Original research article

DIGITAL HEALTH

DIGITAL HEALTH
Volume 11: 1–10

DOI: 10.1177/20552076251335705
journals.sagepub.com/home/dhj

Sage

Developing and validating a machine learning-based model for predicting in-hospital mortality among ICU-admitted heart failure patients: A study utilizing the MIMIC-III database

De Su[1,*], Jie Zheng[1,*], Yue-kai Shao[1], Jun-ya Liu[1], Xin-xin Liu[1], Kun Yu[1], Bang-hai Feng[2], Hong Mei[1] and Song Qin[1]

Abstract

Background: Although the assessment of in-hospital mortality risk among heart failure patients in the intensive care unit (ICU) is crucial for clinical decision-making, there is currently a lack of comprehensive models accurately predicting their prognosis. Machine learning techniques offer a powerful means to identify potential risk factors and predict outcomes within multivariable clinical data.

Methods: This study, based on the MIMIC-III database, extracted demographic characteristics, vital signs, laboratory test values, and comorbidity information of heart failure patients using structured query language. LASSO regression was employed for feature selection, and various machine learning algorithms were utilized to train models, including logistic regression (LR), random forest (RF), and gradient boosting (GB), among others. An ensemble learning model based on a soft voting mechanism was constructed. Model performance was evaluated using accuracy, recall, precision, F1 score, and AUC values through cross-validation and on an independent test set.

Results: In five-fold cross-validation, the soft voting ensemble learning model demonstrated the best overall performance, with accuracy and AUC values both at 0.86. Additionally, RF and GB models also performed well, with RF achieving an accuracy of 0.79 and an AUC of 0.79 on the independent test set, while the GB model achieved an accuracy of 0.77 and an AUC of 0.79. In contrast, other models such as LR, SVM, and KNN exhibited poorer performance in terms of accuracy and AUC values, indicating the significant advantage of ensemble methods in handling complex clinical prediction tasks.

Conclusion: This study demonstrates the potential of machine learning models, particularly ensemble learning models based on soft voting mechanisms, in predicting in-hospital mortality risk among heart failure patients in the ICU. The overall performance of the ensemble learning model confirms its effectiveness as an adjunct clinical decision-making tool. Future research should further optimize the models and validate them in a broader patient population to enhance their practical utility and accuracy in real clinical settings.

Keywords

Heart failure, intensive care unit, in-hospital mortality risk, machine learning

Received: 24 December 2024; accepted: 1 April 2025

[1]Department of Critical Care Medicine, Affiliated Hospital of Zunyi Medical University, Zunyi, Guizhou, P.R. China
[2]Department of Critical Care Medicine, Zunyi Hospital of Traditional Chinese Medicine, Zunyi, Guizhou, P.R. China

*These authors contributed equally to this work.

Corresponding author:
Song Qin, Department of Critical Care Medicine, Affiliated Hospital of Zunyi Medical University, 149 Dalian Street, Zunyi, Guizhou 563000, P.R. China.
Email: 15285230903@163.com

Introduction

Heart failure (HF) is one of the heart diseases with high morbidity and mortality worldwide, especially in the intensive care unit (ICU), due to the complexity of the disease and rapid changes, its mortality is higher. Heart failure (HF) is a world-class problem with an estimated prevalence of 26 million worldwide.[1] And this number is still increasing year by year.[2,3] ICU plays an important role in the treatment of critically ill patients, especially the treatment and nursing of patients with heart failure is more critical,[4] but the prognosis of patients with heart failure in ICU is still not optimistic.[5,6] Therefore, early mortality detection of patients is necessary, and accurate prediction of in-hospital mortality of these patients is of great significance for optimizing treatment, resource allocation and improving patient prognosis. In recent years, with the rapid development of big data and machine learning technology, data mining and prediction models based on Electronic Health Records (EHRs) have been widely used in the medical field.[7–10] In recent years, machine learning and model prediction have gradually developed, and the establishment of predictive models based on big data can help doctors diagnose diseases more accurately, formulate personalized treatment plans, evaluate treatment effects, predict patients' prognosis, and help doctors adjust treatment plans.[11–13] MIMIC-III (Medical Information Mart for Intensive Care III) is a publicly available critical care database that contains a large amount of clinical data and provides a valuable resource for researchers.[14] At present, the research on the prediction scheme and clinical diagnosis strategy based on MIMI-III mainly covers various important diseases related to ICU, such as the construction of the death risk prediction model of septic shock patients based on supervised machine learning algorithm,[15] and the use of coagulation and heparin in the prediction and classification of sepsis survival by machine learning.[11] Studies on serum lactate levels, the SOFA score, and the accuracy of the qSOFA score have been conducted to predict adult mortality in sepsis patients.[16] We can find that most studies are about sepsis and its complications, while the prediction of in-patient mortality of ICU patients with heart failure is relatively rare. One study built a prediction model of the death risk of ICU patients with heart failure through a large amount of data from the MIMIC-III database.[17] This study aimed to develop and validate a model for predicting in-hospital mortality risk among heart failure (HF) patients admitted to the intensive care unit (ICU).

Method

Data acquisition

Data source and extraction method. This study utilized data from the MIMIC-III database (https://datadryad.org/stash/dataset/doi:10.5061/dryad.0p2ngf1zd). The database, maintained by the Computational Physiology Laboratory at the Massachusetts Institute of Technology (MIT), contains extensive medical records for critically ill patients.[17] The author (YS) obtained access to the database (certificate number 59828695). The Institutional Review Boards (IRBs) at Beth Israel Deaconess Medical Center (BIDMC) approved this investigation and waived the requirement for informed consent from patients.

We extracted data, including demographic characteristics, vital signs, and laboratory test results, from the database by executing structured query language (SQL) queries using PostgreSQL (version 9.6). Specifically, the data were obtained from the following tables: ADMISSIONS (admission records), PATIENTS (patient demographics), ICUSTAYS (ICU admission information), D_ICD_DIAGNOSIS (ICD-9 diagnosis code directory), DIAGNOSIS_ICD (diagnosis codes), LABEVENTS (laboratory events), D_LABIEVENTS (laboratory item directory), CHARTEVENTS (nursing records), D_ITEMS (item directory), NOTEEVENTS (medical notes), and OUTPUTEVENTS (output records).

Data content and processing. In this study, we focused on several categories of data: demographic characteristics (including age, gender, race, weight, and height at admission), vital signs (heart rate [HR], systolic blood pressure [SBP], diastolic blood pressure [DBP], mean blood pressure, respiratory rate, temperature, pulse oximetry [SPO2], and urine output within the first 24 hours), comorbidities (hypertension, atrial fibrillation, ischemic heart disease, diabetes, depression, iron deficiency anemia, hyperlipidemia, chronic kidney disease [CKD], chronic obstructive pulmonary disease [COPD]), and laboratory test values (including hematocrit, red blood cell count, mean corpuscular hemoglobin [MCH], mean corpuscular hemoglobin concentration [MCHC], mean corpuscular volume [MCV], red cell distribution width [RDW], platelet count, white blood cell count, neutrophils, eosinophils, lymphocytes, prothrombin time [PT], international normalized ratio [INR], NT-proBNP, creatine kinase, creatinine, blood urea nitrogen [BUN], glucose, potassium, sodium, calcium, chloride, magnesium, anion gap, bicarbonate, lactate, blood pH, arterial carbon dioxide pressure, and left ventricular ejection fraction [LVEF]). For variables with multiple measurements, we calculated the average for analysis. Comorbidities were identified based on ICD-9 codes, laboratory test values covered data throughout the ICU stay, while demographic characteristics and vital signs recorded data within the first 24 hours of each admission. The primary outcome measure of the study was in-hospital mortality rate, i.e., the survival status of patients at discharge.

Data preprocessing

Before model training, the dataset underwent cleaning processes. The first step of cleaning involved removing data rows lacking target labels (i.e., in-hospital mortality status), which were not helpful for subsequent analysis and model training. To ensure that the dataset only contained features relevant to the predictive model, we removed non-predictive features such as patient identifiers (IDs) and grouping information, which had no direct impact on the prediction results and could potentially cause model overfitting.

For missing data, the K-nearest neighbors (KNN) algorithm was used to estimate missing values. This method estimates the values of missing data points by searching for similar cases (i.e., sample points closest in feature space) and using the corresponding values of these cases. This approach is based on the assumption of similarity between neighboring samples and is applicable to both continuous and categorical data in this study.

After handling missing values, the data underwent normalization to eliminate the influence of different scales, thus improving the stability and convergence speed of the model. The Min-Max Scaling method was employed, which scales all feature values to a range between 0 and 1. This step is particularly important for distance-based algorithms such as the K-nearest neighbors algorithm, as it ensures that all features have equal importance when calculating distances.

Key feature selection

To effectively identify the most relevant features associated with in-hospital mortality risk among heart failure patients, this study employed the LASSO (Least Absolute Shrinkage and Selection Operator) logistic regression model. LASSO introduces a tuning parameter (λ) to shrink regression coefficients, simultaneously achieving variable selection and complexity adjustment, which results in some coefficients being shrunk to zero, thus achieving feature selection. We determined the optimal λ value through 10-fold cross-validation, which not only effectively evaluates the robustness of the model but also prevents overfitting.

Construction of machine learning prediction models

After feature selection, this study employed various machine learning models to train the prediction model for in-hospital mortality risk among heart failure patients. The dataset was divided into training and testing sets in a 5:5 ratio. To assess the performance of each model, we applied five-fold cross-validation to the training set to reduce variance across different datasets.

To address potential class imbalance in the dataset, we first used RandomUnderSampler to randomly undersample the majority class, followed by SMOTE to oversample the minority class, aiming to improve data imbalance. After data resampling, we retrained and evaluated the models using the same machine learning models and five-fold cross-validation. This step aimed to optimize the model's generalization ability and improve prediction accuracy for the minority class.

To further enhance the predictive performance of the model, this study introduced an ensemble learning model based on the soft voting mechanism (Soft-Voting Classifier Model). Within the framework of ensemble learning, we selected three top-performing base models as candidates: logistic regression, random forest, and gradient boosting models. The soft voting mechanism aggregates the predictions of different models by weighted averaging of their probability predictions rather than simple majority voting, leveraging the information from each base model's probability estimates to provide more accurate predictions. In our model, the soft voting mechanism adjusts the weights of each base model to optimize the overall prediction accuracy, with the weight of each base model determined based on its performance in cross-validation.

To highlight the superiority of this model, we compared it with mainstream models such as logistic regression (LR), K-nearest neighbors (KNN), support vector machine (SVM), decision tree (DT), random forest (RF), gradient boosting (GB), linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), Gaussian Naive Bayes (GNB), and MLP. We also applied five-fold cross-validation to this model and evaluated it on an independent test set. Finally, by comparing the performance metrics of different models, we identified the most suitable predictive model for this study's purposes.

Model evaluation

Model evaluation is crucial for assessing its performance. We used a range of evaluation metrics to comprehensively assess the model's performance. These metrics include accuracy, recall, precision, F1 score, and the area under the receiver operating characteristic curve (AUC). Accuracy reflects the proportion of correctly predicted instances, recall measures the model's ability to correctly identify positive samples, precision indicates the proportion of predicted positive samples that are actually positive, the F1 score is the harmonic mean of recall and precision, and the AUC value is an important statistical metric for evaluating the model's classification ability. Model evaluation was initially conducted within the framework of five-fold cross-validation to ensure the robustness of the evaluation results. Cross-validation not only provides performance information on different subsets but also helps us adjust model parameters and optimize model performance. Building upon five-fold cross-validation, we further evaluated the model on an independent test set to verify its generalization ability on unknown data.

Statistical analysis

The statistical analysis of this study was conducted on a computing device equipped with an NVIDIA RTX 3070 GPU. All data processing and analysis were performed on the Windows 10 operating system. For data processing, model construction, and statistical analysis, we used Python 3.8, chosen for its powerful library support and wide application in the field of machine learning. The main libraries used included NumPy and Pandas for data manipulation, Scikit-learn for implementing and evaluating machine learning models, Matplotlib and Seaborn for data visualization, and SciPy for more advanced statistical analysis. The deLong test is a non-parametric approach used to compare the areas under two or more receiver operating characteristic (ROC) curves. In the statistical analysis phase, descriptive statistics were used to summarize the main features of the dataset, including mean, standard deviation, and median. To compare differences between different groups, t-tests and ANOVA were used, while the non-parametric Mann-Whitney U test was used for non-normally distributed data. All hypothesis tests were two-tailed. For categorical variables, the chi-square test was used to compare distribution differences between different groups. To evaluate the predictive performance of the models, AUC values and confusion matrices were calculated. The significance level was set at a *P*-value less than 0.05.

Result

Feature selection results

In this study, feature selection for predicting in-hospital mortality risk among heart failure patients was conducted using the LASSO regression model, and the optimal regularization parameter λ value was determined. As shown in Figure 1(a), the relationship curve between cross-validation scores and log(λ) indicates that the score peaks at the optimal λ value (λ=0.32), with a vertical line clearly marking the optimal regularization strength. The selection of this λ value is based on optimizing model performance to ensure that feature selection neither overfits nor underfits. The LASSO coefficient profiles of features, as shown in Figure 1(b), demonstrate that as log(λ) increases, some feature coefficients tend towards zero while others remain non-zero, indicating the importance of these non-zero coefficient features in predicting in-hospital mortality risk among heart failure patients. At the optimal λ value determined through cross-validation, we plotted a vertical line to identify the features selected by the model at this regularization strength.

From the bar chart in Figure 1(c), it can be observed that some features such as creatinine, blood calcium, PCO2, and urea nitrogen have relatively high coefficient values, indicating a strong correlation with the in-hospital mortality risk of heart failure patients. Especially for creatinine and urea nitrogen, as indicators of renal function, they hold significant clinical significance in heart failure management, with elevated levels often indicating worsening patient conditions and poor prognosis. On the other hand, through 3D visualization using principal component analysis (PCA), as shown in Figure 1(d), we can observe the relationships between different features and their distributions in the data. This 3D representation helps identify which features are more important for distinguishing between different patient groups (such as surviving and deceased patients) and also reveals potential complex interactions between features.

Cross-validation results

In the prediction of in-hospital mortality risk among heart failure patients in this study, several machine learning models, including LR, KNN, SVM, DT, RF, GB, LDA, QDA, GNB, and MLP, were employed, and their performance was evaluated through five-fold cross-validation. Table 1 shows the performance results of each model. RF performed the best on almost all evaluation metrics, with accuracy, recall, precision, F1 score, and AUC values reaching 0.86, indicating its high reliability in predicting the mortality risk of heart failure patients under different circumstances. GB closely followed, showing similar high performance with an accuracy of 0.85 and an F1 score of 0.84. LR and LDA exhibited robust performance on all metrics, with accuracy and AUC values of 0.78, while recall, precision, and F1 score also remained consistent, reflecting the balanced and consistent predictive ability of these two models. QDA showed significantly outstanding performance in precision, reaching 0.91, indicating a high proportion of actual positives among predicted positives. However, its recall rate was slightly lower at 0.76, suggesting potential missed positive cases in practical applications. KNN and DT models showed moderate performance, while SVM performed the weakest among these models, especially with poor performance on recall (0.49), indicating deficiencies in identifying all positive samples. Regarding the ensemble learning model based on the soft voting mechanism, this model demonstrated balanced and consistent performance in cross-validation, with an accuracy of 0.86, recall of 0.81, precision of 0.90, F1 score of 0.84, and AUC value of 0.86.

Independent test results

The results of the independent test set provide information about the model's predictive ability in real-world

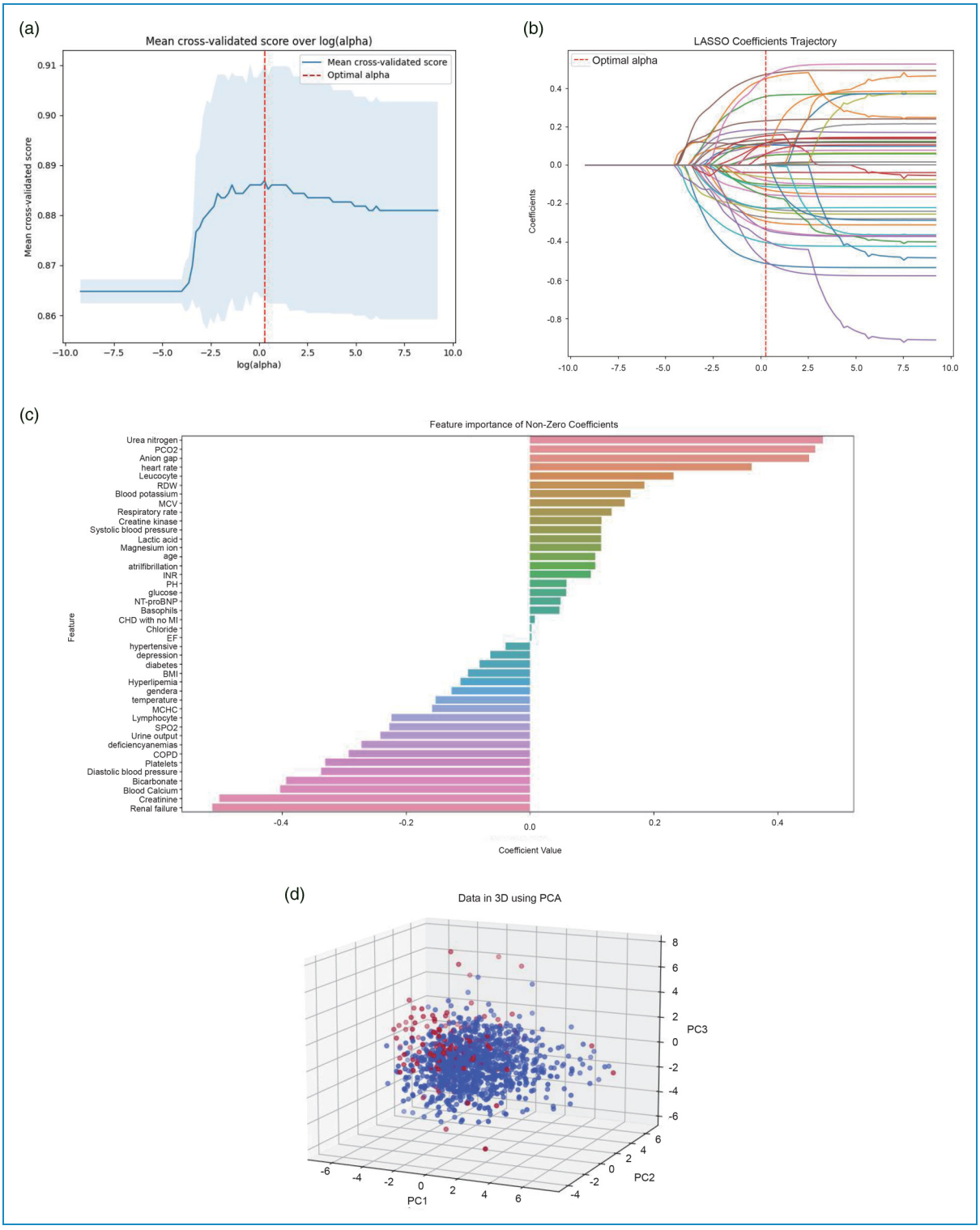

Figure 1. Visualization of the LASSO feature screening process, where (a) represents the relationship curve between the cross-validation score and log(λ); (b) represents the LASSO coefficient profile of the feature; (c) represents the bar chart of the feature coefficient value; (d) represents the 3D visualization of the principal component analysis.

scenarios. The ensemble learning model based on the voting mechanism performed well in cross-validation but showed slightly different performance on the independent test set (corresponding ROC curves of each model are shown in Figure 2, confusion matrices in Figure 3, and model performance metrics in Table 2).

Table 1. Five-fold cross-validation evaluation results of each model.

Model	Accuracy (CV)	Recall (CV)	Precision (CV)	F1 score (CV)	AUC (CV)
Logistic regression	0.78	0.78	0.78	0.78	0.78
K-nearest neighbors	0.70	0.70	0.71	0.70	0.70
Support vector machines	0.63	0.49	0.69	0.57	0.63
Decision tree	0.71	0.72	0.70	0.70	0.71
Random forest	0.86	0.87	0.86	0.86	0.86
Gradient boosting	0.85	0.86	0.84	0.84	0.85
Linear discriminant analysis	0.78	0.77	0.78	0.77	0.78
Quadratic discriminant analysis	0.84	0.76	0.91	0.80	0.84
Soft-voting	0.86	0.81	0.90	0.84	0.86

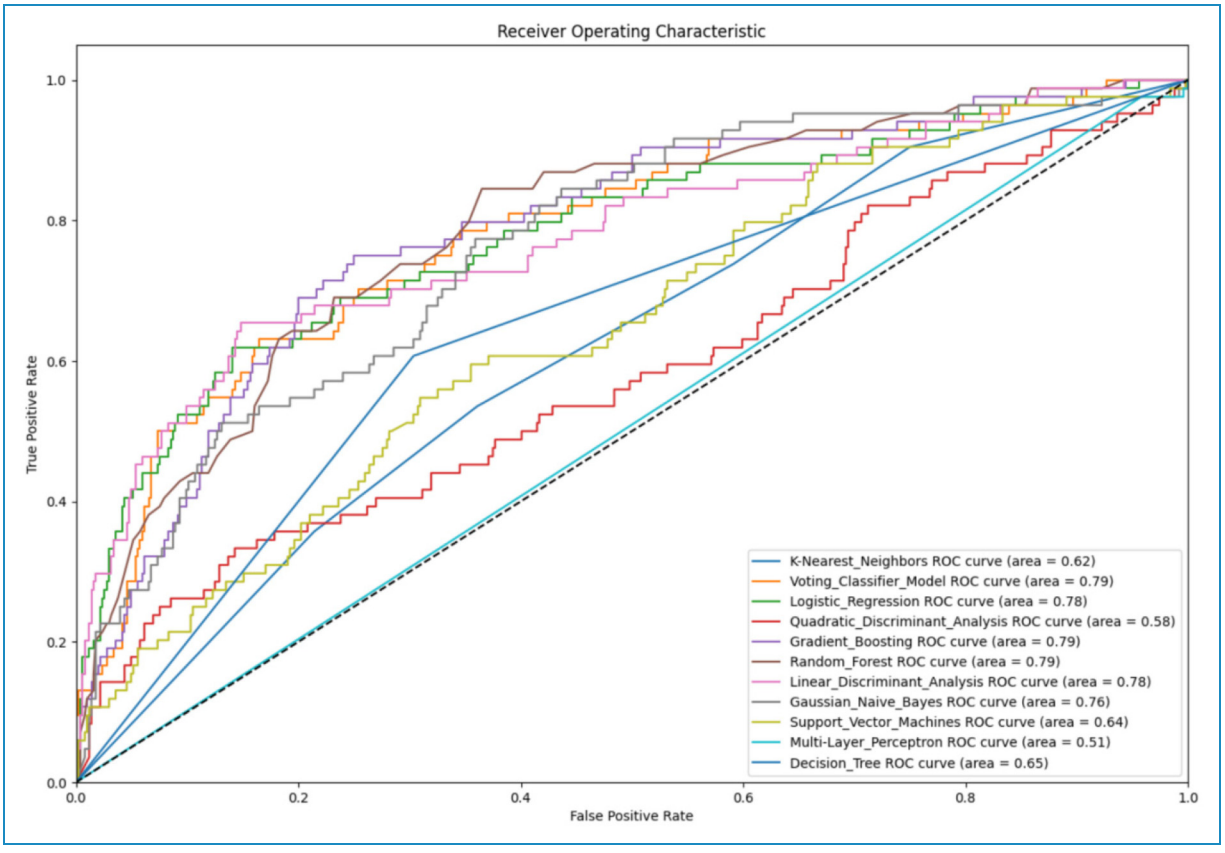

Figure 2. ROC curve of the model in independent testing.

Discussion

In this study, we aimed to develop a machine learning-based model to predict in-hospital mortality risk among heart failure patients admitted to the intensive care unit (ICU). Through comprehensive data collection and pre-processing, we extracted a series of clinical variables from the MIMIC-III database and used the LASSO

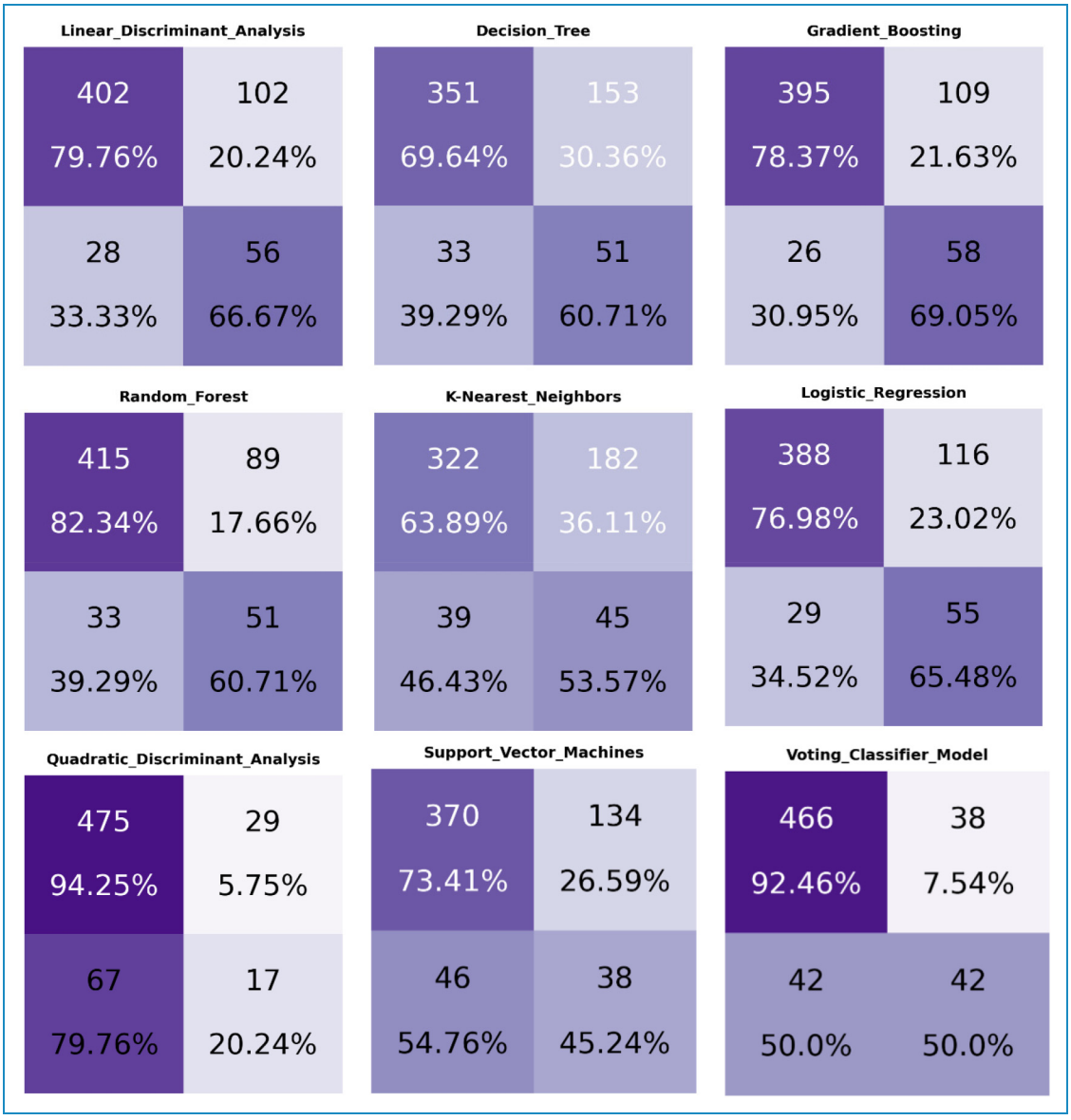

Figure 3. Confusion matrices for each model in independent tests.

regression model to select these features. After feature selection and model building, we employed various algorithms including LR, RF, and GB, among others, to evaluate their performance under five-fold cross-validation, using metrics such as accuracy, recall, precision, F1 score, and AUC values as evaluation criteria. Ultimately, we adopted a self-designed ensemble learning model based on the soft voting mechanism, which demonstrated excellent predictive performance in cross-validation. Although some performance metrics of the model decreased on the independent test set, it still showed potential in practical applications. These results provide valuable insights for risk assessment and management of heart failure patients and indicate directions for further optimization in future research.

The outstanding performance of ensemble learning proposed for predicting in-hospital mortality risk among heart failure patients can be attributed to its integration of the strengths of multiple independent models, mitigating potential shortcomings of individual models. Heart failure is a multifactorial disease, and its clinical manifestations and outcomes may be influenced by numerous physiological parameters. A single predictive model may only capture a certain aspect of the data, making it difficult to fully comprehend these complex interactions. For example, DT may be overly rigid in specific splitting rules, while SVM may not be flexible enough in handling nonlinear data. In contrast, ensemble learning, through the integration of different algorithms such as collective decision-making in RF and progressive improvement in GB, can capture richer

Table 2. Evaluation results of each model in independent tests.

Model	Accuracy	Recall	Precision	F1 score	AUC
Logistic regression	0.75	0.65	0.32	0.43	0.78
K-nearest neighbors	0.62	0.54	0.20	0.29	0.62
Support vector machines	0.69	0.45	0.22	0.30	0.64
Decision tree	0.68	0.61	0.25	0.35	0.65
Random forest	0.79	0.61	0.36	0.46	0.79
Gradient boosting	0.77	0.69	0.35	0.46	0.79
Linear discriminant analysis	0.78	0.67	0.35	0.46	0.78
Quadratic discriminant analysis	0.75	0.65	0.32	0.43	0.78
Soft-voting	0.86	0.50	0.53	0.51	0.79

patterns in patient data. This approach is particularly suitable for complex medical datasets like MIMIC-III, as it contains a large amount of heterogeneity that individual models may struggle to handle effectively. Each individual model in the soft voting ensemble learning model provides unique insights into the mortality risk of specific patient populations, and when combined, their collective action can more comprehensively assess patient risk. Moreover, during training, ensemble learning automatically adjusts the weights of different models through cross-validation, ensuring that models with optimal performance have greater influence in final predictions. This weight allocation process adapts to the clinical data characteristics in this study, especially when dealing with highly individualized and complex clinical conditions such as heart failure.

The machine learning-based predictive model developed in this study has significant implications for clinical practice. Firstly, by accurately predicting the mortality risk of heart failure patients in the ICU, physicians can allocate medical resources more effectively, prioritizing treatment for those at higher risk. This is particularly important in resource-constrained environments, as it can improve the overall operational efficiency of the ICU and may reduce patient mortality through timely interventions. On the other hand, the model can reveal which specific clinical parameters are most relevant to patient mortality risk, helping physicians identify high-risk patients in routine clinical work and take preventive measures early, such as more frequent monitoring or more aggressive treatment strategies. For example, the model emphasizes the importance of Creatinine and Urea nitrogen as indicators of renal function for heart failure patients, implying that renal function monitoring should be an important part of assessing their prognosis. The importance of Blood calcium and PCO2 also highlights the role of electrolyte imbalance and respiratory dysfunction in the mortality risk of heart failure patients. By understanding the key factors influencing prognosis, physicians can better educate patients about the importance of disease management, especially in discharge planning and self-care.

The results of the ensemble model were comparable to or even better than those of the individual models on certain key metrics. When comparing the ensemble model with the previously mentioned optimal individual models—RF and GB, the ensemble model showed a significant advantage in precision, reaching 0.90, while the precision of these two individual models was 0.86 and 0.84, respectively. This high precision indicates that the ensemble model is less likely to misclassify actually surviving patients as deceased, which is crucial for reducing unnecessary medical interventions and psychological burden. Although slightly lower in recall than the RF model, its balance and precision make it a strong candidate model. Additionally, the ensemble model's F1 score and AUC value were equal to or higher than those of the RF and GB models, suggesting competitive performance across key metrics. The F1 score, as the harmonic mean of recall and precision, is an important indicator for evaluating model accuracy and recall ability, especially in datasets with class imbalance. The high AUC value demonstrated by the ensemble model also indicates its good classification ability at different thresholds.

We observed that the accuracy of the ensemble model remained at 0.86 on the independent test set, consistent with its performance in cross-validation. This suggests that the model has good generalization ability and provides accurate predictions for unseen data. However, we noticed that the recall rate decreased from 0.81 in cross-validation to 0.50 in the test set, indicating that the model may miss

half of the positive cases in practical applications. The precision also decreased, from 0.90 in cross-validation to 0.53 in the test set. Nevertheless, with an F1 score of 0.51 on the test set, it still demonstrates better balance compared to individual models such as SVM and DT. Observing the ROC curves further reveals the characteristics of model performance. The AUC value of the ensemble model on the independent test set was 0.79, slightly lower than the 0.86 in cross-validation but still significantly better than most individual models, such as KNN and SVM, confirming its ability to distinguish between mortality risk categories. From these results, it can be seen that although the ensemble learning model showed some decline in certain metrics on the independent test set, it still maintained high accuracy and AUC values overall, indicating good performance in practical applications. However, the decrease in recall suggests the need for further adjustment of the model in clinical applications or in combination with professional judgments from doctors to minimize the risk of missed diagnoses.

Although this study has achieved some success in predicting mortality risk among heart failure patients, there are still some limitations. Firstly, the dataset used, MIMIC-III, although rich in information and multidimensional, comes from a single geographical area and population. Therefore, the generalization ability of the model may be limited to specific populations and medical environments, which may affect the effectiveness of the model in different healthcare systems or populations. The model's performance showed a decrease in recall on the independent test set, suggesting that the model may encounter new challenges when facing real-world data, such as sample imbalance and unseen complexity. Additionally, we rely on the completeness and accuracy of existing features and collected data, and any measurement errors or data entry errors may affect the reliability of the prediction results. Furthermore, although multiple clinical variables were considered in model construction, there may still be potential influencing factors that are not recorded or provided in the database, such as patients' quality of life, mental health status, and genetic background. To overcome these limitations, future research needs to validate the effectiveness of the model in a wider and more diverse population, consider more comprehensive patient information, and continuously optimize the model algorithm to improve its performance and usability in real-world applications. Additionally, strengthening research on model interpretability and conducting prospective studies in actual clinical settings will be crucial steps in further advancing the model from theory to clinical practice.

Conclusion

This study aimed to develop a machine learning-based model to predict in-hospital mortality risk among heart failure patients in the intensive care unit. Utilizing the extensive clinical data in the MIMIC-III database, we employed LASSO regression for feature selection and trained predictive models using various machine learning algorithms. Ultimately, we proposed an ensemble learning model based on the soft voting mechanism, which achieved high accuracy, precision, and AUC values in cross-validation. The results on the independent test set showed that although some performance metrics decreased, the model demonstrated good predictive ability and potential clinical application value overall. However, considering the limitations of the dataset, challenges in model generalization, and implementation issues in actual clinical practice, our research results suggest the need for more rigorous validation and optimization before the model is widely applied in clinical settings.

ORCID iD

Song Qin https://orcid.org/0000-0002-3321-297X

Statements and declarations

Ethical considerations

This research was conducted in compliance with the Helsinki Declaration's guidelines. Approval for using the MIMIC-III database was obtained from the IRBs of both MIT and BIDMC. The ethical approval previously granted for the MIMIC database covers the data used in this study, obviating the need for further ethical approval or informed consent.

Author contributions/CRediT

DS, YS, and SQ designed this study. YS performs data extraction. JL, BF, and HM performed all data analysis and chart preparation. XL, JZ, KY, and SQ drafted the initial manuscript. SQ reviewed and revised the manuscript. All authors read and approved the final manuscript.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Guizhou Provincial Department of Science and Technology, Guizhou Provincial Health Commission Science and Technology Fund Project, Zunyi Science and Technology Bureau Science and Technology Fund Project (grant numbers: ZK- 2022- 660, ZK-2023-544, ZK-2024-299, gzwkj2024-310, 2023, No. 221, 2023, No. 199).

Conflicting interests

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

References

1. Ambrosy AP, Fonarow GC, Butler J, et al. The global health and economic burden of hospitalizations for heart failure: lessons learned from hospitalized heart failure registries. *J Am Coll Cardiol* 2014; 63: 1123–1133.

2. Heidenreich PA, Albert NM, Allen LA, et al. Forecasting the impact of heart failure in the United States: a policy statement from the American Heart Association. *Circulation Heart Failure* 2013; 6: 606–619.
3. Arruda VL, Machado LMG, Lima JC, et al. Trends in mortality from heart failure in Brazil: 1998 to 2019. *Rev Bras Epidemiol* 2022; 25: E220021.
4. Metkus TS, Lindsley J, Fair L, et al. Quality of heart failure care in the intensive care unit. *J Card Fail* 2021; 27: 1111–1125.
5. Li J, Liu S, Hu Y, et al. Predicting mortality in intensive care unit patients with heart failure using an interpretable machine learning model: retrospective cohort study. *J Med Internet Res* 2022; 24: e38082.
6. Kanwar MK, Everett KD, Gulati G, et al. Epidemiology and management of right ventricular-predominant heart failure and shock in the cardiac intensive care unit. *European Heart Journal Acute Cardiovascular Care* 2022; 11: 584–594.
7. Sanchez-Pinto LN, Luo Y and Churpek MM. Big data and data science in critical care. *Chest* 2018; 154: 1239–1248.
8. Carra G, Salluh JIF, da Silva Ramos FJ, et al. Data-driven ICU management: using big data and algorithms to improve outcomes. *J Crit Care* 2020; 60: 300–304.
9. Pirracchio R, Cohen MJ, Malenica I, et al. Big data and targeted machine learning in action to assist medical decision in the ICU. *Anaesth Crit Care Pain Med* 2019; 38: 377–384.
10. Alkhachroum A, Kromm J and De Georgia MA. Big data and predictive analytics in neurocritical care. *Curr Neurol Neurosci Rep* 2022; 22: 19–32.
11. Guo F, Zhu X, Wu Z, et al. Clinical applications of machine learning in the survival prediction and classification of sepsis: coagulation and heparin usage matter. *J Transl Med* 2022; 20: 65.
12. Zou Y, Zhao L, Zhang J, et al. Development and internal validation of machine learning algorithms for end-stage renal disease risk prediction model of people with type 2 diabetes mellitus and diabetic kidney disease. *Renal Fail* 2022; 44: 562–570.
13. Wang R, Dai W, Gong J, et al. Development of a novel combined nomogram model integrating deep learning-pathomics, radiomics and immunoscore to predict postoperative outcome of colorectal cancer lung metastasis patients. *J Hematol Oncol* 2022; 15: 11.
14. Zeng Z, Deng Y, Li X, et al. Natural language processing for EHR-based computational phenotyping. *IEEE/ACM Trans Comput Biol Bioinf* 2019; 16: 139–153.
15. Xie Z, Jin J, Liu D, et al. Constructing a predictive model for the death risk of patients with septic shock based on supervised machine learning algorithms. *Zhonghua wei Zhong Bing ji jiu yi xue* 2024; 36: 345–352.
16. Liu Z, Meng Z, Li Y, et al. Prognostic accuracy of the serum lactate level, the SOFA score and the qSOFA score for mortality among adults with sepsis. *Scand J Trauma Resusc Emerg Med* 2019; 27: 51.
17. Li F, Xin H, Zhang J, et al. Prediction model of in-hospital mortality in intensive care unit patients with heart failure: machine learning-based, retrospective analysis of the MIMIC-III database. *BMJ open* 2021; 11: e044779.