University Library

A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Song, L;Aickelin, U;Fazio, TN;Sharma, A;Kouhounestani, M;Plumb, S;Putland, MJ

Title:
Developing interpretable machine learning models to predict length of stay and disposition decision for adult patients in emergency departments

Date:
2025-06-26

Citation:
Song, L., Aickelin, U., Fazio, T. N., Sharma, A., Kouhounestani, M., Plumb, S. & Putland, M. J. (2025). Developing interpretable machine learning models to predict length of stay and disposition decision for adult patients in emergency departments. BMJ Health and Care Informatics, 32 (1), https://doi.org/10.1136/bmjhci-2024-101152.

Persistent Link:
https://hdl.handle.net/11343/357157

License:
CC BY-NC

BMJ Health & Care Informatics

Developing interpretable machine learning models to predict length of stay and disposition decision for adult patients in emergency departments

Long Song ,[1] Uwe Aickelin ,[1] Timothy N Fazio ,[2,3,4,5] Abhishek Sharma ,[4] Mojgan Kouhounestani ,[1] Samantha Plumb,[6] Mark John Putland [7,8]

To cite: Song L, Aickelin U, Fazio TN, *et al*. Developing interpretable machine learning models to predict length of stay and disposition decision for adult patients in emergency departments. *BMJ Health Care Inform* 2025;**32**:e101152. doi:10.1136/bmjhci-2024-101152

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/bmjhci-2024-101152).

Received 30 May 2024
Accepted 06 May 2025

Check for updates

© Author(s) (or their employer(s)) 2025. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ Group.

For numbered affiliations see end of article.

Correspondence to
Dr Mark John Putland;
Mark.Putland@mh.org.au

ABSTRACT

Objective Machine learning (ML) models have emerged as tools to predict length of stay (LOS) and disposition decision (DD) in emergency departments (EDs) to combat overcrowding. However, site-specific ML models are not transferable to different sites. Our objective was to develop interpretable ML models to predict LOS and DD at specific time points, all while establishing a transparent data analysis framework. This framework was designed to be easily adapted by other institutions for the development of their own ML models.

Methods We analysed data from 297 392 ED visits of patients aged 18 and above at a quaternary hospital between 30 June 2019 and 31 December 2022. Eight ML algorithms were evaluated, and ultimately, twelve lasso models built from 21 features were trained to predict four outcomes of LOS and DD at three time points post-triage. Hold-out testing and cross-validation were conducted for these models.

Results The area under the curve values were 0.862/0.868/0.878 for binary LOS predictions at 10, 60 and 120-minute time points and 0.839/0.851/0.863 for binary DD predictions. The accuracies were 60.2%/60.7%/61.9% for ternary LOS predictions and 61.5%/62.3%/63.4% for ternary DD predictions.

Conclusions Interpretable ML models demonstrated outstanding performances in predicting both LOS and DD. The transparent data analysis framework can be easily adapted by other institutions.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Early predictions of patients' length of stay (LOS) and disposition decision (DD) can help alleviate emergency department (ED) overcrowding.

⇒ Previous studies have developed models for predicting LOS or DD, but none have provided a transparent data analysis framework for other institutions to adapt.

WHAT THIS STUDY ADDS

⇒ Our study developed 12 interpretable models for predicting both binary and ternary LOS and DD at three time points.

⇒ The transparent data analysis framework provided in our study can be easily adapted by other institutions for developing their own machine learning (ML) models.

⇒ A novel approach was devised to effectively address the U-shaped correlation issue in developing generalised linear models.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Our models can improve ED efficiency by predicting decisions clinicians are likely to make or supporting them in making prompt decisions.

⇒ The transparent data analysis framework provides a blueprint that can be applied to other studies using ML algorithms for predictions in the ED.

INTRODUCTION

Efficient patient flow through emergency departments (EDs) is crucial for patient safety.[1 2] While it is well established that the primary barrier to efficient flow is the availability of hospital beds for emergency admissions,[3 4] improvements can also be made in the efficiency of decision-making and referral processes within the ED, as well as in the allocation of beds to ED patients.[5] To address this quality improvement objective, recent years have seen the development of two types of machine learning (ML) models using data extracted from electronic patient records (EPR). Some studies have focused on developing ML models for predicting length of stay (LOS),[6–11] aiming to identify key factors influencing LOS to aid in managing ED overcrowding. Additionally, real-time display of individual LOS predictions could potentially enhance patient satisfaction.[9] Other studies have constructed ML models for early predictions of disposition decision (DD),[5 12–19] aiming to improve patient flow and minimise waiting time.

BMJ Health & Care Informatics: first published as 10.1136/bmjhci-2024-101152 on 26 June 2025. Downloaded from https://informatics.bmj.com on 4 August 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

In healthcare environments, ensuring model interpretability is of paramount importance to establish trust and mitigate biases.[20] Furthermore, in ED settings, it is recognised that site-specific models are not transferable to different hospitals, and multihospital global models do not perform as well as site-specific models.[14 21] When a model trained at one hospital is applied at another, its performance significantly diminishes. There is no one-size-fits-all model that works for all ED settings. For a framework to be adaptable by other institutions for developing their own site-specific models, transparency throughout the entire data analysis process is crucial. Previous studies have not provided a transparent data analysis framework for building interpretable ML models to predict LOS or DD. Common gaps included the omission of essential details on categorical feature encoding,[5] handling of missing or outlier values,[13] feature selection[6] and individual prediction methodologies.[5 6] These gaps hindered other institutions from adapting these frameworks effectively. Therefore, this study aimed to achieve the following objectives:

- Develop interpretable ML models with strong performances for predicting both binary and ternary LOS and DD at 10, 60 and 120 min time points post-triage.
- Demonstrate a transparent data analysis framework.

METHODS

Study design and settings

The hospital, located in Melbourne, Victoria, Australia, serves as a major quaternary adult facility providing various medical, surgical, mental health and sub-acute services, including a major trauma service. A de-identified dataset was obtained from the EPR, covering ED visits from 30 June 2019 to 31 December 2022. Each ED visit was associated with demographic, clinical and administrative data, including flowsheet features recorded at different time points during the visit. Excluded from the study were cases involving patient death in the ED, patients leaving without being seen, those leaving after clinical advice or individuals under 18 years of age at presentation. Additionally, to reduce the potential disruptions caused by the provision of a COVID-19 testing service through the ED on the final models, ED visits solely for COVID-19 diagnosis were also excluded. All data analysis was conducted using Python on the University of Melbourne's high-performance computing system, Spartan.

Outcome definitions and prediction time points

Two primary outcomes were defined, each accompanied by a corresponding secondary outcome. The first primary outcome determined whether the LOS exceeded 4 hours, covering both the ED and subsequent hospital stay. The associated secondary outcome categorised LOS as: (1) within 4 hours; (2) between 4 and 24 hours; (3) exceeding 24 hours. The other primary outcome determined the DD for patients, whether they were discharged or required further hospital care. The related secondary outcome indicated if the patient was: (1) discharged; (2) admitted to short stay; (3) admitted to an inpatient bed or transferred out. Predictions were made at three time points: 10, 60 and 120 min after the end of triage, aligning with key stages of patient assessment: completion of triage (10 min), consultation with a doctor (60 min) and further observation (120 min). In total, 12 models were needed to predict the four outcomes at the three time points.

Data pre-processing

The raw data was split into modelling and test data, each undergoing pre-processing and processing separately (figure 1A). Given the retrospective nature of the study, we ensured that only data available before the prediction time points were used, preventing the creation of predictive models with artificially inflated performances. For more details on data pre-processing, see online supplemental appendix A.

Subsequently, three separate datasets were generated for the 10, 60 and 120 min time points, each forming the foundation for building four unique models. For simplicity, we primarily focused on the model for predicting binary LOS at the 120-minute time point. This specific example served to illustrate the consistent approach applied across all 12 models.

Data processing

In the modelling data, features were divided into two main types: categorical and numeric. To illustrate the processing of categorical features, consider the feature 'Intravenous catheter site'. Univariate association analysis revealed that the ED visits with missing 'Intravenous catheter site' were less likely to have LOS>4 hours compared with those with recorded values (table 1). Consequently, missing values for 'Intravenous catheter site' were treated as a distinct category. Because one-hot encoding would create 13 new features, it was considered undesirable for this study. Instead, each value of 'Intravenous catheter site' was encoded with the logit value of its probability associated with the binary LOS outcome. This target encoding approach[22] effectively replaced a categorical feature with a single numerical feature, thereby simplifying data processing.

Numeric features in the modelling data underwent two distinct processing methodologies. Features with 25 or fewer unique values were transformed into categorical features and treated accordingly. For other numeric features, consider 'Systolic blood pressure' as an example. Univariate association analysis revealed a U-shape correlation: patients with very low or very high systolic blood pressures were more likely to have LOS more than 4 hours (figure 2). A novel approach, combined with target encoding, was devised to address the issues of U-shaped correlations, missing values and outliers. All 'Systolic blood pressure' values were segmented into 25 intervals with nearly equal membership and converted to the logit value of the associated probability of its interval with the binary LOS (see online supplemental appendix B). Acknowledging weaker correlation, missing values were

BMJ Health & Care Informatics: first published as 10.1136/bmjhci-2024-101152 on 26 June 2025. Downloaded from https://informatics.bmj.com on 4 August 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

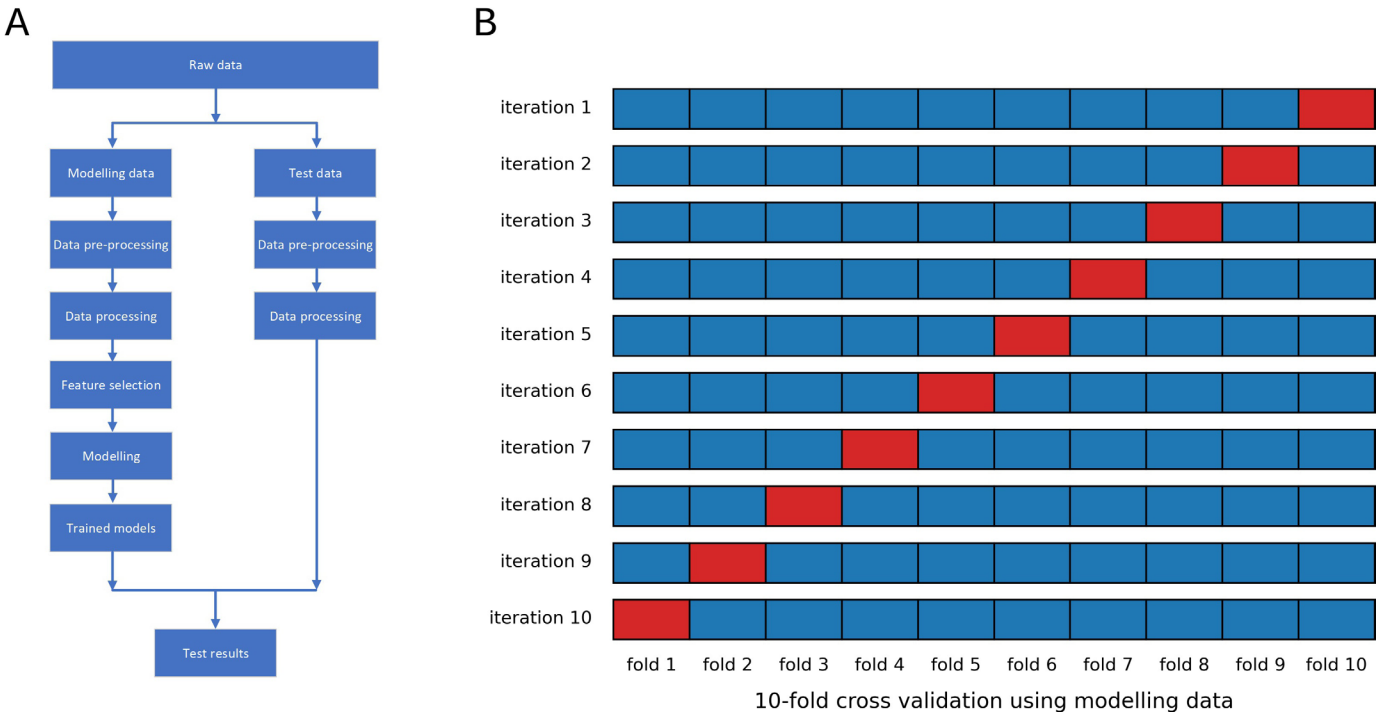

Figure 1 The diagram of data analysis. (A) Modelling and testing; (B) 10-fold cross validation.

treated as a separate interval, rather than being replaced with the mean or median value as in other studies.[5 12 23] Furthermore, outliers were included in the first and last intervals.

Subsequently, standardisation was applied to the modelling data, with each feature normalised by subtracting its mean and dividing by its SD.

Feature selection

To develop an ML model suitable for clinical settings, it is vital to limit the number of features and address multicollinearity—a phenomenon where correlated features within a regression can pose challenges. In this study, feature selection was executed by building a lasso model[24] on modelling data. This approach addressed

Table 1 The univariate association between the categorical feature 'Intravenous catheter site' and the binary LOS outcome in the modelling data for the 120-minute time point

Value	Count	Percentage	Count of LOS>4 hours	Probability of LOS>4 hours	Logit(prob)
Missing	94 884	54.8127	45 934	0.48	−0.06
Arm	82	0.0474	71	0.87	1.86
Cubital fossa	49 877	28.8130	44 434	0.89	2.1
Hand	9496	5.4857	8664	0.91	2.34
Wrist	2807	1.6215	2616	0.93	2.62
Lower leg	16	0.0092	15	0.94	2.71
Forearm	11 914	6.8825	11 229	0.94	2.8
Upper arm	1766	1.0202	1669	0.95	2.85
Dorsal arch	2138	1.2351	2044	0.96	3.08
Ankle	25	0.0144	24	0.96	3.18
Foot	96	0.0555	95	0.99	4.55
Neck	3	0.0017	3	1	4.6
Scalp	1	0.0006	1	1	4.6
Upper leg	1	0.0006	1	1	4.6

Set logit(prob) as logit(0.99) if prob≥0.99, and logit(0.01) if prob≤0.01.
LOS, length of stay.

BMJ Health & Care Informatics: first published as 10.1136/bmjhci-2024-101152 on 26 June 2025. Downloaded from https://informatics.bmj.com on 4 August 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

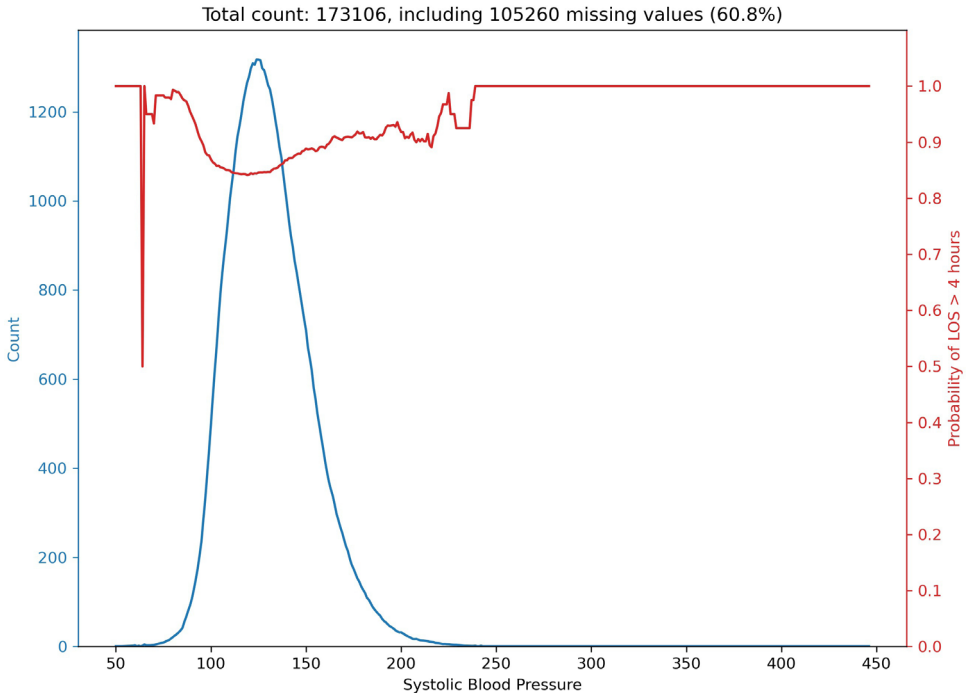

Figure 2 The distribution of feature 'Systolic blood pressure' and its associated probability with binary length of stay (LOS) outcome in the modelling data for the 120-minute time point. Both the count and probability lines were smoothed.

multicollinearity by automatically selecting one of the correlated features and discarding the others. To maintain simplicity, the relevant binary outcome was always used for feature selection. Although lasso-based feature selection reduced features significantly, it still left too many for practical use. After testing various feature sets, we selected the top 21 features from the coefficient table for further model development, balancing model performance with simplicity.

Modelling and validation

Eight predictive models were trained using the modelling data: logistic, ridge, lasso, elastic net, decision tree, random forest, xgboost and ensemble. The ensemble model combined logistic, decision tree and xgboost models. Subsequently, the test data underwent dedicated pre-processing and processing (figure 1A). Derived from the modelling data, the feature encoding and standardisation rules, along with the selected features, were applied to the test data. If a categorical feature value in the test data was not present in the modelling data, it was substituted with the most frequent value of the feature from the modelling data. Each of the eight trained models was then applied to the processed test data, providing predicted probabilities for each ED visit. For binary outcomes, the area under the curve (AUC) values were calculated for model comparison. For ternary outcomes, the models were multi-class classifiers that predicted probabilities for each of the three possible outcome classes. The predicted class was determined by selecting the one with the highest probability, and accuracy for model comparison was then calculated based on this predicted class. The 95% CIs were calculated by using 1000 bootstrapped samples.

Additionally, 10-fold cross-validation was conducted using only the modelling data (figure 1B). The cross-validation results were used to identify the optimal cut-off value corresponding to the highest F1 score for binary outcome predictions.

RESULTS

Study participants and pre-processed datasets

The raw data, extracted from the EPR, consisted of 297 392 ED visits involving 181 407 unique patients and encompassed 170 features. A subset of 196 714 ED visits from randomly selected 150 000 patients was designated as the modelling data. The remaining 100 678 ED visits, involving 61 407 patients, were held out as the test data.

After pre-processing the modelling data, three separate datasets were created for the 10, 60 and 120-minute time points post-triage. These pre-processed datasets comprised 173 193, 173 108 and 173 106 ED visits, with 41, 63 or 68 features, respectively. Across all datasets, approximately 32.5% of visits reported LOS≤4 hours, 34.2% reported LOS between 4 and 24 hours and 33.3% reported LOS>24 hours. DD was distributed as follows: 43.2% 'Discharged', 25.9% 'Short stay' and 31.0% 'Admitted or transferred'.

Processed data and selected features

In the pre-processed modelling data for the 120-minute time point, 7 111 591 cells had observed values, while 4 659 617 were missing, resulting in a 39.6% missingness rate (see online supplemental appendix C). Subsequently, encoding tables were generated, such as table 1 (also see online supplemental appendix B), along with

BMJ Health & Care Informatics: first published as 10.1136/bmjhci-2024-101152 on 26 June 2025. Downloaded from https://informatics.bmj.com on 4 August 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

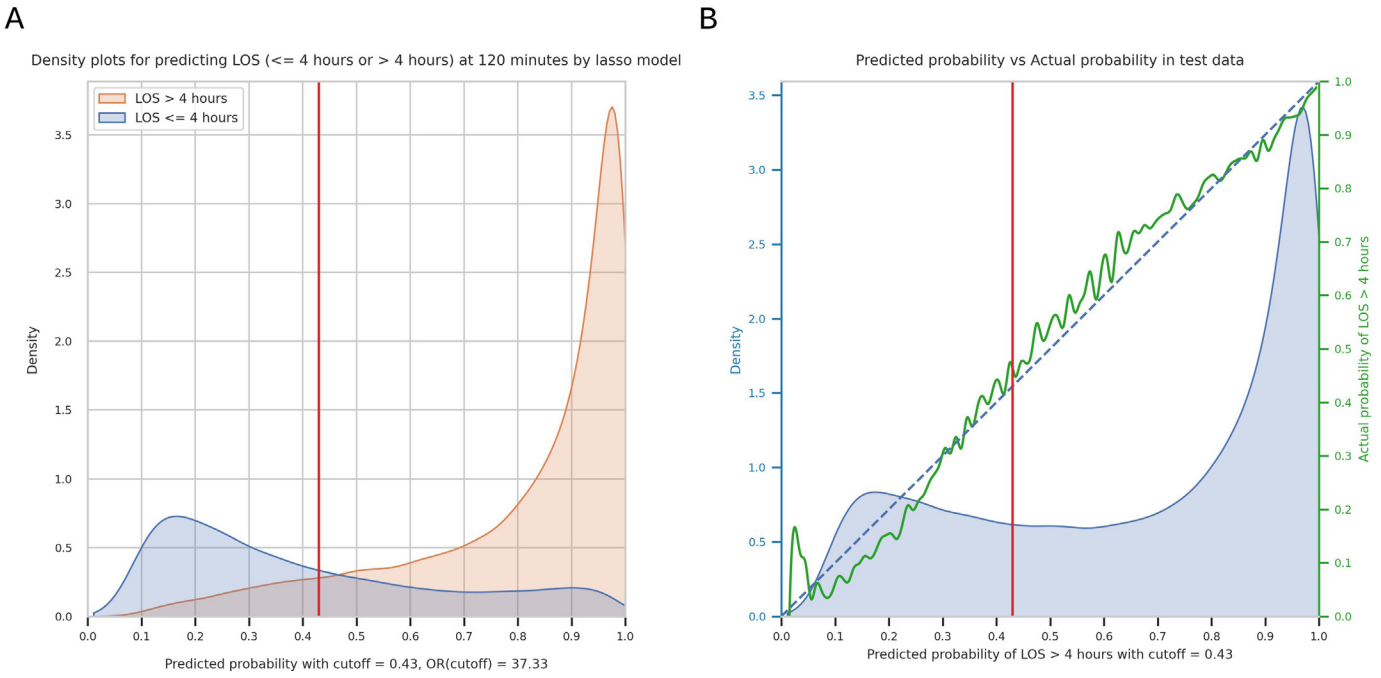

Figure 3 Predicted probability with test data for predicting binary length of stay (LOS) at 120 min by the lasso model. (A) Density plots for cases (LOS>4 hours) and controls (LOS≤4 hours); (B) density plot and line curve between actual probability and predicted probability.

data standardisation tables (see online supplemental appendix D). Moreover, coefficient tables were generated by building lasso models for feature selection (see online supplemental appendix E). For predicting LOS at the 120-minute time point, 21 out of 68 features were initially excluded as their coefficients were reduced to zero. Among the remaining 47 features, a total of 21 were selected to build the final models.

Model evaluation

We evaluated eight predictive models through hold-out testing and cross-validation. Given the minor differences in performances and significant advantages in interpretability, we chose the interpretable lasso for the final 12 models (see online supplemental appendix F). For more details on model evaluation, see online supplemental appendix A.

Interestingly, the AUC or accuracy exhibited a slight but consistent improvement as the time point increased, suggesting the potential for accurate predictions as early as 10 min post-triage. From the lasso model predicting binary LOS at the 120-minute time point, with cut-off 0.429, testing results revealed an AUC of 0.878 (95% CI 0.875 to 0.88), sensitivity of 91.4% (95% CI 91.2% to 91.7%), precision of 83.6% (95% CI 83.3% to 83.9%). Notably, the OR between the right and left sides of this cut-off was remarkably high at 37.33, indicating the lasso model's ability to make highly accurate predictions (figure 3A). Importantly, a substantial portion of predicted probabilities closely aligned with the actual probabilities (figure 3B).

Model application and feature importance

In table 2, we demonstrated the application of predicting a patient's binary LOS at the 120-minute time point post-tirage using the lasso model. The model used 21 features, and the process unfolded as follows:

- Raw values for 11 features were collected for the ED visit, with another five features having missing values.
- Raw values underwent pre-processing, with five novel features created (see online supplemental appendix A).
- During the data processing step, all pre-processed values were converted into numeric form, following the predefined data encoding and standardisation rules.
- Leveraging the calculated coefficient values, we computed the final predicted probability of 'LOS>4 hours', which in this case amounted to 0.973, using the lasso formula:

$$p = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots)}}$$

- Finally, given the predefined cut-off value of 0.429, the LOS for this patient was predicted as 'LOS>4 hours'. In reality, this patient's LOS extended beyond 40 days in the retrospective dataset.

The coefficient values indicate the importance of the features in the model. For instance, in the 120-minute model predicting binary LOS, the feature 'Order count' emerged as the most influential, with the highest coefficient of 0.614. This implies that for every one-unit increase in the processed value of 'Order count', the odds of 'LOS>4 hours' increase by approximately 85% (calculated as $e^{0.614}-1$). Consider two patients, A and B, with identical

BMJ Health & Care Informatics: first published as 10.1136/bmjhci-2024-101152 on 26 June 2025. Downloaded from https://informatics.bmj.com on 4 August 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Table 2 An example of predicting a patient's binary LOS at the 120-minute time point using the lasso model

Feature	Coefficient	Raw value	Pre-processed value	Processed value	Contribution
Intercept	1.36			1	1.36
Intravenous catheter site	0.549	Cubital fossa	Cubital fossa	0.916	0.502
Intravenous catheter size (gauge)	0.398	20	20	0.938	0.373
Arrival transport mode	0.332	Road ambulance service	Road ambulance service	1.089	0.362
Problem count	0.127		20	2.367	0.3
Order count	0.614		48	0.482	0.296
Triage target time compliance	0.179	0	No	1.486	0.266
Orientation	0.129	Posterior	Posterior	1.998	0.258
Patients waiting	0.158		10	1.606	0.254
Intravenous catheter site assessment	−0.46			−0.479	0.221
Postcode	0.255	3046	3046	0.315	0.08
Age	0.358	61	61	0.211	0.075
At risk of, or has, postural hypotension, syncope or dizziness	−0.024			−0.651	0.016
Triage complaint	0.191	Fever/infection	Fever/infection	−0.068	−0.013
Referred by	0.149	Self, family, friends	Self, family, friends	−0.126	−0.019
Oxygen delivery at triage	0.152			−0.217	−0.033
Glasgow coma score	0.161	15.0_level	15.0_level	−0.215	−0.035
Mobility	0.144			−0.655	−0.094
Arrival hour	0.19		15	−0.544	−0.103
Heart rate	0.166			−0.799	−0.133
Triage category	0.163	4—semi urgent	4—semi urgent	−0.923	−0.151
Average waiting time	0.315		28.5	−0.618	−0.194
Sum					**3.588**
Predicted probability					**0.973**

In the raw data, five features—'Intravenous catheter site assessment', 'At risk of, or has, postural hypotension, syncope or dizziness', 'Oxygen delivery at triage', 'Mobility' and 'Heart rate'—had missing values. During data pre-processing, five novel features—'Problem count', 'Order count', 'Patients waiting', 'Arrival hour' and 'Average waiting time'—were created based on existing data. All numbers in the table were rounded.
LOS, length of stay.

feature values except for 'Order count'. Patient A has 0 orders, while patient B has 48 orders. The processed values for 'Order count' are −1.081 and 0.482, respectively. The odds of 'LOS>4 hours' for patient B increase by approximately 161% (calculated as $e^{0.614\times(0.482+1.081)}-1$) compared with patient A. Similarly, 'Age' possesses the fifth largest coefficient of 0.358 (in terms of absolute value). Suppose all other features remain the same; moving from 51 years old to 61 years old, the odds increase by 12%, and from 61 years old to 71 years old, the odds increase by 35%.

DISCUSSION

The models developed in our study are intended to complement the expertise of experienced clinicians, enhancing efficiency by predicting decisions which clinicians are likely to make or supporting them in making decisions more promptly. Therefore, interpretability is crucial to establish trust and to mitigate potential biases in the models.[20] Many previous studies built black box models using ML algorithms like gradient boost, random forest and deep learning,[8 10 12 14–16] lacking effective interpretability. In contrast, our models prioritised interpretability, enabling clinicians to cross-reference predictions against their own judgments, particularly when there are disparities, thereby enhancing clinical decision-making.

Our study emphasised meticulous data pre-processing and processing to construct interpretable models with robust performance. Unlike most existing studies,[5 6 11]

BMJ Health & Care Informatics: first published as 10.1136/bmjhci-2024-101152 on 26 June 2025. Downloaded from https://informatics.bmj.com on 4 August 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

we included all available features at the beginning. After feature selection, 21 features were used in building the models, rather than just a few. A key contribution was the implementation of target encoding and a method for handling missing categorical and numeric values, which differed from conventional methods that replace missing values with the most frequent values or the mean or median.[5 12 23] Moreover, while maintaining interpretability, our study is the first to effectively address the challenging U-shaped correlation issue, a factor overlooked by previous studies in developing generalised linear models in EDs.[5 11 13 17] Due to these rigorous data pre-processing and processing efforts, the evaluation of eight ML algorithms revealed minor performance differences between interpretable lasso models and other black-box models, allowing us to select lasso for the final 12 models.

Using a model trained on data from one site at new sites comes at the cost of reduced accuracy, and global models trained on aggregate data from all sites are less accurate at individual sites than site-specific models.[14 21] This phenomenon is frequently attributed to variations in operational practices, local patient populations and local care protocols among hospitals.[9 14] Even within the same hospital, evaluating a model with later data can result in reduced accuracy due to concept drift—changes in data and model requirements over time.[25] Previous studies lacked a transparent data analysis framework for developing ML models for LOS and DD predictions, with unclear procedures and critical omissions hindering the replication of these analyses using alternative datasets. Our study addresses these gaps by detailing the steps for data pre-processing, processing, modelling and validation. It represents the first to demonstrate a transparent data analysis framework for developing interpretable ML models with robust performance in predicting LOS and DD in EDs. This framework enables easy adaptation by other institutions using their own datasets, even if their features are different from ours. While it also makes it possible to address the concept drift through drift adaptation,[26] resolving this issue in detail is beyond the scope of this study.

The models for predicting LOS and DD in the ED identified several key variables, which align well with current clinical practice. Notably, the feature 'Average waiting time' ranked among the top 10 features for all three models predicting binary LOS but was not significant for any model predicting binary DD. This suggests that waiting time influences a patient's LOS but does not impact the DD made by clinicians. Additionally, models predicting binary LOS and DD at 120 min post-triage shared 8 out of 10 top features: 'Age', 'Arrival transport mode', 'Intravenous catheter site', 'Intravenous catheter site assessment', 'Intravenous catheter size (gauge)', 'Order count', 'Postcode' and 'Triage complaint'. This indicates a natural association between LOS and DD. In simple terms, after accounting for variations in waiting time, a patient whose care is completed within 4 hours might be best discharged directly from the ED, one who requires 4–24 hours of care might be well suited to a short stay unit and one needing over 24 hours of care may be best served on an inpatient ward.

Using predicted probabilities empowers clinicians to optimise resource allocation and enhance patient flow efficiency. For instance, consider a patient highlighted in table 2 with a predicted probability of exceeding a 4-hour LOS at 0.973, notably surpassing the threshold of 0.429. An EPR marker indicating a high probability of LOS exceeding 4 hours can effectively direct a clinician's attention to the right patient and relevant data through personalised interpretations, facilitating timely and instinctive DD. This approach is particularly valuable for communicating decisions to both patients and colleagues. Similarly, predicted probabilities of DD can provide valuable insights to aid decision-making processes. For more discussion, see online supplemental appendix A.

Limitations

Our study is subject to several limitations. First, while our data pre-processing and processing procedures were rigorous, they were not exhaustive. For example, some potentially useful features, such as 'ICD10Code', a hierarchical categorical feature with thousands of unique values that could capture disease types and severity, were not incorporated. Second, during data processing, extreme (invalid) outliers and real outliers are diminished in the first and last intervals. However, some real outliers may be powerful indicators of severe disease. Third, the performances of predictions for ternary LOS and DD may not be optimal. Fourth, our models make predictions at fixed time points (10, 60 and 120 min post-triage), which may limit their flexibility. Predictions at any time during the ED stay could potentially improve adaptability and precision. Finally, subgroup analyses related to age, ethnicity and specific diagnosis have not been conducted yet. These limitations present opportunities for further exploration in future research.

CONCLUSION

In conclusion, our study has developed 12 interpretable ML models for predicting binary and ternary LOS and DD at 10, 60 and 120-minute time points post-triage. Furthermore, we have demonstrated a transparent data analysis framework, designed to be readily adapted by other institutions seeking to develop their own interpretable ML models.

Author affiliations

[1]School of Computing and Information Systems, The University of Melbourne, Melbourne, Victoria, Australia
[2]EMR Team, The Royal Melbourne Hospital, Parkville, Victoria, Australia
[3]Department of Medicine, Melbourne Medical School, The University of Melbourne, Melbourne, Victoria, Australia
[4]Health Intelligence Unit, The Royal Melbourne Hospital, Parkville, Victoria, Australia
[5]Clinical Informatics Centre, The Royal Melbourne Hospital, Parkville, Victoria, Australia
[6]The Royal Melbourne Hospital, Parkville, Victoria, Australia

BMJ Health & Care Informatics: first published as 10.1136/bmjhci-2024-101152 on 26 June 2025. Downloaded from https://informatics.bmj.com on 4 August 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

[7]Department of Emergency Medicine, The Royal Melbourne Hospital, Parkville, Victoria, Australia
[8]Department of Critical Care, Melbourne Medical School, The University of Melbourne, Melbourne, Victoria, Australia

Acknowledgements The authors thank Lauren Grundy, reports analyst at the Royal Melbourne Hospital Health Intelligence Unit, for her support in preparing the raw data and her assistance throughout the study. This research was supported by The University of Melbourne's Research Computing Services and the Petascale Campus Initiative.

Contributors MJP conceived the study and obtained research funding. LS, UA, TNF, SP and MJP participated in study design. TNF and AS acquired the data. LS, UA and MK analysed the data, and all authors assisted with the interpretation of the data. LS and MJP drafted the manuscript, and all authors reviewed the results and approved the final version of the manuscript. MJP is the guarantor of the study.

Funding This project was supported by a Royal Melbourne Hospital Health Service Improvement Grant. The University of Melbourne provided research infrastructure support. The sponsors had no role in the study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval Ethics approval was granted by the Royal Melbourne Hospital Human Research Ethics Committee (HREC/86079/MH-2022).

Provenance and peer review Part of a Topic Collection; Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. The data that support the findings of this study are available on reasonable request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs
Long Song http://orcid.org/0000-0001-8494-4364
Uwe Aickelin http://orcid.org/0000-0002-2679-2275
Timothy N Fazio http://orcid.org/0000-0003-1700-2355
Abhishek Sharma http://orcid.org/0009-0006-7591-4757
Mojgan Kouhounestani http://orcid.org/0000-0001-7935-6410
Mark John Putland http://orcid.org/0000-0002-1994-252X

REFERENCES

1 Richardson DB. Increase in patient mortality at 10 days associated with emergency department overcrowding. *Med J Aust* 2006;184:213–6.
2 Jones S, Moulton C, Swift S, *et al*. Association between delays to patient admission from the emergency department and all-cause 30-day mortality. *Emerg Med J* 2022;39:168–73.
3 Forster AJ, Stiell I, Wells G, *et al*. The effect of hospital occupancy on emergency department length of stay and patient disposition. *Acad Emerg Med* 2003;10:127–33.
4 Paling S, Lambert J, Clouting J, *et al*. Waiting times in emergency departments: exploring the factors associated with longer patient waits for emergency care in England using routinely collected daily data. *Emerg Med J* 2020;37:781–6.
5 Barak-Corren Y, Israelit SH, Reis BY. Progressive prediction of hospitalisation in the emergency department: uncovering hidden patterns to improve patient flow. *Emerg Med J* 2017;34:308–14.
6 Gurazada SG, Gao SC, Burstein F, *et al*. Predicting Patient Length of Stay in Australian Emergency Departments Using Data Mining. *Sensors (Basel)* 2022;22:4968.
7 Khanna S, Boyle J, Good N, *et al*. New emergency department quality measure: from access block to National Emergency Access Target compliance. *Emerg Med Australas* 2013;25:565–72.
8 Gill SD, Lane SE, Sheridan M, *et al*. Why do "fast track" patients stay more than four hours in the emergency department? An investigation of factors that predict length of stay. *Emerg Med Australas* 2018;30:641–7.
9 Rahman MA, Honan B, Glanville T, *et al*. Using data mining to predict emergency department length of stay greater than 4 hours: Derivation and single-site validation of a decision tree algorithm. *Emerg Med Australas* 2020;32:416–21.
10 Chrusciel J, Girardon F, Roquette L, *et al*. The prediction of hospital length of stay using unstructured data. *BMC Med Inform Decis Mak* 2021;21:351.
11 Perdahl T, Axelsson S, Svensson P, *et al*. Patient and organizational characteristics predict a long length of stay in the emergency department - a Swedish cohort study. *Eur J Emerg Med* 2017;24:284–9.
12 Kishore K, Braitberg G, Holmes NE, *et al*. Early prediction of hospital admission of emergency department patients. *Emerg Med Australas* 2023;35:572–88.
13 Dinh MM, Russell SB, Bein KJ, *et al*. The Sydney Triage to Admission Risk Tool (START) to predict Emergency Department Disposition: A derivation and internal validation study using retrospective state-wide data from New South Wales, Australia. *BMC Emerg Med* 2016;16:46.
14 Barak-Corren Y, Chaudhari P, Perniciaro J, *et al*. Prediction across healthcare settings: a case study in predicting emergency department disposition. *NPJ Digit Med* 2021;4:169.
15 Barak-Corren Y, Agarwal I, Michelson KA, *et al*. Prediction of patient disposition: comparison of computer and human approaches and a proposed synthesis. *J Am Med Inform Assoc* 2021;28:1736–45.
16 Tahayori B, Chini-Foroush N, Akhlaghi H. Advanced natural language processing technique to predict patient disposition based on emergency triage notes. *Emerg Medicine Australasia* 2021;33:480–4.
17 Cameron A, Rodgers K, Ireland A, *et al*. A simple tool to predict admission at the time of triage. *Emerg Med J* 2015;32:174–9.
18 Sun Y, Heng BH, Tay SY, *et al*. Predicting hospital admissions at emergency department triage using routine administrative data. *Acad Emerg Med* 2011;18:844–50.
19 Peck JS, Benneyan JC, Nightingale DJ, *et al*. Predicting emergency department inpatient admissions to improve same-day patient flow. *Acad Emerg Med* 2012;19:E1045–54.
20 Rudin C. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nat Mach Intell* 2019;1:206–15.
21 Walker K, Jiarpakdee J, Loupis A, *et al*. Emergency medicine patient wait time multivariable prediction models: a multicentre derivation and validation study. *Emerg Med J* 2022;39:386–93.
22 Pargent F, Pfisterer F, Thomas J, *et al*. Regularized target encoding outperforms traditional methods in supervised machine learning with high cardinality features. *Comput Stat* 2022;37:2671–92.
23 Xie F, Zhou J, Lee JW, *et al*. Benchmarking emergency department prediction models with machine learning and public electronic health records. *Sci Data* 2022;9:658.
24 Tibshirani R. Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society Series B* 1996;58:267–88.
25 Akhlaghi H, Freeman S, Vari C, *et al*. Machine learning in clinical practice: Evaluation of an artificial intelligence tool after implementation. *Emerg Med Australas* 2024;36:118–24.
26 Lu J, Liu A, Dong F. Learning under Concept Drift: A Review. *IEEE Trans Knowl Data Eng* 2019;31:2346–63.

BMJ Health & Care Informatics: first published as 10.1136/bmjhci-2024-101152 on 26 June 2025. Downloaded from https://informatics.bmj.com on 4 August 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.