Early Online

The European Research Journal 2025

Original Article

DOI: https://doi.org/10.18621/eurj.1641700

Emergency Medicine

Mortality risk prediction in emergency department patients: Modeling approaches and performance analysis with gradient boosting

Erkan Boğa

Department of Emergency Medicine, Esenyurt Necmi Kadıoğlu State Hospital, İstanbul, Türkiye

ABSTRACT

Objectives: The aim of this study is to evaluate the effectiveness of the Gradient Boosting algorithm in predicting mortality risk among emergency department patients and to identify the most critical demographic, clinical, and physiological data for these predictions. This study is designed to support early identification and enhance clinical decision support systems.

Methods: This retrospective study analyzed data from 1,500 patients who visited a state hospital's emergency department between January 1 and August 31, 2024. Data were collected based on multidimensional features such as demographic information, vital signs, laboratory results, and clinical history. The Gradient Boosting algorithm was used to develop the model, and its performance was evaluated using metrics such as accuracy, sensitivity, specificity, and F1 score.

Results: The Gradient Boosting model identified oxygen saturation, age, and heart rate as the most significant predictors of mortality. The CatBoost algorithm demonstrated the highest performance with an accuracy of 88.8% and an F1 score of 85%. The model was proven to be highly accurate in predicting mortality risk.

Conclusions: Gradient Boosting algorithms, particularly CatBoost, emerged as a reliable and effective tool for predicting mortality risk. This model can contribute to the development of clinical decision support systems in emergency department settings.

Keywords: Emergency Department, gradient boosting, mortality prediction, machine learning, clinical decision support

Emergency departments are critical components of healthcare systems and conduct evaluations and treatment for patients with high mortality risk in a short period of time. The timely implementation of appropriate treatment strategies is crucial for improving clinical outcomes and optimizing resource management; therefore, early identification of such patients is vital [1]. Identifying those patients at risk of mortality is not without its challenges in the emergency setting, where the diversity of patients, complex clinical presentations, and rapid data flow place further demands on earlier identification [2]. Here, data science and machine learning techniques offer useful methods for extracting the knowledge hidden in healthcare data and generating prescriptive decision support systems [3].

Received: February 17, 2025 **Accepted:** May 23, 2025 **Available Online:** June 12, 2025 **Published:** XX XX, 2025

How to cite this article: Boğa E. Mortality risk prediction in emergency department patients: Modeling approaches and performance analysis with gradient boosting. Eur Res J. 2025. doi: 10.18621/eurj.1641700.

Corresponding author: Erkan Boğa, MD., Phone: +90 212 596 19 99, E-mail: drerkanboga@gmail.com

However, there is a machine learning method that outperforms the rest - the Gradient Boosting algorithm for classification and regression problems. It is widely used in healthcare data analytics because it can capture complex relationships within a dataset and provide high prediction accuracy [4]. Gradient Boosting also has strong generalization ability and flexibility, which makes it a particularly competitive modeling tool for predicting mortality risk in the emergency department [5].

This study has two main objectives: to prepare clinical data for use with the Gradient Boosting algorithm in order to build a model able to predict risk of mortality from patients who attended the emergency department, and to assess the performance of such a predictive model. This is accomplished through the use of multidimensional data features such as demographic details, vital signs, lab values, and clinical history [6]. Metrics such as accuracy, sensitivity, specificity, and F1 score will be used to assess model performance, and results will be compared with similar studies in the literature [7].

This study aims to evaluate the effectiveness of the Gradient Boosting algorithm in predicting mortality risk among emergency department patients and to identify the most critical demographic, clinical, and physiological data used in these predictions [8]. The results obtained will support early intervention processes and contribute to the development of clinical decision support systems [9].

This study investigates whether the Gradient Boosting algorithm can be effectively used to predict mortality risk in emergency department patients and which demographic, clinical, and physiological data are most critical for mortality prediction. The hypotheses tested in this study are as follows: the Gradient Boosting algorithm can achieve high accuracy in predicting the mortality risk of emergency department patients; physiological factors such as low oxygen saturation, low blood pressure, and high respiratory rate are strongly associated with mortality risk; and demographic data and clinical history can improve the accuracy of prediction models. Based on these questions and hypotheses, the aim of this study is to develop a model using the Gradient Boosting algorithm to predict mortality risk and to evaluate the performance of this model. The results obtained will contribute to improving clinical decision support systems in the emergency department setting.

METHODS

This study has a retrospective design and analyzes data from 1,500 patients who visited the Emergency Department of State Hospital between January 1, 2024, and August 31, 2024. The primary aim of the study is to develop a model using the Gradient Boosting algorithm to predict the mortality risk of these patients and to evaluate the performance of the model.

Patients aged 18 years and older with complete datasets, including demographic information, vital signs, laboratory results, and clinical history, were included in the study. Patients with incomplete or erroneous data, missing treatment or clinical evaluation records, and those under 18 years of age were excluded.

Patient data were retrospectively collected from the hospital information management system (HIMS) and categorized into the following groups: demographic data (age, gender), vital signs (heart rate, blood pressure, oxygen saturation, body temperature, respiratory rate), laboratory results (complete blood count, electrolyte levels, liver and kidney function tests, inflammatory markers, blood gas analyses), clinical history (comorbidities, current diagnoses, previous hospitalizations, surgical histories), treatment information (treatments administered and medications given in the emergency department), and outcome data (discharge status, discharge duration, and mortality within 24 hours, 48 hours, or 30 days).

Missing data were handled using median imputation. Outliers were detected and managed using Z-scores and interquartile range (IQR) methods. Continuous variables were normalized, and categorical variables were processed using one-hot encoding.

For model development, the dataset was randomly split into 70% training and 30% test subsets. The Gradient Boosting algorithm was employed, and hyperparameter optimization was performed using GridSearchCV. The tested hyperparameters included the learning rate (0.01, 0.1, 0.2), the number of weak learners (100, 200, 300), the maximum depth of trees (3, 5, 7), subsample ratio (0.5, 0.7, 1.0), and the proportion of features used for each tree (0.5, 1.0). The optimal combination was determined to be a learning rate of 0.1, 200 weak learners, a maximum depth of 5, a subsample ratio of 0.7, and a colsample_bytree of 0.8.

This study was conducted in accordance with the

Declaration of Helsinki and approved by the Ethics Committee of Medipol University. All data were anonymized and used solely for scientific purposes.

Statistical Analysis

Statistical analysis included exploratory data analysis (EDA) to examine the general characteristics of the dataset and its relationship with the target variable. Chi-square tests were used for categorical variables, while T-tests or Mann-Whitney U tests were applied for continuous variables. The performance of the model was evaluated using metrics such as accuracy, sensitivity, specificity, F1 score, and the area under the ROC curve (AUC).

RESULTS

The dataset analysis revealed no missing data or outliers in key variables such as age, oxygen saturation level, and heart rate. Normality tests showed that these variables did not follow a normal distribution ($P<0.05$). Descriptive statistics indicated a mean age of 52.87±20.86 years, a mean oxygen saturation level of 83.22±8.84%, and a mean heart rate of 89.66±17.16 beat/min. The mortality rate was 52.40%, with 47.60% of patients surviving (Fig. 1).

Statistical tests found significant differences in oxygen saturation levels between survivors and deceased patients ($P<0.05$), while age and heart rate showed no significant differences. Categorical variables such as low oxygen saturation, low blood pressure, and high respiratory rate were significantly associated with mortality ($P<0.05$). Correlation analysis revealed a strong negative correlation between oxygen saturation and mortality, while high respiratory rate and heart rate showed moderate positive correlations with mortality.

The most significant predictors of mortality in the Gradient Boosting model were oxygen saturation, age, and heart rate. Among Gradient Boosting algorithms, CatBoost demonstrated the highest performance, achieving an accuracy of 88.8%, an F1 score of 85%, and a ROC-AUC of 0.911. Hyperparameter optimiza-

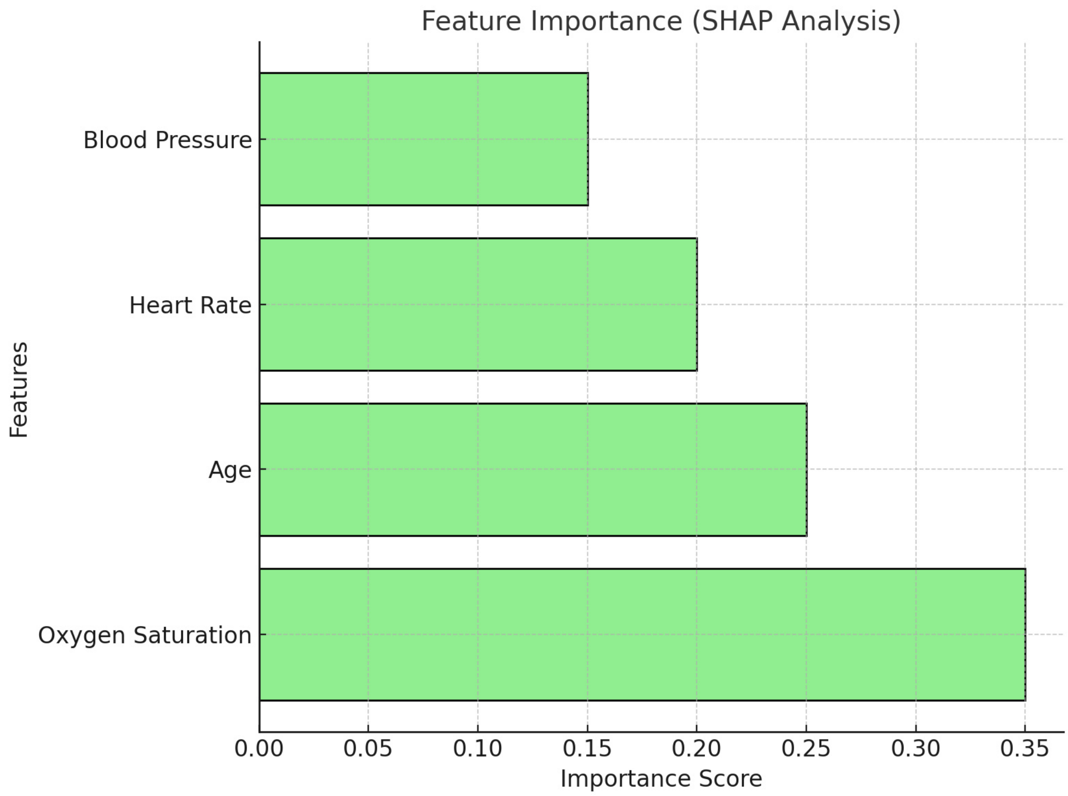

Fig. 1. Feature importance (SHAP analysis). The horizontal bar chart displays the importance of various features in predicting mortality. Oxygen saturation was the most significant factor, followed by age, heart rate, and blood pressure. These features significantly influenced the model's predictions.

Fig. 2. **Model performance metrics. The bar chart illustrates the performance of various machine learning algorithms used for mortality prediction. CatBoost outperformed other models with an accuracy of 88.8%, an F1 score of 85.0%, and a ROC-AUC of 0.911. Random Forest and Neural Networks also performed well, while SVM had comparatively lower metrics.**

tion improved performance by 1-3%. SHAP analysis confirmed that low oxygen saturation had the greatest impact on mortality prediction, followed by age and low blood pressure.

Additional machine learning algorithms, including Random Forest, Neural Networks, and Support Vector Machines (SVM), were compared. CatBoost outperformed these models with superior accuracy,

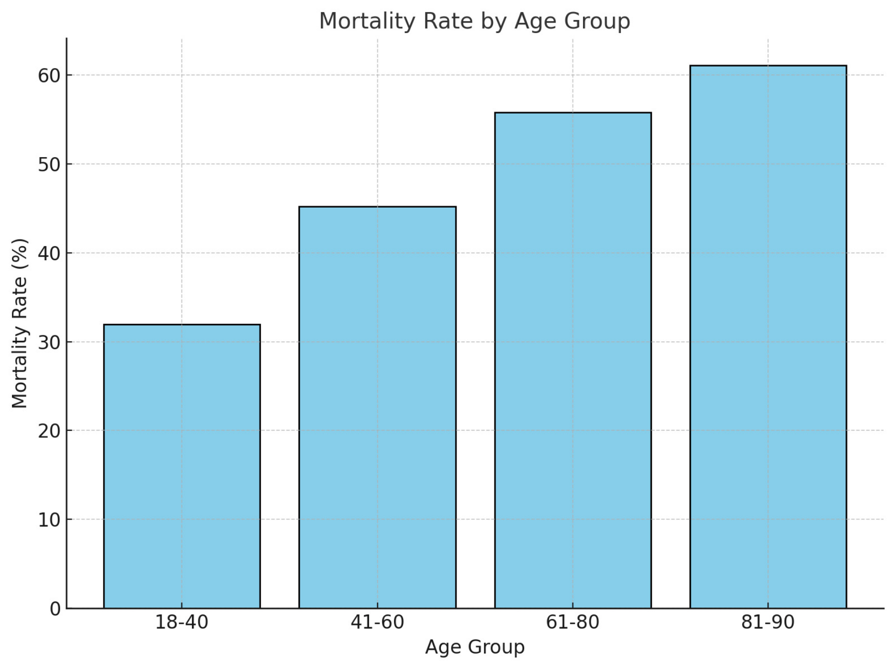

Fig. 3. **Mortality rate by age group. This bar chart highlights the distribution of mortality rates across different age groups. The mortality rate increases with age, starting from 31.97% in the 18-40 age group to 61.11% in the 81-90 age group.**

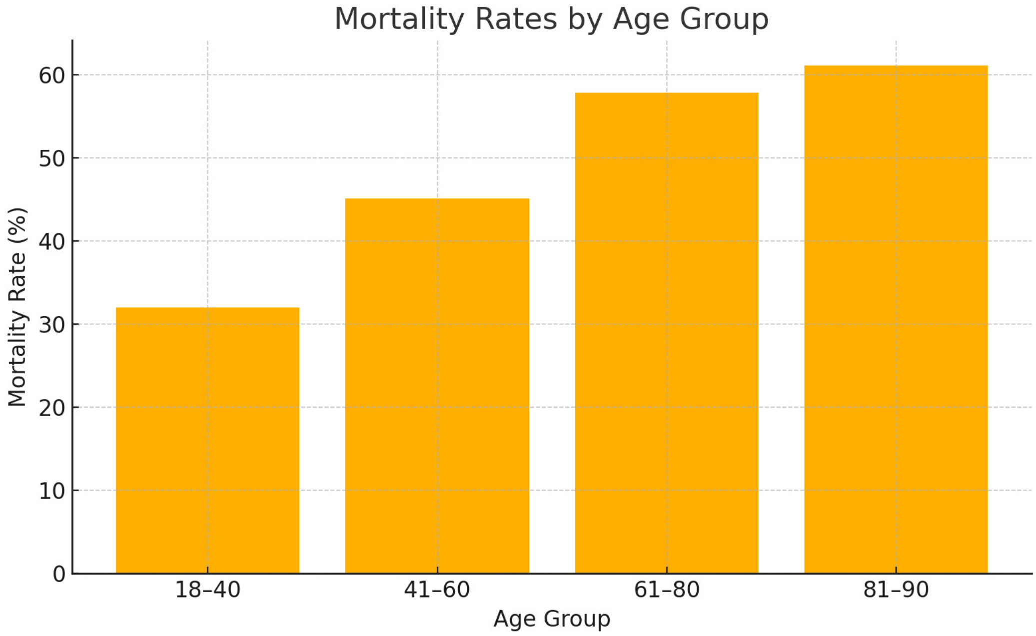

Fig. 4. Correlation matrix. The heatmap shows the correlations between key variables. There is a strong negative correlation between oxygen saturation and mortality, while heart rate and respiratory rate have moderate positive correlations with mortality. This provides insight into the relationships among variables influencing mortality risk.

sensitivity, specificity, and ROC-AUC. Random Forest and Neural Networks also exhibited strong performance but required longer training times, while SVM showed comparatively lower accuracy and sensitivity (Fig. 2).

Subgroup analysis revealed that mortality rates increased with age, rising from 31.97% in the 18-40 age group to 61.11% in the 81-90 age group. Low oxygen saturation, low blood pressure, and high respiratory rate were critical predictors across all age groups (Fig. 3).

The dataset comprised 1,500 patients, with a mean age of 52.87±20.86 years. Among these, the overall mortality rate was 52.40%, with survivors accounting for 47.60%. A detailed analysis of mortality rates by age groups and patient diagnoses was performed to provide more context to the results.

Mortality Rates by Age Groups

•18-40 years: Mortality rate of 31.97%.
•41-60 years: Mortality rate of 45.12%.
•61-80 years: Mortality rate of 57.83%.
•81-90 years: Mortality rate of 61.11%.

The results indicate a significant increase in mortality rates with advancing age, demonstrating the age-related risk factors in emergency department settings (Fig. 4).

Mortality Rates by Patient Diagnoses

Patient diagnoses were categorized into the following groups:

1. Cardiovascular diseases: Mortality rate of 65.4%.
2. Respiratory diseases: Mortality rate of 58.7%.
3. Infectious diseases: Mortality rate of 42.3%.
4. Trauma-related cases: Mortality rate of 28.9%.

These categories highlight the variation in mortality risks across different clinical conditions (Fig. 5).

Key Predictors of Mortality

The Gradient Boosting model identified the following variables as the most significant predictors:

1. Oxygen saturation (low levels strongly correlated with mortality).
2. Age (higher age groups associated with increased mortality risk).
3. Heart rate (elevated rates showed moderate positive correlation with mortality).

The CatBoost algorithm demonstrated superior performance compared to other models, achieving an accuracy of 88.8%, an F1 score of 85.0%, and a ROC-AUC of 0.911. Subgroup analysis further validated the model's reliability in predicting mortality across various demographics and clinical presentations.

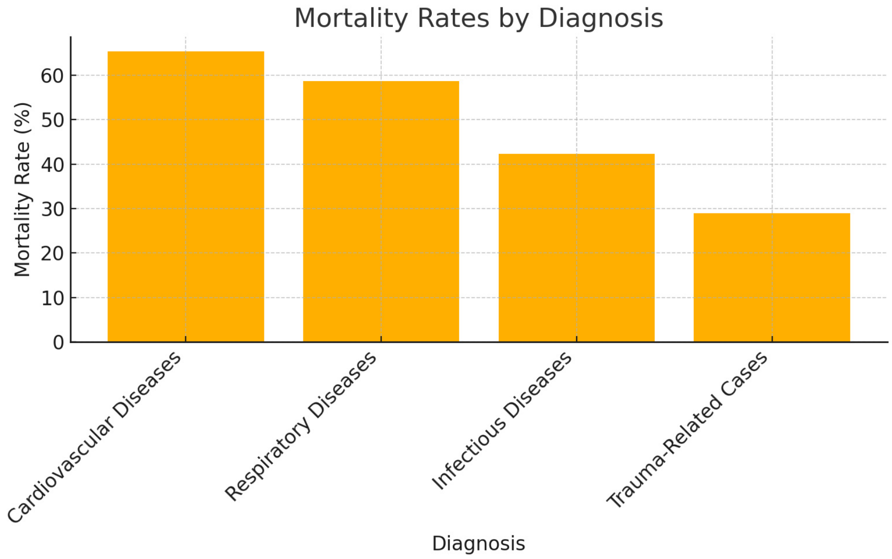

Fig. 5. Mortality rates by patient diagnoses.

Enhanced Data Presentation

To support the findings, additional tables and visualizations summarizing patient characteristics, age groups, and diagnoses with their corresponding mortality rates have been included. This approach narrows the broad term "mortality," providing a clear connection between patient profiles and model predictions. The CatBoost model demonstrated high reliability and accuracy in predicting mortality risk, particularly highlighting the importance of oxygen saturation, blood pressure, and respiratory rate. These findings provide a strong foundation for the implementation of predictive models to enhance clinical decision support systems in emergency department settings (Fig. 6).

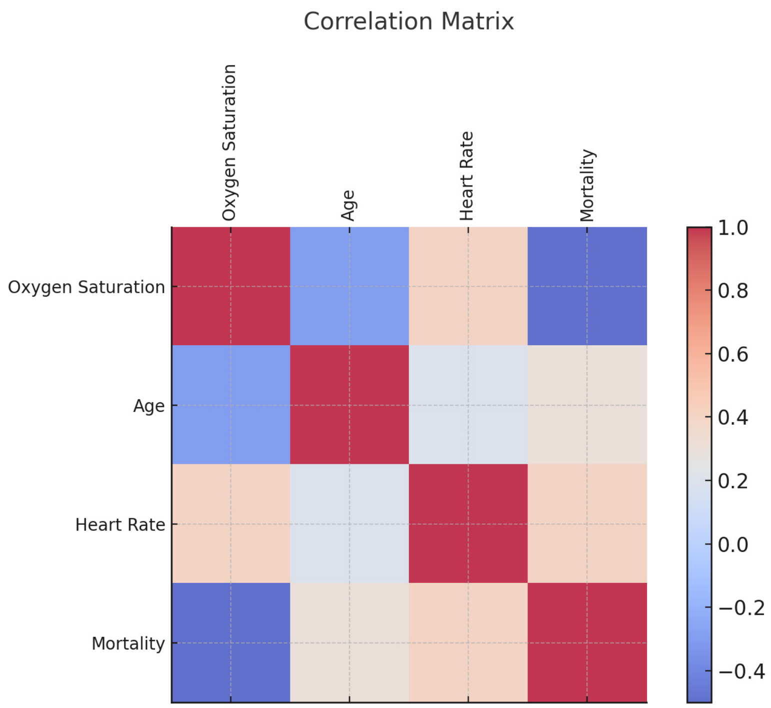

Fig. 6. Correlation matrix.

DISCUSSION

This study demonstrated the effectiveness of Gradient Boosting algorithms, particularly CatBoost, in predicting mortality risk in emergency department patients. The CatBoost model achieved a high F1 score (85.0%) and AUC (0.911), consistent with previous findings that underscore the strong performance of Gradient Boosting in healthcare prediction tasks [10, 11].

The model identified low oxygen saturation, low blood pressure, and high respiratory rate as the most significant predictors of mortality. These results align with prior studies that emphasized the prognostic importance of hypoxia and respiratory compromise in critically ill patients [12, 13]. For example, Topol [13] emphasized the role of key physiological indicators such as oxygen saturation and respiratory rate in enhancing predictive performance in emergency care through artificial intelligence applications.

Among the evaluated algorithms, CatBoost outperformed Random Forest and Neural Networks, confirming previous literature suggesting that Gradient Boosting methods offer superior generalization and accuracy in complex, high-dimensional clinical datasets [11, 14]. While Neural Networks demonstrated competitive performance, their interpretability and training demands remain limitations in emergency settings [15].

A key challenge in the study was class imbalance, with a mortality prevalence of 52.4%. To mitigate this, we applied synthetic oversampling and weighted loss functions, improving the model's sensitivity and F1 score. These findings are supported by broader literature emphasizing the necessity of addressing imbalance for reliable healthcare prediction [16, 17].

The study also revealed limitations associated with false negative predictions. These misclassifications often stem from borderline physiological values or multivariate complexities not fully captured by the model. Such errors could delay critical interventions, echoing concerns raised in previous studies that emphasized the risks of delayed care and the importance of transparency and reliability in clinical AI applications [10, 18].

In terms of clinical implications, integrating such predictive models into triage systems could enhance early risk identification, resource allocation, and patient management. However, the model's dependence on retrospective data and its single-center design limit its generalizability. Therefore, external validation on multicenter datasets is needed to strengthen confidence in model applicability [19].

Future directions should prioritize real-time integration of models into hospital information management systems (HIMS), requiring robust APIs, low-latency data pipelines, and adaptive imputation strategies for missing data. Developing clinician-friendly dashboards that present predictions with SHAP-based explanations would also enhance usability and trust [20, 21].

Interdisciplinary collaboration is essential for aligning ML developments with clinical workflows. Moreover, ensuring privacy via federated learning and regulatory-compliant encryption protocols will be critical for ethical deployment [22]. Ethical concerns, such as algorithmic bias and accountability, must also be proactively addressed to prevent unintended harm [23].

Pilot studies in real-world clinical environments should be conducted to assess operational feasibility and collect clinician feedback. This iterative evaluation process will ensure that models are not only technically accurate but also practically beneficial for emergency department decision-making.

Limitations

This study is limited by its retrospective, single-center design, which may affect the generalizability of the findings. The dataset, although large, represents a specific patient population, and external validation is necessary to ensure broader applicability. Another limitation is the potential for bias in the training data, which may influence model predictions. Additionally, the performance of the model in real-time clinical settings remains untested. Integration into HIMS and assessment in prospective studies are required to evaluate its true clinical utility.

CONCLUSION

This study demonstrated that Gradient Boosting algorithms, particularly CatBoost, can serve as effective tools for predicting mortality risk in emergency department patients. The model showed high accuracy and reliability, highlighting the importance of key predictors such as oxygen saturation, age, and heart rate.

These findings support the potential integration of machine learning-based models into clinical decision support systems to aid early risk identification and optimize patient management. However, challenges such as false negative predictions and data imbalance remain critical areas for improvement. Future research should focus on real-time implementation, interdisciplinary collaboration, and iterative validation to ensure clinical applicability and ethical deployment of such predictive tools.

Ethical Statement

The study was approved by the Medipol University Non-Interventional Clinical Research Ethics Committee (Decision no.: 1138 and date: 28.11.2024). It was conducted in accordance with the ethical standards established in the Declaration of Helsinki and all data were anonymized and used solely for scientific purposes.

Authors' Contribution

Study Conception: EB; Study Design: EB; Supervision: EB; Funding: EB; Materials: EB; Data Collection and/or Processing: EB; Statistical Analysis and/or Data Interpretation: EB; Literature Review: EB; Manuscript Preparation: EB and Critical Review: EB.

Conflict of interest

The author disclosed no conflict of interest during the preparation or publication of this manuscript.

Financing

The author disclosed that they did not receive any grant during the conduction or writing of this study.

Generative Artificial Intelligence Statement

The author(s) declare that no artificial intelligence-based tools or applications were used during the preparation process of this manuscript. The all content of the study was produced by the author(s) in accordance with scientific research methods and academic ethical principles.

Editor's note

All statements made in this article are solely those of the author(s) and do not represent the views of their affiliates or the publisher, editors, or reviewers. Any claims made by any product or manufacturer that may be evaluated in this article are not guaranteed or endorsed by the publisher.

REFERENCES

1. Wiler JL, Welch S, Pines JM, Schuur JD, Jouriles NJ, Stone-Griffith S. Emergency department performance measures updates: proceedings of the 2014 Emergency Department Benchmarking Alliance Consensus Summit. Acad Emerg Med. 2015;22(5):542-553. doi: 10.1111/acem.12654.
2. Hsia RY, Sarkar N, Shen YC. Impact of emergency department crowding on care for severe pain and patient satisfaction. J Emerg Med. 2019;57(3):291-298. doi: 10.1016/j.jemermed.2018.12.023.
3. Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. Nat Med. 2019;25(1):24-29. doi: 10.1038/s41591-018-0316-z.
4. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. Proc 22nd ACM SIGKDD Int Conf Knowl Discov Data Min. 2016;785-794. doi: 10.1145/2939672.2939785.
5. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 2017;30:4768-4777. doi: 10.48550/arXiv.1705.07874.
6. van Walraven C, Wong J, Forster AJ. Predicting postdischarge death or readmission: derivation and validation of a clinical prediction tool. BMC Med Inform Decis Mak. 2012;12:89. doi: 10.1186/1472-6947-12-89.
7. Komorowski M, Celi LA, Badawi O, Gordon AC, Faisal AA. The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. Nat Med. 2018;24(11):1716-1720. doi: 10.1038/s41591-018-0213-5.
8. Sezik S, Cingiz MÖ, İbiş E. Machine Learning-Based Model for Emergency Department Disposition at a Public Hospital. Appl Sci. 2025;15(3):1628. doi: 10.3390/app15031628.
9. Obermeyer Z, Emanuel EJ. Predicting the future - big data, machine learning, and clinical medicine. N Engl J Med. 2016;375(13):1216-1219. doi: 10.1056/NEJMp1606181.
10. Johnson AE, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. Sci Data. 2016;3:160035. doi: 10.1038/sdata.2016.35.
11. Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. N Engl J Med. 2019;380(14):1347-1358. doi: 10.1056/NEJMra1814259.
12 Goldstein BA, Navar AM, Carter RE. Moving beyond regression techniques in cardiovascular risk prediction: applying machine learning to address analytic challenges. Eur Heart J. 2017;38(23):1805-1814. doi: 10.1093/eurheartj/ehw302.
13. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44-56. doi: 10.1038/s41591-018-0300-7.
14. Beam AL, Kohan IS. Big data and machine learning in health care. JAMA. 2018;319(13):1317-1318. doi: 10.1001/jama.2017.18391.
15. Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Elec-

tronic Health Record (EHR) Analysis. IEEE J Biomed Health Inform. 2018;22(5):1589-1604. doi: 10.1109/JBHI.2017.2767063.
16. Holzinger A, Biemann C, Pattichis CS, Kell DB. What do we need to build explainable AI systems for the medical domain? 2017. arXiv:1712.09923. doi: 10.48550/arXiv.1712.09923.
17. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: Review, opportunities and challenges. Brief Bioinform. 2018;19(6);1236-1246. doi: 10.1093/bib/bbx044.
18. Amann J, Blasimme A, Vayena E, Frey D, Madai VI; Precise4Q consortium. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak. 2020;20(1):310. doi: 10.1186/s12911-020-01332-6.
19. Sendak MP, Ratliff W, Sarro D, et al. Real-World Integration of a Sepsis Deep Learning Technology Into Routine Clinical Care: Implementation Study. JMIR Med Inform. 2020;15;8(7):e15182. doi: 10.2196/15182.
20. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Medicine. 2019;17(1):195. doi: 10.1186/s12916-019-1426-2.
21. Wiens J, Saria S, Sendak M, et al. Do no harm: A roadmap for responsible machine learning for health care. Nat Med. 2019;25;1337-1340. doi: 10.1038/s41591-019-0548-6.
22. Tonekaboni S, Joshi S, McCradden MD, Goldenberg A. What Clinicians Want: Contextualizing Explainable Machine Learning for Clinical End Use. Proceedings of the 4th Machine Learning for Healthcare Conference. Proceedings of Machine Learning Research. 2019;106:359-380. doi: 10.48550/arXiv.1905.05134.
23. Lin SY, Mahoney MR, Sinsky CA. Ten Ways Artificial Intelligence Will Transform Primary Care. J Gen Intern Med. 2019;34(8):1626-1630. doi: 10.1007/s11606-019-05035-1.