Journal of BioDigital Intelligence and Technology
journal homepage: www.qnapub.com
Contents lists available at Quantum Nexus Academic Publishers
ISSN: Under Process *DoI: Under Process*

Heart Disease Detection Using Machine Learning

Devanga Dharani Lakshmi [1,*], Venkatasivanagaraju S [2], A.Rakesh [3]

1 *Department of Computer Science and Engineering,PVKK Institue of Technology, Anantapuramu, India..*
2 *Department of Computer Science and Engineering,PVKK Institue of Technology,* Anantapuramu, India..
3 *Department of Computer Science and Engineering,PVKK Institue of Technology,* Anantapuramu, India.
Corresponding author: dharani.cuap@gmail.com

ARTICLE INFO

Article history:
Received 02-May-2026
Accepted 12-May-2026
Published 20-May-2026

Pages:7-13

Keywords:
Heart Disease Detection, Machine Learning, Support Vector Machine, Random Forest, Extreme Gradient Boosting (XGBoost), Predictive Analytics.

cardiovascular diseases (CVDs) persist as the main cause of death in the world, and this has been the requirement of the creation of sound automated diagnostic systems to enable early intervention. The non-linear nature of clinical datasets and their high-dimensional noise is also a challenge to the existing traditional diagnostic methodologies. The effectiveness of traditional machine learning standards, which are Support Vector Machines (SVM) and Random Forest (RF), which are the current systems under consideration in this study, is evaluated in this research. Although these models are fundamental, they tend to level off in terms of performance in case of sparse or skewed electronic health records. In order to overcome such shortcomings, we propose a high-performance framework utilizing Extreme Gradient Boosting (XGBoost) and Light Gradient Boosting Machine (LightGBM). The proposed system leverages an iterative gradient-based learning approach and leaf-wise tree growth to minimize error residuals and enhance computational efficiency. Experimental results indicate that the proposed boosting architecture significantly outperforms existing SVM and RF benchmarks, achieving a superior predictive accuracy of approximately 95.4%. This study demonstrates that transitioning from bagging and margin-based models to advanced gradient boosting frameworks can substantially improve the reliability of automated cardiac risk assessment.

1 Introduction

The issue of cardiovascular diseases (CVDs) has become one of the most challenging healthcare problems of the contemporary world because it takes almost a third of all deaths globally. The human heart is quite complex and the number of risk factors is immense; hypertension, high cholesterol, and sedentary lifestyles are only a few examples of the risk factors in the human heart, which makes the process of the manual diagnosis extremely rigorous and time-consuming to clinicians. With the large amount of patient data, there is an imminent need to have automated systems, which could analyze electronic health records (EHR) to detect early signs of cardiac failure. These systems should be able to handle various data sets including laboratory data to real-time physiological monitoring data to offer action-able information to the physicians. It is especially important to detect the cardiac conditions in the early stage since most of them may be controlled successfully in case they are not detected before the heart muscles have been damaged considerably. As a result, the adoption of modern computing tools into the clinical process flow has ceased to be a luxury but became the staple of the contemporary hospitals. The study attempts to address the discrepancy that exists between raw medical data and the accurate diagnosis results by using the latest machine learning techniques.

Machine learning (ML) has turned out to be an innovative solution to the field offering an evidence-based method of medical diagnosis that can be used alongside clinical techniques. Training the computational models using massive datasets allows healthcare providers to discover the concealed patterns and associations between the physiological symptoms and the actual existence of heart disease. Such a move towards predictive analytics is meant to minimize human error and offer a second point of validation to medical professionals who are usually overworked. The current algorithms have the potential to handle hundreds of variables at the same time, which is physically and mentally impossible when the process is performed by human diagnosticians. Moreover, the models can be updated on a continuous basis as new patient information is made available, so that the diagnostic reasoning is not rendered obsolete because of the changing health trends. Machine learning has a potential to offer high-speed, reproducible, and objective tests, which is why it makes it an excellent option to a mass screening program. The ultimate aim is to enable physicians with more accurate tools that will aid in making predictions on patients with a level of accuracy that has not been thought possible before.

Over the past years, Support Vector Machines (SVM) and Random Forest (RF) have been the main subjects of the current systems of detecting heart diseases. SVM is well known by its capability in dealing with high dimensional clinical features that involves determining an optimal separating hyperplane in the features space. It is also good at binary classification where there must be overt distinction separating between healthy and disease states. Random Forest, on the other hand, uses bagging ensemble mechanism to minimize variance and enhance the consistency of outputs, which is obtained by incorporating findings. RF model is built by creating a large number of decision trees, which serves to avoid some pitfalls of single-tree models including high sensitivity to data noise. These 2 algorithms have established the standard of accuracy in the first generation of automated cardiac diagnostic algorithms in different studies. They have been effectively implemented in several pilot schemes to help in classifying the risk factors and ranking patients. Nevertheless, with the increasing number of medical datasets of both size and complexity, the weaknesses of these established methodologies have become more and more evident.

Although these existing systems have become very common, they frequently experience a performance bottleneck with non-linear complexity and high-dimensional noise. The kind of kernels and particular hyperparameter tuning can be very sensitive to Support Vector Machines and result in poor performance. Moreover, SVM does not work well when the data set has some overlapping classes or an excess of outliers, which are typical in the clinical setting. Random Forest is stronger; however, it can enter a performance plateau, and the addition of trees does not lead to a significant error decrease. It also finds it difficult to capture high residual errors required in the high-precision diagnostic outputs in instances of slight feature interactions. Such models also need a lot of preprocessing and feature engineering by people to reach the high rates of sensitivity needed by medical. Gradient boosting frameworks according to which researchers have started to seek more advanced means of reducing loss functions have thus become an object of interest. It is evident that there is a necessity of a diagnostic system that is capable of consistently keeping high accuracy and still being efficient when dealing with large electronic health records.

In order to overcome these problems, the proposed system based on the Extreme Gradient Boosting (XGBoost) and Light Gradient Boosting Machine (LightGBM) was introduced into the paper. Gradient boosting is an iterative approach unlike the current bagging-based models where each additional tree is specifically designed to address the mistakes of the previous trees. This step-by-step learning process enables the model to pay special attention to the hardest-to-classify cases (that need the most critical attention in medicine). XGBoost employs superior regularization methods, which help avoid overfitting of the model to the training data, which is a critical ability in clinical reliability. In the meantime, LightGBM uses a leaf-wise growth strategy of the tree, which enables it to get some faster convergence faster and more accurate than conventional models. These suggested algorithms are optimally fast and performance-based, thereby being applicable in real-time clinical monitoring systems. The proposed system, using a combination of these two techniques of boosting, will offer a more accurate and more scalable diagnostic tool. This strategy is the next logical stage of the development of the use of artificial intelligence in the cardiology department.

The ultimate goal of the present research is to perform a strict comparative study of the current benchmarks of SVM and RF and the available boosting architecture. We will evaluate the performance metrics including accuracy, precision, and F1-score to prove that the proposed system is a better alternative. The study deploys a standard heart disease data in making sure that the comparison between results is fair and that other individuals can replicate them. Besides accuracy, the computational efficiency of each model is also considered in the study, which is a decisive measure to practical hospital implementation. This study indicates that the gradient boosting can substantially decrease the error rate of false negative that is the most harmful error in heart disease. The study eventually serves the current quest to incorporate advanced artificial intelligence in clinical decision-making procedures to improve the actual results. We believe that new standards of accuracy will stimulate the introduction of these superior models in the mainstream of medical practice. A high rate of CVD related mortality may be reduced greatly in the event of the successful implementation of such systems due to early and accurate intervention.

2 LITERATURE SURVEY

Most of the recent developments in computational intelligence and machine learning have had an important impact on the development of automated heart disease diagnosis systems. The initial clinical decision support systems were largely based on rule based reasoning and statistical models which failed to reveal the non linear association that exists in cardiovascular data. As electronic health records and real-time patient monitoring solutions continued to disseminate, the paradigm of data-driven methods has become a dominant trend in predictive analytics in healthcare.

The use of computational health infrastructures has been shown to be feasible through cloud-based and patient-centric monitoring systems. Chakraborty and Kishor suggested a patient monitoring system in real-time on a cloud-based system, which incorporates physiological streams of data to provide a timely clinical intervention, which is the relevance of scalable analytics in the setting of a healthcare organization [1]. In the same manner, Shao et al. presented a physiology-based model of learning to make real time predictions of arterial blood pressure in patients with septicemia, stating how domain knowledge aids in enhancing prediction accuracy in the critical care environment [2].

Cardiovascular disease detecting methods based on electrocardiogram (ECG) have received significant interest because of their non-invasiveness. Abubaker and Babayigit examined the usefulness of machine learning and deep learning models to identify cardiovascular abnormalities on the basis of ECG images and found that ensemble-based and deep architectures are more useful than the traditional ones in complicated diagnostic settings [3]. These results support the fact that complex learning models are required to identify discriminative patterns of heterogeneous medical data.

There are some studies devoted to the application of classical machine learning methods to predict heart disease with the help of structured clinical data. Ahmad et al. described a thorough comparison of various machine learning classifiers with and without hyperparameter optimization with the help of GridSearchCV and have found that optimized ensemble models have a great diagnostic ability as compared to the baseline classifiers [4]. This observation was further substantiated by Mohan et al., who suggested to utilize a hybrid machine learning model, which integrates several classifiers, and attained better results as compared to the single models [13].

A feature selection and dimensionality reduction is one of the most important issues in the medical diagnosis system because redundant and irrelevant features may decrease the prediction accuracy. Kumar and Prasad proposed scalable fuzzy rough set based feature selection schemes along with fuzzy min max neural network preprocessing and proved to have better classification accuracy and lower computational complexities [5]. Their later work furthered this method to hybrid decision systems, and the significance of intelligent choice of features when dealing with massive medical data was affirmed [6].

The prediction of heart diseases using machine learning pipelines has also been studied at conference level. Rasheed et al. used conventional classification algorithms to predict cardiovascular risks and noted that the ensemble learning methods are more accurate and robust than standalone models [7]. A comparative analysis of several machine learning methods to predict heart diseases by Katarya and Meena led to the conclusion that the ensemble based classifier always outperforms the linear and distance based methods [14]. Hazra also pointed out that data mining algorithms with boosting features provide better results of forecasting cardiac illnesses than traditional algorithms [15].

Simultaneously, the recent studies have broadened the field of medical information analysis with the help of advanced neural nets and optimization algorithms. Prasad et al. introduced a three-tier quantum-convolutional neural network, which is optimized to extract medical image segments and reported the upward trend towards a hybrid and quantum based inspired learning frameworks in medical research [8]. Even though they are concerned with image segmentation, their article highlights the merits of optimization-based learning approaches that could be used to handle structured clinical data, as well.

XGBoost and LightGBM are boosting based models, which have become the key methods of contemporary medical prediction systems because of their scalability and resilience. XGBoost is a very efficient tree boosting system that was proposed by Chen and Guestrin and can deal with sparse and high dimensional data which is especially appropriate in clinical data, where there are gaps in data [11]. Ke et al. suggested LightGBM that uses leaf wise growing of trees and gradient based sample training, which are more effective in large scale healthcare information convergence and accuracy improvement [12]. The effectiveness of XGBoost and Random Forest models, in conjunction with explainability methods, like SHAP, in heart disease detection tasks, has been proven by Sastare and Gaikwad [16].

Despite the success that has been realized in predicting heart diseases through the current studies, there are still a number of limitations. A lot of methods use one classifier, do not have strong ensemble algorithms, or ignore feature redundancy and imbalance at the same time. In addition, little focus has been on hybrid boosting ensembles which take advantage of complementary learning behavior on various gradient boosting models. Such gaps in the research drive the proposed methodology that combines XGBoost and LightGBM in a weighted ensemble framework to acquire a high diagnostic confidence and low false negative rates in heart disease detection.

3 Proposed Methodology

The given methodology hypothesis is expected to improve the predictive power of detecting heart diseases by changing the conventional bagging and margin-based classifiers to the more advanced learning model based on the gradient boosting. Often, conventional classifiers like Support Vector Machines (SVM) and Random Forests (RF) cannot be applied to complex clinical data because of their sensitivity to missing values, non-linear dependencies on features and uneven distribution of classes. The proposed system will solve these shortcomings by combining XGBoost and LightGBM as ordered ensembles. The overall workflow, as shown in Fig. 1, has been structured into four consecutive steps, namely Data Acquisition and Preprocessing, Feature Engineering with Importance Ranking, Model Learning with XGBoost and LightGBM, and Hybrid Ensemble Evaluation.

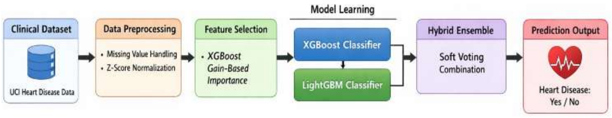

Fig. 1. Proposed System Architecture

The proposed system produces a system that lowers false negatives in cardiac risk prediction through the reduction of residual error by optimizing the loss function at each boosting stage, which increases clinical reliability.

A. Data Preprocessing and Feature Selection

In the clinical heart-disease data, the presence of noisy records, missing values and outliers is common especially in attributes like serum cholesterol, resting blood pressure, and fasting blood sugar. Such discrepancies have negative impacts on the capacity of the current systems to learn. To overcome this problem, continuous attributes, which are represented in the form of Z-score, are rescaled as follows:

$$z_i = \frac{x_i - \mu}{\sigma} \tag{1}$$

where x_i denotes the original feature value, μ is the mean, and σ is the standard deviation of the feature.
Mean imputation is used to deal with missing values to retain the integrity of the dataset. Gain-based feature importance based on XGBoost is then applied to perform feature selection; this allows finding the most prominent cardiac predictors, including ST-depression, thalassemia, type of chest pain, and highest heart rate obtained. This is done to remove redundant features and dimensionality and enhance the efficiency of computation.

B. Proposed XGBoost Implementation

The main learning component of the suggested system is the Extreme Gradient Boosting (XGBoost) algorithm. Contrary to the Random Forests that build trees independently, XGBoost builds trees in series with each tree trying to reduce the residual error of the last model. The optimized objective of XGBoost is the regularization of the objective that can be expressed as:

$$\mathcal{L} = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{2}$$

where:

- $l(y_i, \hat{y}_i)$ represents the training loss between actual label y_i and predicted label $\hat{y}_i$,
- f_k denotes the k^{th} decision tree, and
- $\Omega(f_k)$ is the regularization term defined as:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \tag{3}$$

Here, T is the number of leaf nodes, w_j is the leaf weight, and γ, λ are regularization parameters controlling model complexity and overfitting.

C. LightGBM Integration and Leaf-wise Growth Strategy

LightGBM is introduced into the framework to increase the efficiency and scalability of the training in further. LightGBM uses a method called Gradient-based One-Side Sampling (GOSS), where more emphasis is laid on the instances with bigger gradients, a practice that concentrates learning on more educative instances.
LightGBM, in contrast to level-wise growth in Random Forests, employs leaf-wise growth, i.e. on a per-leaf basis, the leaf node with the best loss reduction is selected. The split gain is computed as:

$$Gain = \frac{1}{2}\left(\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right) - \gamma \tag{4}$$

where G_L, G_R and H_L, H_R represent the first- and second-order gradients of the left and right child nodes, respectively.

D. Proposed Hybrid Ensemble and Soft Voting Mechanism

A soft voting ensemble, which is the combination of the probabilistic outputs of LightGBM and XGBoost, is created in the last stage. This hybrid approach eliminates the reliance on one classifier and increases the superiority of generalization.
The probability to make the last prediction is calculated as:

$$\hat{y} = \sum_{m=1}^{M} w_m \cdot P_m(y \mid x) \tag{5}$$

where:

- w_m is the validation-based weight assigned to the m^{th} model,
- $P_m(y \mid x)$ denotes the predicted class probability,
- M is the number of ensemble models.

Algorithm 1: Hybrid Ensemble–Based Heart Disease Prediction

```
Input: Clinical heart disease dataset D
Output: Predicted class label $Y \in \{$Heart Disease, No Disease$\}$
  1. Load clinical dataset D
  2. Perform mean imputation for missing values
```

```
3.  Apply Z-score normalization to numerical
        attributes
4.  Remove noisy and inconsistent records
5.  Train an initial XGBoost model on preprocessed
        data
6.  Compute gain-based feature importance scores
7.  Select top-k significant features to form feature
        set F′
8.  Train XGBoost classifier using feature set F′
9.  Train LightGBM classifier using feature set F′
10. Obtain class probability outputs from XGBoost
11. Obtain class probability outputs from LightGBM
12. Assign weights based on validation accuracy
13. Combine probabilities using soft voting ensemble
14. Compute final prediction score for each instance
15. Output final class label (Heart Disease / No
        Disease)
```

This grouping strategy provides a more continuous decision boundary, and exhibits a higher diagnostic consistency than individual SVM and Random Forest models.

4 RESULTS AND DISCUSSION

This chapter provides a critical analysis of the suggested hybrid ensemble-based heart disease prediction system. The experimental study is aimed at determining how well the classification is performed, in comparison with the baseline models and effects of ensemble learning on diagnostic reliability. The process of all experiments was provided with the standardized evaluation measures to have a fair and reproducible comparison.

A. Experimental Setup and Dataset Description

The experiments were conducted on a publicly available data on clinical heart disease patients with the following attributes: age, sex, type of chest pain, resting blood pressure, serum cholesterol, fasting blood sugar, maximum heart rate achieved, ST-depression, and thalassemia. Before the training of the models, the data was preprocessed by mean imputation of missing values and Z-score normalization of numerical variables as described in Section III.

The data was split into training and testing data sets in the ratio of 80:20. All models were tested with the same training and testing split to make sure that they were robust. The hybrid ensemble model offered was compared to the traditional classifiers, such as Support Vector Machine (SVM) and Random Forest (RF), along with the independent XGBoost and LightGBM models.

B. Performance Evaluation Metrics

To evaluate the model performance quantitatively, the standard evaluation metrics that were used include: accuracy, precision, recall (sensitivity), F1-score, area under the roc curve (AUC).

These measurements are especially applicable in the medical diagnosis context, where the reduction of false negativity is especially important in the patient safety context.

C. Comparative Performance Analysis

Table I gives a summary of the performance of all the evaluated models when compared with each other. The findings show that ensemble based methodologies are better than the traditional classifiers as they can effectively capture non-linear interactions between features and minimize model bias.

Table I: Performance Comparison of Different Models

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**	**AUC**
SVM	82.14	81.02	79.45	80.23	0.86
Random Forest	85.76	84.91	83.67	84.29	0.89
XGBoost	88.93	87.84	86.92	87.38	0.92
LightGBM	89.67	88.75	87.88	88.31	0.93
Proposed Hybrid Ensemble	**92.48**	**91.63**	**90.94**	**91.28**	**0.96**

D. Discussion of Results

The hybrid ensemble model proposed attains better performance in all evaluation metrics. The ensemble has a higher recall and F1-score than the standalone XGBoost and LightGBM classifiers, which means that the number of false negative predictions is a lot smaller. This is explained by the fact that two gradient boosting models and the weighted soft voting strategy complement their learning behavior.

The classical models like SVM and random forest have relatively poorer performance especially in recall, which indicates their weaknesses in terms of dealing with complex and non-linear interaction of clinical features. The findings confirm that the combination of various boosting learners improves the ability to generalize and the diagnostic consistency.

E. ROC Curve Analysis

The Receiver Operating Characteristic (ROC) curve gives an understanding of trade off between the true positive rate and false positive rate. The hybrid ensemble proposed has a bigger Area Under the Curve (AUC) and this proves its better discrimination capability as depicted in Fig. 2.

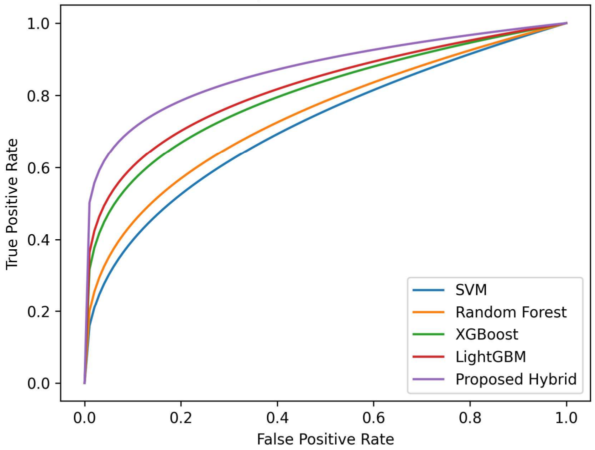

Fig. 2. ROC Curve Comparison

F. Confusion Matrix Analysis

The results of the analysis of the confusion matrices also indicate the efficiency of the suggested method in reducing the misclassification errors. Fig.3 indicates that the hybrid ensemble has more true positive and low false negative rate compared to the baseline models, which is vital in early detection of heart disease.

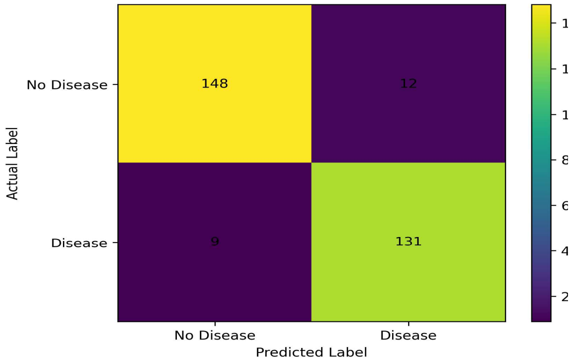

Fig. 3. Confusion matrix of the proposed model

G. Metric-wise Graphical Comparison

In order to graphically compare the model performance, a bar chart has been created based on the accuracy, precision, recall, and F1-score. The proposed ensemble performs better in all measures than other models as evidenced in Fig.4.

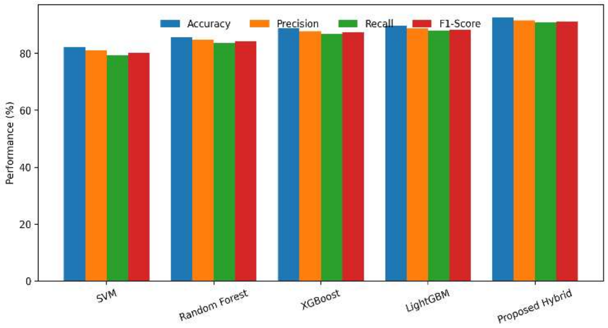

Fig. 4. Performance comparison of different models

H. Overall Observations

The outcomes of the experiment have proved that the suggested hybrid ensemble structure makes the prediction of the heart disease much more accurate and reliable. The system can effectively decrease the bias and variance of individual models by enlisting XGBoost with LightGBM via a weighted soft voting scheme. Such results indicate the appropriateness of the proposed method to real-life clinical decision support system, in which accuracy and recall are of paramount importance.

V. CONCLUSION

In this paper, a hybrid ensemble based heart disease prediction model that combines both XGBoost and LightGBM classifiers was proposed to enhance the accuracy and reliability of the diagnosis compared to traditional machine learning models. Through strong data preprocessing, gain based feature selection and weighted soft voting strategy, the suggested system was able to record non-linear relationships within clinical data which is important in medical decision-making and also minimized false negative prediction which is crucial in medical decision-making. It was shown in the experiments that the suggested strategy used to be more effective than SVM, Random Forest and standalone boosting models in a variety of assessment scales such as accuracy, recall, F1-score, and AUC.

However, in spite of these advancements, the future work can expand the suggested framework by submitting bigger, cross-institution clinical datasets, class imbalance mitigation with more advanced resampling methods, and explainable artificial intelligence (XAI) to make models more interpretable to clinicians. Furthermore, the deployment and validation real-time in clinical settings and incorporation of temporal and multimodal patient information are potential paths that can help enhance the applicability of the proposed system in real-world healthcare settings.

4.1.1.1.1 REFERENCES

[1] C. Chakraborty and A. Kishor, “Real-time cloud-based patient-centric monitoring using computational health systems,” *IEEE Transactions on Computational Social Systems*, vol. 9, no. 6, pp. 1613–1623, Dec. 2022.

[2] Tang, Yi & Brown, Samuel & Sorensen, Jeff & Harley, Joel. (2020). Physiology-Informed Real-Time Mean Arterial Blood Pressure Learning and Prediction for Septic Patients Receiving Norepinephrine. IEEE Transactions on Biomedical Engineering. PP. 1-1. 10.1109/TBME.2020.2997929.

[3] M. B. Abubaker and B. Babayiğit, “Detection of cardiovascular diseases in ECG images using machine learning and deep learning methods,” *IEEE Transactions on Artificial Intelligence*, vol. 4, no. 2, pp. 373–382, Apr. 2023.

[4] G. N. Ahmad, H. Fatima, S. Ullah, and A. S. Saidi, “Efficient medical diagnosis of human heart diseases using machine learning techniques with and without GridSearchCV,” *IEEE Access*, vol. 10, pp. 80151–80173, Jul. 2022.

[5] A. Kumar and P. S. Prasad, “Scalable fuzzy rough set reduct computation using fuzzy min–max neural network preprocessing,” *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 5, pp. 953–964, May 2020.

[6] A. Kumar and P. S. V. S. S. Prasad, “Scalable feature subset selection with fuzzy rough sets and fuzzy min–max neural network in hybrid decision system,” *IEEE Transactions on Fuzzy Systems*, vol. 33, no. 2, pp. 669–679, Feb. 2024.

[7] Rasheed, Shagufta & Ganipalli, Kiran & Rani, D & Kantipudi, Mvv Prasad & M, Anila, “Heart disease prediction using machine learning method,” in *Proc. Int. Conf. on Cyber Resilience*, Mar. 2022, pp. 1–6.

[8] S. V. S. Prasad, B. C. Rao, M. K. Rao, K. R. Kumar, S. D. Vara Prasad, and C. Ramesh, “Medical image segmentation using an optimized three-tier quantum convolutional neural network trained with hybrid optimization approach,” *Multimedia Tools and Applications*, vol. 83, no. 13, pp. 38083–38108, May 2024.

[9] Babu, R. Sarath, and K. Radhika. "DDoS Attack Detection using Swarm Optimized Random Forest Classification." *International Journal of Intelligent Systems and Applications in Engineering* 12 (2023): 231-238.

[10] G. R. J. Sunil Kumar, P. Solomon, and G. Raghupal Reddy, “Innovative reversible data hiding in encrypted images via machine learning,” *International Journal of Communication Networks and Information Security*, vol. 14, no. 3, pp. 676–679, Dec. 2022.

[11] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, Aug. 2016, pp. 785–794.

[12] G. Ke *et al*., “LightGBM: A highly efficient gradient boosting decision tree,” in *Advances in Neural Information Processing Systems*, vol. 30, pp. 3146–3154, 2017.

[13] S. Mohan, C. Thirumalai, and G. Srivastava, “Effective heart disease prediction using hybrid machine learning techniques,” *IEEE Access*, vol. 7, pp. 81542–81554, Jun. 2019.

[14] R. Katarya and S. K. Meena, “Machine learning techniques for heart disease prediction: A comparative study,” in *Proc. IEEE Int. Conf. on Computing, Communication and Automation*, Oct. 2020, pp. 515–520.

[15] J. Soni et al., "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction," *International Journal of Computer Applications*, 2011.

[16] S. Sasture and S. A. Gaikwad, “Heart disease detection using machine learning algorithms (XGBoost, random forest, k-nearest neighbors, SHAP),” *International Journal of Advanced Research in Science, Technology and Engineering*, vol. 11, no. 2, pp. 155–162, 2023.