores the importance of tailoring healthcare resources to address the specific needs of women.

The analysis depicted in Figure 2 unveils several significant patterns within ED visits. Firstly, it is evident that people younger than 14 years exhibited the highest incidence of visits associated with Viral infections, Abdominal pain, Asthma, Nausea, vomiting and injuries. In contrast, seniors aged 64 years and older had the highest frequency of ED visits relating to chest pain, urinary tract infections, lower back pain and syncope. Furthermore, the younger population, particularly those aged between 14 and 24, predominantly sought medical attention for issues such as suicidal ideation, open wounds, finger lacerations and ankle sprains and strains. These observations underscore distinct health concerns across different age groups in accessing emergency care services.

2.4 Factors Correlated with ED Visits

As shown in Figure 3, the feature most strongly correlated with the visitor count for ED visits is the admission year, with a correlation coefficient of 0.88. Interestingly, the top five most common ICD-10 codes with the highest correlation coefficient with visitor count range between 0.34 and 0.72, with lower-ranked codes exhibiting notably higher correlation coefficients. In addition, a distinct correlation pattern exists between age groups and the number of ED patients. At the same time, we can see a positive correlation between increasing age and ED visitation for the majority of age groups; individuals younger than 50 years demonstrate a negative correlation.

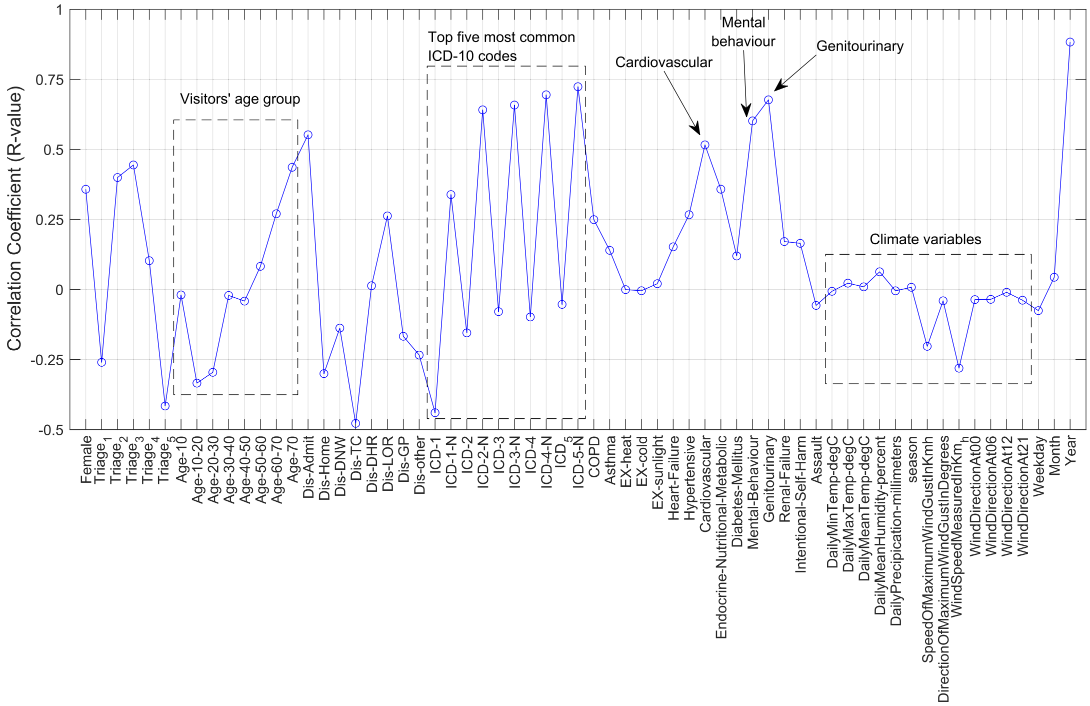


Fig. 3. The correlations between the ED visitors' number and wide range of features, including patient demographic, medical information, admission details and climate variables.

This finding suggests that younger individuals are less likely to visit the ED due to better overall health, a lower prevalence of chronic conditions and fewer medical emergencies. Strong positive correlations are observed for several chronic disorder categories, with cardiovascular, mental behavioural and genitourinary disorders showing correlation coefficients of 0.51, 0.60 and 0.68, respectively. Among the 13 climate-related variables analysed, average wind speed exhibits the strongest negative correlation with ED visit volume, with an R-value of −0.28. In contrast, other weather-related variables show minimal or non-substantive associations with ED visitation in this dataset.

The correlation analysis (see Supplementary Figure 4) provides valuable insight into the relationship between climate variables and ED visitation. In particular, daily temperature shows the strongest correlation with age groups of patients in the ED, with a negative correlation coefficient of −0.30 for individuals under 10 years of age and approximately −0.10 for other age groups. Additionally, rising temperatures are associated with an increase in genitourinary-related ED visits, with a correlation coefficient of 0.10. Humidity also appears to play a role, particularly affecting children under 10, with a correlation of 0.15, and showing a 0.10 correlation with asthma-related ED presentations. Seasonal variation can also be a factor, with the variable 'season' exhibiting a correlation of 0.20 with ED visits among children under 10, underscoring the importance of season-specific healthcare strategies for younger populations. Among all climate variables, wind gust speed demonstrates the strongest negative association with ED visitation, with a correlation coefficient of −0.20, indicating that higher wind speeds are associated with reduced ED attendance. Furthermore, wind speed shows negative correlations with both mental health and genitourinary conditions, underscoring the complex and multi-faceted influence of environmental factors on health outcomes in the ED context.

Table 2. Statistical Results of ED Visitor Prediction for Standard Recurrent Networks and Small and Large Multi-Layer LSTM, GRU and BiLSTM Models and CNN

LSTM									
Metric	RMSE	MAE	R	MSLE	MEDAE	EVS	MAPE	MDA	RSE
Min	26.137	21.429	0.707	0.012	19.182	0.500	8.514	0.675	0.811
Max	37.354	32.528	0.772	0.026	31.770	0.595	12.914	0.696	1.657
Mean	30.571	25.773	0.741	0.017	23.990	0.547	10.230	0.683	1.124
STD	3.7E+00	3.6E+00	2.3E-02	4.6E-03	4.1E+00	3.5E-02	1.4E+00	7.2E-03	2.8E-01

BiLSTM									
Metric	RMSE	MAE	R	MSLE	MEDAE	EVS	MAPE	MDA	RSE
Min	24.092	19.528	0.741	0.010	17.264	0.549	7.781	0.687	0.689
Max	28.949	24.124	0.781	0.015	21.794	0.609	9.582	0.709	0.995
Mean	26.079	21.428	0.762	0.012	19.256	0.579	8.522	0.698	0.810
STD	1.4E+00	1.3E+00	1.1E-02	1.3E-03	1.3E+00	1.6E-02	5.0E-01	7.3E-03	8.6E-02

GRU									
Metric	RMSE	MAE	R	MSLE	MEDAE	EVS	MAPE	MDA	RSE
Min	24.170	19.772	0.703	0.010	17.169	0.494	7.901	0.661	0.694
Max	33.822	28.779	0.764	0.020	27.443	0.579	11.394	0.698	1.358
Mean	27.705	22.996	0.745	0.014	20.921	0.552	9.140	0.682	0.927
STD	3.8E+00	3.6E+00	1.8E-02	4.1E-03	4.1E+00	2.4E-02	1.4E+00	1.1E-02	2.6E-01

CNN									
Metric	RMSE	MAE	R	MSLE	MEDAE	EVS	MAPE	MDA	RSE
Min	22.414	18.107	0.629	0.009	16.003	0.396	7.439	0.616	0.597
Max	36.234	30.794	0.676	0.023	29.537	0.453	12.134	0.644	1.559
Mean	29.304	24.201	0.655	0.015	21.900	0.427	9.623	0.626	1.036
STD	3.9E+00	3.6E+00	1.7E-02	4.1E-03	3.9E+00	2.2E-02	1.3E+00	8.3E-03	2.7E-01

s-LSTM									
Metric	RMSE	MAE	R	MSLE	MEDAE	EVS	MAPE	MDA	RSE
Min	21.636	17.566	0.775	0.008	15.194	0.583	7.098	0.702	0.556
Max	22.269	18.121	0.781	0.008	15.529	0.590	7.298	0.711	0.589
Mean	22.004	17.893	0.777	0.008	15.393	0.585	7.218	0.707	0.575
STD	2.0E-01	1.7E-01	1.5E-03	1.4E-04	1.0E-01	2.2E-03	6.2E-02	2.4E-03	1.0E-02

s-BiLSTM									
Metric	RMSE	MAE	R	MSLE	MEDAE	EVS	MAPE	MDA	RSE
Min	22.690	18.430	0.773	0.009	15.796	0.586	7.393	0.700	0.611
Max	23.532	19.189	0.777	0.009	16.725	0.594	7.674	0.705	0.658
Mean	23.145	18.833	0.775	0.009	16.270	0.590	7.540	0.702	0.636
STD	2.4E-01	2.1E-01	1.5E-03	1.8E-04	3.0E-01	2.7E-03	7.9E-02	1.6E-03	1.3E-02

s-GRU									
Metric	RMSE	MAE	R	MSLE	MEDAE	EVS	MAPE	MDA	RSE
Min	21.441	17.399	0.770	0.008	15.173	0.576	7.039	0.693	0.546
Max	23.411	19.173	0.782	0.009	16.666	0.597	7.687	0.705	0.651
Mean	22.299	18.165	0.775	0.008	15.751	0.585	7.317	0.698	0.591
STD	5.8E-01	5.2E-01	3.4E-03	4.2E-04	5.0E-01	6.8E-03	1.9E-01	4.5E-03	3.1E-02

l-LSTM									
Metric	RMSE	MAE	R	MSLE	MEDAE	EVS	MAPE	MDA	RSE
Min	21.823	17.732	0.775	0.008	15.325	0.583	7.156	0.704	0.566
Max	22.461	18.295	0.780	0.009	15.671	0.590	7.359	0.711	0.599
Mean	22.037	17.923	0.777	0.008	15.464	0.585	7.229	0.707	0.577
STD	1.9E-01	1.6E-01	1.5E-03	1.3E-04	9.7E-02	2.2E-03	5.8E-02	2.3E-03	9.8E-03

l-BiLSTM									
Metric	RMSE	MAE	R	MSLE	MEDAE	EVS	MAPE	MDA	RSE
Min	21.954	17.806	0.783	0.008	15.257	0.593	7.169	0.706	0.572
Max	22.240	18.057	0.783	0.008	15.439	0.594	7.257	0.707	0.587
Mean	22.062	17.901	**0.783**	0.008	15.336	0.594	7.202	0.707	0.578
STD	7.9E-02	7.0E-02	2.0E-04	5.6E-05	5.8E-02	4.6E-04	2.4E-02	3.2E-04	4.2E-03

l-GRU									
Metric	RMSE	MAE	R	MSLE	MEDAE	EVS	MAPE	MDA	RSE
Min	21.133	17.147	0.776	0.008	14.796	0.580	6.957	0.693	0.530
Max	22.185	18.077	0.779	0.008	15.630	0.590	7.284	0.699	0.585
Mean	21.679	**17.646**	0.778	0.008	15.275	0.585	7.137	0.696	0.558
STD	3.1E-01	2.7E-01	9.1E-04	2.1E-04	2.6E-01	2.8E-03	9.7E-02	2.0E-03	1.6E-02

EVS, Explained Variance Score; MAE, Mean Absolute Error; MSLE, Mean Squared Logarithmic Error. The bold number shows the first rank in each metric.

2.5 Exogenous Variables: Identification and Analysis

The climate variables used in this study were sourced from the Australian Bureau of Meteorology [45], with readings recorded every 3 hours. These variables include daily minimum, maximum and average temperatures (in degrees Celsius), daily mean precipitation (in millimetres), maximum wind gust speed (in kilometres per hour) and its direction (in degrees), as well as the average daily wind speed and wind direction recorded at four distinct time offsets. The data were collected from a weather station in Tuggeranong (Isabella Plains) AWS, the closest meteorological station to Canberra Hospital, covering the period from 1996 to 2022. The ranges and specifics of these climate variables are summarised in Table 2. Supplementary Figure 6 shows the wind direction, speed and frequency with 3-hour intervals from midnight to 9:00 PM.

In order to extend the analysis related to climate variables, Figure 4 depicts the relationship between climate variables and the top 10 most common diagnoses based on ICD-10 codes at Canberra Hospital's ED.

Several notable patterns emerge from this analysis. Viral infections show a positive correlation with both seasonal variation and humidity, with correlation coefficients of 0.17 and 0.10, respectively. In contrast, viral infections exhibit negative correlations with temperature and wind speed, exceeding −0.20 and −0.15, respectively. These findings suggest that the likelihood of ED admissions due to viral infections increases as temperatures decline. Regarding chest pain, a positive correlation of 0.13 is observed with higher humidity levels. Furthermore, the risk of chest pain presentations appears to increase by 0.18 as wind speed decreases, highlighting the potential influence of environmental factors on cardiovascular-related symptoms.

The analysis presented in Figure 5 reveals a complex and nonlinear relationship between climate parameters and the volume of ED visitors. Notably, Figure 5(c) illustrates a noteworthy trend where elevating wind speeds (Km/h) correspond to a reduction in the influx of individuals seeking care at the ED. This unexpected inverse correlation suggests that stronger wind conditions have a mitigating effect, potentially reducing the demand for emergency medical services during such periods. Conversely, despite findings reported in some prior studies, the current data does not reveal a substantial relationship between temperature fluctuations and ED visitation rates.

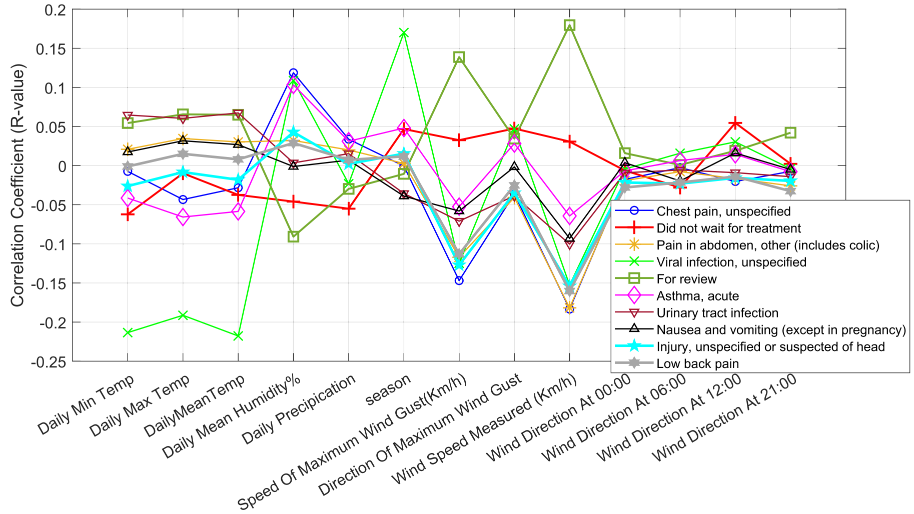
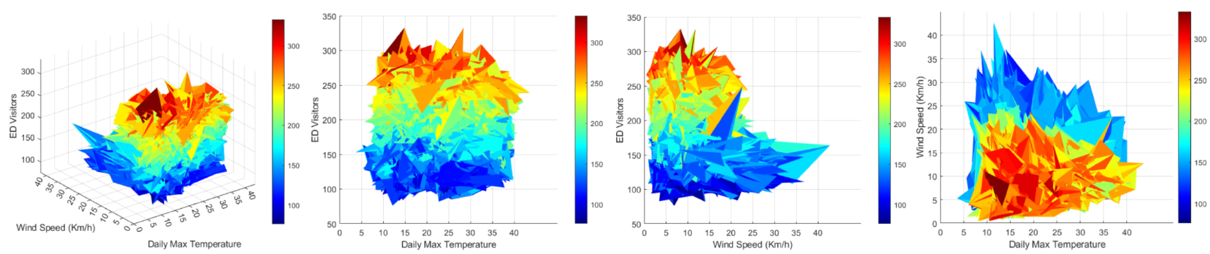


Fig. 4. The correlations between climate variables and top 10 most common diagnoses based on ICD-10 code.



(a) (b) (c) (d)

Fig. 5. (a) A 3D correlation between number of ED visitors, wind speed (Km/h) and daily maximum temperature. (b) Correlation between ED visitors number and daily maximum temperature. (c) Correlation between ED visitors number and wind speed. (d) Correlation between daily maximum temperature and wind speed.

A comprehensive analysis of the correlation between wind directions and ED visitor numbers has revealed distinct patterns. Specifically, findings indicate a notable surge in the volume of ED visitors when winds blow from the South and Southeast directions. This observed relationship suggests a potential influence of these wind patterns on health outcomes within the community, potentially leading to higher incidents of health-related emergencies (see Supplementary Figure 5). Furthermore, we analyse a collection of time attributes inherent to ED patient arrival dates, including year, month, day of the month and day of the week. Such temporal attributes are significant in that they reveal seasonal, cyclical and long-term patterns that significantly influence the rates of ED visitation. For example, seasons during flu season or holidays represent times of high patient volume. Additionally, incorporating temporal patterns enables the identification of short-term fluctuations (e.g., weekday vs. weekend effects) and long-term trends (e.g., increases in population or shifts in service demand), which is essential for healthcare system demand forecasting and resource allocation optimisation.

3 Methods and Materials

In this section, we introduce and describe the technical details of MLs, DLs and ensemble methods developed and compared in this study, as well as our proposed solution to improve the prediction accuracy of ED visitors.

3.1 Predictive Modelling of ED Visitors

In this study, we developed an extensive ML-based framework to provide accurate predictions for daily ED visitor crowding. As the nature of the ED data is a time series, we initiated the implementation using popular sequential models [20], including LSTM, **bidirectional LSTM (BiLSTM)** and GRU. Two versions of these sequential models were developed: a light (one learning layer) and a stacked model (two and five learning layers). This is mainly because testing both a single sequential layer and multiple layers in a prediction task is essential for understanding the tradeoffs between model complexity, accuracy and generalisation abilities. As LSTM and GRU are not effective in extracting spatial patterns (multivariate time series with spatial relationships), we added a convolutional layer to the LSTM and GRU models to make a hybrid Convolutional LSTM [72] or a hybrid CNN-LSTM architecture.

Other well-known ML-based models for forecasting the number of ED visitors are NNs such as MLP and **Dense Neural Networks (DNN)**. The primary reason for applying NN models is their ability to model complex, nonlinear relationships in data [52]. In the following, Decision trees, RF and ExT models are commonly used in predicting ED visitor numbers due to their flexibility, interpretability and ability to model complex relationships among the features. Our dataset is a combination of various features, including demographics, climate, diagnosis and calendar, which makes it heterogeneous data. Therefore, to address this type of data, we implemented and tested the effectiveness of six ensemble models [13], including AdaBoost, XGBoost [36], LGB [23], CatB, **Gradient Boosting Regression (GBR)** [79] and **Histogram-Based Gradient Boosting Regression (HGBR)**. The configurations of the ML-based models applied in this study are listed in Supplementary Table 1.

3.1.1 *Meta-Learning Models.*

Meta-ensemble learning modelling, a sophisticated form of stacking ML methodology, is designed to foster fit diversity by amalgamating predictions from a variety of methods [16]. This approach involves training a meta-learner, often referred to as the second-level learner, to amalgamate predictions from first-level learners, also known as base models. By integrating the strengths of diverse base models (sub-learners), meta-ensemble models aim to enhance predictive accuracy and robustness [65]. While the conventional stacking technique employs a two-level hierarchy with level-0 and level-1 models, the methodology can be extended to include multiple layers, incorporating various level-1 models and a single level-2 model for improved prediction amalgamation. The core premise of stacking lies in leveraging the distinct strengths of individual models to mitigate weaknesses, culminating in superior performance compared to standalone models.

Moreover, the versatility afforded by stacking extends to enabling a flexible and tailored approach to model selection [5]. Rather than relying on a single model, practitioners can leverage a spectrum of base models tailored to address specific facets of the problem at hand. For instance, a RF model may capture nonlinear relationships, a linear regression model can elucidate linear dependencies, and an MLP can unravel intricate feature interactions. Through the synergistic integration of predictions from these specialised models, stacking facilitates the creation of a holistic and refined predictive model that excels in capturing the complexities inherent in diverse datasets. The multi-layered meta-ensemble modelling paradigm not only amplifies predictive accuracy and robustness but also serves as a versatile and effective strategy for model selection, offering a comprehensive approach to addressing the intricacies of ML tasks.

3.2 Proposed Meta-Learning GBR Model

3.2.1 *Problem Formalisation.*

In the context of a dataset that encompasses various examples related to ED attendance, the primary learning task at hand involves the intricate process of constructing predictive models aimed at forecasting future attendance patterns with the utmost precision and reliability. In a standard ED forecasting scenario that occurs at a specific time denoted as t, a comprehensive set of input variables represented as $\vec{A}(t) = [a_1(t), \ldots, a_N(t)]$ is meticulously provided, which is then paired with the corresponding output or target variable $z(t)$, representing the number of ED attendances, thus forming a cohesive pair denoted as

Table 3. Statistical Results of ED Visitor Prediction for Ensemble and Deep Learning Models

MLP										DNN									
Metric	RMSE	MAE	R	MSLE	MEDAE	EVS	MAPE	MDA	RSE	Metric	RMSE	MAE	R	MSLE	MEDAE	EVS	MAPE	MDA	RSE
Min	24.566	19.896	0.669	0.010	19.896	0.437	7.944	0.233	0.717	Min	19.691	15.730	0.758	0.007	15.730	0.569	6.486	0.682	0.461
Max	27.351	22.303	0.725	0.013	22.303	0.525	8.864	0.532	0.889	Max	23.281	18.788	0.770	0.009	18.788	0.586	7.537	0.695	0.644
Mean	25.353	20.457	0.695	0.011	20.457	0.478	8.141	0.339	0.764	Mean	20.957	16.854	0.764	0.008	16.854	0.578	6.870	0.689	0.523
STD	7.7E-01	7.1E-01	1.5E-02	8.2E-04	7.1E-01	2.3E-02	2.8E-01	8.8E-02	4.8E-02	STD	1.0E+00	8.7E-01	3.9E-03	7.1E-04	8.7E-01	5.5E-03	2.9E-01	3.9E-03	5.3E-02
RF										GBR									
Metric	RMSE	MAE	R	MSLE	MEDAE	EVS	MAPE	MDA	RSE	Metric	RMSE	MAE	R	MSLE	MEDAE	EVS	MAPE	MDA	RSE
Min	32.342	26.642	0.327	0.018	26.642	0.091	10.624	0.614	1.243	Min	20.318	16.226	0.789	0.007	16.226	0.551	6.761	0.706	0.491
Max	36.273	29.876	0.492	0.023	29.876	0.236	11.863	0.647	1.563	Max	20.359	16.251	0.790	0.007	16.251	0.552	6.769	0.708	0.493
Mean	34.061	28.051	0.415	0.020	28.051	0.171	11.161	0.631	1.380	Mean	20.332	16.237	0.789	0.007	16.237	0.552	6.764	0.707	0.491
STD	1.3E+00	1.1E+00	5.4E-02	1.6E-03	1.1E+00	4.7E-02	4.0E-01	1.1E-02	1.0E-01	STD	1.3E-02	7.4E-03	2.4E-04	7.6E-06	7.4E-03	3.9E-04	2.5E-03	6.5E-04	6.4E-04
ExT										CatB									
Metric	RMSE	MAE	R	MSLE	MEDAE	EVS	MAPE	MDA	RSE	Metric	RMSE	MAE	R	MSLE	MEDAE	EVS	MAPE	MDA	RSE
Min	23.137	18.857	0.736	0.009	18.857	0.456	7.735	0.708	0.636	Min	18.717	14.476	0.834	0.006	14.476	0.638	5.932	0.728	0.416
Max	23.828	19.434	0.747	0.010	19.434	0.474	7.943	0.728	0.675	Max	19.889	15.594	0.845	0.007	15.594	0.654	6.333	0.757	0.470
Mean	23.560	19.211	0.742	0.010	19.211	0.462	7.865	0.717	0.660	Mean	19.339	**15.038**	**0.839**	0.006	15.038	0.644	6.129	0.741	0.445
STD	2.1E-01	1.9E-01	3.2E-03	1.6E-04	1.9E-01	5.5E-03	6.9E-02	6.0E-03	1.2E-02	STD	3.6E-01	3.5E-01	3.5E-03	2.1E-04	3.5E-01	5.6E-03	1.3E-01	8.3E-03	1.7E-02
AdaB										XGB									
Metric	RMSE	MAE	R	MSLE	MEDAE	EVS	MAPE	MDA	RSE	Metric	RMSE	MAE	R	MSLE	MEDAE	EVS	MAPE	MDA	RSE
Min	22.914	18.637	0.672	0.009	18.637	0.396	7.684	0.651	0.624	Min	29.735	24.452	0.473	0.015	24.452	0.140	9.896	0.375	1.051
Max	25.034	20.530	0.720	0.011	20.530	0.472	8.431	0.686	0.745	Max	31.312	25.865	0.592	0.017	25.865	0.244	10.382	0.488	1.165
Mean	23.827	19.395	0.698	0.010	19.395	0.434	7.994	0.670	0.675	Mean	30.435	25.068	0.541	0.016	25.068	0.187	10.120	0.435	1.101
STD	6.3E-01	5.8E-01	1.5E-02	5.0E-04	5.8E-01	2.2E-02	2.2E-01	1.1E-02	3.6E-02	STD	5.8E-01	5.5E-01	3.5E-02	5.4E-04	5.5E-01	3.9E-02	1.7E-01	3.0E-02	4.2E-02
LGB										HGBR									
Metric	RMSE	MAE	R	MSLE	MEDAE	EVS	MAPE	MDA	RSE	Metric	RMSE	MAE	R	MSLE	MEDAE	EVS	MAPE	MDA	RSE
Min	21.324	16.988	0.801	0.008	16.988	0.574	6.916	0.725	0.540	Min	20.309	16.220	0.789	0.007	16.220	0.551	6.758	0.706	0.490
Max	21.418	17.080	0.806	0.008	17.080	0.585	6.969	0.734	0.545	Max	20.359	16.251	0.790	0.007	16.251	0.553	6.769	0.708	0.493
Mean	21.368	17.033	0.804	0.008	17.033	0.580	6.942	0.730	0.543	Mean	20.334	16.238	0.789	0.007	16.238	0.552	6.764	0.707	0.491
STD	3.7E-02	3.6E-02	2.0E-03	3.5E-05	3.6E-02	3.9E-03	1.8E-02	2.6E-03	1.9E-03	STD	1.4E-02	1.0E-02	3.2E-04	8.7E-06	1.0E-02	4.9E-04	3.7E-03	7.0E-04	7.0E-04

The bold numbers shows the first rank in each metric.

$\langle \vec{A}(t), z(t) \rangle$. For analytical purposes of this study, time is conveniently expressed in weekly intervals. However, it is also essential to note that alternative time units, such as days or months, could also be seamlessly applied without disrupting the methodology. The examples utilised in this research are systematically structured time series, which can be denoted as $\left\langle \vec{A}\left(t_0\right), z\left(t_0\right)\right\rangle, \ldots, \left\langle \vec{A}\left(t_L\right), z\left(t_L\right)\right\rangle$, commencing from the initial time point t_0 and extending to the final time point t_L.

The input variables denoted as $a_i(t)$, where i ranges from 1 to N, represent a variety of exogenous factors that influence ED attendance patterns. Within the scope of this study, these input variables encompass a diverse range of elements, including climatic conditions, demographic characteristics, ICD-10 diagnosis codes and various calendar characteristics, as outlined in Tables 2 and 3. The calendar features in question consist of nominal variables that serve to identify specific temporal attributes such as year, month, day of the week, day of the month and the corresponding season. On the other hand, climate data are made up of numerical variables that encapsulate daily averages of critical factors such as temperature, humidity levels, wind speed and wind direction, all of which play a crucial role in the influence of health outcomes. Furthermore, the demographic information incorporated into the analysis comprises numerical variables related to the characteristics of patients visiting the ED, including vital statistics such as age, sex and race, which collectively contribute to a comprehensive understanding of attendance trends in the emergency healthcare setting.

Given that ED attendance can be quantified numerically, it becomes evident that the model undergoing training can be expressed as a function denoted by $f(\cdot)$, which is specifically designed to predict the output variable $z(t)$ at a future time t (representing the ED attendances the next day t, where t exceeds t_L). This predictive output is defined as $\hat{z}(t) = f(\cdot)$. The inputs that are inputted in this function consists not only of the current array of variables represented as $a_1(t), \ldots, a_N(t)$ but also include historical ED attendance figures $z(t - d_j)$, which are associated with various delays d_j, where the index j falls within the range of $[1, \ldots, S]$, and importantly, d_S remains less than t_L. It is assumed that these previous attendance values are readily available and known when forecasting future attendance for the specified time t. The overarching framework that guides this prediction process is encapsulated in the formulation provided in Equation (1), which serves as a foundational element of the model:

$$\hat{z}(t) = f\left(a_1(t), a_2(t), \ldots, a_N(t), z\left(t-d_1\right), z\left(t-d_2\right), \ldots, z\left(t-d_S\right)\right). \tag{1}$$

The specific modelling technique significantly influences the intricacy of defining $f(\cdot)$. In this particular scenario, we delve into the realm of Meta-learning, which is characterised as a supervised learning approach that trains the function through the utilisation of historical data, comprising a sequence of examples that are structured as $\left\langle \vec{A}(t_0), z(t_0)\right\rangle, \ldots, \left\langle \vec{A}(t_L), z(t_L)\right\rangle$. The primary goal of this process is to minimise the difference that exists between the actual observed values $z(t)$ and the predicted values $\hat{z}(t)$ derived from known examples used in the training phase. A crucial challenge addressed by Meta-learning is the risk of overfitting, a condition in which the model $f(\cdot)$ may become excessively attuned to the particular training dataset, resulting in sub-par performance when confronted with new and previously unseen cases. To combat this issue, the methodology incorporates various strategies aimed at ensuring that the model maintains a strong capacity for generalisation when faced with novel scenarios, thereby achieving an effective balance between flexibility and robustness that is essential for reliable predictions.

3.2.2 Meta-Learning GBR. To develop a robust and effective predictive model for forecasting ED visitor numbers, we propose a novel meta-learning GBR approach to reduce bias and variance in prediction errors. In the first phase, we designed a comprehensive comparative framework to identify the best sub-learners, evaluating a total of 23 different methods, including ML techniques, sequential deep learning models [38] and ensemble models. Statistical analysis of the results demonstrated that (see Tables 2 and 3) CatB and LGB significantly outperformed other models in predicting ED attendance. In the second phase, we selected a diverse set of predictive models to enhance the variety of learning mechanisms, thereby improving the overall performance of the meta-learning ensemble. In particular, the inclusion of RF and ExT models contributed to a marked improvement in prediction accuracy, highlighting the importance of model diversity in reducing error and improving robustness.

After selecting the sub-models, our next focus was on identifying the best-performing main learner. Choosing an effective meta-learner is crucial, as it synthesises the outputs of the base learners into a final prediction. To ensure diversity and robustness, we evaluated several master-learners, including linear regression, MLP, SVM, AdaBoost and XGBoost. Among these, MLP consistently outperformed the other algorithms in our experiments. As the main learner, MLP's role is to learn the optimal method for weighting the predictions of the base learners. It captures the relationship between the base learners' outputs and the target variable, uncovering patterns and dependencies that enhance predictive accuracy. Through the use of backpropagation, the standard algorithm for training MLP, it fine-tunes the combination of base learners' predictions, resulting in a more accurate and robust final prediction. This highlights MLP's effectiveness in meta-learning frameworks for improving overall model performance. After applying **Recursive Feature Elimination (RFE)** to identify the optimal sub-features [56], we developed a fast and efficient hyper-parameter optimisation method. This method combines DE [46] with the NM algorithm, providing a robust approach to fine-tuning model parameters (see Section 3.4). By employing this adaptive technique, we successfully optimised the hyper-parameters for four sub-learners and the MLP. This hybrid strategy optimises the search process for optimal hyper-parameters, ensuring improved model performance while maintaining computational efficiency, especially in complex and high-dimensional feature spaces. The technical details of the proposed Meta-learning model can be seen in Figure 6.

Ensembling multiple distinct learners to achieve performance gains is a challenging and computationally demanding task. It requires not only selecting models that individually perform well, but also ensuring that their prediction errors are sufficiently decorrelated to yield meaningful gains when aggregated. Designing an ensemble in which all components contribute positively to the final prediction without redundancy or overfitting is a significant methodological contribution, particularly in real-world, high-noise settings like ED forecasting. The details of the implementation setup can be seen in Supplementary Section A.1.

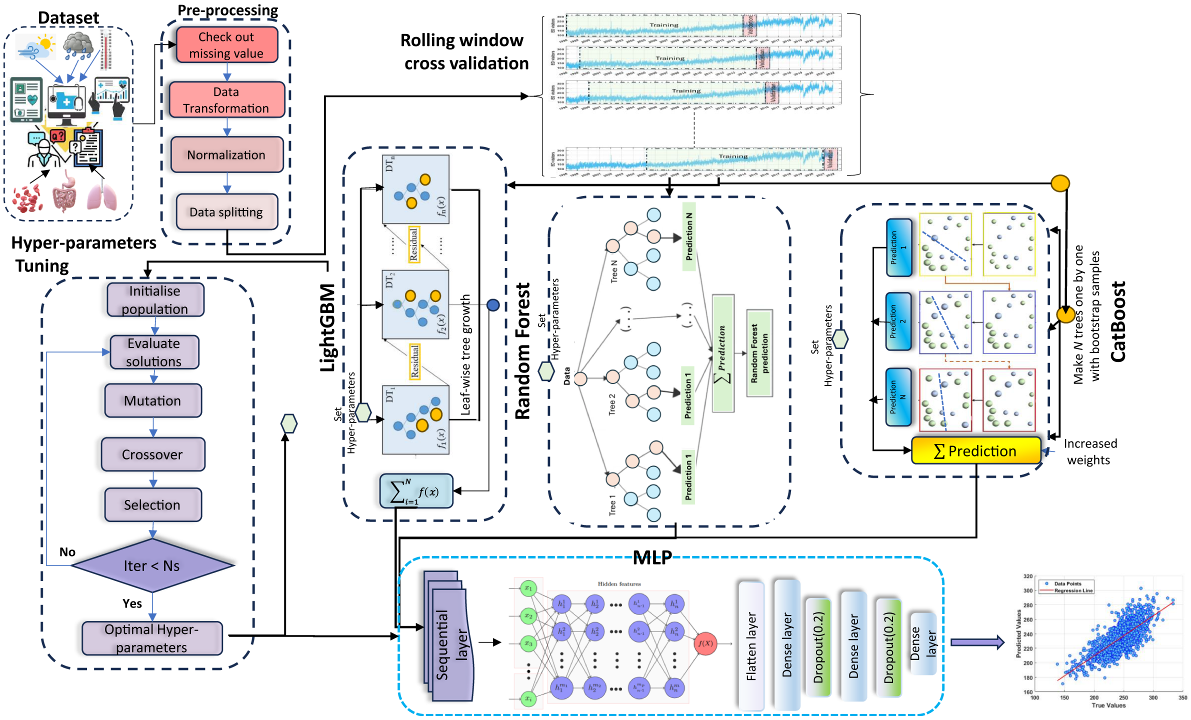


Fig. 6. The landscape of the proposed Meta-learning ensemble architecture.

3.3 Feature Selection

To optimise the selection of sub-features and enhance the performance of the proposed Meta-learner model, we employed the RFE technique [75]. RFE is a highly effective method for feature selection, systematically identifying and removing irrelevant or redundant features. By iteratively eliminating the least important features, RFE simplifies the model while maintaining or improving predictive accuracy. This reduction in feature space helps mitigate overfitting, ensuring the model generalises unseen data better. Additionally, RFE improves the interpretability of the model by focusing on the most significant predictors. Its flexibility and compatibility with a wide range of ML algorithms further reinforce its utility in building robust and efficient predictive models. In order to implement RFE, we used Scikit-learn (an open-source ML library), which was developed based on Python and the RFE class from [59].

3.4 DE Hyper-Parameters Tuning

Adjusting hyper-parameters in ML is incredibly crucial due to its direct impact on the model's efficacy and capacity to generalise. In this research, we demonstrated the adaptability of the XGBoost model by utilising a powerful optimisations technique, DE [61] alongside NM [10], to meticulously adjust four hyper-parameters. These parameters, including tree depth, count of estimators, learning rate and sub-sample rate, were fine-tuned using both accuracy and AUC as evaluative metrics, showcasing the model's outstanding performance. DE represents an innovative approach within evolutionary algorithms, integrating differential vectors into a triangular search pattern, and has emerged as a highly respected population-based optimisation strategy, frequently leveraged to address a broad spectrum of real-world challenges marked by noise, volatility and multimodality. Among DE's various facets, the mutation operator is a pivotal component significantly contributing to the algorithm's effectiveness. Numerous mutation operators have been proposed in the DE landscape, each presenting unique convergence and exploration potential rates. A notable mutation strategy, $DE/best/1/bin$, is celebrated for its rapid convergence, especially in the context of unimodal challenges. Yet, its performance tends to falter when confronting the hurdles of local optima, often leading to premature convergence as it navigates through the complexities of multi-modal problem environments. Equation (2) mathematically articulates this mutation scheme, encapsulating its fundamental operational characteristics within DE:

$$\text{DE/best/1/bin } : \vec{T}_{k,g} = \vec{S}_{\text{best },g} + \zeta \cdot \left(\vec{S}_{r_1,g} - \vec{S}_{r_2,g} \right), \tag{2}$$

where $\vec{T}_{k,g}$ represents a vector that captures the nuances of three distinct solutions, namely $\vec{S}_{\text{best}}$, $\vec{S}_{r_1}$ and $\vec{S}_{r_2}$. Transitioning to another crucial evolutionary component in DE, we unearth the significant influence of crossover. The primary technique utilised for crossover is the binomial method, defined by a structure that incorporates the trial vector, referred to as $\vec{U}$, alongside the crossover rate, indicated as C_r, which varies between 0 and 1 as shown:

$$\vec{U}^{g}_{k,j} = \begin{cases} \vec{T}^{g}_{k,j}, & \text{if (rand} \le C_r) \text{ or } (j = sn), \\ \vec{S}^{g}_{k,j}, & \text{otherwise.} \end{cases} \quad j = 1, 2, \ldots D \tag{3}$$

In the crossover phase, a designated number of candidates is chosen, denoted sn. This leads to the formation of an innovative solution through a balanced amalgamation of the parent and progeny elements, thereby continuing the evolutionary journey of optimisation, as illustrated:

$$\vec{T}^{g+1}_{k} = \begin{cases} \vec{U}^{g}_{k}, & \text{if } \left(f\left(\vec{U}^{g}_{k}\right) \le f\left(\vec{T}^{g}_{k}\right) \right) \rightarrow \text{Minimisation} \\ \vec{T}^{g}_{k}, & \text{otherwise.} \end{cases} \tag{4}$$

The integration of DE alongside the NM algorithm provides a remarkable array of benefits by effectively harnessing the strengths inherent in both global and local optimisation methodologies. While DE excels at conducting an extensive exploration of the global search landscape of hyper-parameters, accurately identifying and locating the most promising regions that may yield optimal configurations for the predictor, NM steps in with its remarkable efficiency to carry out the intricate and crucial task of local optimisation, thereby honing in on the solution with precision. The technical details of this adaptive hyper-parameters optimisation (**Differential Nelder–Mead Optimisation (DNO)**) can be seen in Algorithm 1.

3.5 XAI: Experiments and Visualisation

While AI models typically have good prediction accuracy due to their intricate and complex architectures, two critical aspects are generally overlooked: the explainability and interpretability of their predictions [14]. Understanding how and why the models predict specific outcomes is crucial for establishing trust, accountability and informed decision-making in practical applications.

In the context of AI-based models, complex models such as deep learning and Ensemble methods—which include techniques like XGBoost [71]—are frequently categorised as black-box models due to their inherently opaque nature and the lack of transparency surrounding their internal workings. The intricate architectures of these advanced models create significant challenges in discerning the specific contributions that individual features make toward the final prediction and elucidating the complicated interactions between these various features. For example, within the realm of deep learning models [41], it presents a daunting challenge to trace the precise pathways through which particular input features exert influence on the resulting output features, leading to a veil of uncertainty that can obscure the understanding of the model's behaviour. The endeavour to develop practical and comprehensible outcomes that are also explainable yields considerable advantages, including enhanced decision-making capabilities and the vital opportunity to uncover potential biases or sources

Algorithm 1: DNO

procedure DNO ($solution_0 = h_1, h_2, ..., h_{N_h}$)
Initialisation
  $N_p = 15, \zeta = 0.5, C_r = 0.5, Max_{iter} = 100$ ▷ Initialise DE parameters
  $Pop_0 = Gen(solution_i, i \in [1 : N_h])$ ▷ Generate initial population randomly
  $fit = Eval(Pop_0)$ ▷ Evaluate initial population
Differential Evolution
  for i in $[1, .., Max_{iter}]$ **do** ▷ Main loop of hyper-parameters tuning
    $r_1, r_2, r_3 = randperm(Pop_i)$ & $r_{best} = Best(Pop_i)$
    $P^1 = Pop(r_1), P^2 = Pop(r_2), P^3 = Pop(r_3), Gbest = Pop(r_{best})$
    for j in $[1, .., N_h]$ **do** ▷ Mutation and crossover loop
      if $c_r < rand()$ **then**
        $Tsol_j = Gbest_j + \zeta \times (P^1_j - P^2_j) + \zeta \times (P^3_j - Gbest_j)$
      else
        $Tsol_j = Pop^i_j$
      end if
      $fit_T = Eval(Tsol)$
      if $fit_T > fit_j$ **then** ▷ Maximise accuracy of Meta-ED model
        $fit_j = fit_T, Pop_j = Tsol$
      end if
    end for
    $\Delta P = Max(Pop_i) - Max(Pop_{i-1})$ ▷ Compute the optimisation progress
Nelder-Mead
    if $\Delta P < \Lambda$ **then** ▷ Run local search
      $< Sol_{nm}, fit_{nm} >$=Nelder-mead($Max(Pop_i)$)
      $Best_{pop} = Sol_{nm}, Best_{fit} = fit_{nm}$
    end if
  end for
  $Best_{solution} = Best_{pop_{Max_{iter}}}$
  $\langle fit_{best} \rangle = Train(Meta - ED(Best_{solution}))$
  return $Best_{solution}, fit_{best}$ ▷ Optimal hyper-parameters
end procedure

of discrimination that may exist within the model, thereby allowing for these issues to be addressed in a thoughtful and effective manner.

In this study, we developed comprehensive global and local interpretation frameworks to dissect the predictive behaviour of the proposed Meta-learning model. Global interpretation provides a high-level understanding of the model's overall dynamics, shedding light on the relationships and interactions between features and determining their relative contributions to the model's performance [32]. We utilised advanced techniques such as feature importance analysis and partial dependence plots to quantify feature relevance and visualise the interactions between predictors. These methods offer a holistic view of how key variables influence the model's output across the entire dataset. Conversely, local interpretation delves into the rationale behind individual predictions [68], providing case-specific explanations for why the model suggests certain outcomes. For this, we employed the SHAP approach [35], a sophisticated and widely used tool for isolating the contribution of each feature to a specific prediction and formulated as follows:

$$\Omega_i(h, a) = \sum_{z' \subseteq a'} \frac{|z'|! \left(N_f - |z'| - 1\right)!}{N_f!} \left[h_a(z') - h_a(z' \backslash i)\right], \tag{5}$$

Table 4. Statistical Results of ED Visitor Prediction for Three Types of the Proposed Meta-ED Models

Meta-ED-MLP									
Metric	RMSE	MAE	R	MSLE	MEDAE	EVS	MAPE	MDA	RSE
Min	15.969	12.016	0.850	0.005	12.016	0.684	5.129	0.738	0.303
Max	19.598	15.886	0.866	0.007	15.886	0.738	6.943	0.757	0.456
Mean	17.366	13.566	0.856	0.006	13.566	0.714	5.915	0.748	0.360
STD	1.1E+00	1.1E+00	4.2E-03	6.9E-04	1.1E+00	1.6E-02	5.1E-01	5.5E-03	4.6E-02
Meta-ED-LR									
Metric	RMSE	MAE	R	MSLE	MEDAE	EVS	MAPE	MDA	RSE
Min	21.853	17.709	0.831	0.009	17.709	0.677	7.766	0.725	0.567
Max	24.911	21.135	0.847	0.011	21.135	0.713	9.206	0.754	0.737
Mean	23.545	19.581	0.836	0.010	19.581	0.691	8.564	0.746	0.659
STD	8.1E-01	9.0E-01	4.9E-03	5.9E-04	9.0E-01	9.8E-03	3.8E-01	6.7E-03	4.5E-02
Meta-ED-Ridge									
Metric	RMSE	MAE	R	MSLE	MEDAE	EVS	MAPE	MDA	RSE
Min	22.655	18.755	0.825	0.009	18.755	0.669	8.184	0.731	0.610
Max	24.239	20.439	0.847	0.011	20.439	0.707	8.938	0.757	0.698
Mean	23.680	19.708	0.835	0.010	19.708	0.687	8.632	0.746	0.667
STD	4.8E-01	5.0E-01	6.7E-03	4.1E-04	5.0E-01	1.2E-02	2.2E-01	8.0E-03	2.7E-02

where Ω and h are the Shapley value of ith features and black-box model, and a is the sample, $|z'|$ refers to the subset of sample count which is not equal to zero in z'. Additionally, N_f is the number of features that play the role of inputs in the first layer. By leveraging this method, we were able to illuminate the inner workings of the model's decision-making process at a granular level, providing actionable insights into its behaviour on a case-by-case basis. The combination of these global and local interpretation methods ensures a well-rounded and transparent understanding of both the general and instance-specific behaviour of the Meta-learning model, enhancing its interpretability and trustworthiness.

4 Results and Discussions

4.1 Fundamental Prediction Results

In this study, we compare the performance of 10 sequential deep learning models in predicting daily ED visitor numbers (the details of evaluation metrics can be seen in Supplementary Section A.4) and assess the impact of the number of layers on model outcomes using nine metrics. Table 2 presents the detailed statistical prediction results of the 10 time-series models, including LSTM, BiLSTM and GRU. To evaluate the performance of each model, we employed 10-fold cross-validation and reported the statistics, including minimum, maximum, average and STD. Additionally, we evaluated the stacked architectures of these three models, comparing versions with two layers (denoted as s-LSTM) and larger models with four learning layers (denoted as l-LSTM). The results indicated that BiLSTM achieved the best accuracy (R-value) compared to LSTM and GRU. Furthermore, increasing the number of learning layers resulted in only a marginal improvement in average prediction accuracy (around 2%). Due to the increased training time, this additional complexity is not recommended if the sample size is below 10,000.

Table 4 presents a detailed analysis of the performance metrics associated with two distinct NN models, namely the MLP and the DNN, alongside two tree-based models, which include RF and ExT, as well as an impressive collection of six ensemble models, all evaluated across a total of nine different assessment metrics that provide a comprehensive view of their efficacy. Upon careful examination of the prediction results, it becomes clear that the standout winner among these models was none other than CatB, which achieved an accuracy rate of 0.839 and a **Mean Absolute Error (MAE)** of 15.04, showcasing its remarkable capabilities in the realm of predictive modelling. This considerable performance can likely be attributed to CatB's innovative ordered boosting methodology, specifically designed to minimise overfitting by ensuring that any potential leakage of target information is reduced, thereby enhancing the overall integrity of the model. The ordered boosting

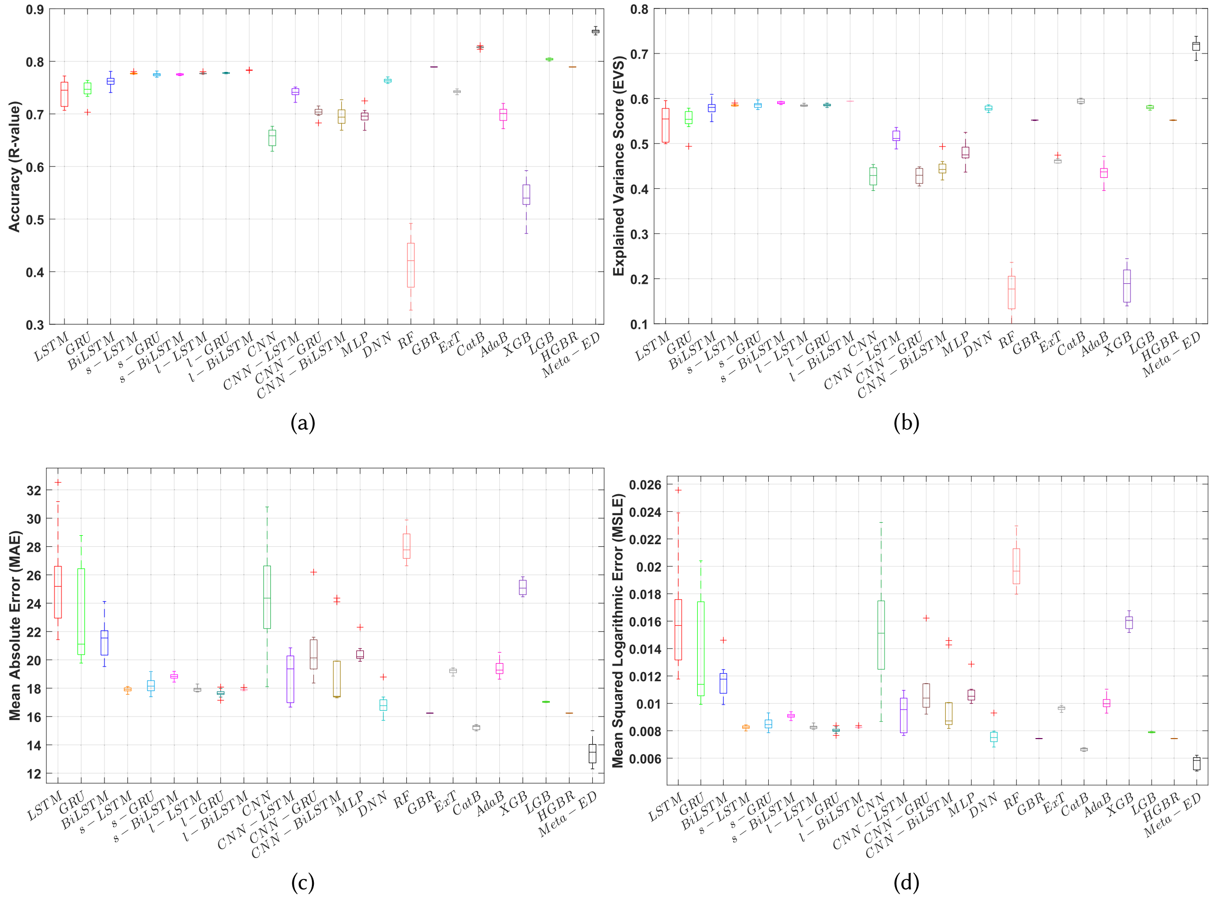


Fig. 7. The statistical results for the proposed hybrid model's performance compared with 23 ML, DL and ensemble models in predicting the number of ED visitors (daily) based on collected data from Canberra Hospital.

technique employed by CatB adeptly addresses this significant issue, resulting in a model that not only produces more accurate predictions but also demonstrates a higher level of reliability, particularly in contexts where either the size or quality of the data poses challenges to effective analysis. Furthermore, it is worth noting that the DNN outperformed its MLP counterparts significantly, a distinction that can be attributed to their superior network depth, which enables them to effectively capture intricate relationships within the data, along with their advanced capabilities in feature extraction and a more robust application of regularisation techniques that help prevent overfitting and enhance model performance.

Additionally, we evaluated the effectiveness of three different master-learners—MLP, linear regression and Ridge regression—on the proposed Meta-ED prediction model. The relevant results are in Table 4. We can see that a combination of MLP with four sub-learners (Meta-ED-MLP) outperformed the other two models by 2% on average in terms of accuracy.

The analysis conducted on 23 deep learning and ML models to predict the daily number of ED visitors is detailed in Figure 7. Each model was trained and tested in 10 independent iterations, culminating in a comprehensive evaluation of its performance as showcased within the depicted boxplot. In terms of prediction accuracy, as

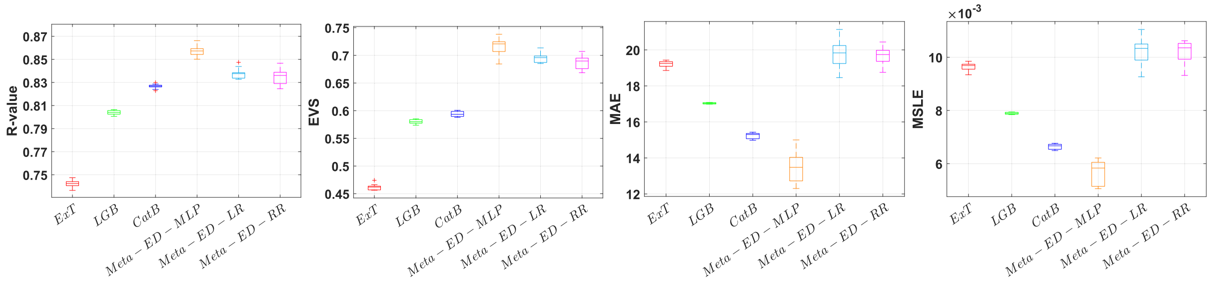


Fig. 8. The statistical prediction results for the proposed Meta-learning models compared with their components (ExT, LGB and CatB) in predicting ED visitor number (daily) based on collected data from Canberra Hospital.

indicated by the R-value (refer to Figure 7(a)), the standout performer was the CatB ensemble model, exhibiting an impressive average accuracy of 84%. Following closely behind, LGB and GBR secured the second and third positions with 80% and 79% accuracy rates, respectively. In evaluating predictive models using the **Explained Variance Score (EVS)** as the second metric (refer to Figure 7(b)), which quantifies the predictability of the dependent variable from the independent variables, significant insights emerged. The EVS metric spans a range from negative infinity to 1, where a score of 1 signifies excellent prediction accuracy, while lower scores denote decreasing predictive precision. Remarkably, CatB emerged as the top performer, showcasing an impressive average EVS of 0.64, surpassing other models by a notable margin. Specifically, CatB's EVS outperformed LGB by 10.3%, underscoring its superior ability to capture and explain the variance in the dataset, thereby significantly enhancing its predictive power in this analytical context. In the assessment of MAE and **Mean Squared Logarithmic Error (MSLE)**, where lower values signify superior performance and reduced training errors, noteworthy observations were made. Illustrated in Figure 7(c) and (d), CatB particularly outshone all other 22 ML models in these metrics, reaffirming its exceptional performance. This unique advantage of CatB, which can be attributed to its utilisation of an advanced variant of GBR that incorporates a more potent form of regularisation, is truly enlightening. This regularisation technique mitigates overfitting, enhancing the model's generalisation capabilities. Such expertise is particularly advantageous in scenarios with limited data, such as our dataset, which comprises 8,600 samples. CatB's ability to prevent overfitting and bolster generalisation performance sets it apart, leading to superior predictive accuracy and model robustness.

To evaluate the benefits of the proposed Meta-ED model, which incorporates various master-learners (MLP, LR, RR) against individual ensemble models such as CatB, LGB and ExT, we conducted a comparative experiment by running each method 10 times. The results, displayed in Figure 8, demonstrate that Meta-ED-MLP outperformed other Meta models as well as ensemble models in terms of R-value, EVS, MAE and MSLE on average.

To facilitate a comparative visualisation of the proposed Meta-ED learner against other widely used ensemble models, Figure 9 is presented. The results clearly demonstrate that the Meta-ED model excels in capturing the complex and dynamic patterns of ED visitor numbers, as evidenced by testing data from 2019 to 2022. The zoomed-in segments of the overall line chart further highlight the Meta-ED model's superior accuracy in predicting various fluctuation patterns in visitor numbers. Besides, both CatB and LGBoost exhibit comparable performance, effectively predicting the upward and downward trends in ED crowding. However, they do not match the precision of the Meta-ED model.

The results presented in Table 5 indicate a comprehensive comparison of the proposed Meta-ED model against 23 other ML models evaluated using the Friedman test. Notably, the Meta-ED model attained the highest average performance rank based on the correlation coefficient (R-value), highlighting its superior predictive capability compared to the other evaluated models. The p-value results further substantiate this finding, demonstrating

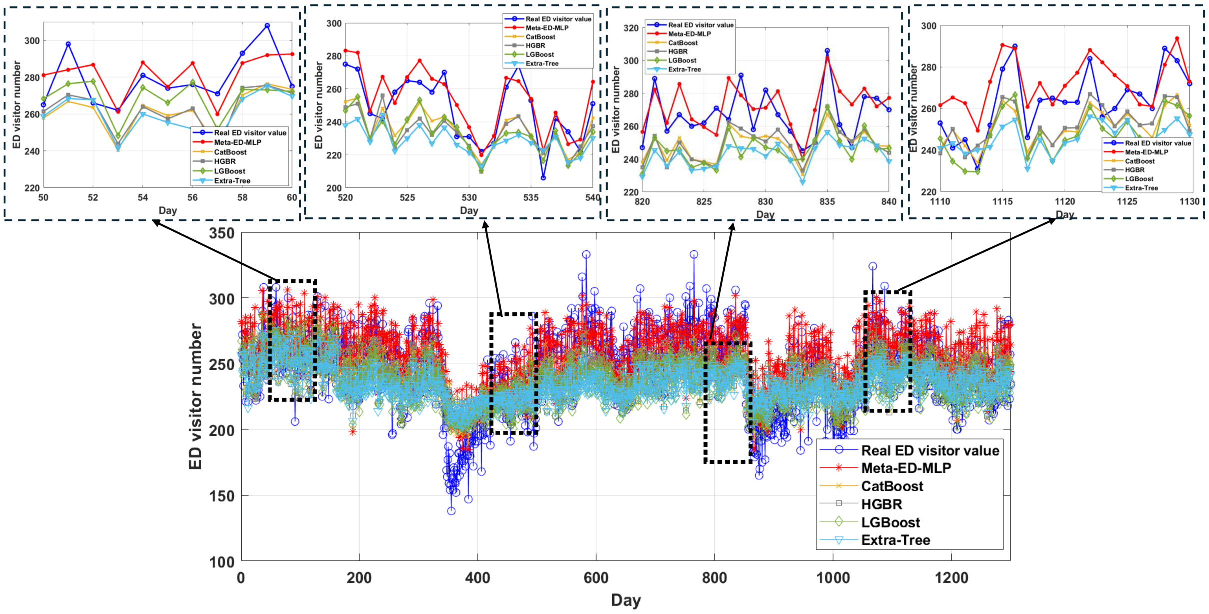


Fig. 9. Performance comparison between the proposed Meta-ED and other well-known ensemble models with the real ED visitor number. The test range data span from 1 January 2019 (Day 0) to December 2022 (Day 1200).

Table 5. Comparative Analysis of 23 ML Models and the Proposed Meta-ED Model Using Friedman Test: p-Values and Average Ranks

Model	LSTM	GRU	BiLSTM	s-LSTM	s-GRU	s-BiLSTM	l-LSTM	l-GRU
Avg-Rank	14.6	15.4	12.9	9	9.9	10.1	8.7	7.9
p-value	2.03E-13	3.05E-15	9.01E-18	3.55E-24	8.07E-23	2.12E-24	3.18E-24	2.09E-24
Model	l-BiLSTM	CNN	CNN-LSTM	CNN-GRU	CNN-BiLSTM	MLP	DNN	RF
Avg-Rank	6	21.8	15.7	18.8	19.5	19.9	12.5	24
p-value	5.62E-24	8.54E-21	5.04E-21	6.28E-23	1.09E-17	1.36E-19	1.78E-23	1.02E-17
Model	GBR	ExT	CatB	AdaB	XGB	LGB	HGBR	Meta-ED-MLP
Avg-Rank	4.5	15.6	2	19.7	23	3	4.5	1
p-value	3.14E-23	6.57E-26	1.44E-15	1.34E-19	5.09E-19	2.32E-20	3.17E-23	N/A

that the performance of the Meta-ED model is statistically significantly different from that of the other models. This suggests that the improvements offered by the Meta-ED model are not only substantial but also reliable, highlighting its potential as a robust solution for the targeted prediction tasks. The combination of top-ranking performance and significant p-values positions the Meta-ED model as a leading choice in the evaluated set of ML models.

4.2 Prediction Results without Admission Year Feature

Given the strong correlation between the year of admission and the number of ED visitors, as illustrated in Figure 3, we evaluated the impact of excluding this feature on the predictive accuracy of the proposed Meta-ED model and four other high-performing models. Figure 10 and Supplementary Table 7 demonstrate that removing the admission year from the training dataset led to a decrease in prediction accuracy by 3.21%, 7.69%, 7.69%, 7.32% and 2.79% for ExT, LGB, CatB, HGBR and Meta-ED, respectively. The Meta-ED model demonstrates greater

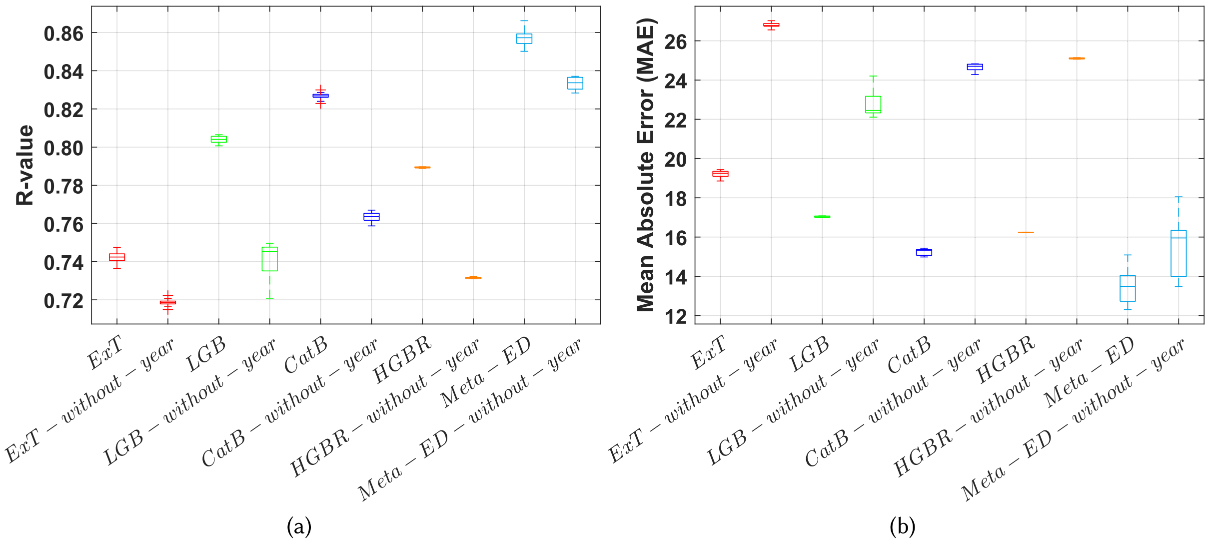


Fig. 10. A performance comparison of models trained with the inclusion of the Year feature versus those trained without it.

resilience to the removal of the admission Year feature due to its ensemble structure, which integrates multiple models with varying dependencies on input features, and the adaptive capacity of its meta-learner to redistribute predictive emphasis among robust components.

4.3 Transformer Comparison Results

To rigorously analyse the performance of the proposed Meta-ED model, we conducted a comparative study with several state-of-the-art transformer-based and attention-augmented models, like a standard Transformer (encoder–decoder), self-attention Transformer and TabNet with attention. The encoder–decoder Transformer with three layers and four attention heads achieved an R-value of 0.34 and an MAE of 24.6 visits per day, indicating a moderate ability to represent the complex, heterogeneous character of ED data. The self-attention Transformer, however, was stronger with an R-value of 0.685 and an MAE of 16.1, capitalising on its superiority in capturing temporal dependencies more effectively. TabNet-attention, designed for tabular data with sparsity and interpretability of feature choice, produced a higher R-value of 0.70 and an MAE of 17.3 but once more fell short in capturing the high-order interactions between demographic, clinical and climate-based features. In comparison to them, the Meta-ED model performed significantly better on all three alternatives, with an average R-value of 0.856 and an MAE of 13.5. This remarkable performance demonstrates the robustness of Meta-ED's ensemble model, which leverages the strengths of heterogeneous sub-learning entities and a meta-entity, assisted by optimal feature selection and hyperparameter tuning. The results highlight the model's capacity to generalise effectively across varied inputs and demonstrate its practical applicability for predictive and scalable ED visit forecasting. Details of the comparative results are summarised in Table 6.

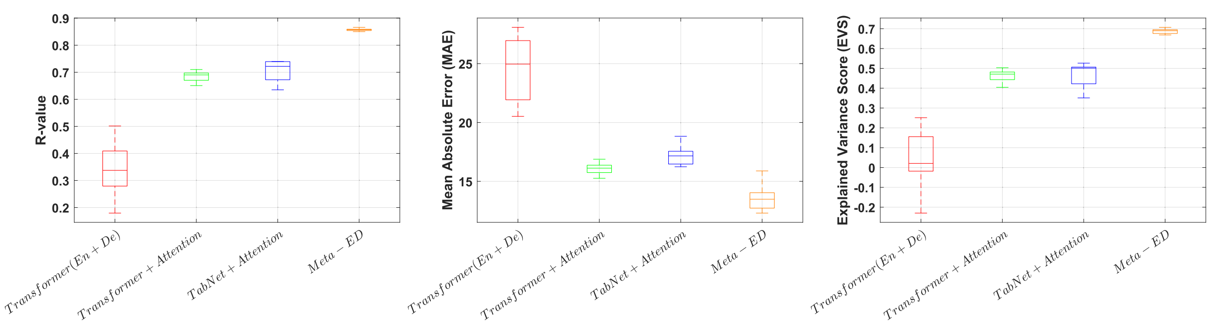
Figure 11 presents a comparative analysis of the proposed Meta-ED model and three transformer-based models for predicting ED visit volumes, evaluated using three metrics: R-value, MAE and EVS. As illustrated, Meta-ED consistently outperforms the transformer models across all three evaluation criteria.

4.4 Feature Selection Results

We employed the widely used and efficient RFE method to identify the optimal subset of practical features for prediction. Two key factors were considered during the feature selection phase: the number of features and the

Table 6. Comparative Analysis of Three Transformer Learning Models in Predicting the ED Visitor Number

TabNet-attention									
Metrics	RMSE	MAE	R-value	MSLE	MEDAE	EVS	MAPE	MDA	RSE
Min	2.08E+01	1.62E+01	6.35E-01	8.17E-03	1.62E+01	3.51E-01	7.11E+00	6.65E-01	5.15E-01
Max	2.38E+01	1.88E+01	7.49E-01	1.11E-02	1.88E+01	5.57E-01	8.16E+00	7.07E-01	6.73E-01
Mean	2.19E+01	1.73E+01	7.03E-01	9.06E-03	1.73E+01	4.66E-01	7.51E+00	6.90E-01	5.73E-01
STD	1.09E+00	8.46E-01	4.40E-02	9.71E-04	8.46E-01	6.82E-02	3.77E-01	1.60E-02	5.77E-02
Transformer-attention									
Metrics	RMSE	MAE	R-value	MSLE	MEDAE	EVS	MAPE	MDA	RSE
Min	1.92E+01	1.53E+01	6.51E-01	6.48E-03	1.53E+01	4.04E-01	6.41E+00	6.67E-01	5.10E-01
Max	2.15E+01	1.69E+01	7.10E-01	8.11E-03	1.69E+01	5.03E-01	7.21E+00	7.19E-01	6.36E-01
Mean	2.04E+01	1.61E+01	6.85E-01	7.29E-03	1.61E+01	4.61E-01	6.82E+00	6.91E-01	5.75E-01
STD	6.28E-01	4.78E-01	1.65E-02	4.60E-04	4.78E-01	2.69E-02	2.33E-01	1.38E-02	3.53E-02
Transformer (encoder–decoder)									
Metrics	RMSE	MAE	R-value	MSLE	MEDAE	EVS	MAPE	MDA	RSE
Min	2.53E+01	2.05E+01	1.80E-01	1.17E-02	2.05E+01	−2.30E-01	8.70E+00	5.59E-01	7.62E-01
Max	3.42E+01	2.81E+01	5.02E-01	2.10E-02	2.81E+01	2.51E-01	1.15E+01	6.24E-01	1.39E+00
Mean	3.00E+01	2.46E+01	3.44E-01	1.64E-02	2.46E+01	4.29E-02	1.02E+01	5.90E-01	1.08E+00
STD	3.10E+00	2.66E+00	9.25E-02	3.30E-03	2.66E+00	1.34E-01	9.57E-01	1.80E-02	2.20E-01

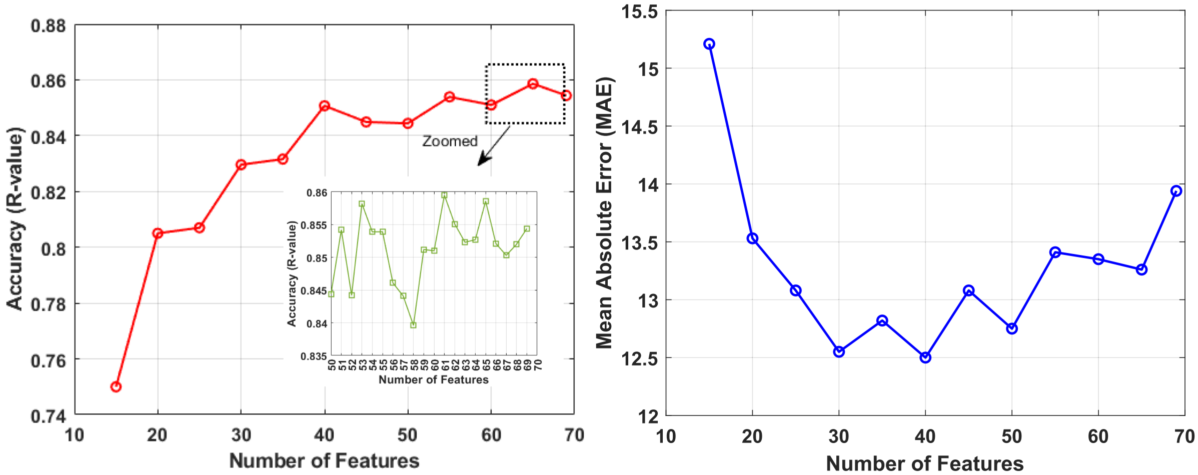


(a) (b) (c)

Fig. 11. The statistical prediction results compare three transformer-based models with the proposed Meta-ED model using three evaluation metrics: (a) R-value, (b) MAE and (c) EVS.

specific subset of features chosen. We conducted a comprehensive evaluation of the performance of RFE with varying feature counts, ranging from 15 to 69 (all available features). Figure 12 illustrates the accuracy of these feature selection experiments and MAE. The results indicate that the optimal number of features, in terms of accuracy, is 61. However, the lowest error was observed with 30 and 40 features when focusing on the validation of MAE.

Table 7 shows that the Meta-ED analyses encompass a series of distinct perspectives on ED visitor prediction and the 'Impact' column quantifies the reduction in prediction accuracy observed after the removal of specific features from the Meta-ED model. Meta-ED-1 delves into forecasting without temporal markers like year, month and day. Conversely, Meta-ED-2 explores predictions void of disposition parameters, condensing the intricate variability into a more streamlined outcome. In a similar vein, Meta-ED-3 studies performance without age group influences, offering a purer insight into prediction dynamics. Moving along, Meta-ED-4 isolates the impact of 12 prevalent ICD10 diagnoses from the equation, refining the predictive landscape. Furthermore, Meta-ED-5 extends this observation by excluding the top 5 daily ICD10 occurrences, challenging conventional forecasting norms.



(a) (b)

Fig. 12. The performance of RFE in optimising the subset of features based on the Meta-ED model.

Table 7. Effect of Feature Selection on Meta-ED Model Performance in Predicting Daily ED Visitor Numbers

Model	Details	Accuracy	Impact (%)
Meta-ED-1	Without temporal features (year, month and day)	45.9%	86.97
Meta-ED-2	Without disposition	84.9%	0.96
Meta-ED-3	Without age groups	84.3%	1.66
Meta-ED-4	Without most viral ICD10 features	81.1%	5.76
Meta-ED-5	Without most frequent five ICD10 features (daily)	61.6%	39.25
Meta-ED-6	Without most frequent ICD10 features	76.7%	11.82
Meta-ED-7	Without triage features	83.5%	2.66
Meta-ED-8	Without climate parameters	83.0%	3.25
Meta-ED-9	All features	85.7%	–

Based on the examination, the Meta-ED-6 tests model performance in the absence of the 10 most frequent ICD-10 codes, revealing how predictive accuracy changes under limited diagnostic evidence. Meta-ED-7, in contrast, removes triage-based features and identifies a different factor in predicting performance, demonstrating the model's adaptability under early clinical evaluation conditions that are not typically available. Finally, Meta-ED-8 utilises the complete 69-feature input set, providing a full picture of the complex mix of factors influencing ED visit prediction.

After evaluating the performance of Meta-ED by removing specific features (from Meta-ED-1 to Meta-ED-7 as illustrated in Figure 13), the reduction in prediction accuracy compared to Meta-ED-8 (which utilises the full set of features) is noteworthy, with declines of 87%, 1.0%, 1.66%, 5.76%, 39.25%, 3.25% and 2.66%, respectively. The results indicate that the most critical feature is temporal markers, including year, month and day (Meta-ED-1). The absence of these temporal features severely impacts the model's ability to forecast ED visitors, as it removes essential seasonal trends, such as those associated with flu outbreaks in winter. Additionally, disregarding temporal

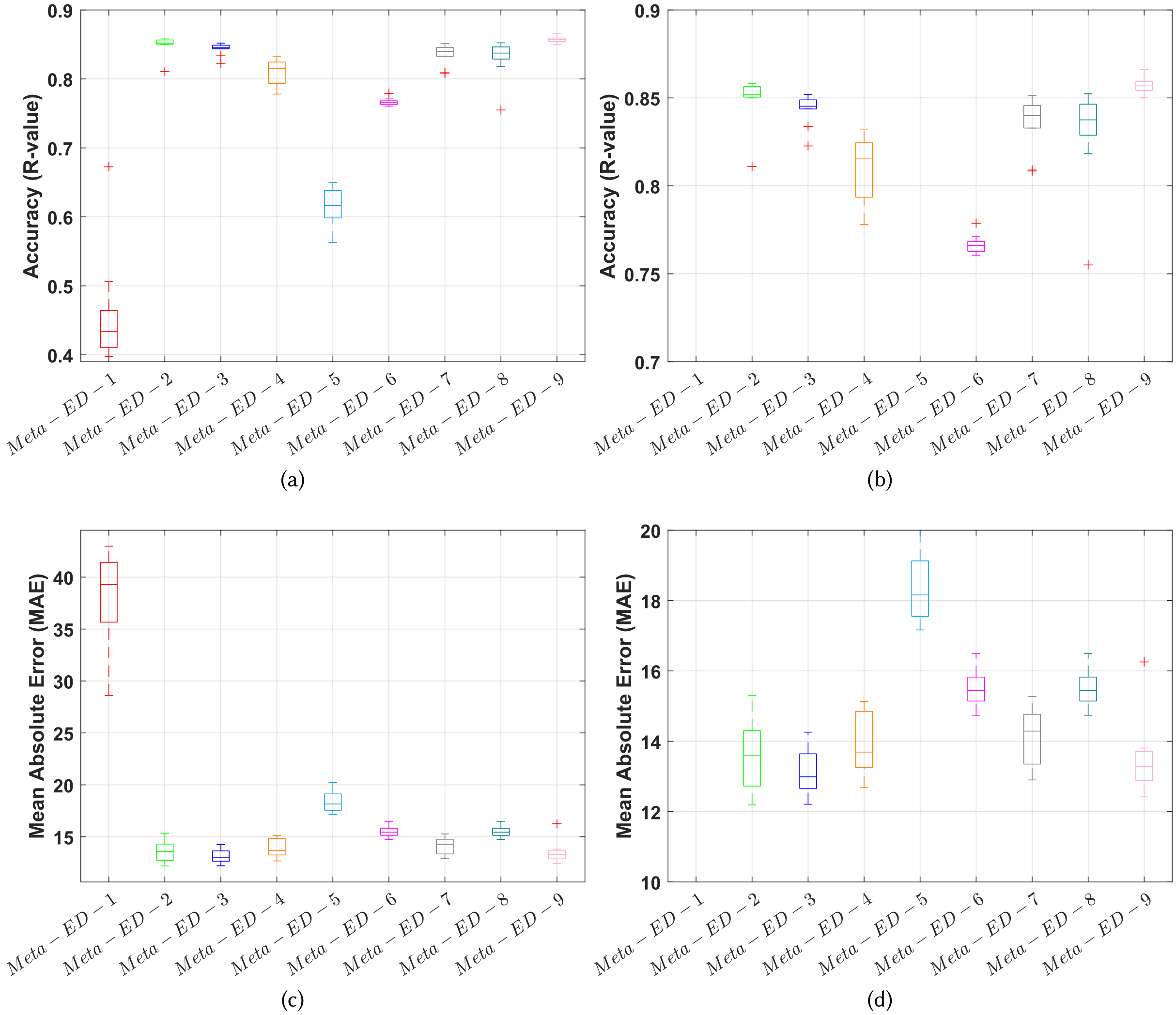
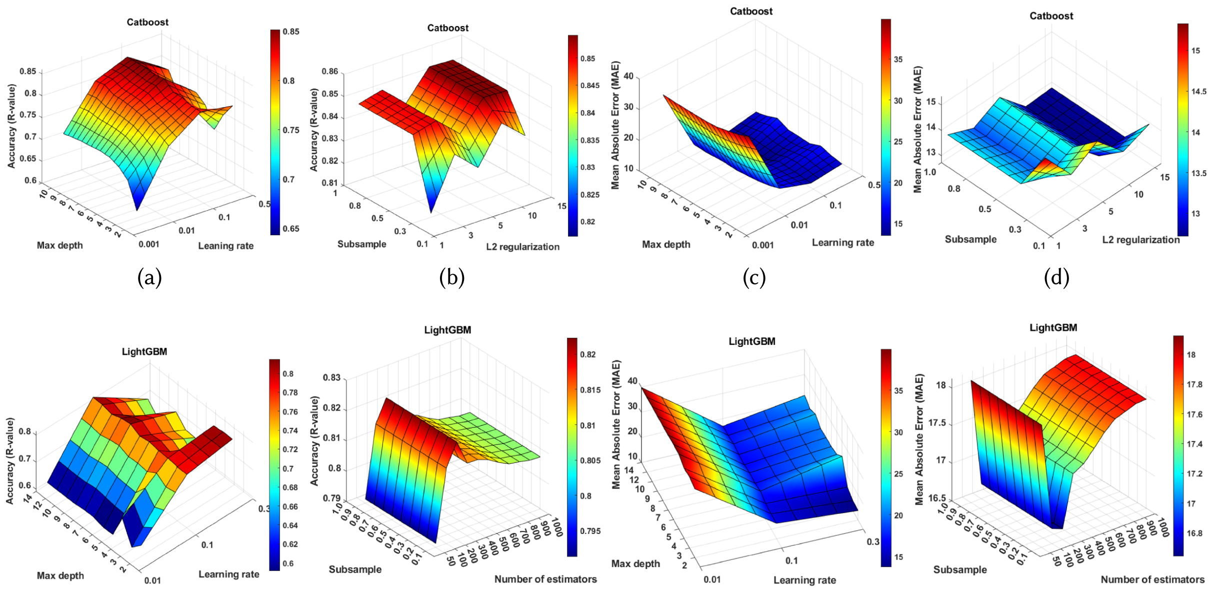


Fig. 13. Impact of feature selection on proposed Meta-ED model performance: (a) accuracy, (b) zoomed version of accuracy, (c) MAE, and (d) zoomed version of MAE.

context may obscure evolving patterns in ED visitation, driven by factors such as population growth and shifts in healthcare accessibility, ultimately removing critical insights essential for accurate and context-aware forecasting.

4.5 Hyper-Parameters Tuning

The fine-tuning of hyper-parameters is essential for effectively training contemporary ML models and has a significant impact on boosting their efficacy. It also aids in navigating the complexities and frameworks of these algorithms [26]. Thus, selecting hyper-parameters judiciously prior to model training is not merely important—it is essential for ensuring optimal performance and generalisability. Techniques that adjust based on data, like cross-validation, are frequently employed to refine and assess models using separate training and testing datasets. Classic hyper-parameter search techniques encompass grid search and random search. Furthermore, recent



(a) (b) (c) (d)

(e) (f) (g) (h)

Fig. 14. Visualising performance landscape: Grid search optimisation of hyper-parameters for CatB and LGB models based on accuracy (R-value) and MAE.

studies have unveiled more sophisticated methodologies, including Bayesian optimisation and meta-heuristic algorithms, which are specifically designed to tackle unique challenges for enhanced efficiency and precision.

To explore the hyper-parameter search space, we selected the two best-performing models: CatB and LGB. Using grid search for CatB, we evaluated four key hyper-parameters: maximum tree depth, learning rate, sub-sample rate and L2 regularisation. The hyper-parameters for LGB were similar, with the exception of L2 regularisation, which was replaced by the number of estimators. These hyper-parameters were chosen to optimise the models' predictive performance by fine-tuning their complexity and learning capabilities. Figure 14 illustrates the performance landscape of CatB and LGB based on accuracy and MAE. As shown in Figure 14(a) and (e), the learning rate and tree depth have a more significant impact compared to other hyper-parameters, leading to an improvement of up to 20% in terms of accuracy.

Furthermore, Supplementary Figure 7 presents a performance comparison of two optimisation methods, NM and DE, for tuning the hyper-parameters of the Meta-ED model. The most notable observation is the superior performance of DE in finding optimal settings compared to the NM method. NM faced premature convergence and became stuck in a local optimum after only a few iterations, lacking an effective strategy to escape these sub-optimal regions. Finally, Supplementary Table 6 presents the optimal hyperparameter configurations identified for the sub-learners: LGB, CatB, RF and ExT.

4.6 XAI and Feature Importance

As illustrated in Figure 15, the most influential feature is the year of attendance in the ED, which has a significant positive impact on predicting visitor numbers. Following this, the features ICD10 Z53.1 (did not wait for treatment) and the five most frequent daily diagnoses (first ICD-10, second ICD-10, etc.) exhibit an impact range of −10 to +10, with lower values indicating a negative contribution to the predictions. Other features have a more modest influence, with impacts ranging between −5 and +5. Furthermore, Figure 16 and Supplementary Figure 8 illustrate

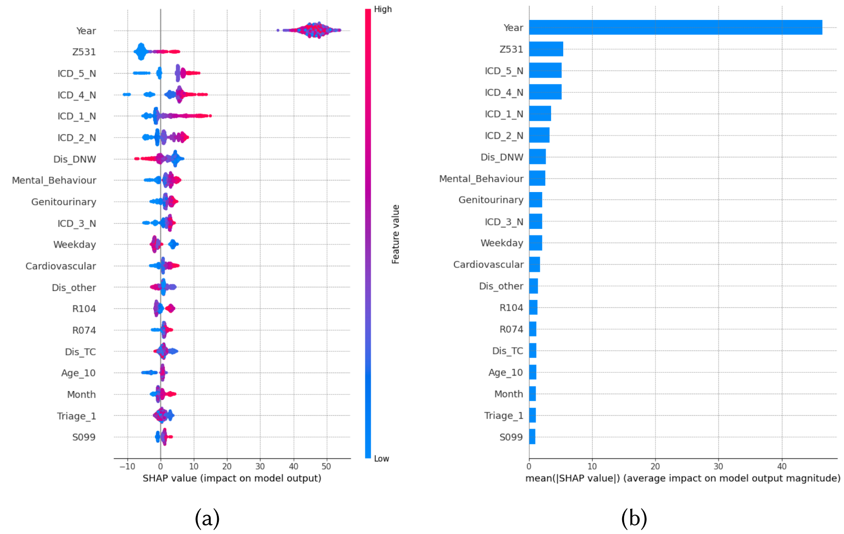
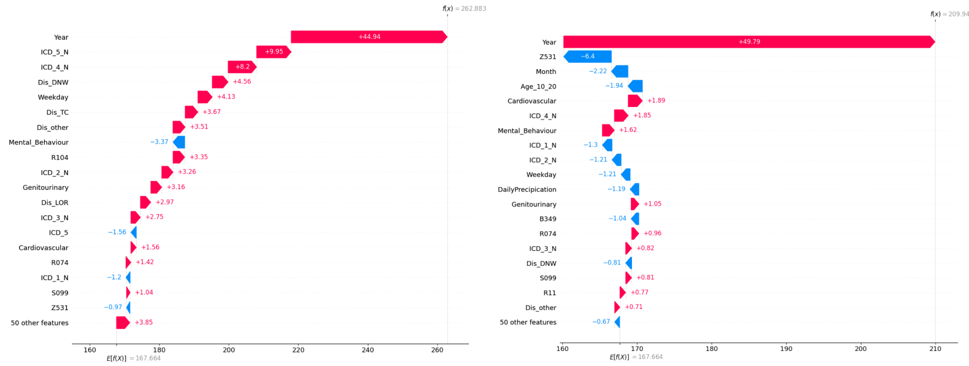


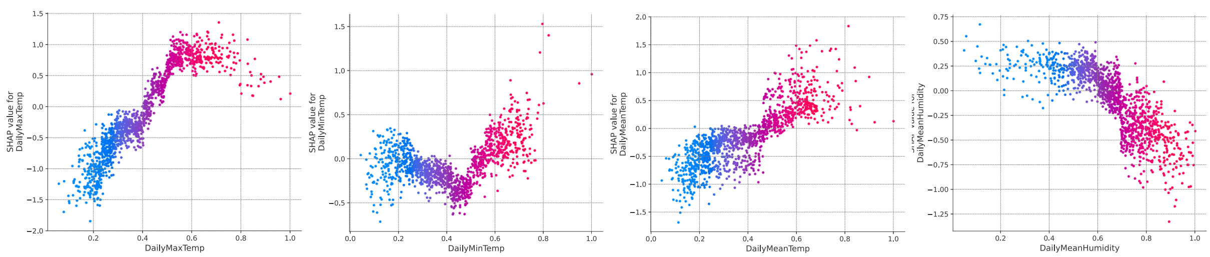
Fig. 15. Summary of SHAP values illustrating the importance of various features in predicting the daily number of ED visitors. (a) The distribution of SHAP values for each feature highlights their respective contributions to model predictions, revealing how each feature influences the expected volume of ED visitors. (b) The accompanying bar plot displays the mean absolute SHAP value for each feature across all instances in the test dataset. Higher mean absolute values indicate greater contributions to predicting ED visitor numbers, allowing for a clear comparison of feature significance in driving patient inflow.

the contributions of various features, which can be either positive or negative, in predicting the number of ED visitors based on four random samples from the test dataset. As anticipated, the Year feature shows the highest positive contribution, ranging from 45 to 50. Furthermore, the fourth and fifth most frequent ICD-10 codes rank as the second most influential features, contributing between 5 and 10 across different samples. The blue-coloured features indicate a negative contribution to the final prediction values, such as ICD-10 code Z53.1. We conducted an analysis to examine how the values of climate features interact to influence the model's output. The results of this analysis are presented in Figure 17. For instance, as illustrated in Figure 17(a), low values of maximum daily temperature are associated with negative SHAP values, suggesting that colder weather may result in fewer ED visits. This phenomenon could be explained by reduced outdoor activity levels or a decrease in heat-related health issues on colder days. Conversely, higher maximum daily temperatures correspond to positive SHAP values, indicating that the predicted number of ED visitors also increases as temperatures rise. This trend may be linked to a rise in heat-related health issues, such as heat exhaustion, dehydration and exacerbations of chronic conditions, which tend to be more prevalent during warmer weather. Furthermore, Supplementary Figure 9 illustrates the relative importance of various climate parameters, including daily precipitation, wind gust speed and direction, as well as general wind speed and direction, in influencing ED visit predictions. The visualisation in Supplementary Figure 10 shows the varying contribution levels of the primary reasons for ED visits. With the exception of the N39.0 ICD-10 code, which represents urinary tract infections, other frequent diagnosis codes within the ICD-10 system positively impact ED visitor numbers by influencing patient volumes. Conversely, a decrease in these rates results in a negative impact. An analysis reveals the differential effects of ICD-10 codes on ED visitor rates ranging from −8 to +6. Building upon the preceding analysis, Z53.1 exhibits the most significant positive (+6) and negative (−8) contributions to the output model. Cardiovascular and genitourinary parameters are closely followed, displaying positive and negative impact values at +5 and −3. Finally, Supplementary Figure 11 highlights the substantial predictive contributions of the five most frequently recorded ICD-10 codes on a daily basis. The



(a) (b)

Fig. 16. The waterfall plot illustrates the contributions of individual features to the model's predictions for ED visitor numbers across four random samples from the test dataset. This visualisation aids in understanding how different features influence the predictions, either increasing or decreasing them relative to a baseline value. Z53.1 (Did not wait for treatment), R10.4 (Pain in abdomen).



(a) (b) (c) (d)

Fig. 17. Evaluating the importance of climate parameters in predicting the daily number of ED visitors using SHAP values.

analysis indicates that higher values of these features are associated with a positive influence on prediction outcomes, whereas lower values correspond to a negative impact.

4.7 Real-Time Implementation and Ethical Considerations

In the interest of the pragmatic applicability of the proposed Meta-ED framework in real-world hospital settings, it is essential to address significant concerns related to actual implementation in real-time, computational efficiency, model flexibility and the ethical issues that arise from the use of AI systems in medicine. This subsection outlines how the proposed methodology can be translated from research to practice, as well as the mechanisms to achieve responsible and efficient deployment.

Real-Time Deployment and Workflow Integration. The Meta-ED model was developed with practicality for deployment in consideration. It can accommodate real-time or near-real-time data pipelines, which are typically available in today's hospital information systems. The end-to-end process can be executed on a daily or hourly basis to forecast short-term ED visits (e.g., 24–72 hours in advance), enabling proactive resource allocation and triage planning. The recommended pipeline for deployment is the following:

— Automated data ingestion and processing, including normalisation, encoding and imputation of missing values, through reusable scikit-learn pipelines.

— Model inference, with pre-trained sub-models and a master-learner, requires low computational overhead.
— Decision-support system integration, where outputs can be delivered through RESTful APIs or hospital dashboards for clinical administrators and operational planners to consume.

Computational Efficiency and Scalability. Although the Meta-ED system consists of an ensemble of four sub-learner models (CatB, LGB, ExT and RF) and an MLP as the meta-learner, the platform is designed for offline training and real-time inference. All computationally intensive operations, such as feature selection and hyper-parameter optimisation, are performed during development with an efficient implementation of the DE and Nelder–Mead local search algorithms. This two-level optimisation not only improves performance but also optimises the model's adaptability to new data distributions across regions. Once trained, the model offers low-latency inference, with the mean prediction time for ED forecasts on a daily basis estimated to be under a few seconds on a standard CPU (Intel i7, 32 GB RAM) without a GPU. This is viable for integration into routine hospital operations without requiring drastic hardware requests. In addition to increasing portability in low-resource environments, future releases of Meta-ED will explore model compression algorithms, including knowledge distillation and pruning, that reduce memory and computational requirements without compromising prediction accuracy.

Ethical and Governance Issues. The utilisation of AI within the clinical setting demands strict adherence to ethical standards, transparency and fairness. In this regard, several provisions were instituted to facilitate reasonable use:

— Explainability was obtained through SHAP values and feature importance scores, providing both global and local insight into the model's decision-making process.
— Fairness testing was carried out to determine potential biases across subgroups (e.g., age group, gender, diagnosis groups). No systematic differences in model residuals were observed in these groups.
— Practices in data governance conformed to national and institutional privacy regulations, and all data were anonymised prior to analysis in accordance with Australian health data protection standards.

It is also noted that the Meta-ED model is intended to supplement, never supplant, clinical decision-making. However, operational deployment is made, and it ought to be accompanied by human oversight, most importantly by clinical leadership and institutional ethics committees, to ensure that AI predictions are used in a manner compliant with patient safety, equity and accountability.

4.8 Limitations and Future Directions

Despite the promising results demonstrated by the Meta-ED framework, certain limitations must be mentioned to put the findings into context and guide future refinements as follows.

Firstly, this research is based on a single retrospective database from Canberra Hospital, spanning 23 years. While the database is large and diverse, the predictive model may be influenced by local clinical practice, patient populations and environmental conditions specific to the Canberra region. Consequently, the generalisability of the model to other geographic regions is a question. External validation using data from other hospitals or health systems is a critical next step to confirm more general applicability. Secondly, the study uses administrative health data and climate data, which may not capture all underlying reasons for ED utilisation, e.g., patient behaviour, health policy or uncoded comorbidities. Future work can incorporate additional contextual data sources, such as electronic health records, mobility data or social determinants of health, to achieve greater explanatory power. Furthermore, though the model architecture and hyper-parameters were search-optimised by a hybrid approach of DE and Nelder–Mead local search, the computational expense can be high with larger datasets or for real-time applications. As part of future work, quicker, parallelisable optimisation approaches will be explored to improve scalability. Finally, while SHAP values and feature importance scores enhance interpretability, particularly for exogenous and demographic features, deep learning components like the MLP meta-learner may still be a 'black box' in complex scenarios. Future work may investigate interpretable surrogates or sparse approximation models for better clinical interpretability.

To enhance the robustness and utility of Meta-ED, we plan the following. External validation using datasets from other Australian states and international institutions to assess cross-regional performance. Also, integration into operational workflows at Canberra Hospital will be used to test real-time forecasting for staffing and triage planning. Model refinement by incorporating new data streams such as real-time syndromic surveillance, flu alerts or ambulance dispatch records. The development of a lightweight version of Meta-ED is essential for resource-constrained settings, utilising dimensionality reduction and model compression techniques. By addressing these limitations and expanding the scope of the evaluation, we aim to further position Meta-ED as a scalable and generalisable solution for ED forecasting in diverse healthcare environments.

5 Conclusions

In this study, we address the limitations of existing AI-driven frameworks by introducing a novel Meta-Learning Gradient Booster (Meta-ED) approach for accurately forecasting daily ED visits. Leveraging a comprehensive dataset of exogenous variables that include socio-demographic characteristics, healthcare service usage, chronic conditions, diagnoses, climate and temporal variables spanning 23 years from Canberra Hospital in the ACT, Australia, Meta-ED aims to enhance predictive accuracy and robustness for ED attendance patterns.

The Meta-ED model comprises four foundational learners—CatB, RF, ExT and LGB—integrated with a reliable master-learner, MLP. By combining the unique capabilities of various base models (sub-learners), Meta-ED strengthens the precision and reliability of its predictions. We evaluate the effectiveness of Meta-ED through a comprehensive comparative analysis involving 23 distinct models, encompassing diverse architectures such as sequential models, NNs, tree-based algorithms and ensemble methods.

The results indicate that Meta-ED significantly outperforms other models, achieving an accuracy of 85.7% (95% CI [85.4%, 86.0%]) across ten evaluation metrics. Compared with leading techniques, including XGBoost, RF, AdaBoost, LGB and ExT, Meta-ED shows substantial accuracy improvements of 58.6%, 106.3%, 22.3%, 7.0% and 15.7%, respectively. Furthermore, Meta-ED outperforms transformer-based and attention-enhanced models on both key evaluation metrics (R-value, MAE) in the context of ED visit prediction. In the meantime, incorporating weather-related features results in a 3.25% improvement in prediction accuracy, emphasising the model's ability to capture seasonal trends that affect patient volumes.

These promising results establish Meta-ED as a foundational model for precise prediction of daily ED visits, demonstrating its potential as a reliable and robust framework for predictive analytics in dynamic healthcare environments. Meta-ED offers a promising approach to improving prediction accuracy, particularly for sequential data characterised by heterogeneous features and limited datasets.

Ethics Approval

This study has received ethical approval from The ACT Health Human Research Ethics Committee (Approval No. 2024/ETH01037).

Data Availability

The ED Presentation dataset utilised in this study is proprietary to Canberra Health Services and the ACT Health Directorate. Data availability for our research is contingent upon accessing anonymised datasets, a process facilitated through coordination with Michael Phipps (michael.phipps@act.gov.au), the esteemed Senior Director of the Canberra Health Service. By engaging with him, researchers can access the requisite anonymised data sets for conducting comprehensive and insightful analyses.

Declaration of Conflicting Interests

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

Acknowledgments

The authors used Grammarly and ChatGPT 4o to assist with language editing and improving the clarity of the manuscript. All content was critically reviewed and approved by the authors to ensure accuracy and integrity.

References

[1] Patrick Aboagye-Sarfo, Qun Mai, Frank M. Sanfilippo, David B. Preen, Louise M. Stewart, and Daniel M. Fatovich. 2015. A comparison of multivariate and univariate time series approaches to modelling and forecasting emergency department demand in Western Australia. *Journal of Biomedical Informatics* 57 (2015), 62–73.

[2] ACT Government. 2023. ACT Health Quarterly Performance Report Quarter 1 and Quarter 2, 2023–24.

[3] Mohamed Afilal, Lionel Amodeo, Farouk Yalaoui, and Frédéric Dugardin. 2019. Forecasting patient flows into emergency services. In *Hospital Logistics and e-Management: Digital Transition and Revolution*. Philippe Blua, Farouk Yalaoui, Lionel Amodeo, Michaël De Block, and David Laplanche (Eds.), Wiley, 85–143.

[4] Emrah Akkoyun, Sebastian T. Kwon, Aybar C. Acar, Whal Lee, and Seungik Baek. 2020. Predicting abdominal aortic aneurysm growth using patient-oriented growth models with two-step Bayesian inference. *Computers in Biology and Medicine* 117 (2020), 103620.

[5] Helena Seabra Almeida, Margarida Sousa, Inês Mascarenhas, Ana Russo, Manuel Barrento, Manuel Mendes, Paulo Nogueira, and Ricardo Trigo. 2022. The dynamics of patient visits to a public hospital pediatric emergency department: A time-series model. *Pediatric Emergency Care* 38, 1 (2022), e240–e245.

[6] Hugo Álvarez-Chaves, Iván Maseda-Zurdo, Pablo Muñoz, and María D. R-Moreno. 2024. Evaluating the impact of exogenous variables for patients forecasting in an emergency department using attention neural networks. *Expert Systems with Applications* 240 (2024), 122496.

[7] Yuval Barak-Corren, Pradip Chaudhari, Jessica Perniciaro, Mark Waltzman, Andrew M. Fine, and Ben Y. Reis. 2021. Prediction across healthcare settings: A case study in predicting emergency department disposition. *NPJ Digital Medicine* 4, 1 (2021), 169.

[8] Francisco M. Caldas and Cláudia Soares. 2022. A Temporal fusion transformer for long-term explainable prediction of emergency department overcrowding. In *Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 71–88.

[9] Emergency Department of Canberra Hospital, Canberra Health Services. 2024. Retrieved July 16, 2024 from https://www.canberrahealthservices.act.gov.au/

[10] Kuo-Hao Chang. 2012. Stochastic Nelder–Mead simplex method—A new globally convergent direct search method for simulation optimization. *European Journal of Operational Research* 220, 3 (2012), 684–694.

[11] Pui Hing Chau, Kevin Ka-Lun Lau, Xing Xing Qian, Hao Luo, and Jean Woo. 2022. Visits to the accident and emergency department in hot season of a city with subtropical climate: Association with heat stress and related meteorological variables. *International Journal of Biometeorology* 66, 10 (2022), 1955–1971

[12] Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.

[13] Anderson Bessa Da Costa, Larissa Moreira, Daniel Ciampi De Andrade, Adriano Veloso, and Nivio Ziviani. 2021. Predicting the evolution of pain relief: Ensemble learning by diversifying model explanations. *ACM Transactions on Computing for Healthcare* 2, 4 (2021), 1–28.

[14] Rudresh Dwivedi, Devam Dave, Het Naik, Smiti Singhal, Rana Omer, Pankesh Patel, Bin Qian, Zhenyu Wen, Tejal Shah, Graham Morgan, et al. 2023. Explainable AI (XAI): Core ideas, techniques, and solutions. *ACM Computing Surveys* 55, 9 (2023), 1–33.

[15] Thomas Falvo, Lance Grove, Ruth Stachura, David Vega, Rose Stike, Melissa Schlenker, and William Zirkin. 2007. The opportunity loss of boarding admitted patients in the emergency department. *Academic Emergency Medicine* 14, 4 (2007), 332–337.

[16] Bin Feng, Zequn Liu, Nanlan Huang, Zhiping Xiao, Haomiao Zhang, Srbuhi Mirzoyan, Hanwen Xu, Jiaran Hao, Yinghui Xu, Ming Zhang, et al. 2024. A bioactivity foundation model using pairwise meta-learning. *Nature Machine Intelligence* 6, 8 (2024), 962–974.

[17] Michael Fralick, Joshua Murray, and Muhammad Mamdani. 2021. Predicting emergency department volumes: A multicenter prospective study. *The American Journal of Emergency Medicine* 46 (2021), 695–697.

[18] Yoav Freund and Robert E. Schapire. 1995. A decision-theoretic generalization of on-line learning and an application to boosting. In *Proceedings of the 2nd European Conference Computational Learning Theory (EuroCOLT '95)*. Springer, 23–37.

[19] Gregory Gafni-Pappas and Mohammad Khan. 2023. Predicting daily emergency department visits using machine learning could increase accuracy. *The American Journal of Emergency Medicine* 65 (2023), 5–11.

[20] Mehak Gupta, Thao-Ly T. Phan, H. Timothy Bunnell, and Rahmatollah Beheshti. 2022. Obesity prediction with EHR data: A deep learning approach with interpretable elements. *ACM Transactions on Computing for Healthcare* 3, 3 (2022), 1–19.

[21] Fouzi Harrou, Abdelkader Dairi, Farid Kadri, and Ying Sun. 2022. Effective forecasting of key features in hospital emergency department: Hybrid deep learning-driven methods. *Machine Learning with Applications* 7 (2022), 100200.

[22] Andrew Fu, Wah Ho, Bryan Zhan Yuan Se To, Jin Ming Koh, and Kang Hao Cheong. 2019. Forecasting hospital emergency department patient volume using internet search data. *IEEE Access* 7 (2019), 93387–93395.

[23] Fan Hou, ZhiXiang Cheng, LuoYao Kang, and Wen Zheng. 2020. Prediction of gestational diabetes based on LightGBM. In *Proceedings of the 2020 Conference on Artificial Intelligence and Healthcare*, 161–165.

[24] Ya-Han Hu, Chun-Tien Tai, Solomon Chih-Cheng Chen, Hai-Wei Lee, and Sheng-Feng Sung. 2017. Predicting return visits to the emergency department for pediatric patients: Applying supervised learning techniques to the Taiwan National Health Insurance Research Database. *Computer Methods and Programs in Biomedicine* 144 (2017), 105–112.

[25] Shancheng Jiang, Qize Liu, and Beichen Ding. 2023. A systematic review of the modelling of patient arrivals in emergency departments. *Quantitative Imaging in Medicine and Surgery* 13, 3 (2023), 1957.

[26] Honghe Jin. 2022. Hyperparameter importance for machine learning algorithms. arXiv:2201.05132. Retrieved from https://arxiv.org/abs/2201.05132

[27] Wang-Chuan Juang, Sin-Jhih Huang, Fong-Dee Huang, Pei-Wen Cheng, and Shue-Ren Wann. 2017. Application of time series analysis in modelling and forecasting emergency department visits in a medical centre in Southern Taiwan. *BMJ Open* 7, 11 (2017), e018628.

[28] Hye Jin Kam, Jin Ok Sung, and Rae Woong Park. 2010. Prediction of daily patient numbers for a regional emergency medical center using time series analysis. *Healthcare Informatics Research* 16, 3 (2010), 158–165.

[29] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. LightGBM: A highly efficient gradient boosting decision tree. In *Proceedings of the Advances in Neural Information Processing Systems*, Vol. 30.

[30] Krishan L. Khatri and Lakshman S. Tamil. 2017. Early detection of peak demand days of chronic respiratory diseases emergency department visits using artificial neural networks. *IEEE Journal of Biomedical and Health Informatics* 22, 1 (2017), 285–290.

[31] Zella King, Joseph Farrington, Martin Utley, Enoch Kung, Samer Elkhodair, Steve Harris, Richard Sekula, Jonathan Gillham, Kezhi Li, and Sonya Crowe. 2022. Machine learning for real-time aggregated prediction of hospital admission for emergency patients. *NPJ Digital Medicine* 5, 1 (2022), 104.

[32] Ambeshwar Kumar, Ramachandran Manikandan, Utku Kose, Deepak Gupta, and Suresh C. Satapathy. 2021. Doctor's dilemma: Evaluating an explainable subtractive spatial lightweight convolutional neural network for brain tumor diagnosis. *ACM Transactions on Multimedia Computing, Communications, and Applications* 17, 3s (2021), 1–26.

[33] Seo-Young Lee, So-Ryoung Lee, Eue-Keun Choi, Kyung-Do Han, Seil Oh, and Gregory Y. H. Lip. 2022. Impact of socioeconomic status on emergency department visits in patients with atrial fibrillation: A nationwide population-based cohort study. *Journal of the American Heart Association* 11, 24 (2022), e027192.

[34] Beatriz López Ibáñez, Ferran Torrent-Fontbona, J. Roman, and José María Inoriza. 2021. Forecasting of emergency department attendances in a tourist region with an operational time horizon. Bohrium.

[35] Scott Lundberg. 2017. A unified approach to interpreting model predictions. arXiv:1705.07874. Retrieved from https://arxiv.org/abs/1705.07874

[36] Baoshan Ma, Fanyu Meng, Ge Yan, Haowen Yan, Bingjie Chai, and Fengju Song. 2020. Diagnostic classification of cancers using extreme gradient boosting algorithm and multi-omics data. *Computers in Biology and Medicine* 121 (2020), 103761.

[37] Peter McKenna, Samita M. Heslin, Peter Viccellio, William K. Mallon, Cristina Hernandez, and Eric J. Morley. 2019. Emergency department and hospital crowding: Causes, consequences, and cures. *Clinical and Experimental Emergency Medicine* 6, 3 (2019), 189.

[38] Gaurav Menghani. 2023. Efficient deep learning: A survey on making deep learning models smaller, faster, and better. *ACM Computing Surveys* 55, 12 (2023), 1–37.

[39] Chua Ming, Geraldine J. W. Lee, Yao Neng Teo, Yao Hao Teo, Xinyan Zhou, Elizabeth S. Y. Ho, Emma M. S. Toh, Marcus Eng Hock Ong, Benjamin Y. Q. Tan, and Andrew F. W. Ho. 2025. Deep learning modelling to forecast emergency department visits using calendar, meteorological, internet search data and stock market price. *Computer Methods and Programs in Biomedicine* 267 (2025), 108808.

[40] Claire Morley, Maria Unwin, Gregory M. Peterson, Jim Stankovich, and Leigh Kinsman. 2018. Emergency department crowding: A systematic review of causes, consequences and solutions. *PloS One* 13, 8 (2018), e0203316.

[41] Sajid Nazir, Diane M. Dickson, and Muhammad Usman Akram. 2023. Survey of explainable artificial intelligence techniques for biomedical imaging with deep neural networks. *Computers in Biology and Medicine* 156 (2023), 106668.

[42] Mehdi Neshat, Nikhil Jha, Michael Phipps, Chris A. Browne, and Walter P. Abhayaratna. 2024. Predicting frequent emergency department visitors using adaptive ensemble learning model. *medRxiv* (2024).

[43] Mehdi Neshat, Michael Phipps, Nikhil Jha, Danial Khojasteh, Michael Tong, and Amir Gandomi. 2024. Effective predictive modeling for emergency department visits and evaluating exogenous variables impact: Using explainable meta-learning gradient boosting. arXiv:2411.11275. Retrieved from https://arxiv.org/abs/2411.11275

[44] Arthur Novaes de Amorim, Rob Deardon, and Vineet Saini. 2021. A stacked ensemble method for forecasting influenza-like illness visit volumes at emergency departments. *PloS One* 16, 3 (2021), e0241725.

[45] Australian Bureau Of Meteorology. 2024. Australian Capital Territory Weather and Warnings. Retrieved July 16, 2024 from https://reg.bom.gov.au/act/?ref=hdr

[46] Karol R. Opara and Jarosław Arabas. 2019. Differential evolution: A survey of theoretical analyses. *Swarm and Evolutionary Computation* 44 (2019), 546–558.

[47] Torben Ostendorf, Michael Bernhard, Thomas Hartwig, Markus Voigt, Thomas Keller, Michael Stumvoll, and André Gries. 2020. Association between rapid weather changes and incidence of chiefly cardiovascular complaints in the emergency department. *The American Journal of Emergency Medicine* 38, 8 (2020), 1604–1610.

[48] César Peláez-Rodríguez, Jorge Pérez-Aracil, Dušan Fister, Ricardo Torres-López, and Sancho Salcedo-Sanz. 2024. Bike sharing and cable car demand forecasting using machine learning and deep learning multivariate time series approaches. *Expert Systems with Applications* 238 (2024), 122264.

[49] César Peláez-Rodríguez, Ricardo Torres-López, Jorge Pérez-Aracil, Nieves López-Laguna, S. Sánchez-Rodríguez, and Sancho Salcedo-Sanz. 2024. An explainable machine learning approach for hospital emergency department visits forecasting using continuous training and multi-model regression. *Computer Methods and Programs in Biomedicine* 245 (2024), 108033.

[50] Junfeng Peng, Chuan Chen, Mi Zhou, Xiaohua Xie, Yuqi Zhou, and Ching-Hsing Luo. 2020. Peak outpatient and emergency department visit forecasting for patients with chronic respiratory diseases using machine learning methods: Retrospective cohort study. *JMIR Medical Informatics* 8, 3 (2020), e13075.

[51] Spyridon Petsis, Areti Karamanou, Evangelos Kalampokis, and Konstantinos Tarabanis. 2022. Forecasting and explaining emergency department visits in a public hospital. *Journal of Intelligent Information Systems* 59, 2 (2022), 479–500.

[52] Oleg S. Pianykh, Steven Guitron, Darren Parke, Chengzhao Zhang, Pari Pandharipande, James Brink, and Daniel Rosenthal. 2020. Improving healthcare operations management with machine learning. *Nature Machine Intelligence* 2, 5 (2020), 266–273.

[53] Liudmila Prokhorenkova, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, and Andrey Gulin. 2018. CatBoost: Unbiased boosting with categorical features. In *Proceedings of the Advances in Neural Information Processing Systems*, Vol. 31.

[54] Alex Pryce, Maria Unwin, Leigh Kinsman, and Damhnat McCann. 2021. Delayed flow is a risk to patient safety: A mixed method analysis of emergency department patient flow. *International Emergency Nursing* 54 (2021), 100956.

[55] Zahra Rahmatinejad, Toktam Dehghani, Benyamin Hoseini, Fatemeh Rahmatinejad, Aynaz Lotfata, Hamidreza Reihani, and Saeid Eslami. 2024. A comparative study of explainable ensemble learning and logistic regression for predicting in-hospital mortality in the emergency department. *Scientific Reports* 14, 1 (2024), 3406.

[56] Beatriz Remeseiro and Veronica Bolon-Canedo. 2019. A review of feature selection methods in medical applications. *Computers in Biology and Medicine* 112 (2019), 103375.

[57] Antonio Jesús Rivera, J. Cobo Muñoz, M. D. Perez-Goody, B. Sáenz de San Pedro, Francisco Charte, D. Elizondo, C. Rodríguez, M. L. Abolafia, A. Perea, and María José del Jesus. 2023. XAIRE: An ensemble-based methodology for determining the relative importance of variables in regression tasks. Application to a hospital emergency department. *Artificial Intelligence in Medicine* 137 (2023), 102494.

[58] A. Saranya and R. Subhashini. 2023. A systematic review of explainable artificial intelligence models and applications: Recent developments and future trends. *Decision Analytics Journal* 7 (2023), 100230.

[59] Scikit-learn. 2024. Recursive Feature Elimination (RFE). Retrieved October 16, 2024 from https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.RFE.html

[60] Abhibhav Sharma and Buddha Singh. 2020. AE-LGBM: Sequence-based novel approach to detect interacting protein pairs via ensemble of autoencoder and LightGBM. *Computers in Biology and Medicine* 125 (2020), 103964.

[61] Rainer Storn and Kenneth Price. 1997. Differential evolution—A simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization* 11 (1997), 341–359.

[62] Vidya K. Sudarshan, Mikkel Brabrand, Troels Martin Range, and Uffe Kock Wiil. 2021. Performance evaluation of emergency department patient arrivals forecasting models by including meteorological and calendar information: A comparative study. *Computers in Biology and Medicine* 135 (2021), 104541.

[63] Yan Sun, Bee Hoon Heng, Yian Tay Seow, and Eillyne Seow. 2009. Forecasting daily attendances at an emergency department to aid resource planning. *BMC Emergency Medicine* 9 (2009), 1–9.

[64] Teo Susnjak and Paula Maddigan. 2023. Forecasting patient flows with pandemic induced concept drift using explainable machine learning. *EPJ Data Science* 12, 1 (2023), 11.

[65] M. Tanveer, T. Goel, R. Sharma, A. K. Malik, I. Beheshti, J. Del Ser, P. N. Suganthan, and C. T. Lin. 2024. Ensemble deep learning for Alzheimer's disease characterization and estimation. *Nature Mental Health* 2, 6 (2024), 665–667.

[66] Ali Cankut Tatlıparmak, Suleyman Alpar, and Sarper Yilmaz. 2023. Factors influencing recurrent emergency department visits for mild acute respiratory tract infections caused by the influenza virus. *PeerJ* 11 (2023), e16198.

[67] Ricardo Torres-López, David Casillas-Pérez, Jorge Pérez-Aracil, Laura Cornejo-Bueno, Enrique Alexandre, and Sancho Salcedo-Sanz. 2022. Analysis of machine learning approaches' performance in prediction problems with human activity patterns. *Mathematics* 10, 13 (2022), 2187.

[68] Bas H. M. Van der Velden, Hugo J. Kuijf, Kenneth G. A. Gilhuijs, and Max A. Viergever. 2022. Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. *Medical Image Analysis* 79 (2022), 102470.

[69] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Proceedings of the Advances in Neural Information Processing Systems*, Vol. 30.

[70] Te-Hao Wang, Jing-Cheng Jheng, Yen-Ting Tseng, Li-Fu Chen, and Jui-Yuan Chung. 2021. National early warning score for predicting intensive care unit admission among elderly patients with influenza infections in the emergency department: An effective disposition tool during the influenza season. *BMJ Open* 11, 6 (2021), e044496.

[71] Niyaz Ahmad Wani, Ravinder Kumar, and Jatin Bedi. 2024. DeepXplainer: An interpretable deep learning based approach for lung cancer detection using explainable artificial intelligence. *Computer Methods and Programs in Biomedicine* 243 (2024), 107879.

[72] Kieran Woodward, Eiman Kanjo, and Athanasios Tsanas. 2024. Combining deep learning with signal-image encoding for multi-modal mental wellbeing classification. *ACM Transactions on Computing for Healthcare* 5, 1 (2024), 1–23.

[73] Cao Xiao, Edward Choi, and Jimeng Sun. 2018. Opportunities and challenges in developing deep learning models using electronic health records data: A systematic review. *Journal of the American Medical Informatics Association* 25, 10 (2018), 1419–1428.

[74] Mai Xu, Tse-Chiu Wong, and Kwai-Sang Chin. 2013. Modeling daily patient arrivals at emergency department and quantifying the relative importance of contributing variables using artificial neural network. *Decision Support Systems* 54, 3 (2013), 1488–1498.

[75] Malik Yousef, Segun Jung, Louise C. Showe, and Michael K. Showe. 2007. Recursive cluster elimination (RCE) for classification and feature selection from gene expression data. *BMC Bioinformatics* 8 (2007), 1–12.

[76] Milad Yousefi, Moslem Yousefi, Masood Fathi, and Flavio S. Fogliatto. 2020. Patient visit forecasting in an emergency department using a deep neural network approach. *Kybernetes* 49, 9 (2020), 2335–2348.

[77] Hyoungju Yun, Jinwook Choi, and Jeong Ho Park. 2021. Prediction of critical care outcome for adult patients presenting to emergency department using initial triage information: An XGBoost algorithm analysis. *JMIR Medical Informatics* 9, 9 (2021), e30770.

[78] Yan Zhang, Jie Zhang, Min Tao, Jian Shu, and Degang Zhu. 2022. Forecasting patient arrivals at emergency department using calendar and meteorological information. *Applied Intelligence (Dordrecht, Netherlands)* 52, 10 (2022), 11232–11243.

[79] Zhendong Zhang and Cheolkon Jung. 2020. GBDT-MO: Gradient-boosted decision trees for multiple outputs. *IEEE Transactions on Neural Networks and Learning Systems* 32, 7 (2020), 3156–3167.

[80] Xinxing Zhao, Joel Weijia Lai, Andrew Fu Wah Ho, Nan Liu, Marcus Eng Hock Ong, and Kang Hao Cheong. 2022. Predicting hospital emergency department visits with deep learning approaches. *Biocybernetics and Biomedical Engineering* 42, 3 (2022), 1051–1065.

[81] Lingling Zhou, Qin Zhu, Qian Chen, Ping Wang, and Hao Huang. 2025. Predicting hospital outpatient volume using XGBoost: A machine learning approach. *Scientific Reports* 15, 1 (2025), 1–13.

Received 15 November 2024; revised 4 June 2025; accepted 21 August 2025