**International Journal of Informatics and Communication Technology (IJ-ICT)**
Vol. 14, No. 1, April 2025, pp. 11~19
ISSN: 2252-8776, DOI: 10.11591/ijict.v14i1.pp11-19

# Enhancing predictive modelling and interpretability in heart failure prediction: a SHAP-based analysis

**Niaz Ashraf Khan[1], Md. Ferdous Bin Hafiz[2], Md. Aktaruzzaman Pramanik[1]**
[1]Department of Computer Science and Engineering, University of Liberal Arts Bangladesh, Dhaka, Bangladesh
[2]Department of Computer Science and Engineering, Southeast University, Dhaka, Bangladesh

**Article Info**

***Article history:***

Received Mar 24, 2024
Revised Aug 14, 2024
Accepted Sep 22, 2024

***Keywords:***

CVD
Ensemble models
Heart failure
Interpretability
SHAP analysis

## ABSTRACT

Predictive modelling plays a crucial role in healthcare, particularly in forecasting mortality due to heart failure. This study focuses on enhancing predictive modelling and interpretability in heart failure prediction through advanced boosting algorithms, ensemble methods, and SHapley Additive exPlanations (SHAP) analysis. Leveraging a dataset of patients diagnosed with cardiovascular diseases (CVD), we employed techniques such as synthetic minority over-sampling technique (SMOTE) and bootstrapping to address class imbalance. Our results demonstrated exceptional predictive performance, with the gradient boosting (GBoost) model achieving the highest accuracy of 91.39%. Ensemble techniques further enhanced performance, with the voting classifier (VC), stacking classifier (SC), and Blending achieving accuracies of 91.00%. SHAP analysis uncovered key features such as time, Serum_creatinine, and Ejection_fraction, significantly impacting mortality prediction. These findings highlight the importance of transparent and interpretable machine learning models in healthcare decision-making processes, facilitating informed interventions and personalized treatment strategies for heart failure patients.

***Corresponding Author:***

Niaz Ashraf Khan
Department of Computer Science and Engineering, University of Liberal Arts Bangladesh
Dhaka, Bangladesh
Email: niaz.ashraf@ulab.edu.bd

## 1. INTRODUCTION

Cardiovascular disease (CVD) stands as the foremost cause of mortality globally, presenting a significant challenge in public health worldwide [1]. In the realm of cardiovascular research, machine learning algorithms have emerged as valuable tools for prediction, offering promising avenues for understanding and addressing CVD-related issues. With projections indicating that CVD will account for approximately 23 million deaths by 2030 [2], the urgency to develop effective predictive models has intensified. Consequently, the integration of artificial intelligence (AI) and extensive datasets in CVD prediction models is increasingly prevalent [3], [4]. To improve mortality forecasting in CVD cases, predictive modeling is crucial in healthcare. However, for these models to be effective, they must also be interpretable and transparent, fostering trust among healthcare professionals and patients. This study aims to enhance heart failure prediction by employing advanced boosting algorithms, ensemble methods, and SHapley Additive exPlanations (SHAP) analysis for better interpretability. Techniques like XGBoost (XGB) and gradient boosting (GBoost), along with ensemble methods such as voting and blending, are utilized to enhance predictive accuracy and gain deeper insights into the factors influencing heart failure mortality. Previous studies have extensively applied machine learning techniques to predict survival in heart failure patients, focusing on identifying critical risk factors. The healthcare sector has seen notable advancements in machine learning within cardiology, underscoring the increasing adoption and effectiveness of these methods. Alotaibi [5], machine learning techniques for heart failure prediction using data from the cleveland clinic foundation were explored. The decision tree algorithm achieved the highest accuracy at 93.19%, followed by support vector machine (SVM) at 92.30%. Weng *et al.* [6] conducted tests on four models using clinical data from over 300,000 homes in the UK, with NN demonstrating the most accurate predictions for CVD on a larger dataset. Dimopoulos *et al.* [7] evaluated k-nearest neighbor (KNN), random forest (RF), and decision tree (DT) models using the ATTICA dataset, showing RF's superiority in combination with the HellenicSCORE tool. Mohan *et al.* [8] proposed a hybrid HRFLM strategy to improve prediction accuracy in IoT applications using machine learning. Yang *et al.* [9] employed logistic regression (LR) to assess cardiovascular risk factors in eastern China. Mamun and Elfouly [10] propose a hybrid 1D CNN using a large online dataset and feature selection algorithms, achieving 80.1% accuracy for non-coronary heart disease (no-CHD) and 76.9% for CHD. Muniasamy *et al.* [11] introduces a deep convolutional neural networks (CNN) model for CHD classification with feature selection via LASSO, achieving 99.36% accuracy. Arabasadi *et al.* [12] combines genetic algorithms and NNs to enhance diagnostic accuracy. Revathi *et al.* [13] presents the OCI-LSTM model, using the salp swarm algorithm and genetic algorithm for early diagnosis of CVD. Hossen *et al.* [14] applies supervised learning, particularly LR, to predict heart disease using the UCI Cleveland database. Bharti *et al.* [15] explores machine learning and deep learning algorithms, achieving high accuracy and integrating these methods with multimedia technology. Alqahtani *et al.* [16] emphasizes the importance of early detection of CVD symptoms and timely intervention. Phasinam *et al.* [17] introduces a machine learning framework for predicting heart disease likelihood, with RF showing the highest accuracy. Dharmendra and Saravanan [18] introduces a categorization algorithm for heart attack prediction, finding SVM superior to DT. Kumar and Rekha [19] proposes a method using deep neural networks for CVD risk prediction on well-defined datasets. Elminaam *et al.* [20] highlights the effectiveness of machine learning, particularly LR and RF, in improving heart disease prediction accuracy. El-Hasnony *et al.* [21] advocate for ML-based systems in heart disease prediction and diagnosis, emphasizing active learning with user-expert feedback. Guleria *et al.* [22] explores ensemble models in the explainable artificial intelligence (XAI) approach from the CVD datasets, employing multiple machine learning models.

In contrast to prior studies, our research addresses the need for enhanced predictive modeling and interpretability in heart failure prediction. and contributes to the field of heart failure prediction by:

- Integration of advanced boosting algorithms, ensemble methods, and SHAP analysis for heart failure prediction.
- Utilization of SHAP analysis to provide insights into mortality prediction factors.
- Comprehensive approach combining global and local SHAP analysis for a holistic understanding of model predictions.
- Achievement of superior predictive performance compared to previous studies.

The technical contributions of our study are outlined in detail in the subsequent chapters. In section 2, we provide an overview of the dataset used in our investigation and the methodology utilized in developing predictive models. Furthermore, in section 3, we present the results of our study, including model performance metrics and insights gained from SHAP analysis. Finally, in section 4, we conclude our findings and discuss their implications for healthcare decision-making processes.

## 2. RESEARCH METHOD

In this study, we analyzed a comprehensive dataset comprising 299 patients diagnosed with CVD. This dataset included essential attributes such as anaemia status, high blood pressure, creatinine phosphokinase levels, and diabetes status, among others, totaling 13 features. These features were meticulously chosen based on their clinical relevance and potential impact on predicting mortality due to heart failure. Additionally, all 299 patients have previous experiences with heart failures categorized into classes III or IV of the New York Heart Association (NYHA) classification system for heart failure stages.

To tackle the inherent class imbalance in the dataset, we utilized two techniques: synthetic minority over-sampling technique (SMOTE) and bootstrapping. SMOTE was employed to generate synthetic samples for the minority class (death event), aiming to balance the class distribution and mitigate potential model bias towards the majority class. Furthermore, bootstrapping was implemented to augment the dataset size. Figure 1 depicts the class distribution within the dataset. Moreover, we balanced the distribution of the target label (death event) to approximately 400 instances for each dataset. Moreover, exploratory data analysis was conducted to gain insights into the dataset's characteristics, including the distribution of numerical and categorical features. Figures 2 illustrate the correlation matrix which shows the relationships between two variables of the dataset.

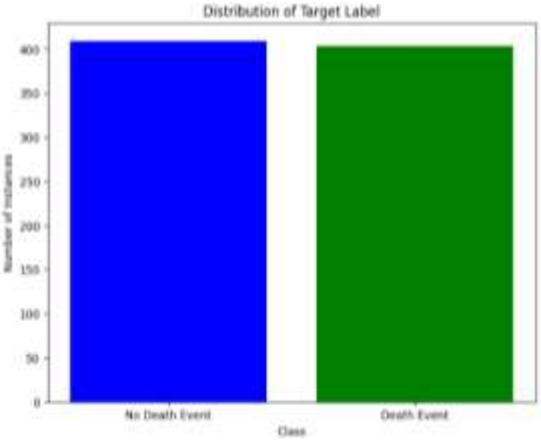


Figure 1. Distribution of target label

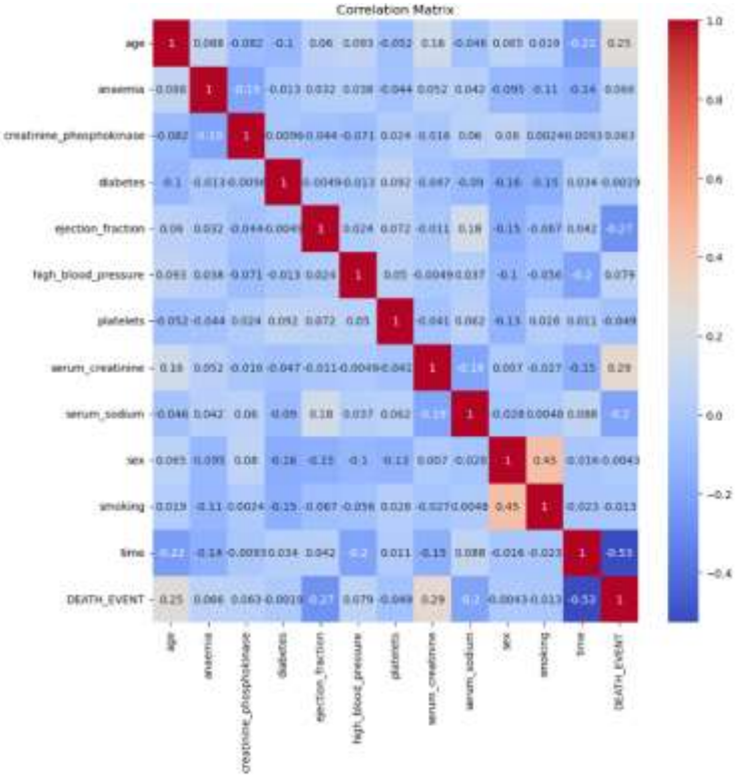


Figure 2. Correlation matrix

### 2.1. Data preprocessing and feature engineering

Our methodology began with meticulous data preprocessing to ensure the dataset's suitability for predictive modelling. Given the prevalence of class imbalance inherent in healthcare datasets, we employed SMOTE and bootstrapping. SMOTE effectively resolved the class disparity issue by generating artificial samples for the underrepresented class. (death event), thereby balancing the distribution of the target label. Concurrently, bootstrapping augmented the dataset's size through resampling with replacement from the original data, enhancing the model's robustness against overfitting.

### 2.2. Model evaluation and selection

Following data preprocessing, we proceeded with model evaluation and selection to identify the most suitable algorithms for predicting death events in heart disease patients. Leveraging a diverse set of machine learning models, including GB, RF, and XGB, we conducted rigorous training and testing procedures on the dataset. A concise explanation of each model is given below.

### 2.2.1. XGBoost

XGB is a sophisticated implementation of gradient enhancement algorithms designed for efficiency. It’s known for its speed and accuracy in handling large datasets. The objective functions’ crucial characteristic lies in their composition of two elements: the training loss and the regularization term.

$$O(\theta) = L(\theta) + \Omega(\theta) \tag{1}$$

In (1), L is the training loss function and Ω is the regularization term.

### 2.2.2. Random forest

RF is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes as the prediction of individual trees. The selection of the root node is based on using the Gini Index, which varies between 0 (indicating perfect purity) and 1 (reflecting higher inequality). The Gini Index calculation can be expressed as (2).

$$\text{Gini Index}=1-[(P)^2+(N)^2] \tag{2}$$

Here in (2), P is the probability of a positive class, and N is the probability of a negative class.

### 2.2.3. GBoost

GBoost is an advanced ensemble learning technique that builds a strong predictive model by sequentially adding weak learners, typically decision trees. In this method, each subsequent model is trained to correct the errors of the previous ones, minimizing a specified loss function. This process results in a robust model that can effectively handle complex patterns in the data, improving the accuracy of predictions.

### 2.2.4. Voting and stacking classifier

VC is an ensemble technique that combines predictions from multiple individual classifiers to determine the final class label through a majority voting mechanism. This approach leverages the strengths of different algorithms to achieve better overall performance. On the other hand, SC utilizes a meta-learner to combine predictions from multiple base classifiers.

## 2.3. SHAP analysis for interpretability

To enhance the interpretability of our models and gain insights into feature importance, we employed SHAP analysis. SHAP provides a unified framework for explaining the output of machine learning models by assigning each feature an importance value for a particular prediction. This technique facilitates the detection of model biases and aids in understanding the reasoning behind individual predictions, making it a valuable tool for validating and refining our models.

$$\phi_i = \frac{1}{N}\sum_{S\subseteq\{1,\dots,M\}\setminus\{i\}} \frac{|S|!(M-|S|-1)!}{M!}[f(S\cup\{i\})-f(S)] \tag{3}$$

Here, in (3),
- $\phi_i$ represents the Shapley value for feature *i*.
- *N* is the number of possible permutations of features.
- *M* is the total number of features.
- *S* is the subset of features excluding feature i.
- $f(S\cup\{i\})$ is the model’s output when including feature *i* in subset *S*.
- *f(S)* is the model’s output without including feature *i* in subset *S*.

## 3. RESULTS AND DISCUSSION

Our assessment included the examination of various evaluation metrics. Here, GBoost stood out as the top performer, achieving an impressive accuracy of 91.39%. The precision and recall scores for both classes for GB, along with the other two models can be found in the accompanying confusion matrix in Figure 3. Figure 3(a) contains the confusion matrix of GBoost while Figure 3(b) contains the confusion matrix of RF and Figure 3(c) contains the confusion matrix of XGB. We further investigated ensemble learning techniques, including the VC and SC, to enhance performance. The VC, which amalgamated predictions from XGB and RF models, attained an accuracy of 91%, while the SC yielded an accuracy of

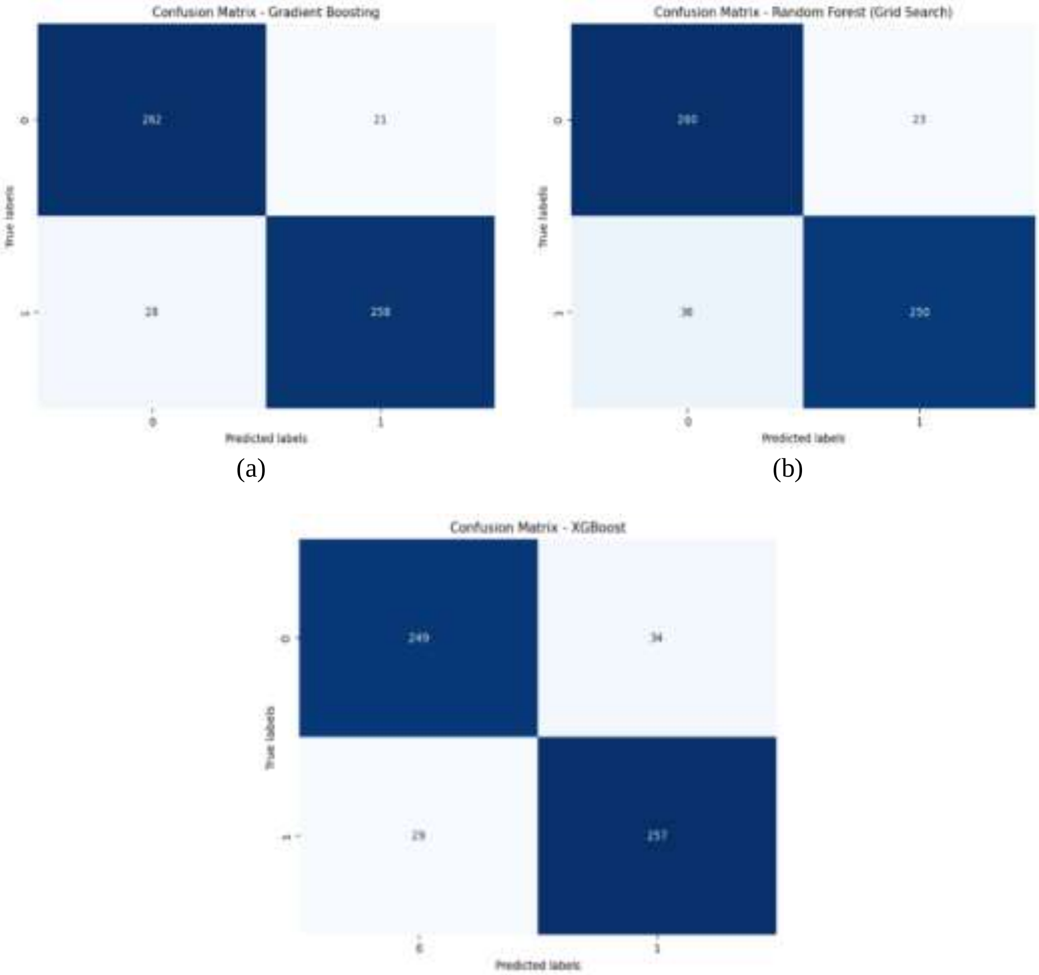
92%. Additionally, combining predictions from XGB and RF resulted in a robust accuracy of 91%. Table 1 presents the performance summary of individual models (GBoost, RF, and XGBoost) in terms of accuracy, precision, and recall, while Table 2 summarizes the performance metrics of VC, SC, and BC in predicting critical health outcomes for CVD patients, confusion matrix for the VC and BC is provided in Figure 4. Figure 4(a) contains the confusion matrix of VC while Figure 4(b) contains the confusion matrix of BC. We also compared with previous work, where our model provides superior performance, as shown in Table 3.

Table 1. Performance summary of the ensemble techniques

| Model | Accuracy (%) | Class 0 | | Class 1 | |
|---|---|---|---|---|---|
| | | Precision | Recall | Precision | Recall |
| GB | 91.39 | 0.90 | 0.93 | 0.92 | 0.90 |
| RF | 90.33 | 0.88 | 0.92 | 0.92 | 0.87 |
| XGB | 88.93 | 0.90 | 0.88 | 0.88 | 0.90 |

Table 2. Performance summary metrics of voting, stacking and bleding classifiers

| Model | Accuracy (%) | Class 0 | | Class 1 | |
|---|---|---|---|---|---|
| | | Precision | Recall | Precision | Recall |
| VC | 91 | 0.87 | 0.95 | 0.95 | 0.86 |
| SC | 92 | 0.90 | 0.94 | 0.94 | 0.89 |
| BC | 91 | 0.87 | 0.95 | 0.95 | 0.86 |



(a)

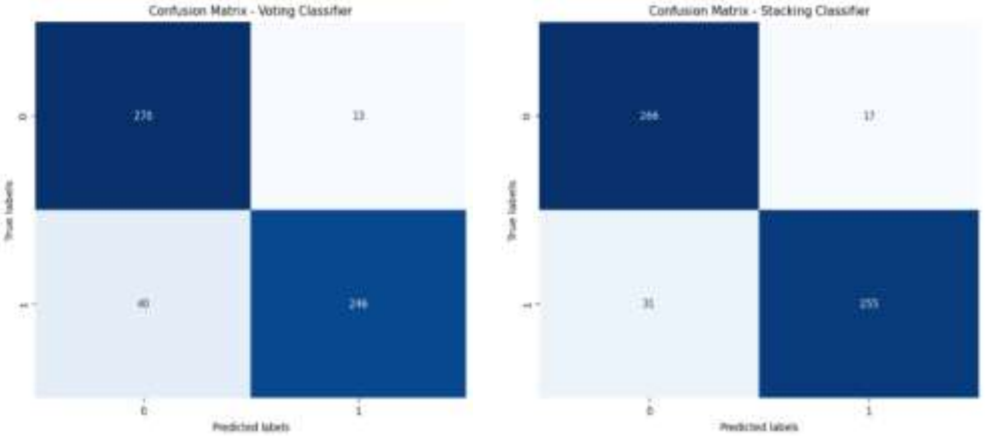
(b)

(c)

Figure 3. Confusion matrix of; (a) GBoost, (b) random forest classifier, and (c) XGBoost

Table 3. Comparative analysis with previous works

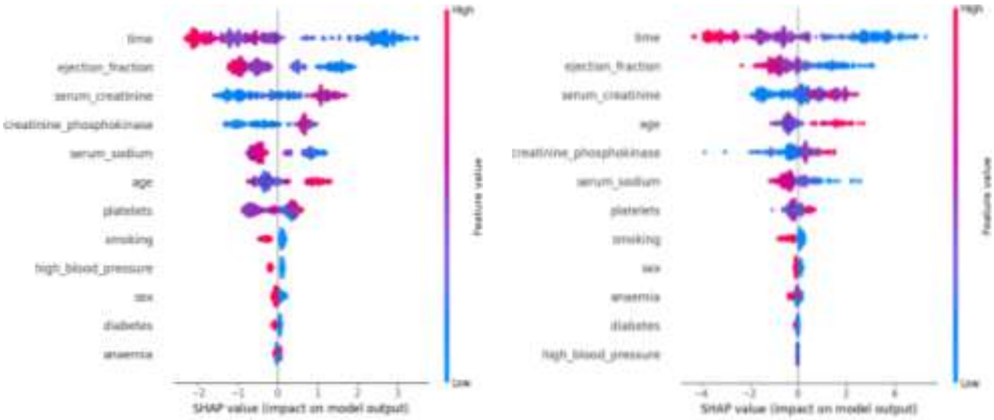
| Paper | Model | Accuracy |
|---|---|---|
| [23] | LR | 83.3% |
| [24] | DT+MRMR+RFE | 80.00% |
| [25] | BRF+CH2 | 76.25% |
| Our work | GB | 91.39% |



(a) (b)

Figure 4. Confusion matrix of (a) voting classifier and (b) stacking classifier

Figure 5 corresponds to the summary plot for the Figure 5(a) XGBoost and Figure 5(b) GBoost model respectively. Figure 6 also provides the SHAP analysis for RF as well. These plots visualize the global feature importance, showing the impact of each feature on model predictions across the dataset. Analyzing the SHAP chart provides three key insights: firstly, features at the top exert the most significant impact on predictions, while those at the bottom have lesser influence. Secondly, the position of each feature bar reflects its range of impact on predictions. Lastly, each dot represents a data point, with the density around a region indicating the contraction of feature values. Among the analyzed features, Time emerges as the most crucial, followed by Serum_creatinine, with high_blood_pressure exhibiting the least influence. Notably, features like time, serum_creatinine, ejection_fraction, platelets, and age demonstrate a right-tailed distribution, indicating higher values significantly impact predictions. Certain features, including time, serum_creatinine, ejection_fraction, and diabetes, exhibit high density within specific value ranges, suggesting concentrated data points. Local SHAP analysis, as depicted in Figure 7, underscores the importance of exploring individual cases to grasp divergences from broader trends identified through global importance analysis. Local SHAP analysis on two random data points with three features providing most impact is shown in Figure 7(a) with Figure 7(b) showing with two features providing the most impact.



(a) (b)

Figure 5. SHAP analysis on (a) XGBoost and (b) GBoost algorithm

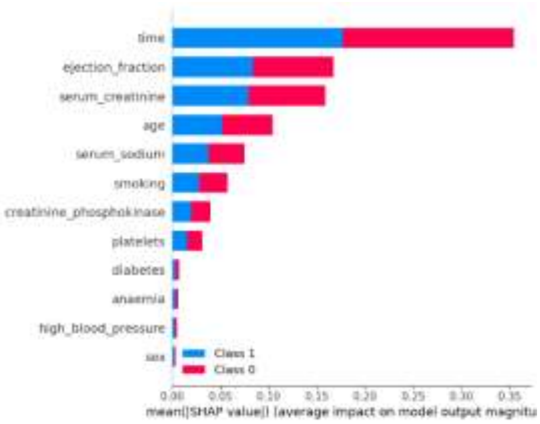
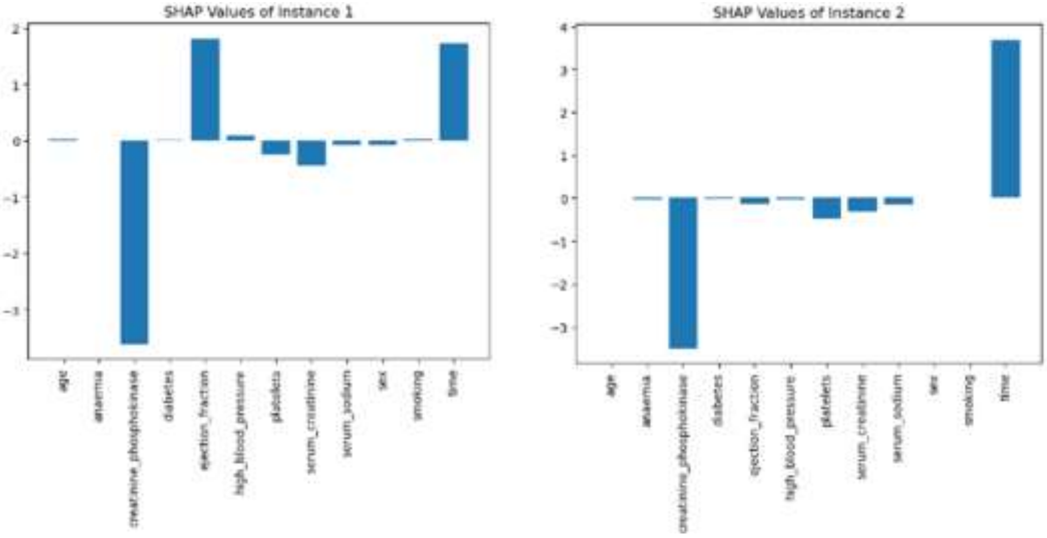


Figure 6. SHAP analysis on random forest classifier



(a) (b)

Figure 7. Local SHAP analysis on two random data points (a) with three features providing most impact and (b) with two features providing most impact

## 4. CONCLUSION

To conclude, this study showcased the effectiveness of advanced boosting algorithms, ensemble methods, and SHAP analysis in improving predictive modelling and interpretability for heart failure prediction. Results demonstrated high predictive performance, with the GBoost model achieving an accuracy of 91.39%. Ensemble techniques further bolstered performance, with accuracies reaching 91.00%. Furthermore, we conducted both local and global SHAP analyses to gain insights into feature importance and individual predictions. The global SHAP analysis provided a comprehensive understanding of the overall impact of features on model predictions, highlighting key factors such as time, Serum_creatinine, and ejection_fraction. On the other hand, local SHAP analysis underscored the importance of exploring individual cases to grasp deviations from broader trends identified through global SHAP analysis. These findings underscore the significance of transparent and interpretable machine learning models in healthcare decision-making processes. By integrating SHAP analysis, we not only achieved exceptional predictive performance but also gained valuable insights into mortality prediction factors. Such insights are pivotal for informed decision-making in clinical settings, facilitating personalized treatment strategies and interventions for CVD patients. Future research could explore additional features and larger datasets to enhance model performance further. Investigating temporal dynamics and integrating patient-specific data may improve predictive accuracy and personalized treatment strategies. Prospective studies validating model performance in clinical settings are essential. Continuous refinement of predictive models and interpretability techniques remains crucial for advancing heart failure management and patient care.

## REFERENCES

[1] S. Aryal, A. Alimadadi, I. Manandhar, B. Joe, and X. Cheng, "Machine learning strategy for gut microbiome-based diagnostic screening of cardiovascular disease," *Hypertension*, vol. 76, no. 5, pp. 1555–1562, Nov. 2020, doi: 10.1161/HYPERTENSIONAHA.120.15885.

[2] M. Juhola, H. Joutsijoki, K. Penttinen, and K. Aalto-Setälä, "Detection of genetic cardiac diseases by Ca2+ transient profiles using machine learning methods," *Scientific Reports*, vol. 8, no. 1, p. 9355, Jun. 2018, doi: 10.1038/s41598-018-27695-5.

[3] B. H. M. van der Velden, H. J. Kuijf, K. G. A. Gilhuijs, and M. A. Viergever, "Explainable artificial intelligence (XAI) in deep learning-based medical image analysis," *Medical Image Analysis*, vol. 79, p. 102470, Jul. 2022, doi: 10.1016/j.media.2022.102470.

[4] J. C. Wolfe, L. A. Mikheeva, H. Hagras, and N. R. Zabet, "An explainable artificial intelligence approach for decoding the enhancer histone modifications code and identification of novel enhancers in Drosophila," *Genome Biology*, vol. 22, no. 1, p. 308, Dec. 2021, doi: 10.1186/s13059-021-02532-7.

[5] F. S. Alotaibi, "Implementation of machine learning model to predict heart failure disease," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 6, pp. 261–268, 2019, doi: 10.14569/ijacsa.2019.0100637.

[6] S. F. Weng, J. Reps, J. Kai, J. M. Garibaldi, and N. Qureshi, "Can machine-learning improve cardiovascular risk prediction using routine clinical data?," *PLoS ONE*, vol. 12, no. 4, p. e0174944, Apr. 2017, doi: 10.1371/journal.pone.0174944.

[7] A. C. Dimopoulos *et al.*, "Machine learning methodologies versus cardiovascular risk scores, in predicting disease risk," *BMC Medical Research Methodology*, vol. 18, no. 1, p. 179, Dec. 2018, doi: 10.1186/s12874-018-0644-1.

[8] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," *IEEE Access*, vol. 7, pp. 81542–81554, 2019, doi: 10.1109/ACCESS.2019.2923707.

[9] L. Yang *et al.*, "Study of cardiovascular disease prediction model based on random forest in eastern China," *Scientific Reports*, vol. 10, no. 1, p. 5245, Mar. 2020, doi: 10.1038/s41598-020-62133-5.

[10] M. M. R. K. Mamun and T. Elfouly, "Detection of cardiovascular disease from clinical parameters using a one-dimensional convolutional neural network," *Bioengineering*, vol. 10, no. 7, p. 796, Jul. 2023, doi: 10.3390/bioengineering10070796.

[11] A. Muniasamy, A. Begum, A. Sabahath, H. Yaqub, and G. Karunakaran, "Coronary heart disease classification using deep learning approach with feature selection for improved accuracy," *Technology and Health Care*, vol. 32, no. 3, pp. 1991–2007, 2024, doi: 10.3233/THC-231807.

[12] Z. Arabasadi, R. Alizadehsani, M. Roshanzamir, H. Moosaei, and A. A. Yarifard, "Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm," *Computer Methods and Programs in Biomedicine*, vol. 141, pp. 19–26, Apr. 2017, doi: 10.1016/j.cmpb.2017.01.004.

[13] T. K. Revathi, S. Balasubramaniam, V. Sureshkumar, and S. Dhanasekaran, "An improved long short-term memory algorithm for cardiovascular disease prediction," *Diagnostics*, vol. 14, no. 3, p. 239, Jan. 2024, doi: 10.3390/diagnostics14030239.

[14] M. D. A. Hossen *et al.*, "Supervised machine learning-based cardiovascular disease analysis and prediction," *Mathematical Problems in Engineering*, vol. 2021, pp. 1–10, Dec. 2021, doi: 10.1155/2021/1792201.

[15] R. Bharti, A. Khamparia, M. Shabaz, G. Dhiman, S. Pande, and P. Singh, "Prediction of heart disease using a combination of machine learning and deep learning," *Computational Intelligence and Neuroscience*, vol. 2021, no. 1, Jan. 2021, doi: 10.1155/2021/8387680.

[16] A. Alqahtani, S. Alsubai, M. Sha, L. Vilcekova, and T. Javed, "Cardiovascular disease detection using ensemble learning," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–9, Aug. 2022, doi: 10.1155/2022/5267498.

[17] K. Phasinam, T. Mondal, D. Novaliendry, C. H. Yang, C. Dutta, and M. Shabaz, "Analyzing the performance of machine learning techniques in disease prediction," *Journal of Food Quality*, vol. 2022, pp. 1–9, Mar. 2022, doi: 10.1155/2022/7529472.

[18] D. Dharmendra and M. S. Saravanan, "Prediction of heart failure using support vector machine compared with decision tree algorithm for better accuracy," in *International Conference on Sustainable Computing and Data Communication Systems, ICSCDS 2022 - Proceedings*, Apr. 2022, pp. 1535–1540, doi: 10.1109/ICSCDS53736.2022.9760989.

[19] A. S. Kumar and R. Rekha, "A dense network approach with gaussian optimizer for cardiovascular disease prediction," *New Generation Computing*, vol. 41, no. 4, pp. 859–878, Nov. 2023, doi: 10.1007/s00354-023-00234-1.

[20] D. A. Elminaam, M. Radwan, N. M. Abdelrahman, H. W. Kamal, A. K. A. Elewa, and A. M. Mohamed, "MLHeartDisPrediction: heart disease prediction using machine learning," *Journal of Computing and Communication*, vol. 2, no. 1, pp. 50–65, Jan. 2023, doi: 10.21608/jocc.2023.282098.

[21] I. M. El-Hasnony, O. M. Elzeki, A. Alshehri, and H. Salem, "Multi-label active learning-based machine learning model for heart disease prediction," *Sensors*, vol. 22, no. 3, p. 1184, Feb. 2022, doi: 10.3390/s22031184.

[22] P. Guleria, P. N. Srinivasu, S. Ahmed, N. Almusallam, and F. K. Alarfaj, "XAI framework for cardiovascular disease prediction using classification techniques," *Electronics (Switzerland)*, vol. 11, no. 24, p. 4086, Dec. 2022, doi: 10.3390/electronics11244086.

[23] D. Chicco and G. Jurman, "Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone," *BMC Medical Informatics and Decision Making*, vol. 20, no. 1, p. 16, 2020, doi: 10.1186/s12911-020-1023-5.

[24] M. A-M. Hasan, J. Shin, U. Das, and A. Y. Srizon, "Identifying prognostic features for predicting heart failure by using machine learning algorithm," in *ACM International Conference Proceeding Series*, Mar. 2021, pp. 40–46, doi: 10.1145/3460238.3460245.

[25] A. M. Qadri, M. S. A. Hashmi, A. Raza, S. A. J. Zaidi, and A. ur Rehman, "Heart failure survival prediction using novel transfer learning based probabilistic features," *PeerJ Computer Science*, vol. 10, pp. 1–30, Mar. 2024, doi: 10.7717/peerj-cs.1894.

## BIOGRAPHIES OF AUTHORS

**Niaz Ashraf Khan** is a dedicated academic professional with over 5 years of experience in the field of Computer Science and Engineering. Currently serving as a Lecturer at the Department of Computer Science and Engineering at the University of Liberal Arts Bangladesh, Niaz holds a Master’s degree in Computer Science and Engineering from North South University, Dhaka, Bangladesh, along with a bachelor’s degree from the same institution. his research interests lie in the intersection of sound signal processing, NLP, machine learning, and deep learning. He can be contacted at email: niaz.ashraf@ulab.edu.bd.

**Md. Ferdous Bin Hafiz** is currently working an Assistant Professor in the Department of Computer Science and Engineering at Southeast University. He completed his BSc. in Computer Science and Information Technology from Islamic University of Technology (IUT), Bangladesh, and obtained a Master of Information Technology from Charles Sturt University, Australia. With almost 6 years of experience, he has previously worked at the University of Liberal Arts Bangladesh (ULAB), East Delta University, and Northern University Bangladesh as a full-time lecturer. His research interests include data science, machine learning, and cyber security. He can be contacted at ferdous.binhafiz@seu.edu.bd.

**Md. Aktaruzzaman Pramanik** is currently working as a Lecturer at the Department of Computer Science and Engineering at the University of Liberal Arts Bangladesh (ULAB). He has completed his Bachelors and Masters from Jahangirnagar University (JU). Before joining ULAB, he worked as a Software Engineer at Fronture Technologies Ltd. for 8 months. Previously he worked as an Assistant Programmer at Bangladesh House Building Finance Corporation (BHBFC) for 1 year and 3 months. He also worked as a full-time Lecturer for 1 year and as a contractual Lecturer for 1 year and 3 months at the Department of Computer Science and Engineering, Daffodil International University. He can be contacted at email: aktaruzzaman.pramanik@ulab.edu.bd.