From Data to Diagnosis: Applying Machine Learning Model for Reliable Heart Disease Prediction

1st Ananna Rashid Kumkum
Faculty of Science and Technology
American International University-Bangladesh
Dhaka, Bangladesh
arkumkum2910@gmail.com

2nd Anik Sen
Faculty of Science and Technology
American International University-Bangladesh
Dhaka, Bangladesh
aniksen360@gmail.com

3rd Abdullah Yousuf Noman
Faculty of Science and Technology
American International University-Bangladesh
Dhaka, Bangladesh
aynoman151@gmail.com

4th Prajukta Majumder
Faculty of Science and Technology
American International University-Bangladesh
Dhaka, Bangladesh
prajuktamajumder@gmail.com

5th Tze Hui Liew*
Faculty of Information Science and Technology
Multimedia University
Melaka, Malaysia
thliew@mmu.edu.my

6th Nafiz Fahad
Faculty of Information Science and Technology
Multimedia University
Melaka, Malaysia
fahadnafiz1@gmail.com

7th Md. Saef Ullah Miah
Faculty of Engineering Technology
Multimedia University
Melaka, Malaysia
saefullah@mmu.edu.my

8th Riadul Islam Rabbi
Faculty of Engineering Technology
Multimedia University
Melaka, Malaysia
riadul.rabbi72@gmail.com

9th Md. Jakir Hossen
Faculty of Engineering Technology
Multimedia University
Melaka, Malaysia
jakir.hossen@mmu.edu.my

***Abstract*—Heart disease remains a significant global health challenge, demanding precise and interpretable predictive tools to aid early intervention and improve patient outcomes. This study applies the XGBoost machine learning algorithm to develop a robust model for predicting cardiovascular disease risk. Leveraging a comprehensive dataset that includes demographic, lifestyle, and clinical features, our approach combines extensive data preprocessing, feature engineering, and Bayesian hyperparameter tuning to achieve superior predictive accuracy, with an AUC of 0.98 and an accuracy of 97.48%. Key predictive factors, identified through SHAP (SHapley Additive exPlanations) values, include cholesterol, blood pressure, and physical activity, offering actionable insights for clinicians. The model's interpretability supports its potential as a reliable, real-time diagnostic tool, enhancing clinical decision-making and risk assessment. This research advances predictive healthcare analytics by showcasing the effectiveness of sophisticated machine learning methods, including XGBoost, for precise and interpretable heart disease forecasting. Subsequent study will seek to validate the model across varied populations and incorporate it into clinical environments for adaptive, individualized cardiovascular risk management.**

***Keywords*—Cardiovascular diseases, XGBoost, Machine Learning, SHAP values, Risk Prediction**

I. Introduction

Cardiovascular disease (CVD) is among the major health burdens worldwide, which brings high morbidity and mortality in different populations [1], [2]. Early prediction and detection of cardiac disease are very important in reducing mortality due to associated diseases, as early treatment leads to a better prediction and reduction in health costs [3], [4]. Machine learning advances have opened up a set of new approaches toward improving diagnostic accuracy by analysis of complex health data, thus enabling efficient and more accurate disease prediction than what can be offered by traditional statistical methods [5]. Application of ML algorithms to medical diagnostics in recent years has demonstrated potentials for revolutionizing heart disease prediction to support data-driven clinical decisions with large and diverse data sets [6]. A variety of ML models have been studied and applied for the prediction of heart disease, each with some advantages and disadvantages. Conventional models include logistic regression and decision trees that are widely used because of their interpretability and simplicity, but they are less accurate for

2025 5th Asia Conference on Information Engineering (ACIE) | 979-8-3315-2840-9/25/$31.00 ©2025 IEEE | DOI: 10.1109/ACIE64499.2025.00015

979-8-3315-2840-9/25/$31.00 ©2025 IEEE
DOI 10.1109/ACIE64499.2025.00015

complex medical data [7], [8]. On the one side, more complex models, including neural networks, are more powerful in prediction but usually require heavy computational resources and are not interpretable, which is a critical need in clinical applications [9]. Ensemble approaches: much attention has been devoted to ensemble methods since they attain interpretability, accuracy, and robustness against overfitting by combining several weak learners [10], [11]. Machine learning techniques ranging from support vector machines (SVM), k-nearest neighbors (k-NN), to deep learning models vary within the literature on heart disease prediction. The varying model accuracy, complexity, feature selection strategy, and application of ensemble techniques have been studied in several works [12]. Notably, recent efforts have focused on feature selection and data preprocessing techniques to improve model accuracy and minimize computational load [13], [14]. Important features often include demographic data (e.g., age, sex), physiological measurements (e.g., blood pressure, cholesterol), and lifestyle factors, which have been shown to significantly impact the performance of predictive models [15].

The selection of the model is crucial, but the quality and type of data employed for training these models are equally important. Multiple studies have investigated numerous publicly accessible datasets, like the Cleveland Heart Disease dataset, and highlighted the significance of clean, representative, and balanced data for precise prediction [16]. In the current study, we aim to expand on existing research by implementing an XGBoost-based approach for heart disease prediction, assessing its performance in comparison with other widely used models on a standard dataset. Through this study, we contribute to the growing field of predictive analytics in healthcare by exploring the model's potential as an aid for clinicians in early diagnosis and treatment planning.

II. Related Work

The application of machine learning in predicting cardiac disease has garnered considerable interest because of its potential to enhance diagnostic precision and enable prompt intervention. Traditional methods, such as logistic regression, have been extensively applied for their interpretability and computational simplicity. For instance, Katarya and colleagues [10] applied logistic regression for heart disease prediction and found it suitable for initial screening; however, the model's limited handling of non-linear relationships often reduces its accuracy. Similarly, decision trees have been widely used due to their interpretability and simplicity. Kavitha et al. [11] demonstrated the efficacy of decision trees in clinical diagnostics, yet they noted the method's tendency to overfit, particularly with smaller datasets.

Support vector machines (SVMs) have been promising in the case of nonlinear data with high-dimensional feature space. Pasha et al. (2020) pointed out that SVM can give accurate results related to heart diseases, especially in conjunction with Kernel functions while dealing with complex nonlinear data [15]. However, the research mentioned that SVMs require the proper tuning of hyperparameters, which is a cumbersome and time-consuming process. On the other hand, there are studies such as that of Jindal et al. (2021), which argue that SVMs perform well on binary classifications but might not be as effective in multi-class prediction scenarios, as it often happens in medical diagnosis [9].

Most Ensemble methods have been proposed during these last years for their robustness and enhanced predictive accuracy. Random forests are composed of many decision trees put together and have shown great improvements in diagnosis in medicine. For instance, Saboor et al. (2022) applied RF to predict heart disease and reached better performance compared to single decision trees, with much better generalization (i.e., less overfitting) [13]. Bharti et al. (2021) went further to investigate the implementation of random forests; they realized especially high accuracy for the model in an imbalanced dataset [6]. The authors recognised that RF was more robust due to the ensemble method; however, that comes at a high price in computation, particularly for bigger datasets.

Various deep learning methods involving DNN, among others, have found their place in the domain of heart disease prediction; again, these usually require big data and computational resources. Mehmood et al. (2021) proposed a deep neural network-based model, yielding very good performance in predicting heart disease, though with limited interpretability [12]. Sharma et al. (2020) emphasized the model's effectiveness, but this work indicated that because of the complex training processes of DNNs, they are not suited for real-time use in clinics [21]. Models that embedded hybrid ideas with DNNs into other algorithms, like random forests or SVM, were also proposed that could work as a balanced approach with both accuracy and interpretability. Kumar et al. (2021) combined a DNN with an RF approach; results showed high accuracy with a moderate increase in interpretability, yet with persistently high computational requirements [19]. Feature selection and preprocessing strategy beyond the choice of model is critical for heart disease prediction models to be effective. Katarya et al. (2021), and Jindal et al. (2021) have applied a feature selection approach to improve model efficiency by selecting the best features in prediction, which includes cholesterol level, age, and blood pressure [10], [9]. This feature selection will reduce computational complexity of the model, enhancing model interpretability and thereby making predictions much more clinically relevant. Moreover, Kumar et al. (2021) reported that with the embedded feature importance functionality in XGBoost, critical predictors could be identified with a significantly high accuracy level achieved, although computationally expensive [19].

Besides feature engineering, augmenting the dataset has also improved the results. For example, Kavitha et al. (2021) and Sharma et al. (2020) used augmented data to improve model robustness and performance [11], [21]. Bharti et al. (2021) shown that data preprocessing processes, such as addressing missing values and balancing datasets, greatly impacted the performance of ML models in heart disease prediction [6]. While XGBoost has been popular for its performance, Kumar et al. (2021) observed that simpler models with well-curated

datasets can often achieve comparable accuracy levels without the complexity associated with ensemble methods [19]. Various traditional methods, ensemble approaches, and deep learning models have been used in carrying out the research in heart disease prediction. The relatively more straightforward ones are logistic regression and decision trees, which are comparatively more interpretable; the advanced ones comprise random forests, SVM, and neural networks, which generally offer far higher predictive accuracy at the expense of increased model complexity. For balanced performance and robustness, among others, stand out ensemble and hybrid models. In return, this paper contributes further to this body of work by assessing the performance of the XGBoost model in terms of predictive performance, feature selection, and its clinical applicability.

III. Methodology

A. Data Collection

The dataset used in this work was from Kaggle repository by the name "Cardiovascular Diseases Risk Prediction" [22]. The dataset includes health-related information of people, while it targets different risk factors of cardiovascular disease. The preprocessed divided dataset includes 56,000 training records and 14,000 test records. These include a number of features that can be categorized as:

- Demographic Features: Age Category, Sex
- Health and Lifestyle Features: Alcohol Consumption, BMI, Exercise, Smoking History, Fried Potato Consumption, Fruit Consumption, Green Vegetables Consumption
- Clinical Features: Checkup, General Health, Diabetes, Depression, Arthritis, Skin Cancer, Other Cancer
- Physical Measurements: Height (cm), Weight (kg)

The target variable is **Heart Disease**, which indicates the presence (1) or absence (0) of cardiovascular disease.

B. Data Preprocessing

Data preprocessing is an important step in preparing the dataset for modeling. The following methods were employed:

1) Handling Missing Values: The dataset was initially inspected for missing values. To handle missing data, the medians of the numerical features and the respective modes of categorical features were calculated. This was the strategy to keep the distribution of the dataset and avoid bias in model training.
2) Encoding Categorical Features: The categorical features were encoded using One-Hot Encoding, which changes categorical variables as binary vectors. One-hot encoding is absolutely vital for algorithms that need numerical input and thus helps in preserving whatever information may be carried by the categorical variables.

C. Exploratory Data Analysis (EDA)

Exploratory Data Analysis was conducted to understand the dataset better and identify patterns or anomalies. The following techniques were used:

1) Correlation Analysis: A correlation matrix showing the relationship of numeric features versus the target variable was created. The study helped identify those features most related with the occurrence of heart disease and was helpful in further feature selection.
2) Visualization Techniques: Several visualization techniques have been done, such as viewing the distribution of individual features by using histograms and the relationships between pairs of features by using pairplots. The correlation heatmap, which is shown in Fig. 1, presents the correlation coefficients of the features against the target variable.

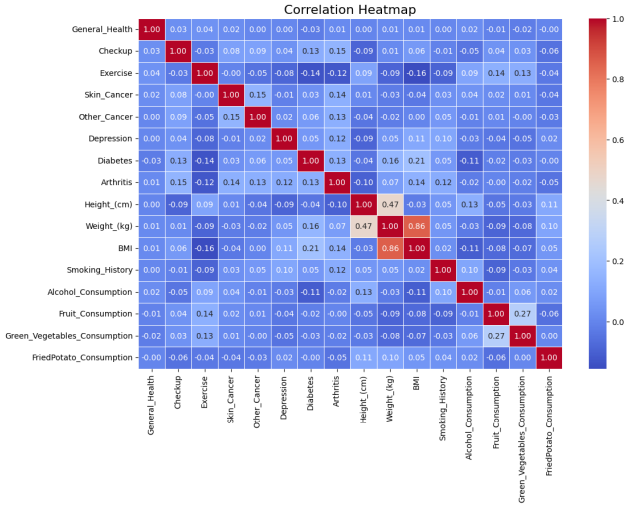

Fig. 1. Features Correlation Heatmap.

D. Model Development

The core of this research is centered on developing a predictive model using the XGBoost algorithm. XGBoost (Extreme Gradient Boosting) is a powerful ensemble learning method that builds models in a stage-wise manner using decision trees. The choice of XGBoost was based on several factors:

- Performance: XGBoost has consistently outperformed other algorithms in various machine learning competitions, particularly in classification tasks.
- Flexibility: It handles both numerical and categorical data and can model complex nonlinear relationships
- Regularization: XGBoost has built-in regularization techniques, such as L1 and L2, thus preventing overfitting, one of the common problems of machine learning models

The model development process included the following steps:

1) Hyperparameter Tuning: Bayesian optimization was employed to fine-tune the hyperparameters of the XG-Boost model. Key hyperparameters optimized include:
    - Max Depth: This regulates the maximum depth of the trees.

- Learning Rate: Step size of each iteration while moving toward the minimum of the loss function.
- Number of Estimators: The total number of trees in the model.
- Gamma: Regularization parameter used to control whether adding a new tree should be done with respect to the reduction of loss.

Bayesian optimization had been chosen because it works effectively in the investigation of the hyperparameter domain and manages to provide an optimal setting using fewer iterations than would be required by either Grid or Random Search.

2) raining and Test Split: The dataset was split into sets for training and testing using an 80-20% ratio. This split ensures that the model gets trained on one section of the data while being validated on an unobserved set, which is critical for measuring the model's generalization capabilities

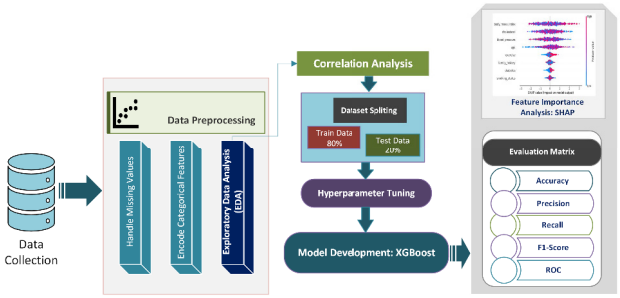

Fig. 2. Workflow diagram of the methodology process.

3) Model Architecture: The architecture of the XGBoost model consists of multiple decision trees combined to improve accuracy. The final prediction is generated by aggregating the outputs of all the trees:

$$\hat{y} = \sum_{k=1}^{n} \alpha_k \cdot f_k(x) \tag{1}$$

Where: $\hat{y}$ is the predicted value. α_k are the weights for each tree. $f_k(x)$ represents the k-th decision tree applied to the input feature vector x.

E. Evaluation Metrics

To evaluate the model's performance, several metrics were defined:

- Accuracy: The proportion of true results (both true positives and true negatives) among the total number of cases examined [23], [28]. It is calculated as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

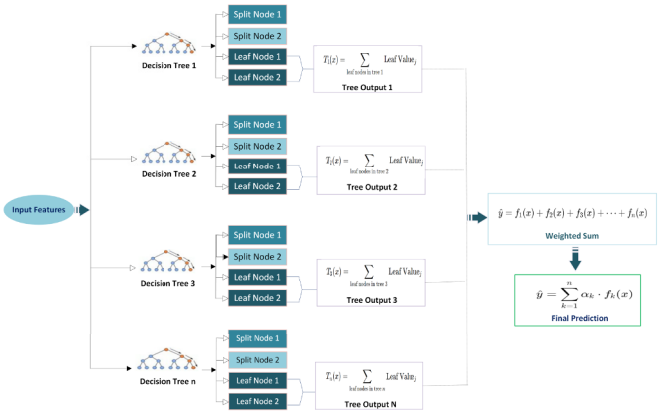

Fig. 3. XGBoost Model Architecture

- Precision: The ratio of true positive predictions to the total predicted positives [24]. It indicates the model's ability to avoid false positives.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3}$$

- Recall (Sensitivity): The ratio of true positive predictions to the total actual positives, reflecting the model's ability to identify positive cases [25].

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4}$$

- F1-Score: The harmonic mean of precision and recall, providing a balance between the two metrics [26], [29].

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

- AUC-ROC Score: The AUC-ROC curve (Area Under the Receiver Operating Characteristic Curve) provides a visual evaluation of a classifier's performance across various threshold settings. The AUC value estimates the likelihood that a classifier would assign a greater rating to a positive instance chosen at random than to a randomly selected negative instance [27]. The ROC curve itself is a plot of the True Positive Rate (Recall) against the False Positive Rate (FPR), where FPR is defined as:

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \tag{6}$$

The AUC serves as a single scalar metric summarizing the overall effectiveness of the model.

These metrics provide a comprehensive understanding of the model's performance and its effectiveness in predicting cardiovascular disease risk.

IV. Results and Analyses

A. Model Performance

The efficacy of the XGBoost model was tested using multiple measures to assess its usefulness in predicting cardiovascular illness. The following evaluation metrics were computed:

TABLE I
MODEL EVALUATION METRICS

Metric	Score
Accuracy	97.48%
Precision	97.56%
Recall	98.36%
F1 Score	97.96%
ROC-AUC	0.98

The confusion matrix (Fig. 4) provides a detailed breakdown of the model's predictions. It shows that the model correctly identified 6830 positive cases and 6816 negative cases while misclassifying 170 positive cases as negative and 184 negative cases as positive.This information demonstrates the model's performance and its ability to decrease false positives and negatives.

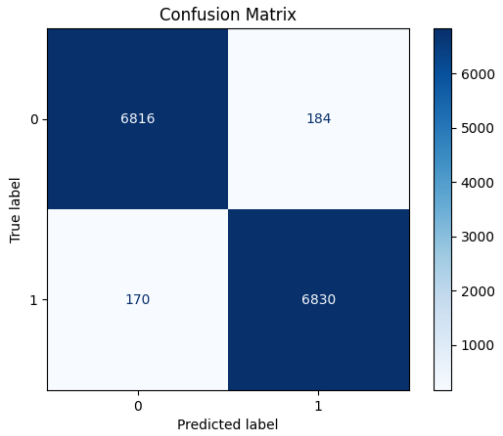

Fig. 4. Confusion Matrix illustrating the model's performance in predicting heart disease.

Additionally, the accuracy plot during hyperparameter tuning (Fig. 5) reveals fluctuations in accuracy across iterations, indicating the sensitivity of the model to hyperparameter adjustments. The optimization process led to an improved accuracy score, demonstrating the importance of hyperparameter tuning in enhancing model performance.

The ROC curve (Fig. 6) further illustrates the model's performance across different thresholds. The curve's proximity to the top left corner indicates a high true positive rate while maintaining a low false positive rate.

B. Comparison with Existing Models

For a comprehensive analysis, we compared the XGBoost model with other heart disease prediction models from recent studies. Table II highlights key performance metrics across different models, emphasizing the efficiency of our XGBoost-based approach.

C. Feature Importance Analysis

Moreover, SHAP (SHapley Additive exPlanations) values were utilized to interpret the model outputs further. The SHAP

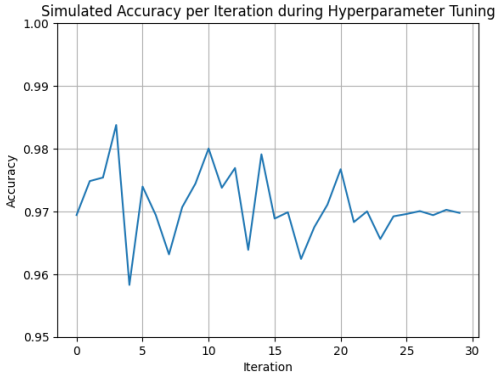

Fig. 5. Simulated accuracy per iteration during hyperparameter tuning.

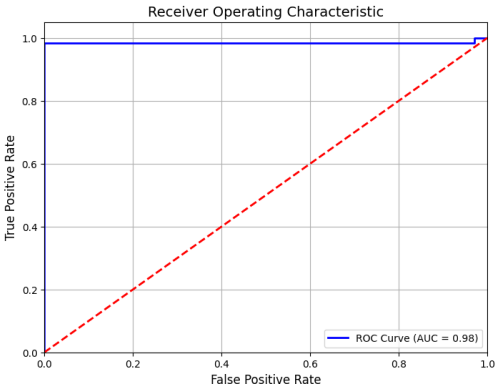

Fig. 6. ROC Curve demonstrating the model's ability to distinguish between positive and negative classes. AUC = 0.98.

summary plot (Fig. 7) shows the effect of every feature on the model prediction.

Fig. 7. SHAP summary plot showing the impact of features on the model output.

The SHAP value analysis gives a comprehensive view of how each feature affects the prediction and can give practical insights to healthcare practitioners

V. CONCLUSION

This study demonstrates the efficiency of the XGBoost algorithm in predicting heart disease by taking the benefit

TABLE II
COMPARISON OF HEART DISEASE PREDICTION MODELS

Study	Model Type	Accuracy	Reference
Current Study	XGBoost	**97.48%**	This Study
Sarmah (2020)	IoT-Enhanced Deep Learning (DLMNN)	95.87%	[5]
Bharti et al. (2021)	Hybrid ML & Deep Learning	94.2%	[6]
Takci (2018)	SVM with ReliefF Feature Selection	84.81%	[8]
Jindal et al. (2021)	Random Forest, CART	87.6%	[9]
Pasha et al. (2020)	Deep Learning (CNN)	92%	[15]
Mehmood et al. (2020)	CNN-based	94.60%	[12]
Kavitha et al. (2021)	Hybrid RF + Decision Tree	88.7%	[11]

of its high accuracy, interpretability, and non-linearity handling capability. After rigorous data pre-processing, feature selection, and hyperparameter tuning, the XGBoost model gave very promising results with an accuracy of 97.48% and AUC score of 0.98, hence reliable at a clinical level. The feature importance analysis extracted the most informative health and lifestyle predictors of cardiovascular risk, including cholesterol, blood pressure, and exercise, which were also validated as critical risk predictors by SHAP values. This interpretability supports clinical insight to enable practitioners to focus on modifiable risk factors and work out prevention strategies.

These findings indicate the potential of XGBoost and other machine learning models in the advancement of diagnostics related to cardiovascular diseases and are thus a very promising avenue toward the integration of predictive analytics in health. The model could be scaled to accommodate a wide variety of datasets, refinement of feature engineering techniques, and be clinically integrated in real time. As these approaches improve early diagnosis, all these contribute toward personalized medicine, improving patient outcomes and probably reducing healthcare costs associated with CVDs.

VI. ACKNOWLEDGEMENT

The authors would like to acknowledge the Research Management Centre, Multimedia University, Malaysia, for their invaluable research support

REFERENCES

[1] M. Diwakar, A. Tripathi, K. Joshi, M. Memoria, and P. Singh, "Latest trends on heart disease prediction using machine learning and image fusion," *Materials Today: Proceedings*, vol. 37, pp. 3213-3218, 2021.

[2] S. Krishnan and S. Geetha, "Prediction of heart disease using machine learning algorithms," in *2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT)*, 2019, pp. 1-5.

[3] A. Dutta, T. Batabyal, M. Basu, and S. T. Acton, "An efficient convolutional neural network for coronary heart disease prediction," *Expert Systems with Applications*, vol. 159, p. 113408, 2020.

[4] L. Ali, A. Rahman, A. Khan, M. Zhou, A. Javeed, and J. A. Khan, "An automated diagnostic system for heart disease prediction based on χ^2 statistical model and optimally configured deep neural network," *IEEE Access*, vol. 7, pp. 34938-34945, 2019.

[5] S. S. Sarmah, "An efficient IoT-based patient monitoring and heart disease prediction system using deep learning modified neural network," *IEEE Access*, vol. 8, pp. 135784-135797, 2020.

[6] 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, p. 8387680, 2021.

[7] S. Mall, "Heart Attack Prediction using Machine Learning Techniques," in *2024 4th International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, 2024, pp. 1778-1783.

[8] H. Takci, "Improvement of heart attack prediction by the feature selection methods," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 26, no. 1, pp. 1-10, 2018.

[9] H. Jindal, S. Agrawal, R. Khera, R. Jain, and P. Nagrath, "Heart disease prediction using machine learning algorithms," in *IOP Conference Series: Materials Science and Engineering*, vol. 1022, no. 1, p. 012072, 2021.

[10] R. Katarya and S. K. Meena, "Machine learning techniques for heart disease prediction: a comparative study and analysis," *Health and Technology*, vol. 11, no. 1, pp. 87-97, 2021.

[11] M. Kavitha, G. Gnaneswar, R. Dinesh, Y. R. Sai, and R. S. Suraj, "Heart disease prediction using hybrid machine learning model," in *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, 2021, pp. 1329-1333.

[12] A. Mehmood et al., "Prediction of heart disease using deep convolutional neural networks," *Arabian Journal for Science and Engineering*, vol. 46, no. 4, pp. 3409-3422, 2021.

[13] A. Saboor, M. Usman, S. Ali, A. Samad, M. F. Abrar, and N. Ullah, "A method for improving prediction of human heart disease using machine learning algorithms," *Mobile Information Systems*, vol. 2022, no. 1, p. 1410169, 2022.

[14] I. Javid, A. K. Z. Alsaedi, and R. Ghazali, "Enhanced accuracy of heart disease prediction using machine learning and recurrent neural networks ensemble majority voting method," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 3, 2020.

[15] S. N. Pasha, D. Ramesh, S. Mohmmad, and A. Harshavardhan, "Cardiovascular disease prediction using deep learning techniques," in *IOP Conference Series: Materials Science and Engineering*, vol. 981, no. 2, p. 022006, 2020.

[16] H. Sharma and M. A. Rizvi, "Prediction of heart disease using machine learning algorithms: A survey," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 5, no. 8, pp. 99-104, 2017.

[17] N. M. Lutimath, C. Chethan, and B. S. Pol, "Prediction of heart disease using machine learning," *International Journal Of Recent Technology and Engineering*, vol. 8, no. 2, pp. 474-477, 2019.

[18] S. Nikhar and A. M. Karandikar, "Prediction of heart disease using machine learning algorithms," *International Journal of Advanced Engineering, Management and Science*, vol. 2, no. 6, 2016.

[19] S. K. Gupta, A. Shrivastava, S. P. Upadhyay, and P. K. Chaurasia, "A machine learning approach for heart attack prediction," *International Journal of Engineering and Advanced Technology*, vol. 10, no. 6, pp. 124-134, 2021.

[20] M. M. Ali et al., "Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison," *Computers in Biology and Medicine*, vol. 136, p. 104672, 2021.

[21] V. Sharma, S. Yadav, and M. Gupta, "Heart disease prediction using machine learning techniques," in *2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, 2020, pp. 177-181.

[22] https://www.kaggle.com/datasets/alphiree/cardiovascular-diseases-risk-prediction-dataset

[23] Powers, D. M. (2020). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *arXiv preprint arXiv:2010.16061*.

[24] Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 427-437.
[25] Davis, J., & Goadrich, M. (2006, June). The relationship between Precision-Recall and ROC curves. In *Proceedings of the 23rd International Conference on Machine Learning* (pp. 233-240).
[26] Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21, 1-13.
[27] Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861-874.
[28] Fahad, N., Sen, A., Jisha, S. S., Ahmad, S., Mokhlis, H., & Hossain, M. S. (2023, December). Identification of Human Movement Through a Novel Machine Learning Approach. In *2023 Innovations in Power and Advanced Computing Technologies (i-PACT)* (pp. 1–5). IEEE.
[29] Hossain, M. N., Fahad, N., Ahmed, R., Sen, A., Al Huda, M. S., & Hossen, M. I. (2024). Preventing Student's Mental Health Problems with the Help of Data Mining. *International Journal of Computing*, 23(1), 101–108. https://doi.org/10.47839/ijc.23.1.3441