J Arch Mil Med. 2023 September; 11(3):e140442. https://doi.org/10.5812/jamm-140442.

Published online 2023 October 4. **Research Article**

Mortality Prediction in Emergency Department Using Machine Learning Models

Sina Moosavi Kashani [1] and Sanaz Zargar Balaye Jame [2, *]

[1]Faculty of Industrial and Systems, Tarbiat Modares University, Tehran, Iran
[2]Department of Management Sciences and Health Economics, Faculty of Medicine, AjA University of Medical Sciences, Tehran, Iran

[*]*Corresponding author*: Department of Management Sciences and Health Economics, Faculty of Medicine, AjA University of Medical Sciences, Tehran, Iran. Email: sanazzargar@gmail.com

Received 2023 August 30; **Accepted** 2023 September 02.

Abstract

Background: Diagnosing patient deterioration and preventing unexpected deaths in the emergency department is a complex task that relies on the expertise and comprehensive understanding of emergency physicians concerning extensive clinical data.
Objectives: Our study aimed to predict emergency department mortality and compare different models.
Methods: During a one-month period, demographic information and records were collected from 1,000 patients admitted to the emergency department of a selected hospital in Tehran. We rigorously followed The Cross Industry Standard Process for data mining and methodically progressed through its sequential steps. We employed Cat Boost and Random Forest models for prediction purposes. To prevent overfitting, Random Forest feature selection was employed. Expert judgment was utilized to eliminate features with an importance score below 0.0095. To achieve a more thorough and dependable assessment, we implemented a K-fold cross-validation method with a value of 5.
Results: The Cat Boost model outperformed Random Forest significantly, showcasing an impressive mean accuracy of 0.94 (standard deviation: 0.03). Ejection fraction, urea (body waste materials), and diabetes had the greatest impact on prediction.
Conclusions: This study sheds light on the exceptional accuracy and efficiency of machine learning in predicting emergency department mortality, surpassing the performance of traditional models. Implementing such models can result in significant improvements in early diagnosis and intervention. This, in turn, allows for optimal resource allocation in the emergency department, preventing the excessive consumption of resources and ultimately saving lives while enhancing patient outcomes.

Keywords: Data Mining, Emergency Department, Ensemble Models, Mortality, Prediction

1. Background

Healthcare has grown into one of the largest industries worldwide, and the Emergency Department (ED) stands out as a crucial department within these services that exhibits significant demands (1). The ED is fully prepared and equipped to deliver comprehensive emergency care to the community during emergencies and non-emergencies. Operating round the clock, 365 days a year, this department operates uniquely, involving multiple interactions and requiring intensive decision-making. These factors can result in interruptions and disruptions within this section (2).

Over the past few decades, overcrowding in hospital ED has become a widespread issue across the globe. The rise in patient numbers and the influx of patients requiring admission have exacerbated this problem. Substantial evidence suggests that overcrowding has detrimental consequences, such as prolonged wait times for critically ill individuals, reduced patient satisfaction, heightened mortality rates, and increased medical errors (3).

Efficiently distributing resources in the healthcare industry and enhancing societal health quality is paramount. However, given the constraints of limited resources, the high expenses associated with healthcare, and the sensitive and complex nature of the field, it has consistently remained contentious (4). Hence, to effectively allocate resources to patients, it is necessary first to diagnose the deterioration of their condition. Conversely, early identification and prevention of untimely deaths using extensive clinical data present a significant challenge for emergency physicians, requiring substantial expertise and precise intuition (5).

Artificial intelligence (AI) refers to the capacity of computer programs to perform tasks or reasoning processes typically associated with human intelligence. Its main focus is on making accurate decisions despite ambiguity, uncertainty, or the presence of large data sets. In the healthcare domain, where extensive amounts of data exist, machine learning (ML) algorithms are utilized for classification purposes ranging from clinical symptoms to imaging features. Machine learning is a methodology that leverages pattern recognition techniques. Within the clinical field, AI has found applications in diagnostics, therapeutics, and population health management. Notably, AI has significantly impacted areas such as cell immunotherapy, cell biology, biomarker discovery, regenerative medicine, tissue engineering, and radiology. The application of ML in healthcare encompasses drug detection and analysis, disease diagnosis, smart health records, remote health monitoring, assistive technologies, medical imaging diagnosis, crowdsourced data collection, and outbreak prediction, as well as clinical trials and research (6).

Several conventional methods are used in clinical settings to assess the condition and predict the mortality risk of intensive care patients. These methods include the Simplified Acute Physiology Score (SAPS II), Sequential Organ Failure Assessment (SOFA), and Acute Physiological Score (APS). They incorporate factors such as age, medical history, vital signs, and laboratory test results. These scoring systems help healthcare professionals determine the severity of a patient's illness and predict life-threatening events like sepsis, cardiac arrest, or respiratory arrest (7). Barboi et al. (8) indicated that ML models exhibit higher accuracy than traditional scoring models. Therefore, clinicians are encouraged to prioritize the selection of models that have undergone more rigorous validation.

Li et al. (5) demonstrated that ensemble models, specifically bagging and boosting, exhibit superior performance compared to single classifiers. By analyzing demographic and laboratory data from 1,114 ED patients, the researchers found that the gradient boosting machine (GBM) model stood out with an impressive accuracy rate of 93.6% in predicting patient mortality.

In a retrospective cohort study conducted by van Doorn et al. (9), the accuracy of predicting patient outcomes in the ED differs when utilizing only laboratory information compared to a combination of laboratory and clinical data. Specifically, the study employed the Extreme Gradient Boosting (XG Boost) model and found that when using solely laboratory information, the accuracy was 82%. However, by integrating both clinical and laboratory data, the accuracy increased to 84%. The study involved 1,344 ED patients.

Klug et al. (10) used variables including age, admission mode, chief complaint, five primary vital signs, and emergency severity index (ESI) to analyze ED patients. By implementing the XG Boost model, the study achieved an impressive accuracy rate of 92%. The ESI is a tool used in ED to assess the severity of a patient's condition and prioritize care accordingly (11).

2. Objectives

This study aimed to accurately predict patients' mortality within the ED while also conducting a comparative evaluation of different models. By achieving high forecasting accuracy, this study aimed to provide doctors and ED specialists with valuable insights to prioritize patients effectively regarding resource allocation.

3. Methods

The Cross Industry Standard Process for Data Mining (CRISP-DM) is a process model designed for data mining that can be applied across various industries. This model encompasses six sequential phases, executed iteratively from understanding the business requirements to the final deployment and implementation of the data mining solution (12).

To conduct our study, we gathered the electronic health records, medical data, and demographic information of 1,000 patients who were admitted to the ED of a hospital in Tehran. The data were retrospectively collected using the Hospital Information System unit during a one-month timeframe.

We initially removed patients with missing data from the study during the data preparation phase. Additionally, we employed the Interquartile Ranges (IQRs) to detect and eliminate outliers. As a result, 200 patients were excluded from the complete dataset. The IQR is a measure of statistical dispersion that quantifies the spread of a data set. It is defined as the difference between the third quartile (Q3) and the first quartile (Q1) in a data set (13). We employed a label encoder for the target column to represent binary categories, where class 0 signifies discharged, and class 1 signifies expired. After considering the research by Newaz et al. (14), which explored the model's accuracy with over-sampling and under-sampling, we concluded that over-sampling would be the most suitable approach for balancing the classes in the target column.

Feature selection is a crucial step in analyzing data as it involves selecting a concise group of pertinent features.

The RF classifier serves as a critical foundation for wrapper algorithms, effectively addressing all significant issues by offering a measure of variable importance (15). To prevent overfitting, RF feature selection was employed. Expert judgment was utilized to eliminate features with an importance score below 0.0095. Subsequently, the models were created using the remaining features. In the modeling phase, the decision was made to use ensemble models due to their relatively good accuracy.

Ensemble models combine multiple models that work together to make predictions. These models can be of the same type or different types, and by leveraging the strengths of each individual model, ensemble models can often outperform any single model. Ensemble models have become popular in various domains, including machine learning and data science because they can improve the overall performance and robustness of a prediction system. They reduce bias and variance, increase model generalization, and mitigate the risk of overfitting. By aggregating the predictions from multiple base models, ensemble models can capture a wider range of patterns and improve the accuracy of predictions (16). The commonly used ensemble techniques are bagging, boosting, and stacking (17):

- Bagging involves training multiple decision trees on various subsets of the same dataset and then averaging their predictions.

- Boosting, on the other hand, works by sequentially adding ensemble members that improve upon the predictions of prior models, ultimately resulting in a weighted average of all predictions.

- Stacking involves training multiple models of different types on the same data and utilizing another model to learn the most effective way to combine these predictions.

The RF algorithm is a widely known supervised ML technique used in both classification and regression problems. This algorithm leverages a collection of decision trees, each trained on different subsets of the dataset, and combines their predictions through averaging to enhance the overall predictive accuracy. This approach, known as bagging, has contributed to the algorithm's popularity. Notably, empirical studies have shown that the Random Forest (RF) classifier outperforms individual classifiers regarding classification rates. Furthermore, it demonstrates shorter training time than Decision Tree and SVM algorithms (18).

Cat Boost (CB) is a GB framework developed by Yandex, a Russian search engine company. It is specifically designed to work with categorical features in the dataset and provides superior performance compared to other traditional gradient-boosting models. Cat Boost can automatically handle categorical features without requiring explicit feature engineering or encoding, making it a convenient choice for working with datasets containing categorical variables. It uses a novel algorithm called "Ordered Boosting" that reduces the impact of the order of categorical features on model performance (19).

Some key features of CB include (20):

- Handling of categorical features
- Improved accuracy
- Fast training time
- Robustness to outliers

Hence, we employed RF and CB models in this study to predict mortality and assess their relative efficacy.

In the evaluation phase, accuracy, precision, recall, and the F1-score are essential criteria for evaluating classification problems. These metrics are calculated as follows (21):

$$Accuracy = \frac{T_P + T_N}{T_P + F_P + F_N + T_N}$$

$$Precision = \frac{T_P}{T_P + F_P}$$

$$Recall = \frac{T_P}{T_P + F_N}$$

$$F_1 - score = 2 \times \frac{(Precision \times recall)}{(Precision + recall)}$$

A true positive (T_P) occurs when both the actual and predicted classes of data points are labeled as 1. Conversely, a true negative (T_N) occurs when both the actual and predicted classes of data points are labeled as 0. On the other hand, a false positive (F_P) happens when the actual class of the data point is 0, but the predicted class is 1. Finally, a false negative (F_N) refers to the scenario where the true class of the data point is 1, but the predicted class is 0.

K-fold cross-validation is a popular technique used in ML to evaluate the performance of a model on a limited dataset. It helps estimate how well the trained model performs on unseen data. In K-fold cross-validation, the dataset is divided into k equal-sized subsets or folds. The model is then trained on k-1 folds and tested on the remaining fold. This process is repeated k times, each time using a different fold as the test set and the remaining folds as the training set. The model's performance is averaged over all k iterations to obtain a more reliable estimate (22). To ensure a more precise assessment, we employed 5-fold cross-validation.

The receiver operating characteristic (ROC) curve visually depicts how well a binary classifier system performs as its threshold for decision-making is adjusted. It is commonly used in data mining and ML to assess the classifier's performance. The area beneath this curve serves as a measure to evaluate the classifier, and a higher area indicates a better-performing model (23).

4. Results

After completing the data preparation phase, 800 patient records were ready for review and model building. The research findings revealed that 63.88% of these patients were men. Additionally, 18.36% of the overall data was observed to reflect cases of patient mortality.

To ensure a more thorough analysis, we decided to separate the numerical features from the binary features. We then analyzed their statistical characteristics separately for two specific scenarios - discharges and deaths. Employing a confidence level of 95%, we conducted t and chi-square tests. This led us to compile Table 1, which contains statistical information related to the numerical features, and Table 2, which shows statistical information related to the binary features.

After employing the RF algorithm to identify the variables with the greatest influence on the outcome variable, the analysis revealed that ejection fraction (EF), UREA, and diabetes mellitus (DM) possessed the highest impact. Among the pool of 46 research variables, we prioritized the first 22 based on expert opinion from the field (Figure 1).

After analyzing the impact of each variable, the performance of RF and CB models was evaluated using 5-fold cross-validation. The results, presented as mean (standard deviation), indicated that CB outperformed the other model in terms of performance (Table 3).

In summary, both models performed well in terms of accuracy, precision, recall, and F1 score. However, CB achieved a slightly higher recall rate (94% vs. 92%) and overall F1 score (94% vs. 93%) than the RF model.

The ROC curves with k-fold cross-validation offer several advantages. It allows for a fairer comparison of model performance, as cross-validation provides more accurate estimates. Additionally, it helps assess the robustness of the model by evaluating its performance across various data subsets, providing a comprehensive understanding of performance across different distributions. The ROC curve also allows for a trade-off analysis between T_P and F_P rates, aiding in generalization assessment for unseen data. Furthermore, calculating the standard deviation or confidence interval of the performance metric provides insight into the reliability and uncertainty associated with the model's predictions. Overall, displaying the ROC curve with k-fold cross-validation provides a more rigorous and comprehensive evaluation of the model's capabilities and limitations. Therefore, the ROC diagram for the CB model with 5-fold cross-validation is depicted in Figure 2.

5. Discussion

The primary objective of this study was to predict the likelihood of mortality among patients in the ED. To achieve this objective, ensemble models were employed, specifically chosen from the bagging models, i.e., the RF and CB models, in the boosting mode.

Based on the research findings, the CB model displayed better performance than the RF, albeit with a minor advantage. However, it is important to consider the unique dataset characteristics and project objectives when selecting between CB and RF classification. To identify the most suitable algorithm, it is advisable to conduct experiments and evaluate the performance of various algorithms on the provided dataset. Based on previous research, it has been consistently demonstrated that ML outperforms traditional scoring methods in terms of performance. Furthermore, recent studies have shown that the highest predictive accuracy achieved so far is 92%. However, upon reviewing the present study, it was observed that analyzing additional patient records and incorporating more variables can enhance the model's accuracy by up to 94%.

The study revealed significant variations in variables, including age, Total Leukocyte Count (TLC), platelet count, and urea levels, between the patients who expired and those who were discharged. Safaei et al. (24) conducted a study similar to ours, where they developed an extremely precise and effective CB model to anticipate mortality after patients were discharged from the ICU. They focused on data collected within the initial 24 hours of hospitalization. The outcomes of their research revealed a range of significant factors, such as age, heart rate, respiration rate, blood urea nitrogen, and creatinine level, which greatly impacted mortality prediction.

Furthermore, to enhance the patient's condition in the ED and ensure the effective allocation of resources, it is advised to perceive the admission and discharge of patients as a cohesive process. Utilizing simulation techniques can aid in refining this process and optimizing the distribution of resources. Hence, forthcoming research should concentrate on augmenting resource efficiency and determining the optimal allocation of resources, prioritizing patients in critical conditions.

Table 1. Numeric Features

Features	Description	Outcome		P-Value (0.05)	*t*-Test
		Discharged (Class 0)	Expired (Class 1)		
Age, y	Patient's age	64.56 ± 12.3	67.28 ± 13.1	0.017 [a]	-2.381
HB	Hemoglobin	11.91 ± 2.2	11.59 ± 2.2	0.127	1.526
TLC	Total leukocytes count	12.27 ± 5.8	17.42 ± 13.6	< 0.001 [a]	-4.481
Platelets	Thrombocytes	242.43 ± 102.6	208.12 ± 120.05	0.002 [a]	3.210
Glucose	Carbohydrate	179.75 ± 88.1	194.20 ± 108.6	0.087	-1.715
Urea	Body waste	57.03 ± 39.7	86.18 ± 53.5	< 0.001 [a]	-6.227
Creatinine	Creatinine	1.56 ± 1.3	1.89 ± 1.05	0.004 [a]	-2.859
BNP	B-type natriuretic peptide	888.32 ± 921.6	1370.18 ± 1203.6	< 0.001 [a]	-4.562
EF	Ejection fraction	37.89 ± 12.6	30.40 ± 8.3	< 0.001 [a]	8.799

[a] Statistical significance at a P-value of 0.05.

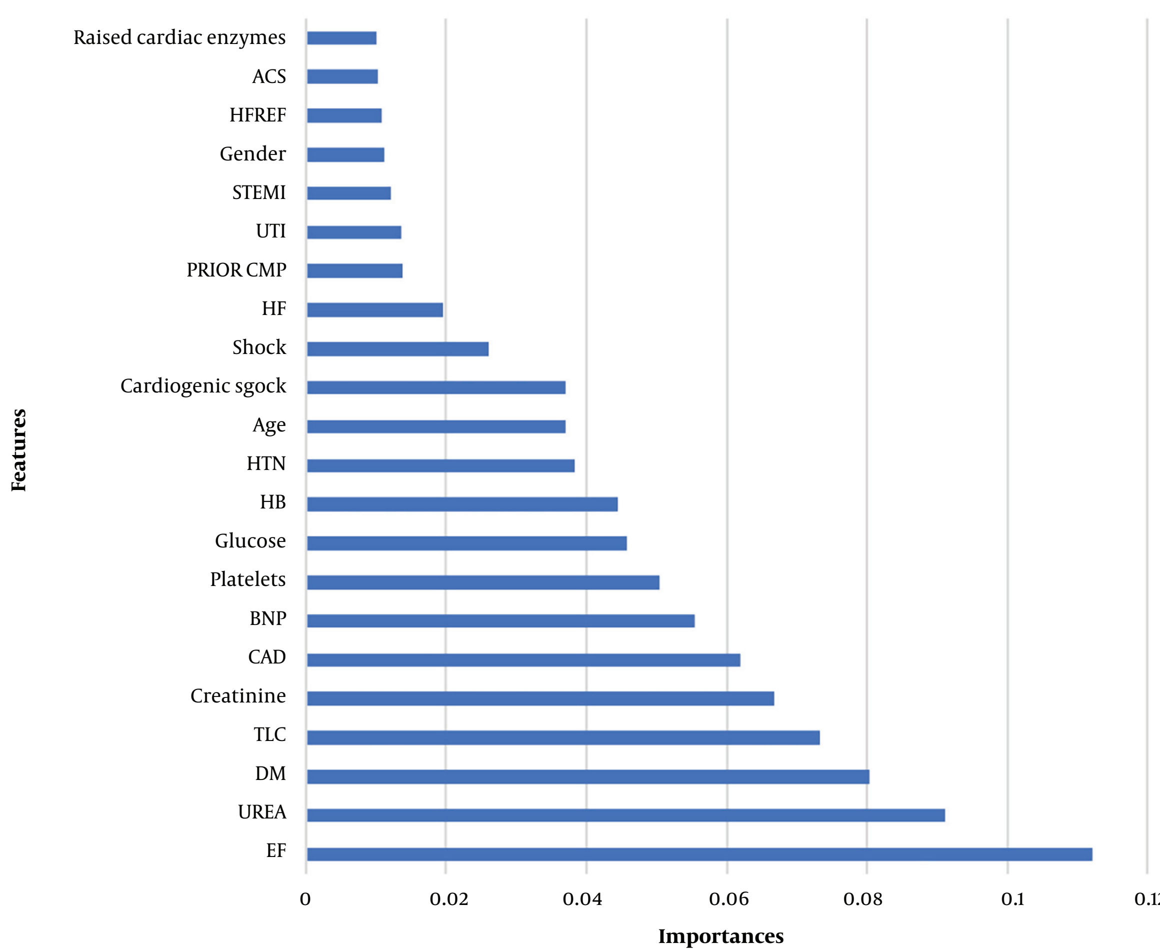

Figure 1. The importance of selected variables

Table 3. Evaluation of Models with 5-Fold Cross-Validation

Models	Accuracy	Precision	Recall	F1 Score
RF	0.93 ± 0.04	0.94 ± 0.01	0.92 ± 0.08	0.93 ± 0.04
CB	0.94 ± 0.03	0.94 ± 0.02	0.94 ± 0.06	0.94 ± 0.04

[a] Values are expressed as mean ± SD.

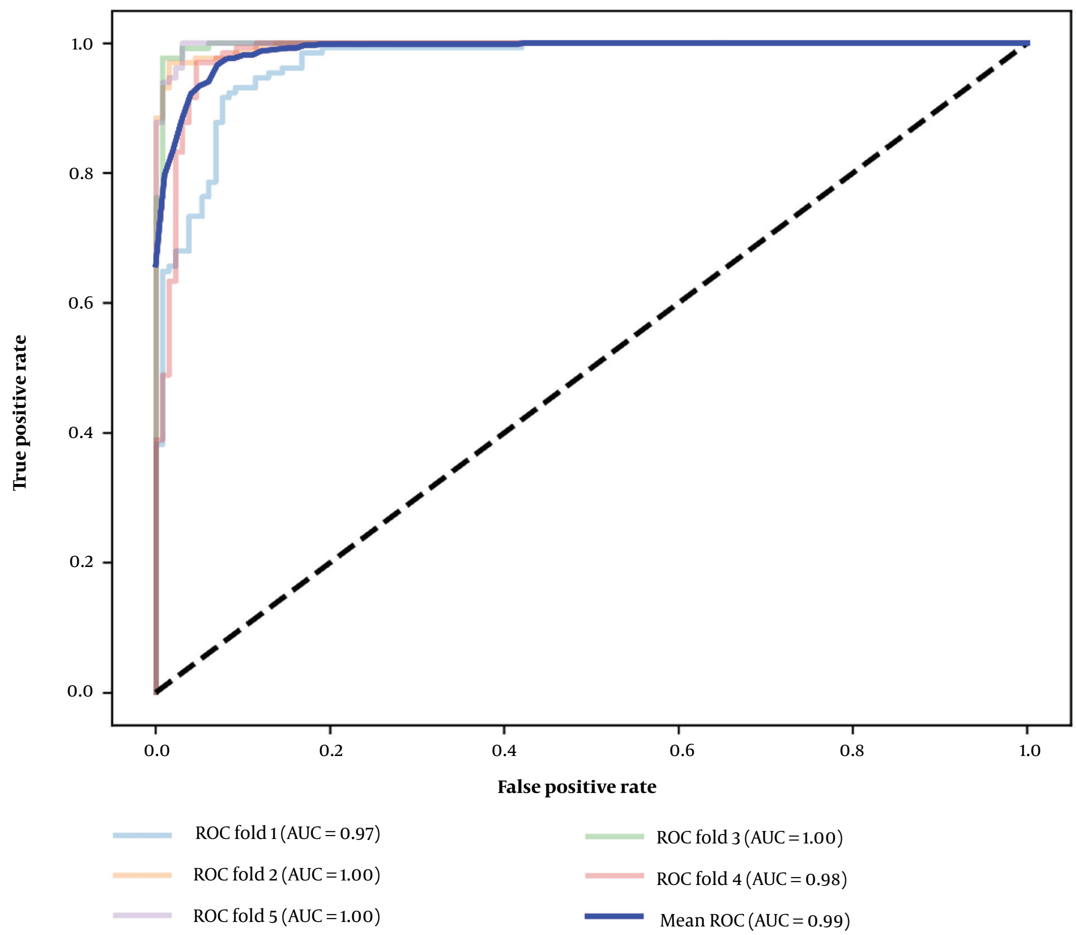

Figure 2. ROC curve for CB model

5.1. Conclusions

This study sheds light on the exceptional accuracy and efficiency of ML in predicting ED mortality, surpassing the performance of traditional models. Implementing such models can result in significant improvements in early diagnosis and intervention. This, in turn, allows for optimal resource allocation in the ED, preventing the excessive consumption of resources and ultimately saving lives while enhancing patient outcomes.

Footnotes

Authors' Contribution: SMK played a role in designing and determining the eligibility criteria, performing the statistical analysis, and conducting a literature review. SZBJ contributed to conceptualizing the study and interpreting the results while overseeing the research process.

Conflict of Interests: We have no conflict of interest.

Ethical Approval: This study, bearing the code IR.MODARES.REC.1400.297 has received approval from the

Research Ethics Committee of Tarbiat Modares University.

Funding/Support: No funding/support was received for this study.

References

1. Sariyer G, Öcal Taşar C, Cepe GE. Use of data mining techniques to classify length of stay of emergency department patients. *Bio-Algorithms Med-Syst*. 2019;**15**(1). https://doi.org/10.1515/bams-2018-0044.
2. Seow E. Leading and managing an emergency department—A personal view. *J Acute Med*. 2013;**3**(3):61–6. [PubMed Central ID: PMC7147188]. https://doi.org/10.1016/j.jacme.2013.06.001.
3. Hsu CM, Liang LL, Chang YT, Juang WC. Emergency department overcrowding: Quality improvement in a Taiwan Medical Center. *J Formos Med Assoc*. 2019;**118**(1 Pt 1):186–93. [PubMed ID: 29665984]. https://doi.org/10.1016/j.jfma.2018.03.008.
4. Keshtkar L, Salimifard K, Faghih N. A simulation optimization approach for resource allocation in an emergency department. *QSci Connect*. 2015;**2015**(1). https://doi.org/10.5339/connect.2015.8.
5. Li C, Zhang Z, Ren Y, Nie H, Lei Y, Qiu H, et al. Machine learning based early mortality prediction in the emergency department. *Int J Med Inform*. 2021;**155**:104570. [PubMed ID: 34547624]. https://doi.org/10.1016/j.ijmedinf.2021.104570.
6. Siddique S, Chow JC. Machine Learning in Healthcare Communication. *Encyclopedia*. 2021;**1**(1):220–39. https://doi.org/10.3390/encyclopedia1010021.
7. Zhang G, Xu J, Yu M, Yuan J, Chen F. A machine learning approach for mortality prediction only using non-invasive parameters. *Med Biol Eng Comput*. 2020;**58**(10):2195–238. [PubMed ID: 32691219]. https://doi.org/10.1007/s11517-020-02174-0.
8. Barboi C, Tzavelis A, Muhammad LN. Comparison of Severity of Illness Scores and Artificial Intelligence Models That Are Predictive of Intensive Care Unit Mortality: Meta-analysis and Review of the Literature. *JMIR Med Inform*. 2022;**10**(5). e35293. [PubMed ID: 35639445]. [PubMed Central ID: PMC9198821]. https://doi.org/10.2196/35293.
9. van Doorn W, Stassen PM, Borggreve HF, Schalkwijk MJ, Stoffers J, Bekers O, et al. A comparison of machine learning models versus clinical evaluation for mortality prediction in patients with sepsis. *PLoS One*. 2021;**16**(1). e0245157. [PubMed ID: 33465096]. [PubMed Central ID: PMC7815112]. https://doi.org/10.1371/journal.pone.0245157.
10. Klug M, Barash Y, Bechler S, Resheff YS, Tron T, Ironi A, et al. A Gradient Boosting Machine Learning Model for Predicting Early Mortality in the Emergency Department Triage: Devising a Nine-Point Triage Score. *J Gen Intern Med*. 2020;**35**(1):220–7. [PubMed ID: 31677104]. [PubMed Central ID: PMC6957629]. https://doi.org/10.1007/s11606-019-05512-7.
11. Platts-Mills TF, Travers D, Biese K, McCall B, Kizer S, LaMantia M, et al. Accuracy of the Emergency Severity Index triage instrument for identifying elder emergency department patients receiving an immediate life-saving intervention. *Acad Emerg Med*. 2010;**17**(3):238–43. [PubMed ID: 20370755]. https://doi.org/10.1111/j.1553-2712.2010.00670.x.
12. Schröer C, Kruse F, Gómez JM. A Systematic Literature Review on Applying CRISP-DM Process Model. *Procedia Comput Sci*. 2021;**181**:526–34. https://doi.org/10.1016/j.procs.2021.01.199.
13. Sullivan JH, Warkentin M, Wallace L. So many ways for assessing outliers: What really works and does it matter? *J Bus Res*. 2021;**132**:530–43. https://doi.org/10.1016/j.jbusres.2021.03.066.
14. Newaz A, Ahmed N, Shahriyar Haq F. Survival prediction of heart failure patients using machine learning techniques. *Inform Med Unlocked*. 2021;**26**:100772. https://doi.org/10.1016/j.imu.2021.100772.
15. Kursa MB, Rudnicki WR. The all relevant feature selection using random forest. *arXiv preprint arXiv:1106.5112*. 2011. https://doi.org/10.48550/arXiv.1106.5112.
16. Sagi O, Rokach L. Ensemble learning: A survey. *WIREs Rev Data Min Knowl Discov*. 2018;**8**(4). https://doi.org/10.1002/widm.1249.
17. Odegua R. An empirical study of ensemble techniques (bagging, boosting and stacking). *Proc Conf Deep Learn*. 2019.
18. Parmar A, Katariya R, Patel V. A Review on Random Forest: An Ensemble Classifier. *ICICI 2018: International Conference on Intelligent Data Communication Technologies and Internet of Things*. Springer; 2019. p. 758–63.
19. Beskopylny AN, Stel'makh SA, Shcherban' EM, Mailyan LR, Meskhi B, Razveeva I, et al. Concrete Strength Prediction Using Machine Learning Methods CatBoost, k-Nearest Neighbors, Support Vector Regression. *Appl Sci*. 2022;**12**(21):10864. https://doi.org/10.3390/app122110864.
20. Ibrahim AA, Ridwan RL, Muhammed MM, Abdulaziz RO, Saheed GA. Comparison of the CatBoost Classifier with other Machine Learning Methods. *Int J Adv Comput Sci Appl*. 2020;**11**(11). https://doi.org/10.14569/ijacsa.2020.0111190.
21. Japkowicz N, Shah M. *Evaluating learning algorithms: a classification perspective*. Cambridge University Press; 2011.
22. Anguita D, Ghelardoni L, Ghio A, Oneto L, Ridella S. The'K'in K-fold Cross Validation. ESANN; 2012. p. 441–6.
23. Fawcett T. An introduction to ROC analysis. *Pattern Recognit Lett*. 2006;**27**(8):861–74. https://doi.org/10.1016/j.patrec.2005.10.010.
24. Safaei N, Safaei B, Seyedekrami S, Talafidaryani M, Masoud A, Wang S, et al. E-CatBoost: An efficient machine learning framework for predicting ICU mortality using the eICU Collaborative Research Database. *PLoS One*. 2022;**17**(5). e0262895. [PubMed ID: 35511882]. [PubMed Central ID: PMC9070907]. https://doi.org/10.1371/journal.pone.0262895.

Table 2. Binary Features

Features and Description	Outcome	P-Value (0.05)	Chi-square Test	
	Discharged (Class 0)	Expired (Class 1)		
Gender			0.332	0.941
Female	241	48		
Male	412	99		
Smoking			0.003 [a]	8.798
No	599	145		
Yes	54	2		
Alcohol			0.001 [a]	10.284
No	593	145		
Yes	60	2		
Diabetes (DM)			< 0.001 [a]	26.713
No	308	104		
Yes	345	43		
Hypertension (HTN)			0.001 [a]	11.970
No	310	93		
Yes	343	54		
Coronary artery disease (CAD)			0.003 [a]	8.798
No	228	74		
Yes	425	73		
Cardiomyopathy (PRIOR CMP)			< 0.001 [a]	24.491
No	463	73		
Yes	190	74		
Chronic kidney disease (CKD)			0.006 [a]	7.675
No	569	115		
Yes	84	32		
Raised cardiac enzymes			0.009 [a]	6.810
No	471	90		
Yes	182	57		
Severe anemia			0.365	0.819
No	642	146		
Yes	182	1		
Anemia			0.091	2.858
No	509	105		
Yes	144	42		
Stable angina			0.243	1.361
No	647	147		
Yes	6	0		
Acute coronary syndrome (ACS)			0.004 [a]	8.798
No	375	65		
Yes	278	82		
St elevation myocardial infarction (STEMI)			0.938	0.006
No	526	118		
Yes	127	29		
Chest pain			0.635	0.225
No	652	147		
Yes	1	0		

Continued on next page

Table 2. Binary Features (Continued)

Heart failure (HF)			< 0.001 [a]	54.185
No	366	33		
Yes	287	114		
HF with reduced ejection fraction (HFREF)			< 0.001 [a]	46.062
No	433	53		
Yes	220	94		
HF with normal ejection fraction (HFNEF)			0.197	1.662
No	584	126		
Yes	69	21		
Valvular heart disease (Valvular)			0.224	1.480
No	620	143		
Yes	33	4		
Complete heart block (CHB)			0.515	0.425
No	637	142		
Yes	16	5		
Sick sinus syndrome (SSS)			0.207	1.590
No	646	147		
Yes	7	0		
Acute kidney injury (AKI)			< 0.001 [a]	23.843
No	442	68		
Yes	211	79		
Cerebrovascular accident infract (CVAI)			0.099	2.726
No	619	144		
Yes	34	3		
CVA BLEED			0.248	1.337
No	652	146		
Yes	1	1		
Atrial fibrillation (AF)			0.789	0.072
No	586	133		
Yes	67	14		
Ventricular tachycardia (VT)			< 0.001 [a]	32.691
No	634	126		
Yes	19	21		
PAROXYSMAL SUPRA VT (PSVT)			0.410	0.678
No	650	147		
Yes	3	0		
Congenital Heart disease (CONGENITAL)			0.207	1.590
No	646	147		
Yes	7	0		
Urinary tract infection (UTI)			< 0.001 [a]	17.654
No	573	146		
Yes	80	1		
Neuro cardiogenic syncope (NCS)			0.243	1.361
No	647	147		
Yes	6	0		
Orthostatic			0.477	0.506
No	638	145		
Yes	14	2		
Infective endocarditis			0.635	0.225

Continued on next page

Table 2. Binary Features (Continued)

No	652	147		
Yes	1	0		
Deep venous thrombosis (DVT)			0.571	0.320
No	645	146		
Yes	8	1		
Cardiogenic shock			< 0.001 [a]	198.708
No	617	74		
Yes	36	73		
Shock			< 0.001 [a]	219.871
No	628	77		
Yes	25	70		
Embolism			0.032 [a]	4.618
No	633	147		
Yes	20	0		
Chest infection			0.274	1.199
No	640	146		
Yes	13	1		

[a] Statistical significance at a P-value of 0.05.