Evaluating the Effectiveness of Feature Selection and Explainable AI in predicting Acute Myocardial Infarction using Machine Learning Models

Yong Wang [1], Jiexuan Shen [1], Shi Li [1], Xinglin Li [1], Grace Ugochi Nneji[1,2,*], and Happy Nkanta Monday[1, 2, *]
[1]Chengdu University of Technology, Chengdu, 610059 China
[2]Intelligent Computing Lab, HACE SOFTTECH, Lagos, 102241 Nigeria
[*]Corresponding author (happy.monday@zy.cdut.edu.cn; grace.nneji@zy.cdut.edu.cn)

Nhồi máu cơ tim cấp tính

2024 5th International Conference on Machine Learning and Computer Application (ICMLCA) | 979-8-3315-3033-4/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICMLCA63499.2024.10753731

***Abstract*-Acute myocardial Infarctions (MIs), commonly known as heart attacks, are critical medical emergencies that often lead to severe complications, including post-myocardial infarction angina (PMIA). Early and accurate prediction of PMIA is crucial for timely intervention and improved patient outcomes. This study focuses on predicting PMIA using machine learning classifiers on the UCI myocardial infarction complication datasets, which includes 1700 patient samples with 111 clinical features. The dataset is balanced using Borderline Synthetic Minority Oversampling (Borderline-SMOTE) and standardized. Twelve classifiers are employed and the Shapley Additive exPlanations (SHAP) are used to interpret the best-performing models, revealing the decision-making processes. The SVM classifier achieved the highest cross-validated accuracy of 98.55%, demonstrating superior predictive capability. SHAP values provided insights into the importance of each feature, enhancing model interpretability and transparency. This approach highlights the significance of feature selection and explainability in building trustworthy AI models for healthcare. The findings contribute to improved diagnosis and management of PMIA, underscoring the ethical adoption of AI in clinical settings.**

Keywords- Myocardial Infarction, PMIA, SHAP, Machine Learning, Borderline-SMOTE, explainability, SHAP value

I. INTRODUCTION

Acute Myocardial Infarctions (MI) [1], commonly known as heart attacks, are serious medical emergencies that frequently occur despite advancements in treatment and prevention measures. Experts have observed that if identified early on, some MI problems might be able to avert deadly consequences. Plaque accumulation in the coronary arteries might burst during a MI, resulting in the creation of a clot that blocks the heart's blood supply and ultimately triggers a heart attack. These attacks could manifest in a variety of ways, ranging from silent, undiagnosed attacks to severe cases that result in unexpected, sudden death. When chest pain examination is handled poorly, people without MI might be hospitalized unnecessarily, while those who do might be ignored.

Numerous machine learning (ML) studies have aimed to predict heart attacks and save lives. Golande et al., [1] reviewed various ML approaches, finding decision tress most effective, especially when combined with other techniques. Nagamani et al. [2] used MapReduce on the Cleveland dataset, achieving a 98% average accuracy, outperforming recurrent fuzzy neural networks. Alotaibi et al., [3] found the Raid Miner program's decision tree approach superior in predicting heart failure using the UCI heart disease dataset. Repaka et al., [4] demonstrated that the random forest (RF) classifier could achieve 98% accuracy on the same dataset. Kim et al., [5] developed an AI system for predicting acute MI prognosis, achieving an AUC of 0.90, but noted limitations due to dataset size and diversity. Ahn et al., [6] created a robust prognostic model for acute MI using random forest (RF) classifiers, also achieving an AUC of 0.90, though their study focused on in-hospital mortality, limiting long-term predictions. El-Baily et al., [7] explored various ML techniques, including logistic regression and neural networks, achieving an AUC of 83.5%, but faced data imbalance challenges.

This research aims to elucidate the results of ML models in the healthcare domain, focusing on predicting post myocardial infarction angina (PMIA). By employing various classifiers to categorize MI symptoms, the study examines the influence of each feature on PMIA prediction at both local and global levels. Given the overlapping characteristics among different ML levels, the key contribution of this work is in explaining the role and impact of each feature to understand the root causes of PMIA. This study underscores the importance of AI decision interpretability, which is vital for building trust among patients and healthcare professionals. The subsequent sections of this paper are structured as follows: Section II offers a detailed explanation of the proposed framework, Section III presents the experimental results, and the paper concludes in Section IV.

II. MATERIALS AND METHODS

In The integration of the ML classifiers with UCI MIs complications dataset could enhance better clarity of the symptoms involved in health to assess both the predictive power and the interpretability of these models to ensure decision-making transparency which will be discussed in this section.

A. Data Collection and Preprocessing

The MIs complications dataset [8] from UCI ML repository includes 1,700 patient samples with 111 clinical phenotypes features. It covers twelve potential complications within three days post-MI, including: atrial fibrillation, supraventricular tachycardia, and post infarction angina. Post myocardial infarction angina (PMIA) is selected as the target variable, y, for the classification task in this study experiment. Preprocessing involves data cleaning, removing features with >90% missing values, MICE interpolation for < 30%

missing mean/mode imputation for 30 -90% missing, data standardization and Borderline-SMOTE for imbalanced datasets. Classifiers parameters were optimized using hyperparameter optimization and random grid search. The models were validated with 5-fold cross validation and an 80:20 train-test split ratio. Figure 1 outlines the study's workflow including the evaluation and the model interpretability.

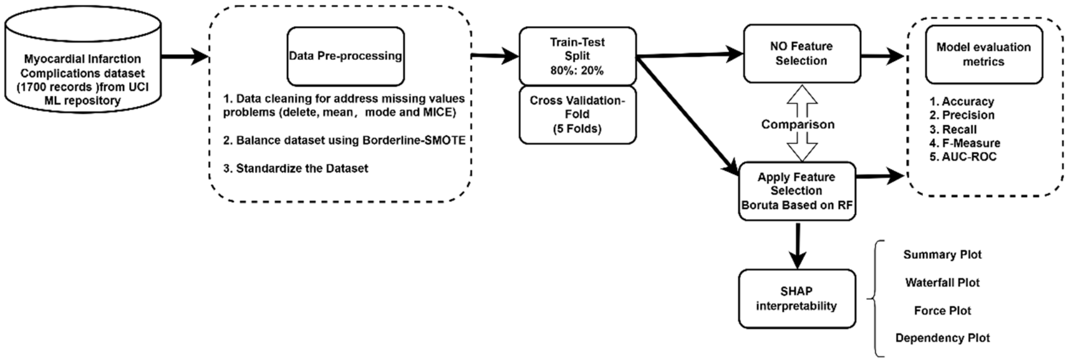


Figure 1. The architectural structure of the proposed PMIA model

B. *Feature Selection*

1) *No Feature selection (NFS)*

After removing features with a missing feature ratio of over 90%, without feature selection, there are still 109 features left for model training and evaluation.

2) *Boruta Feature Selection (BFS)*

Boruta is a comprehensive feature selection method designed to identify all relevant features, not just the most predictive ones. It is mainly used with random forests but can be adapted for other classifiers. The process involves creating shadow features, training a RF, and iteratively comparing feature importance until all significant features are identified. In this study, after 100 iterations, 47 features were confirmed important, 61 were rejected because they were insignificant and 1 remained tentative.

C. *ML Classifiers*

This study uses the following model classifiers; support vector machine (SVM), k-nearest neighbours (K-NN), decision tree (Tree), quadratic discriminant analysis (QDA), extreme gradient boosting (XGBoost), light gradient boosting (LightGBM), voting, bootstrap aggregating (bagging), adaptive boosting (AdaBoost), gradient boosting decision tree (GBDT), random forest (RF) and stacking.

D. *Evaluation Metrics*

These are the evaluation metrics used in this study for PMIA prediction; accuracy (ACC), cross-validation accuracy (CV-ACC) in (Mean ± standard deviation), precision (PREC), recall (REC), f1-score (F1), receiver operating characteristic-area under the curve (ROC-AUC), and confusion matrix.

$$\text{Confusion Matrix} = \begin{bmatrix} \text{TN} & \text{FP} \\ \text{FN} & \text{TP} \end{bmatrix} \quad (1)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (2)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

$$\text{F1-score} = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (5)$$

$$\text{ROC-AUC: } \frac{TP}{TP+FN} \text{ on y-axis } \& \frac{FP}{FP+TN} \text{ x-axis} \quad (6)$$

E. *Model Interpretability*

For model interpretability, SHAP values are used to explain the contribution of each feature to the model's predictions. The plots used to provide insights into how individual features impact the predictions include; summary, waterfall, dependency and the force plots.

III. RESULT ANALYSES

This section will mainly discuss the model performance and the SHAP Interpretability of PMIA.

A. *Evaluation Metrics Performance on different classifiers with and without Feature Selection*

Table 1 and Table 2 compare classifier performances without and with Boruta feature selection (BFS). BFS improves overall accuracy, precision, recall, F1-score and AUC-ROC for most classifiers. Notably, SVM's cross-validated accuracy rose from 97.35% to 98.55% with a short time range of 7.399seconds, nevertheless AUC-ROC of SVM instead dropped from 99.15% without FS to 98.68% when BFS is applied. Overall, the enhancements demonstrate BFS's effectiveness in enhancing classifier performance by identifying and utilizing all relevant features. Additionally, this paper has considered the area under curve, confusion matrix as seen figure 2 and figure 3 and will also explore the SHAP values with the top model –SVM. Figure 2 shows the ROC curve for the SVM model using Boruta Feature selection (BFS), with an impressive AUC of 0.987, indicating excellent model performance. Figure 3 displays the confusion matrix, revealing 326 true negatives, 282 true positives, 3 false negatives, and 10 false negatives. This high accuracy demonstrates the model's strong predictive capability for the prediction of PMIA.

Table 1. Evaluation Metrics Performance of the twelve different classifiers without Feature Selection

Classifiers	CV-ACC	ACC (%)	PREC (%)	REC (%)	F1-S (%)	AUC-ROC (%)	Time (secs)
AdaBoost	97.59 ± 0.62	97.26	97.37	97.26	97.26	98.16	189.174
Bagging	95.34 ± 0.54	95.01	95.08	95.01	95.00	97.17	21.081
GDBT	95.58 ± 0.80	95.65	95.75	95.65	95.64	97.49	103.780
KNN	90.04 ± 1.33	86.47	88.81	86.47	86.38	87.14	0.089
LGB	95.50 ± 1.23	95.81	95.90	95.81	95.81	97.12	4.314
QDA	75.98 ± 2.61	70.21	80.77	70.21	68.14	72.12	0.560
RF	96.06 ± 0.53	96.14	96.28	96.14	96.13	97.70	8.971
Stacking	97.11 ± 0.74	96.78	96.78	96.78	96.78	98.86	58.034
SVM	97.35 ± 1.31	97.26	97.32	97.26	97.26	99.15	12.389
Tree	90.84 ± 1.35	87.92	88.00	87.92	87.93	89.26	0.450
Voting	97.59 ± 0.67	96.46	96.49	96.46	96.46	98.86	97.714
XGB	95.42 ± 1.23	95.01	95.10	95.01	95.00	96.67	2.076

Table 2. Evaluation Metrics Performance of the twelve different Classifiers with BFS

Classifiers	CV-ACC	ACC (%)	PRE (%)	REC (%)	F1-S (%)	AUC-ROC (%)	Time (secs)
AdaBoost	97.35 ± 0.83	97.26	97.37	97.26	97.26	98.00	235.609
Bagging	95.42 ± 0.48	94.69	94.75	94.69	94.68	97.18	19.435
GDBT	95.10 ± 0.82	95.49	95.57	95.49	95.48	97.45	152.394
KNN	88.35 ± 0.67	87.60	89.68	87.60	87.53	88.22	0.305
LGB	95.66 ± 1.33	95.01	95.04	95.01	95.00	96.92	4.067
QDA	86.83 ± 0.96	85.83	86.54	85.83	85.82	93.83	0.496
RF	95.66 ± 0.59	96.14	96.28	96.14	96.13	97.66	16.060
Stacking	98.07 ± 0.78	97.75	97.79	97.75	97.74	99.03	39.033
SVM	98.55 ± 0.48	97.91	97.93	97.91	97.90	98.68	7.399
Tree	91.16 ± 1.32	87.12	87.23	87.12	87.13	88.06	0.535
Voting	97.91 ± 0.69	97.42	97.42	97.42	97.42	98.94	111.417
XGB	95.42 ± 1.00	94.85	94.95	94.85	94.84	96.54	1.792

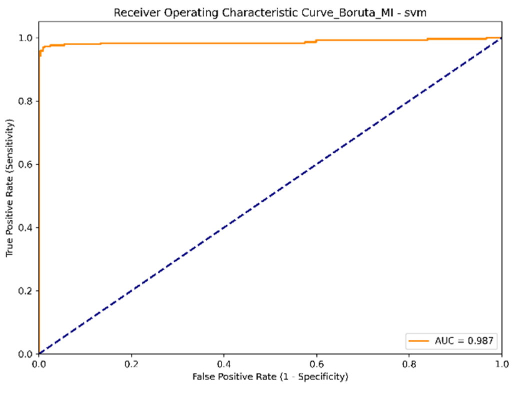


Figure 2. ROC Curve of SVM using BFS

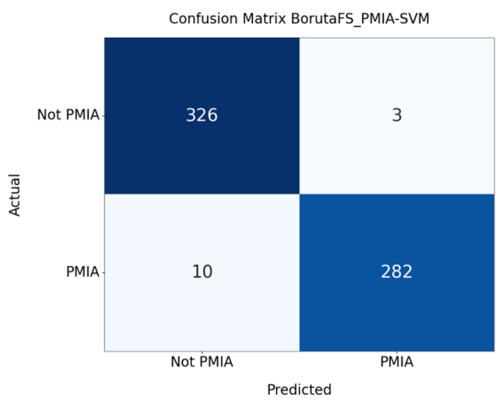


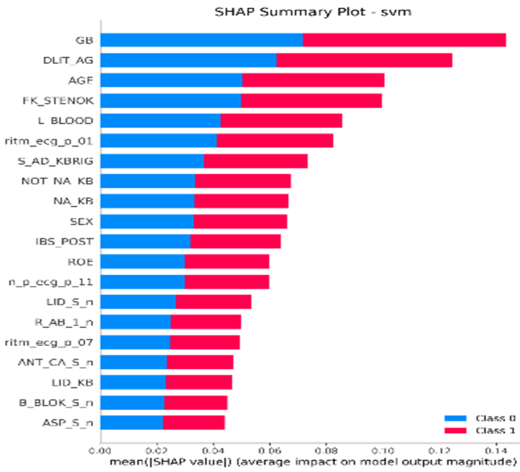
Figure 3. Confusion Matrix of SVM using BFS

B. *SHAP Interpretability Discussion*

This section will discuss the different SHAP plots; summary, waterfall, force and dependency plots.

1) *SHAP Summary Plot (Bar) and Dependency Plot Interpretability*

Figure 4(a) presents the SHAP summary plot, which ranks features by their mean SHAP values, indicating their importance. Key features such as GB, DLT_AG, and AGE show significant contributions to the model's predictions, with higher SHAP values reflecting greater influence on the prediction outcomes. Figure 4(b) shows the dot plot, illustrating how individual feature values influence the SVM model's predictions. Each dot represents a sample, coloured by feature value (red for high, blue for low). Positive SHAP values indicate a higher likelihood of predicting PMIA, and vice versa. Higher values of features like GB and DLT_AG strongly push predictions towards PMIA.



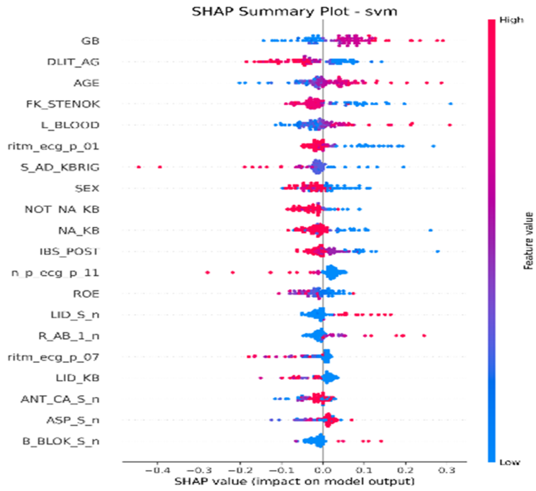
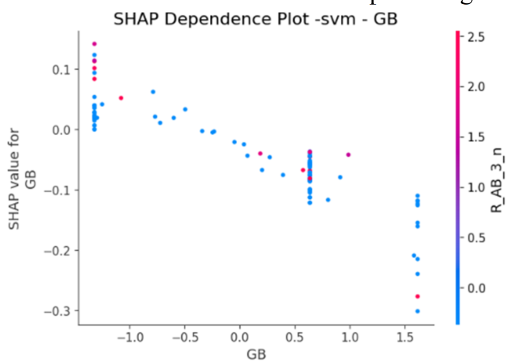


Figure 4 (a) SHAP summary plot of SVM using BFS (b) dot summary plot of SVM using BFS

2) SHAP Waterfall Plot Interpretability

Figure 5(a) represents the dependence plot for feature GB, showing its impact on the SVM model's predictions. Higher GB values generally decrease the SHAP value, reducing the likelihood of PMIA whereas the SHAP waterfall plot in figure 5(b) details the contribution of each feature to the model's prediction. Features like IBIS_POST and DLT_AG significantly influence the production outcome, with an overall model output of 98.7%



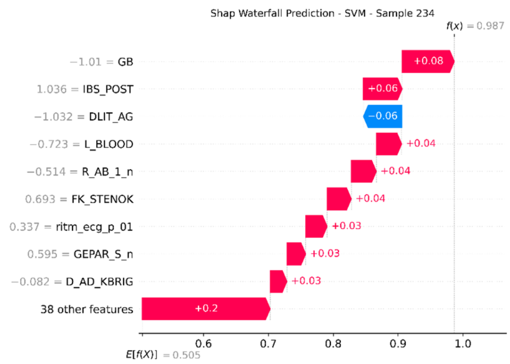


Figure 5(a) SHAP dot plot of SVM using BFS (b) SHAP waterfall plot of SVM using BFS

3) SHAP Force Plot Interpretability

Figure 6 displays the SHAP force plot for the SVM using BFS showing the contribution of each feature to the model's prediction. The base value is 0.5, representing a neutral prediction. Key features like GEPAR_S (value = 0.6) and IBS_POST (value=1.04) increase the prediction towards 1.00, while features like GB (value= -1.01) and TIME_B (value = -1.31) decrease it. This highlights how each feature's value pushes the prediction higher (red) or lower (blue), resulting to a final output of 1.00.

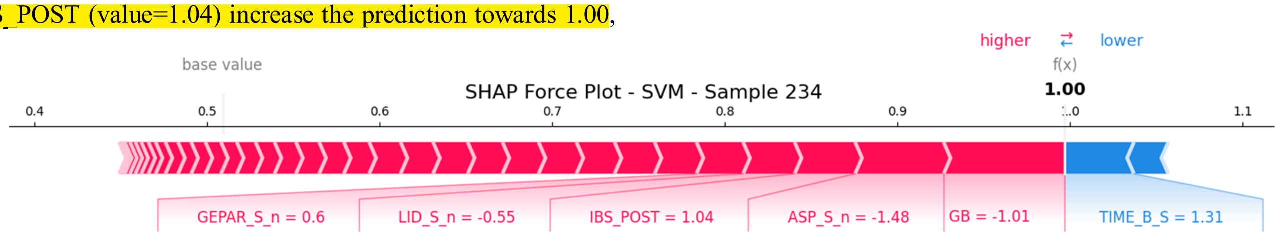


Figure 6. SHAP Waterfall Plot interpretability of SVM using BFS

IV. COMPARISON ANALYSIS

This paper emphasizes the interpretability of black-box models; specifically utilizing SHAP plots to elucidate the contributions of features in the SVM model when applied to the UCI myocardial infarction dataset. As of the writing of this paper, no existing literature focuses on the interpretability aspect of models using this dataset, thereby presenting a novel approach in our study. Consequently, there are no directly comparable works in terms of interpretability focus with the same dataset. Nonetheless, our model demonstrates robust performance in terms of evaluation metrics when compared to other studies [1] - [7]. Additionally, those studies lack an emphasis on model interpretability, making their performance less justifiable without insights into feature contributions and decision-making processes.

V. CONCLUSION AND FUTURE WORK

This study enhances the interpretability of ML models for predicting PMIA by explaining feature impacts both globally and locally. The SVM classifier achieved the best performance in all metrics at 97.91% with a short training time of 7.399 seconds. By demystifying AI's black-box nature, this approach fosters trust among practitioners and patients. Future work will explore integrating deep neural networks and analysing additional clinical data for further improve healthcare outcomes.

REFERENCES

[1] Golande A. and Kumar P., 'Heart Disease Prediction Using Effective Machine Learning Techniques', Int. J. Recent Technol. Eng., 2019, 8(1), 944-950.

[2] Nagamani T., Logeswari S., and Gomathy B., 'Heart Disease Prediction using Data Mining with Mapreduce Algorithm', Int. J. Innov. Technol. Explor. Eng., 2019, 8(3), 208-213.

[3] Alotaibi F S., 'Implementation of Machine Learning Model to Predict Heart Failure Disease', Int. J. Adv. Comput. Sci. Appl. 2019, 10(6), 261-265.

[4] Repaka A. N., Ravikanti S. D. & Franklin R. G, 'Design And Implementation Heart Disease Prediction Using Naives Bayesian', Proc. Int. Conf. Trends Electron. Informat, 2019, 292-297.

[5] Kim, M. et al. 'A reliable and interpretable AI system for predicting acute myocardial infarction prognosis.' Journal of the American Medical Informatics Association, 2024, 31(7), 1540-1550

[6] Ahn, J.H. et al. 'Robust prognostic prediction model developed with integrated biological markers for acute myocardial infarction', PLOS ONE, 2024, 19(3), e0277260.

[7] El-Bialy R. et al. 'Acute myocardial Infarction: Prediction and Patient Assessment through different ML techniques', International Journal of Intelligent Systems and Applications in Engineering, 2023, 11(1), 102 – 110.

[8] Golovenkin,S.E., Shulman,V.A., Rossiev,D.A., Shesternya,P.A., Nikulina,S.Yu., Orlova,Yu.V., and Voino-Yasenetsky,V.F.. (2020). 'Myocardial infarction complications. UCI Machine Learning Repository', https://doi.org/10.24432/C53P5M.