Received: April, 2024 | Revised: October, 2024 | Accepted: November, 2024

ORIGINAL RESEARCH

Using Machine Learning and Explainable AI for Early and Accurate Detection and Diagnosis of Heart Disease

Mekhala Mariam Mary[1], Md. Faysal Ahamed[2]

[1] Department of Computer Science and Engineering, International Standard University, Dhaka, Bangladesh

[2] Department of Computer Science and Engineering, RUET, Rajshahi, Bangladesh

Abstract

Heart disease is a growing concern worldwide, disrupting blood circulation and physiological functions. Accurate diagnosis is crucial, and non-invasive computational techniques based on machine learning are favored. This study presents a new predictive system for detecting and diagnosing cardiovascular disease, emphasizing explainable artificial intelligence. The model's intelligent decision-making process promotes confidence and dependability among medical professionals. This research investigates various classification algorithms based on machine learning, seamlessly integrated with explainable AI, to offer clear diagnostic insights. Principal Component Analysis (PCA) algorithm refines the data by eliminating redundant and extraneous information. The system's effectiveness is thoroughly assessed using critical metrics such as accuracy, precision, recall, F1-score, the area under the curve (AUC), and the Lime curve. Additionally, the system's robustness and generalizability are comprehensively evaluated using the K-fold cross-validation method. By dividing the dataset into 'K' subsets and designating one for validation and the other for training purposes, this approach improves performance in various scenarios and guarantees a reliable evaluation of efficacy. By integrating explainable AI, K-fold cross-validation, and Lime explanation, this system can become an indispensable instrument for medical professionals, offering an intelligent, streamlined, and precise pathway to heart disease diagnosis.

1 | INTRODUCTION

Heart disease is a globally acknowledged and fatal condition that significantly impacts humans worldwide. This condition occurs when the heart is unable to provide enough blood for important physiological functions. The obstruction and constriction of coronary arteries, which provide food to the heart, are often the main cause. The United States has been identified as the primary country with the highest incidence of heart disease, affecting a significantly larger share of its population [1]. The main symptoms of cardiovascular illness include weariness, swelling in the extremities, and physical weakness, along with other related signs. Cardiovascular disease vulnerability can be exacerbated by personal lifestyle choices such as smoking, unhealthy eating habits, high blood pressure, lack of physical activity, and poor fitness levels. Heart illness encompasses various conditions, with coronary artery disease (CAD) being the most common subtype, capable of causing chest discomfort, strokes, and myocardial infarction [2].

Heart illness can manifest in several forms such as cardiovascular disease (CVD), congenital heart disease (CHD), cardiac rhythm disturbances, and congestive heart failure. Conventional diagnostic methods previously utilised for heart disease diagnosis are now recognised as intricate. In less developed areas, diagnosing and treating patients can be particularly difficult because of the scarcity of medical equipment and skilled staff. Therefore, an accurate and timely diagnosis of cardiac disease is crucial to prevent further decline in the patient's health. Regrettably, there is a global increase in significant heart disease, affecting both developed and developing nations [3]. In 2016, the World Health Organisation (WHO) stated that Cardiovascular Disease (CVD) was responsible for around 30% of all worldwide deaths, resulting in approximately 17.90 million fatalities. Every year in Pakistan, 0.2 million individuals

succumb to heart disease, with a rising trend in these numbers. The European Society of Cardiology (ESC) approximates that 26.5 million persons currently have cardiac disease, and 3.8 million new cases are diagnosed annually. Approximately 50-55% of individuals diagnosed with heart disease pass away during the initial one to three years, resulting in a notable 4% of yearly healthcare funds being allocated to heart disease treatment [4].

In the past, doctors utilised invasive methods to identify heart problems. This entailed reviewing the patient's medical history, doing a physical examination, and assessing symptoms related to the disease. Angiography was one of the most precise techniques, however it was costly and had other drawbacks, including side effects and the need for considerable technical expertise. Traditional procedures were susceptible to human error and time-consuming. Moreover, disease diagnosis was excessively expensive and required significant computational resources.

Researchers set out to develop a number of intelligent healthcare systems that could detect cardiac problems without resorting to intrusive procedures in an effort to overcome the drawbacks of these techniques. Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Naïve Bayes (NB), and Decision Tree (DT) were among the predictive machine learning algorithms employed by these systems. The result is a lower death rate for people suffering from heart disease. Researchers frequently use the Cleveland heart disease dataset in their scholarly publications.

The main contributions of our proposed methodology:

- The classifier performance evaluation has encompassed a comprehensive assessment across the entire feature space, focusing primarily on performance evaluation metrics, notably accuracy.
- Our investigation has extended to appraise the classifier performances within carefully selected feature spaces, thoughtfully curated through applying previously outlined diverse feature selection algorithms.
- This research provides valuable recommendations regarding the compatibility of specific feature selection algorithms with classification algorithms, serving as a blueprint for constructing an advanced intelligence system for diagnosing heart disease patients.

2 | RELATED WORK

Specialists who focused on detecting cardiac disease have made significant advancements in the field. Robert et al. [5] implemented a logistic regression classification algorithm, resulting in a classification accuracy rate of 77.1%. Similarly, Wankhade et al. [6] used a multi-layer perceptron (MLP) classifier to achieve an 80% accuracy rate in the diagnosis process. Allahverdi et al. [7] incorporated neural networks into a framework of artificial neural networks to develop a classification system for cardiac disease, resulting in an impressive accuracy rate of 82.4%. Yar et al. [8] employed NB and DT to diagnose and forecast cardiac disease, achieving commendable outcomes with NB accuracy of 82.7% and DT accuracy of 80.4%. Oyedodum et al. [8] proposed a three-phase artificial neural network (ANN) system for predicting cardiac disease. Das et al. [9] introduced an ensemble-based predictive model for the prognosis of cardiac disease that utilizes ANNs. Paul et al. [10] investigated the adaptive fuzzy ensemble method for predicting cardiovascular disease. Tomov et al. [11] proposed a deep neural network methodology that resulted in advantageous outcomes in predicting cardiac disease. Manogaran et al. [12] introduced a hybrid recommendation system for diagnosing cardiac disease with encouraging results. In their study, Alizadehsani et al. [13] created a non-invasive model to forecast coronary artery disease, which exhibited exceptional performance metrics and accuracy. Amin et al. [14] introduced a machine learning-based hybrid system designed to detect cardiac disease with an impressive accuracy rate of 86.0%. Mohan et al. [15] incorporated Random Forest (RF) into their intelligent system alongside a linear model, resulting in a classification accuracy of 88.7%. Liaqat et al. [16] developed an expert system for predicting cardiac disease using stacked SVMs, achieving a remarkable classification accuracy of 91.11% by employing specific features.

The study introduces a sophisticated medical decision system designed to diagnose heart illness using advanced machine learning algorithms. We want to identify the best effective model for early detection of heart disease with the highest level of accuracy. We employed four feature selection techniques, with Principal Component Analysis (PCA) being the most prominent, to selectively include essential features that exhibit a strong association and effectively represent the core essence of our intended objective. Each computational and processing action was meticulously executed utilising the Anaconda Integrated Development Environment (IDE). We proficiently implemented the classifiers using Python and incorporated necessary packages and libraries such as pandas, NumPy, matplotlib, sci-kit learn (sklearn), seaborn, and Lime.

3 | Methodology

The process begins with the dataset undergoing pre-processing, which includes steps like scaling, standardization, and normalization. Then, the dataset is visualized to understand its characteristics better. The next phase is feature extraction, where Principal Component Analysis (PCA) might be used to reduce the dimensionality of the data. The processed data is then divided into a training set, which comprises 80% of the data, and a testing set, which consists of the remaining 20%. These sets are used

to train and evaluate various machine learning models formance of these models is assessed using metrics like confusion matrices, accuracy, precision, recall, F1-score, and ROC-AUC. Moreover, the use of K-fold cross-validation with (k = 5) for model validation, which is a method to assess how the results of a statistical analysis will generalize to an independent dataset.

Finally, the flowchart indicates that predictions will be made to determine whether heart disease is 'Not Found (0)' or 'Found (1)' and mentions the use of LIME (Local Interpretable Model-agnostic Explanations) plotting for explainable AI (XAI), which helps in understanding the predictions made by the machine learning models.

4 | DATASET OVERVIEW

The dataset used in this research is collected from Kaggle platform, the dataset is also known as Heart Disease Dataset [15]. Altogether, the data was the combination of four different databases, but only Cleveland data used in this experiment. It is an open dataset, having several attributes, but for this experiment only fourteen attributes are selected as described and suggested by different scholars that selected 14 attributes, most useful to predict the heart disease in a patient [17,18]. In addition, the database file contains the record of 303 patients. The dataset attributes and their values are presented in Table 1.

4.1 | Dataset Preparation

The following dataset preparation steps were taken:

Data scaling: each column of the data set had varying scales, so it was necessary to scale the data to have the same scale of 0 to 1 for each column. The following formula was used:

$$x_{scaled} = \frac{x - x_{min}}{(x_{max} - x_{min})}$$

Data standardization: for n dataset, all the column's mean and variance were converted to 0 and 1 respectively using the following formula:

$$x_{stand} = \frac{x - \mu}{\sigma^2}$$

Here μ and σ^2 represents the average and variance of each row respectively.

Data normalization: The Euclidean norm of each row was converted to 1 using the following formula.

$$x_{norm} = \frac{x}{||x||}$$

Where ‖x‖ represents the Euclidean length or the L2 norm

Table 1: Heart Disease Dataset's Attributes

Attribute	Code given	Note	Attributes Values
age	Age	in years	Numeric
sex	Sex	1 = male; 0 = female	Binary
chest pain type	level of pain	0,1,2,3	4 values
resting blood pressure	trestbps	in mm Hg	in mm Hg
serum cholesterol	cholesterol	in mg/dl	Numeric
fasting blood sugar	fbs	> 120 mg/dl	Numeric
resting electrocardiographic results	restecg	0,1,2	3 values
maximum heart rate achieved	thalach	71–202	Numeric
exercise induced angina	exang	0,1	Binary
oldpeak = ST	oldpeak	depression	Numeric
the slope of the peak exercise ST segment	slope	0,1,2	3 values
number of major vessels flourosopy	ca	0,1,2,3	4 values
defect: normal; fixed; reversible; nonreversible	thal	0,1,2,3	4 values
class	target	0,1	Binary

4.2 | Dataset Visualization

Data preprocessing is an essential step used to clean the data and make it useful for any experiment associated with machine learning or data mining. In this study, multiple preprocessing steps were applied on the selected dataset.

Firstly, the size of the dataset was found not enough for the implementation of machine learning approaches. As described by the size of the dataset for machine learning implementation may create biasness and would also affect the results generated through machine learning models. Therefore, for each attribute using minimum and maximum values, the random number generation technique applied to generate random values for each column. This helped us to enhance the capacity of the data, which has created a positive impact on the performance of the classifier as can be seen in the results section. Thus, the data has increased the volume by three times.

Secondly, using rapid miner, data cleaning step applied to find out missing values and noisy data values. The data has some missing values which have been imputed using K-Nearest Neighbor (KNN) method. KNN method has proved to be a useful method for missing data imputation. In addition, the outlier detection methods are used to estimate the noise in the data. The data has not found noisy values and no outlier is detected in the dataset. The outlier detection is applied using rapid miner's operator with distances method. To check the other discrepancies in the dataset, data discretization, transformation and the beginning techniques were applied as well.

4.3 | Correlation Matrix

Figure 1 presents the correlational matrix. Each cell in the table shows the correlation between two variables. The value is in the range of -1 to 1. If two variables have a high positive correlation (closer to 1), it means that they tend to increase or decrease together. A high negative correlation (closer to - 1) indicates that one variable increases when the other decreases. A correlation close to 0 implies little or no linear relationship between the variables.

In this matrix, each row and column represents a different variable, such as 'age', 'sex', 'cp' (chest pain type), 'trestbps' (resting blood pressure), 'chol' (serum cholesterol), etc., with 'target' likely being the variable we are trying to predict. The colors provide a visual representation of the correlation values, where red denotes positive correlation and blue denotes negative correlation. For instance, 'cp' has a positive correlation of 0.43 with 'target', suggesting that as 'cp' increases, the likelihood of the 'target' variable also increases. Conversely, 'thalach' (maximum heart rate achieved) has a negative correlation of -0.42 with 'target', indicating that higher values of 'thalach' are associated with a decrease in the 'target' variable. The diagonal is filled with 1s because each variable is perfectly correlated with itself. This matrix is useful for identifying relationships between variables that might warrant further analysis or for informing feature selection in machine learning models.

4.4 | Distribution Plots

Figure 2 presents the distribution plotting. Histograms are used to visualize the distribution of a single numerical variable by dividing the data into bins or intervals and showing the frequency (count) of data points within each bin.

From our description, the histograms include variables such as 'age', 'trestbps' (resting blood pressure), 'chol' (serum cholesterol), 'thalach' (maximum heart rate achieved), and 'oldpeak' (ST depression induced by exercise relative to rest). Overlaid on each histogram is a line that likely represents the kernel density estimate (KDE), which is a smoothed version of the histogram and provides an estimate of the probability density function of the variable. The 'age' histogram, for instance, would show how many individuals fall within certain age ranges, while the 'trestbps' histogram would show the distribution of resting blood pressure readings across the population in the dataset. The shape of each histogram can give us insights into the characteristics of the distribution, such as whether it is symmetric, skewed, has a single peak (unimodal), or more than one peak (multimodal).

The distribution of 'chol' would tell us how serum cholesterol levels are spread among the individuals, which is useful for identifying common cholesterol levels or outliers. The 'thalach' histogram reflects the distribution of maximum heart rates, and 'oldpeak' shows the distribution of ST depression levels.

5 | FEATURE EXTRACTION (PCA)

PCA (Principal Component Analysis) is a dimensionality reduction technique commonly used in the analysis of tabular data. It helps in simplifying complex datasets by transforming them into a new set of variables (principal components) that capture the most significant variations in the data. Here's a brief explanation of PCA for tabular data and its mathematical formula:

- Dimension Reduction: PCA is employed to reduce the number of features (columns) in tabular data while retaining the essential information. It does this by identifying linear combinations of the original features that explain the most variance in the data.
- Orthogonal Components: PCA ensures that the new variables (principal components) are orthogonal to each other, meaning they are uncorrelated. This property helps in reducing multicollinearity in the data.
- Variance Maximization: The principal components are ordered by the amount of variance they explain in the original data. The first principal component explains the most variance, the second explains the second most, and so on.

.

Figure 1: Histogram plotting.

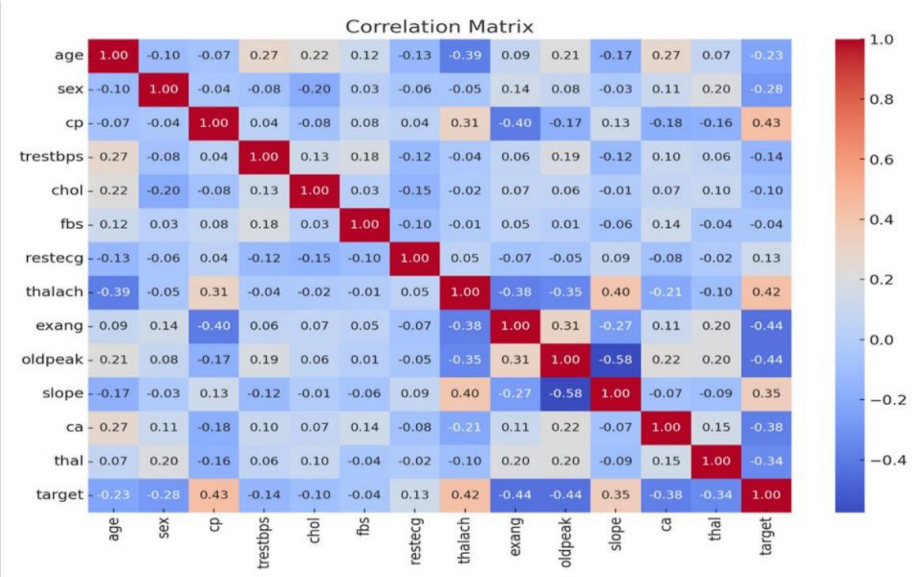


Figure 2: Distribution Plotting.

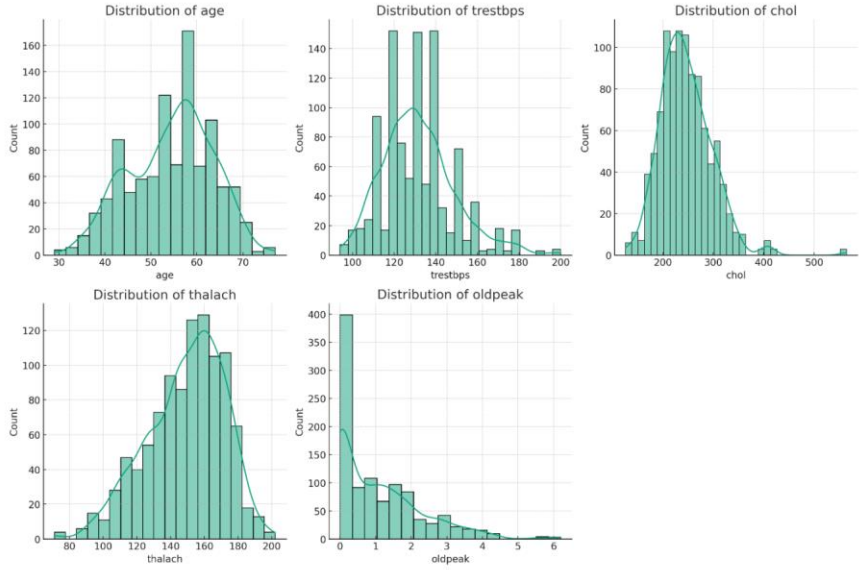
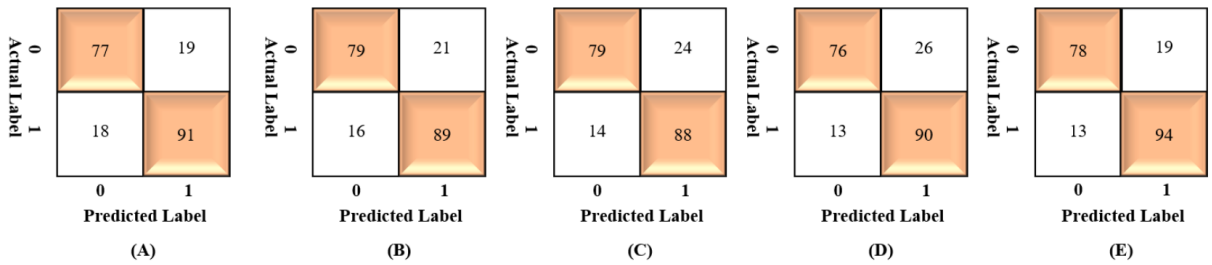


Figure 3: Confusion matrices for SVM classifier, (A) Fold 1, (B) Fold 2, (C) Fold 3, (D) Fold 4, and (E) Fold 5.



Data Projection: After computing the principal components, we can project the original data onto these components to obtain a reduced-dimensional representation of the data.

Kernel principal components analysis was used to non-linearly reduce the dimension of the data, thus extracting features with highest variance. The idea of the principal components analysis is the following:

At first the covariance matrix of the data is computed as:

$$C = \frac{1}{N} \sum x_i x_i^T$$

The full principal components analysis can then be written as follows:

$$T = XW$$

Where W is a $p \times p$ matrix of weights whose columns are the eigenvectors of the covariance matrix C

6 | RESULTS & DISCUSSIONS

6.1 | SVM Classifier

Figure 3 presents confusion matrices. Support Vector Machine (SVM) classifier is evaluated across five folds, a common practice in cross-validation. This method tests the model's performance on different subsets of the dataset to ensure generalizability. The accuracy of the model across these folds varies slightly, ranging from 80.98% to 84.31%, with an average accuracy of 82.13%. This relatively narrow range, along with a standard deviation of 1.15%, suggests that the SVM classifier maintains consistent performance across different data partitions.

In terms of precision, which measures the ratio of correctly predicted positive observations to total predicted positives, the classifier shows a slightly better ability to correctly identify Class 0 instances (mean precision of 0.84) compared to Class 1 (mean precision of 0.81). The precision values for Class 0 fluctuate between 0.81 and 0.86, while for Class 1, they range from 0.78 to 0.83.

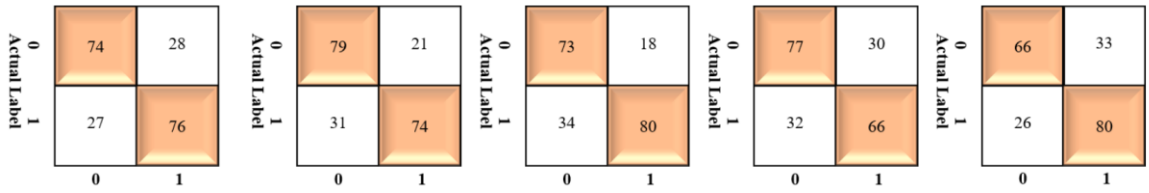
Regarding recall, which is the ratio of correctly predicted positive observations to all observations in the actual class, the classifier performs better in identifying Class 1 instances, with a mean recall of 0.86, compared to 0.78 for Class 0. The recall for Class 0 lies between 0.75 and 0.80, and for Class 1, it is between 0.83 and 0.88. This indicates that the SVM model is generally more efficient at recognizing all relevant instances of Class 1. The F1-score, a balance between precision and recall, also reflects this trend.

The mean F1-scores are 0.81 for Class 0 and 0.83 for Class 1, suggesting a balanced performance, especially for Class 1, which has slightly higher scores ranging from 0.82 to 0.85 compared to Class 0's range of 0.80 to 0.83. Lastly, the 'Support' column shows the number of actual occurrences of each class in the dataset, indicating a small imbalance in the last fold with 204 instances compared to 205 in the first four folds. Overall, the SVM classifier demonstrates stable and balanced performance, with a slight inclination towards better identifying and classifying instances of Class 1.

6.2 | K-Nearest Neighbor (KNN) Classifier

Figure 4 presents confusion matrices. In a 5-fold cross-validation evaluation of a K-Nearest Neighbors (KNN) classifier, several key performance metrics were assessed. Table 4 presents the evaluation scores. On average, the classifier achieved an accuracy of 72.68%, with a standard deviation of 1.93%, indicating a moderate level of variability in its performance across different folds. When looking specifically at Class 0 and Class 1 metrics, the classifier demonstrated an average precision of 0.71 and 0.75, respectively, showing a slightly better precision for Class 1. The average recall for Class 0 was 0.74, while for Class 1, it was 0.71, suggesting better recall for Class 0. Both classes had similar average F1-Scores of approximately 0.72, striking a balance between precision and recall. It's noteworthy that individual folds exhibited some variability in performance, with Fold 2 achieving the highest accuracy at 74.63%, while Fold 4 had the lowest at 69.76%. Class-specific metrics also displayed variations across different folds. To ensure more consistent performance, further analysis and potential model tuning may be required. The AUC values for each fold range from 0.84 to 0.88, which indicates a good level of predictive accuracy for the classifier. Figure 17 presents the ROC-AUC scores on 5-fold scores.

Figure 4: Confusion matrices for KNN classifier, (A) Fold 1, (B) Fold 2, (C) Fold 3, (D) Fold 4, and (E) Fold 5.



7 | CONCLUSIONS

Cardiovascular disease is a chronic condition that can be highly debilitating and potentially fatal. The number of cases is rising in both developed and developing countries and timely identification and treatment are crucial to reducing the damage they cause. To address this issue, we developed an intelligent predictive system that uses modern machine-learning algorithms to diagnose and forecast cardiac disease. Our research paper outlines the training and testing of our procedure utilizing an optimal feature set. We combined four classification algorithms (SVM, KNN, LR and LDA), PCA feature selection techniques, and K-fold cross-validation to identify the most pertinent features. This approach helped us to reduce classification errors and optimize the feature space. We utilized various evaluation metrics, such as confusion matrices, accuracy, precision, recall, F1-score, AUC scores, and ROC curves, to assess the effectiveness of these classification algorithms. Our results showed that the RF and DT classifiers achieved the highest classification accuracies of 99.71% and 99.41%, respectively, when the complete feature set was utilized. We found that the RF classifier, when combined with the relief feature selection algorithm, achieved exceptional performance while maintaining the interpretability of the model through explainable AI. In the future, we plan to incorporate additional optimization techniques, algorithms for feature selection, and classification algorithms to improve the performance of the predictive system in diagnosing cardiac disease.

8 | REFERENCES

[1] A.L. Bui, T.B. Horwich, G.C. Fonarow, Epidemiology and risk profile of heart failure, Nat. Rev. Cardiol. 8 (2011) 30–41.

[2] M. Durairaj, N. Ramasamy, A comparison of the perceptive approaches for preprocessing the data set for predicting fertility success rate, Int. J. Control Theory Appl. 9 (2016) 255–260.

[3] P.A. Heidenreich, J.G. Trogdon, O.A. Khavjou, J. Butler, K. Dracup, M.D. Ezekowitz, E.A. Finkelstein, Y. Hong, S.C. Johnston, A. Khera, Forecasting the future of cardiovascular disease in the United States: a policy statement from the American Heart Association, Circulation. 123 (2011) 933–944.

[4] J. Lee, J. Park, S. Yang, H. Kim, Y.S. Choi, H.J. Kim, H.W. Lee, B.-U. Lee, Early seizure detection by applying frequency-based algorithm derived from the principal component analysis, Front. Neuroinform. 11 (2017) 52.

[5] R. Detrano, A. Janosi, W. Steinbrunn, M. Pfisterer, J.-J. Schmid, S. Sandhu, K.H. Guppy, S. Lee, V. Froelicher, International application of a new probability algorithm for the diagnosis of coronary artery disease, Am. J. Cardiol. 64(1989) 304–310.

[6] M. Gudadhe, K. Wankhade, S. Dongre, Decision support system for heart disease based on support vector machine and artificial neural network, in: 2010 Int. Conf. Comput. Commun. Technol., IEEE, 2010: pp. 741–745.

[7] H. Kahramanli, N. Allahverdi, Design of a hybrid system for the diabetes and heart diseases, Expert Syst. Appl. 35 (2008) 82–89.

[8] Y. Muhammad, M. Tahir, M. Hayat, K.T. Chong, Early and accurate detection and diagnosis of heart disease using intelligent computational model., Sci. Rep. 10 (2020) 19747. https://doi.org/10.1038/s41598-020-76635-9.

[9] R. Das, I. Turkoglu, A. Sengur, Effective diagnosis of heart disease through neural networks ensembles, Expert Syst. Appl. 36 (2009) 7675–7680.

[10] A.K. Paul, P.C. Shill, M.R.I. Rabin, K. Murase, Adaptive weighted fuzzy rule-based system for the risk level assessment of heart disease, Appl. Intell. 48 (2018) 1739–1756.

[11] N.-S. Tomov, S. Tomov, On deep neural networks for detecting heart disease, ArXiv Prepr. ArXiv1808.07168. (2018).

[12] G. Manogaran, R. Varatharajan, M.K. Priyan, Hybrid recommendation system for heart disease diagnosis based on multiple kernel learning with adaptive neuro-fuzzy inference system, Multimed. Tools Appl. 77 (2018) 4379–4399.

[13] R. Alizadehsani, M.J. Hosseini, A. Khosravi, F. Khozeimeh, M. Roshanzamir, N. Sarrafzadegan, S. Nahavandi, Non-invasive detection of coronary artery disease in high-risk patients based on the stenosis prediction of separate coronary arteries, Comput. Methods Programs Biomed. 162 (2018) 119–127.

[14] A.U. Haq, J.P. Li, M.H. Memon, S. Nazir, R. Sun, A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms, Mob. Inf. Syst. 2018 (2018) 1–21.

[15] S. Mohan, C. Thirumalai, G. Srivastava, Effective heart disease prediction using hybrid machine learning techniques, IEEE Access. 7 (2019) 81542–81554.

[16] L. Ali, A. Niamat, J.A. Khan, N.A. Golilarz, X. Xingzhong, A. Noor, R. Nour, S.A.C. Bukhari, An optimized stacked support vector machines based expert system for the effective prediction of heart failure, IEEE Access. 7 (2019) 54007–54014.

[17] S. Palaniappan, R. Awang, Intelligent heart disease prediction system using data mining techniques, in: 2008 IEEE/ACS Int. Conf. Comput. Syst. Appl., IEEE, 2008: pp. 108–115.

[18] L. Ali, A. Niamat, N.A. Golilarz, A. Ali, X. Xingzhong, An Expert System Based, (2019).

Declaration of interest:

We, the authors of this research manuscript, declare that we have no financial interest. We have provided written consent to publish the paper in this journal.

To cite this article: Ahmed M. F., Mary M. M., and Salam T., (2024). Using Machine Learning and Explainable AI for Early and Accurate Detection and Diagnosis of Heart Disease. *Journal of Engineering and Technology (JET)*, Vol:01, Issue:01, page:37:44, ISUCRDP, Dhaka