Journal of Research in Science and Engineering (JRSE)
ISSN: 1656-1996 Volume-8, Issue-6, June 2026
DOI: 10.66069/ojspub.16560604

An Interpretable XGBoost-Based Mortality Prediction Model for ICU Heart Failure Patients: Leveraging the eICU Database and SHAP Analysis

Jianghai Lan, Zhen Liao, Xu Zhou

School of Automation and Information Engineering, Sichuan University of Science & Engineering,
YiBing 644000, China

Abstract: ***Artificial intelligence, particularly deep learning, is increasingly deployed across medical domains, driven by big data labeling, enhanced computing power, and cloud storage. While AI promises to address persistent healthcare challenges—including diagnostic errors, treatment inaccuracies, resource waste, workflow inefficiencies, inequities, and limited patient-clinician interaction—its application in critical care remains under rigorous investigation. This study developed an interpretable mortality prediction model for intensive care unit (ICU) patients with heart failure, utilizing the freely accessible eICU Collaborative Research Database (eICU-CRD). The extreme gradient boosting (XGBoost) algorithm was applied, with model interpretability enhanced via the SHapley Additive exPlanations (SHAP) method to identify and quantify key prognostic factors. The proposed framework offers a transparent, data-driven tool for risk stratification and clinical decision support in heart failure management.***

Keywords: Artificial Intelligence, XGBoost, SHAP, mortality prediction, heart failure, ICU, eICU database**.**

1. Introduction

In recent years, artificial intelligence has been widely used to explore early warning predictors of many diseases. Given the inherent power of harnessing machine learning algorithms to capture non-linear relationships, more researchers are advocating the use of new predictive models based on machine learning to support appropriate treatment for patients, rather than traditional disease severity classification systems such as SOFA, APACHE II, or SAPS II. Although a large number of predictive models have shown good performance in studies, evidence of their application in clinical Settings and interpretable risk prediction models that contribute to disease prognosis remain limited. MIT researchers have developed a machine learning model that can divide patients into subgroups based on health status to better predict a patient's risk of dying during an ICU stay. The technique outperformed "global" mortality prediction models and revealed differences in the performance of these models in specific patient subpopulations.

In intensive care units, patients present with a variety of health conditions, and emergency triage relies heavily on clinical judgment. ICU staff perform many physiological tests, such as blood tests and checking vital signs, to determine if a patient is at immediate risk of death without aggressive treatment.

A number of machine learning models have been developed in recent years to help predict patient mortality in the ICU, based on various health factors during their stay. However, these models have performance flaws. One common type of "global" model is trained on a single large patient population. These may be effective on average, but less effective in some patient subgroups. On the other hand, another type of model analyzes different subpopulations, for example, those grouped by similar conditions, patient age, or hospital department, but typically has limited data for training and testing.

In a paper presented at the recent Proceedings of Knowledge Discovery and Data Mining conference, MIT researchers describe a machine learning model that is the best of both worlds: It trains specifically for patient subgroups, but also shares data across all subgroups for better predictions of effects. By doing so, the model can better predict a patient's risk of dying during the first two days in the ICU compared to rigorous global models and other models.

The model first processed physiological data in the electronic health records of previously admitted ICU patients, some of whom died during their hospitalization. In doing so, it learned high predictors of mortality, such as low heart rate, high blood pressure, as well as various lab test results, high glucose levels and white blood cell counts in the first few days, and divided patients into subgroups based on their health status. Given a new patient, the model can look at that patient's physiological data from the first 24 hours and, by analyzing these patient subpopulations, understand the likelihood that the new patient will die within the next 48 hours.

In addition, the researchers found that the model also highlighted differences in the performance of global models in predicting mortality in patient subgroups through subpopulation-specific assessments (testing and validation). This is important information for developing models that can more precisely study specific patients.

2. Related Work

2.1 Clinicians and Artificial Intelligence

Artificial Intelligence (AI) is changing the way we live and work at an alarming rate. The application field of artificial intelligence is very wide, and we will talk about the impact of AI on the medical field today, and from this, we can see a grand trend. On June 30 this year, a technology company in Chengdu held the country's first artificial intelligence (AI) doctors and human doctors consistency evaluation.

In the course of eight hours of consultation, Dr. AI and 10 doctors from the Department of cardiology, gastroenterology, respiratory medicine, endocrinology, nephrology, orthopedics and urology of West China Hospital of Sichuan University jointly consulted more than 100 patients. The accuracy of AI doctors was scored by 7 experts and professors. In the end, the human doctor scored 7.5 points and the AI doctor scored 7.2 points. The AI doctors and the top three attending doctors had a 96% agreement in the score results.

Artificial intelligence is more and more closely related to clinical medicine, and its application scope covers many fields such as medical image analysis, disease diagnosis and prediction, and personalized treatment plan design. Through deep learning algorithms and big data analysis, artificial intelligence can provide faster and more accurate diagnostic results, assist doctors to develop more effective treatment plans, and help understand patients' conditions and predict disease trends through interpretive analysis, providing support for clinical decision-making.

2.2 Machine Learning Predictive Models

Machine learning is an artificial intelligence technique that enables computers to learn from data and make decisions autonomously. Predictive models are an important application of machine learning, which can predict future outcomes based on historical data. Python is a popular programming language, and its ease of use and strong library support make it the language of choice for machine learning and predictive models. Machine learning can be divided into three types: supervised learning, unsupervised learning and reinforcement learning. Supervised learning requires labeling of training data, such as classification and regression problems; Unsupervised learning does not require labels, such as clustering and dimensionality reduction problems; Reinforcement learning requires interaction in the environment to obtain reward or punishment signals.

A predictive model is a special type of machine learning model that aims to predict future outcomes based on historical data. Forecasting models can be divided into two types: time series forecasting and non-time series forecasting. Time series forecasting needs to consider the influence of time order, such as ARIMA and LSTM; Non-time series forecasting does not require consideration of time order, such as random forests and support vector machines.

Machine learning and predictive models in Python rely heavily on the Scikit-learn library. Scikit-learn is an open source machine learning library that provides many commonly used algorithms and tools, as well as a concise API interface.

1. General process of Logistic regression

Collect data: Collect data using any method.

Prepare data: Because distance calculations are required, the data type is numeric. In addition, structured data formats are best.

Analyze data: Use any method to analyze data.

Training algorithms: Most of the time will be spent on training, and the purpose of training is to find the best classification regression coefficient.

Test the algorithm: Once the training step is complete, the classification will be fast.

Using algorithms: First, we need to input some data and convert it into the corresponding structured value; Then, based on the trained regression coefficient, these values can be simple regression calculation to determine which category they belong to; After that, we can do some other analysis on the output category.

2. Advantages of Logistic regression:

The realization is simple, easy to understand and implement;

The calculation cost is not high, the speed is fast;

Storage resources are low. Procedure

logistic regression models directly classify the possibility of classification without assuming that the data meets a certain classification model

logistic regression can handle sparse high-dimensional features (cases with few non-zero values)

Not only can you predict the class of sample, but you can also predict the approximate probability of predicting a certain class

3. Logistic regression points:

Easy to underfit;

The classification accuracy may not be high;

4. Other:

The purpose of Logistic regression is to find the best fitting parameters of a nonlinear function Sigmoid, which can be solved by an optimization algorithm. Improved some optimization algorithms, such as sag. It can complete parameter updates as soon as new data arrives, without the need to re-read the entire data set for batch processing. An important problem in machine learning is how to deal with missing data. There is no standard answer to this question, depending on the requirements in the actual application. Several solutions exist, each with its own advantages and disadvantages. We need to look at the data, which is a parameter of Sklearn, in order to achieve a better classification effect.

Plot the logarithmic likelihood function:

$$\ln L(w,b) = \sum^{M} \{y_i \cdot \ln[P(y_i = 1 | x_i)] + (1 - y_i) \ln[1 - P(y_i = 1 | x_i)]\} \tag{1}$$

2.3 Prediction Model Function

Predictive models are generally presented in mathematical form to accurately characterize and express quantitative relationships between the values of input and output variables. It can be subdivided into regression prediction model and classification prediction model, which are suitable for regression problem and classification problem respectively

(1) Regression prediction model:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p + \varepsilon \quad (2)$$

(2) Classification prediction model:

$$log(\frac{P}{1-P}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p + \varepsilon \quad (3)$$

(3) Regression equation for prediction:

$$\hat{y} = \hat{\beta_0} + \hat{\beta_1} X_1 + \hat{\beta_2} X_2 + ... + \hat{\beta_p} X_p \quad (4)$$

$$Logit\hat{P} = log(\frac{\hat{P}}{1-\hat{P}}) = \hat{\beta_0} + \hat{\beta_1} X_1 + \hat{\beta_2} X_2 + ... + \hat{\beta_p} X_p \quad (5)$$

(4) Other forms of prediction model:

Inference rule: The value rule between input variable and output variable is reflected in the form of logical judgment

Inference rule sets and graphical representation:

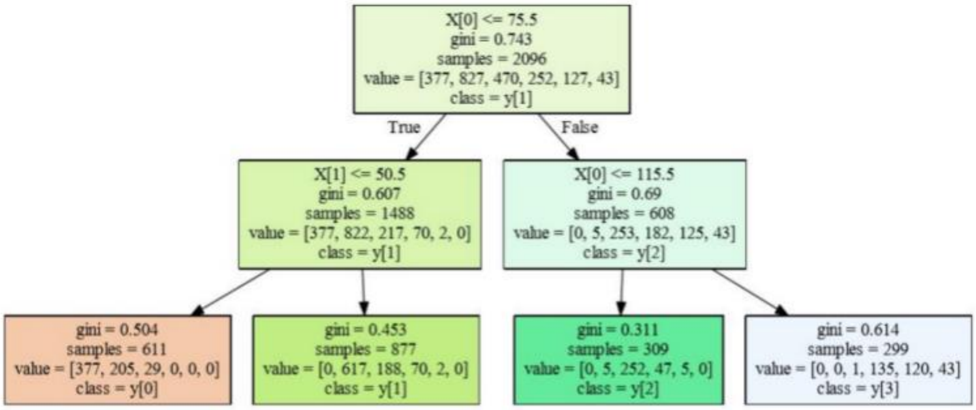

Figure 1: Machine learning predictive model architecture diagram

3. Methodology

3.1 Data Source and Object

Ecicu-crd (version 2.0) is a publicly available multicentre database containing de-identification data related to more than 200,000 ICU admissions from 208 hospitals in the United States during 2014-2015. It details all patients, demographic data, vital sign measurements, diagnostic information, and treatment information.

Therefore, all patients in the IECU-CRD database were considered. Inclusion criteria were as follows :(1) Patients were coded according to the ninth and tenth revisions of the International Classification of Diseases; (2) When admitted to the ICU within 24 hours, the diagnostic priority label is "major"; (3) ICU stay longer than 1 day; (4) The patient is 18 years old or above. Patients with more than 30% missing values were excluded.

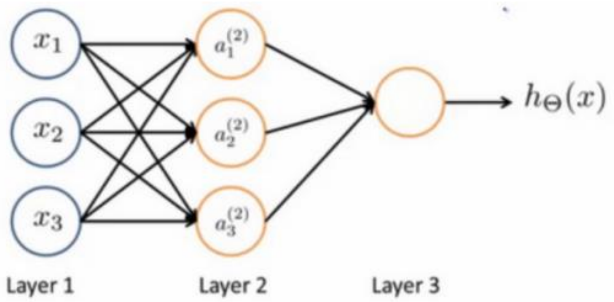

Figure 2: Network structure diagram

3.2 Experimental Process

Data preparation: CT scan images and corresponding labels are obtained using a publicly available medical image dataset, such as LUNA16, where labels indicate the location and size of nodules.

Data preprocessing: CT images are preprocessed, including grayscale, normalization and cropping, to improve the effect and speed of network training.

3.3 Predictive Variables and Data Processing

The predicted outcome of the study was the probability of in-hospital mortality, defined as the patient's condition at the time of discharge. Based on previous studies and expert opinion, demographics, comorbidities, vital signs, and laboratory test results, the following tables for the IECU-CRD were used: "Diagnosis," "intake," "laboratory," "patient," and "care" In addition to demographic characteristics, other variables were collected within the first 24 hours of each ICU admission. In addition, to avoid overfitting, the minimum absolute contraction and selection operator (LASSO) is used to select and filter the variables.

Data missing variables are common in IECU-CRD. However, an analysis that ignores missing data can produce biased results. Therefore, we use multiple interpolation for missing data. All selected variables contain < 30% missing values.

Categorical variables were expressed as totals and percentages, and $\chi 2$ test or Fisher exact test (expected frequency < 10) were used to compare differences between groups. Continuous variables are expressed as the median and

IQR, and Wilcoxon rank sum tests are used when comparing the two groups.

Four machine learning models - XGBoost, logistic Regression (LR), Random Forest (RF), and support vector Machine (SVM) - are used to develop predictive models. The predictive performance of each model was assessed by the area under the subject operating characteristic curve. In addition, we calculated scores for accuracy, sensitivity, positive predictive value, negative predictive value, and F1 prediction specificity fixed at 85%. In addition, in order to evaluate the utility of the decision model by quantifying the net benefits under different threshold probabilities, decision curve analysis (DCA) is performed.

3.4 Experimental Result

Of the 17,029 HF patients with eICU CRD, a total of 2,798 adult patients diagnosed with primary HF were included in the final cohort of this study. The patient screening process is shown in Figure 1. The dataset was randomly divided into two parts: 70% (n=1958) of the data was used for model training and 30% (n=840) of the data was used for model validation. The LASSO regularization process produced 24 potential predictors based on a training dataset of 1,958 patients for model development. Patients in the non-survivor group were older than those in the survivor group (P<.001). The hospital mortality rate was 9.96% (195/1958) in the training dataset and 10% (84/840) in the test dataset. Table 1 shows the comparison of predictors between survivors and non-survivors during hospitalization.

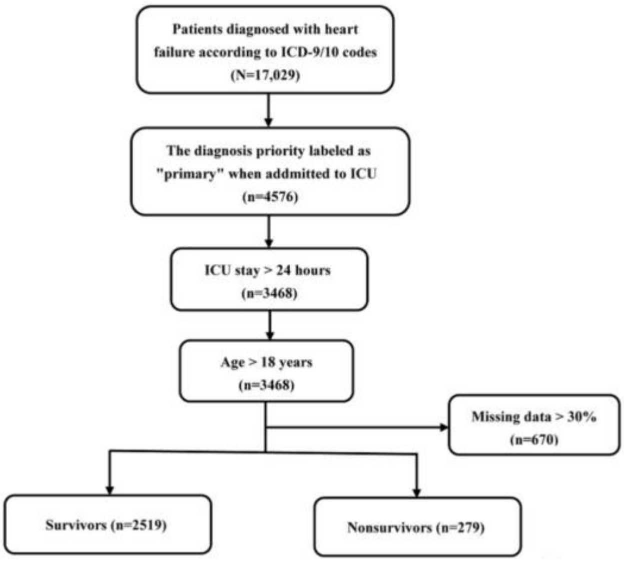

Figure 3: Flowchart of patient selection, ICD: International Classification of Diseases
ICU: intensive care unit.

In the training data set, XGBoost, LR, RF and SVM models were established, and the test data set obtained an AUC of 0.824, 0.800, 0.779 and 0.701 respectively (Table 2 and Figure 2). In comparison, XGBoost had the highest predictive performance among the four models (AUC=0.824, 95%CI 0.7766-0.8708), while SVM had the worst generalization ability (AUC=0.701, 95%CI 0.6433-0.7582). DCA was performed on four machine learning models in the test dataset to compare the net benefit of the best model and clinical decision alternatives. Clinical net benefit is defined as the minimum probability of the disease occurring when further intervention is required. The graph measures the net benefit for different threshold probabilities. The orange line in Figure 3 indicates the assumption that all patients received the intervention, while the yellow line indicates that no patients received the intervention. Due to the heterogeneity of the study population, the treatment strategy provided by any of the four machine learning-based models was superior to the default strategy of treating all patients or none. The net benefit of the XGBoost model exceeds that of other machine learning models at a threshold probability of 10% to 28% (Figure 2).

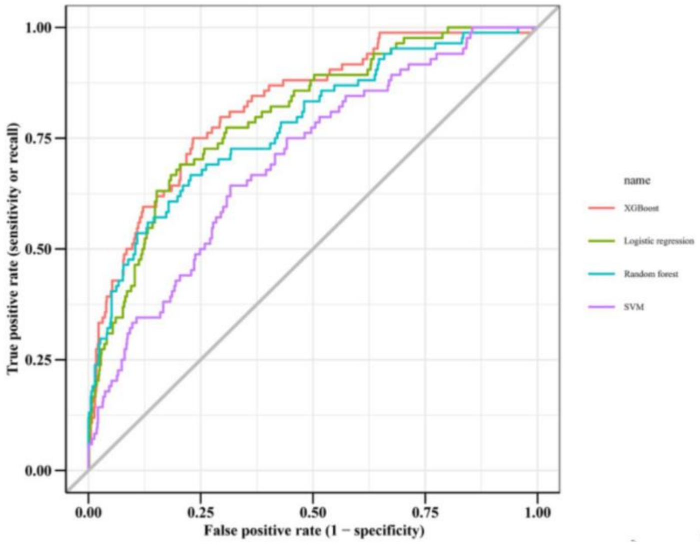

Figure 4: Variables chart

The SHAP algorithm is used to obtain the importance of each predictor to the predicted results of the XGBoost model. The importance of variables chart lists the most important variables in descending order (Figrue 4). Mean blood urea nitrogen (BUN) has the strongest predictive value for all predicted levels, followed by age factors, mean noninvasive systolic blood pressure, urine volume, and maximum respiratory rate. In addition, to detect positive and negative relationships between predicted values and target outcomes, SHAP values were applied to reveal mortality risk factors. As shown in Figure 5, the horizontal position shows whether the effect of the value is associated with higher or lower predictions, and the color shows whether the variable is high (red) or low (blue) for the observation; We can see that an increase in average BUN has a positive effect and pushes the prediction toward mortality, while an increase in urine volume has a negative effect and pushes the prediction toward survival.

4. Conclusion

In conclusion, the study has some limitations. First, our data was extracted from a publicly available database, and some variables were missing. For example, we intend to include more predictors that may influence in-hospital mortality, such as prothrombin time and brain natriuretic peptide and lactate; However, the missing value exceeds 70%. Second, all the data came from ICU patients in the United States, so the applicability of our model to other populations is unclear. Third, our mortality prediction model is based on available data within 24 hours of each ICU admission, which may ignore subsequent events that alter prognosis and contribute to confounding factors to some extent. Fourth, the applicability of the developed XGBoost model in clinical practice may not be very effective due to the lack of an external validation cohort. We are currently collecting data from patients with heart failure in the intensive care unit of West China Hospital of Sichuan University. Although we have obtained some preliminary data, external prospective validation is not feasible due to the limited sample size.

The study used machine learning techniques to make significant advances in the field of medical prediction, successfully predicting risk mortality in ICU patients with central force failure by developing interpretable prediction models. This achievement is not only expected to improve the accuracy and efficiency of clinical decision-making, but also provides strong support for the widespread application of artificial intelligence in the medical field. In the future, with the continuous progress of technology and the accumulation of data, artificial intelligence is expected to play a more important role in clinical medicine, providing better medical services and treatment programs for doctors and patients.

Acknowledgement

I would like to sincerely thank Professor Huiying Weng for his outstanding contributions and profound insights in the field of machine learning. His literature and core theories provide valuable theoretical support and inspiration for the research of this paper. In particular, I would like to thank him for his article "The Importance of AI Algorithm Combined with Tunable LCST Smart Polymers in Biomedical "Applications" provides important reference materials for this research, and provides key guidance for the theoretical framework and practical application of my paper. In addition, I would like to thank Professor Huiying Weng for his guidance and support in academic exchange and research cooperation. Thank you for your valuable advice and encouragement, which have played an immeasurable role in promoting my research work. Finally, I would like to express my sincere thanks and high respect to Professor Huiying Weng again, and look forward to more opportunities for academic exchanges and cooperation in the future.

References

[1] Akbar, A., Peoples, N., Xie, H., Sergot, P., Hussein, H., Peacock IV, W. F., & Rafique, Z. . (2022). Thrombolytic Administration for Acute Ischemic Stroke: What Processes can be Optimized?. McGill Journal of Medicine, 20(2).

[2] Chen, Wangmei, et al. “Applying Machine Learning Algorithm to Optimize Personalized Education Recommendation System”. Journal of Theory and Practice of Engineering Science, vol. 4, no. 01, Feb. 2024, pp. 101-8, doi:10.53469/jtpes.2024.04(01).14.

[3] “Implementation of Computer Vision Technology Based on Artificial Intelligence for Medical Image Analysis”. International Journal of Computer Science and Information Technology, vol. 1, no. 1, Dec. 2023, pp. 69-76, https://doi.org/10.62051/ijcsit.v1n1.10.

[4] Dong, Xinqi, et al. “The Prediction Trend of Enterprise Financial Risk Based on Machine Learning ARIMA Model”. Journal of Theory and Practice of Engineering Science, vol. 4, no. 01, Jan. 2024, pp. 65-71, doi:10.53469/jtpes.2024.04(01).09.

[5] “A Deep Learning-Based Algorithm for Crop Disease Identification Positioning Using Computer Vision”. International Journal of Computer Science and Information Technology, vol. 1, no. 1, Dec. 2023, pp. 85-92, https://doi.org/10.62051/ijcsit.v1n1.12.

[6] K. Jin, Z. Z. Zhong and E. Y. Zhao, "Sustainable Digital Marketing Under Big Data: An AI Random Forest Model Approach," in IEEE Transactions on Engineering Management, vol. 71, pp. 3566-3579, 2024, doi: 10.1109/TEM.2023.3348991.

[7] Jin, Keyan. "Impacts of Word of Mouth (WOM) on E-Business Online Pricing." JGIM vol.31, no.3 2023: pp.1-17. http://doi.org/10.4018/JGIM.324813

[8] Du, S., Li, L., Wang, Y., Liu, Y., & Pan, Y. (2023). Application of HPV-16 in Liquid-Based thin Layer Cytology of Host Genetic Lesions Based on AI Diagnostic Technology Presentation of Liquid. Journal of Theory and Practice of Engineering Science, 3(12), 1-6.

[9] Xin, Q., He, Y., Pan, Y., Wang, Y., & Du, S. (2023). The implementation of an AI-driven advertising push system based on a NLP algorithm. International Journal of Computer Science and Information Technology, 1(1), 30-37.

[10] He, Yuhang, et al. “Intelligent Fault Analysis With AIOps Technology”. Journal of Theory and Practice of Engineering Science, vol. 4, no. 01, Feb. 2024, pp. 94-100, doi:10.53469/jtpes.2024.04(01).13.

[11] Chen , J., Xiong, J., Wang, Y., Xin, Q., & Zhou, H. (2024). Implementation of an AI-based MRD Evaluation and Prediction Model for Multiple Myeloma. Frontiers in Computing and Intelligent Systems, 6(3), 127-131. https://doi.org/10.54097/zJ4MnbWW

[12] Tan, Kai, et al. “Integrating Advanced Computer Vision and AI Algorithms for Autonomous Driving Systems”. Journal of Theory and Practice of Engineering Science, vol. 4, no. 01, Jan. 2024, pp. 41-48, doi:10.53469/jtpes.2024.04(01).06.

[13] “Exploring New Frontiers of Deep Learning in Legal Practice: A Case Study of Large Language Models”. International Journal of Computer Science and Information Technology, vol. 1, no. 1, Dec. 2023, pp. 131-8, https://doi.org/10.62051/ijcsit.v1n1.18.