



Karbala International Journal of Modern Science

Manuscript 3354

ELM and LIGHTGBM: A Hybrid Machine Learning Technique with Intelligent IOT to Predict the cardiovascular disease

Gorapalli Srinivasa Rao

G Muneeswari

Follow this and additional works at: <https://kijoms.uokerbala.edu.iq/home>



Part of the [Biology Commons](#), [Chemistry Commons](#), [Computer Sciences Commons](#), and the [Physics Commons](#)



ELM and LIGHTGBM: A Hybrid Machine Learning Technique with Intelligent IOT to Predict the cardiovascular disease

Abstract

Cardiologists can more accurately classify a patient's condition by performing an accurate diagnostic and prognosis of cardiovascular disease (CVD). The clinical diagnosis, and therapies processes within the medical field have been substantially accelerated by ML-based approaches enabled by IoT-based systems. This structure is based on IoT-based system with enabled ML approach. This study investigates an approach known as ensemble categorization, which enhances the precision of weak algorithms by integrating multiple classifiers. For effective CVD classification, we utilized Ensemble learning machine (ELM) and Light GBM. The appropriate traits are chosen to speed up the categorization process using the Gorilla Troops Optimizer technique. The investigation findings demonstrate that ensemble techniques are beneficial in improving the predictive ability of weak categorizers.

Keywords

Internet of Things (IoT); Machine Learning (ML); cardiovascular (CVD); Ensemble learning machine (ELM); Light GBM; Gorilla Troops Optimizer

Creative Commons License



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

RESEARCH PAPER

ELM and LIGHTGBM: A Hybrid Machine Learning Technique With Intelligent IOT to Predict the Cardiovascular Disease

Gorapalli S. Rao^{*}, Ganesapandian Muneeswari

School of Computer Science and Engineering, VIT-AP University, Amaravathi, India

Abstract

Cardiologists can more accurately classify a patient's condition by performing an accurate diagnostic and prognosis of cardiovascular disease (CVD), which enables them to offer the most appropriate care. With considerable outcomes, the Internet of Things (IoT) and machine learning (ML) have recently shifted the focus for usage in CVD. The methods based on machine learning made possible by IoT-based systems have significantly accelerated clinical diagnosis, continuous tracking, and therapy processes in the medical field while also reducing the workload associated with healthcare procedures and the cost of caregivers and physicians treating patients. Therefore, this research demonstrates a platform that uses information from wearable systems with sensors attached to patients' bodies to detect different illnesses in real-time. This framework uses an enabled machine learning approach on an Internet of Things system. This paper examines an ensemble classification method that combines multiple classifiers to improve the accuracy of weak algorithms. For efficient CVD classification, we combine two machine learning methods: Light GBM and Ensemble Learning Machine (ELM). To expedite the process of categorization, the Gorilla Troops Optimizer technique is used to select the relevant traits. These characteristics mean that reducing computational effort is necessary to increase categorization effectiveness. The study's findings demonstrate that ensemble techniques can improve the efficiency of weak categorizers. Up to 7% more accuracy can be achieved with inadequate classifiers when using ensemble categorization. The method's efficiency was further enhanced by adopting particular features, and the outcomes demonstrated an interesting rise in prediction reliability.

Keywords: Internet of Things (IoT), Machine learning (ML), Cardiovascular (CVD), Ensemble Learning Machine (ELM), Light GBM, Gorilla Troops Optimizer

1. Introduction

Automatic heart disease detection constitutes one of the most significant and complex health challenges in the world [1]. Specifically in adults and elderly persons, heart disease damages the body by impairing blood vessel function and leading to coronary artery infections. The World Health Organization (WHO) estimates cardiovascular diseases cause more than 18 million deaths annually [2,3]. In addition, the United States spends \$1 billion per day on treating diseases in the heart [4]. Heart illnesses, including stroke, heart attack, and hypertension, are the leading causes of death in

the United States. So, it is crucial to identify heart illness early before a heart attack or stroke occurs to successfully treat those with cardiovascular disease [5–7].

Age, obesity, high blood pressure, alcohol consumption, sex, cholesterol, smoking, poor diet, inactivity, and family history are all thought to increase the chance of developing heart disease [8,9]. Some risk factors can be changed. Along with the elements mentioned above, lifestyle choices like eating patterns, inactivity, and obesity are also regarded as significant risk factors [10,11]. Coronary, congestive, angina pectoris, cardiomyopathy, congenital, and myocarditis are just a few of the

Received 23 August 2023; revised 18 April 2024; accepted 21 April 2024.
Available online 27 May 2024

* Corresponding author.

E-mail addresses: srinivasarao.22phd7042@vitap.ac.in (G.S. Rao), muneeswari.g@vitap.ac.in (G. Muneeswari).

<https://doi.org/10.33640/2405-609X.3354>

2405-609X/© 2024 University of Kerbala. This is an open access article under the CC-BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

several types of heart disorders that might exist. It is difficult to manually estimate the chance of acquiring heart disease based on risk factors [12–14]. Nevertheless, machine learning techniques help to predict the outcomes from the existing data.

Recently, the healthcare sector has embraced machine learning to make medical decisions. Several MLAs, including the random forest (RF) [15], the chaos firefly algorithm, the multilayer perceptron (MLP), the logistic regression (LR) [16], and the backpropagation neural network (BPNN) have been successfully used as tools for evaluating options for the forecasting of CVD based on personal information [17]. Research has also demonstrated the benefits of using hybrid models that successfully predict cardiac disease. Examples are majority voting using naïve Bayes (NB), RF, Bayes net (BN), MLP, two stacked SVMs, and RF with a linear model. However, in machine learning, anomalies and unbalanced data may show up and affect how effectively the model for forecasting works [18].

Consequently, preventing significant complications requires early and prompt identification of heart issues. A precise method can easily forecast cardiac illness, allowing us to hope for early diagnosis while also assisting the medical industry in saving more lives through prediction [19]. The current approach to diagnosing the cardiovascular system is pricy and complicated. Each individual wants to have their heart condition examined repeatedly, which requires much time. Due to the need to include multiple variables, the evaluation procedure is tedious and costly [20]. This is particularly true in underdeveloped nations with limited access to diagnostic tools, medical professionals, and other resources.

As a result, we plan to incorporate an efficient and flexible clinical decision support network model for CVD diagnosis into the recommended approach. Initially, data is gathered in an Internet of Things environment. One method for predicting cardiovascular disease is the ensemble classification approach. It reduces down on computation time and improves efficiency even further.

1.1. Novel contribution

The primary essential contribution of this paper is as follows:

- In Pre-processing, first, clean the data. Data transformation is then normalized using the min–max normalization method. Then the SMOTE and Tomek methods are used to

increase the minority samples to overcome the data imbalance problem.

- The Gorilla Troops optimizer algorithm selects the essential features by eliminating unnecessary features.
- Hybrid the Ensemble Learning Machine and LightGBM techniques to predict cardiovascular disease.

1.2. Paper outline

The format of this study is as follows. Phase 2 examines existing research in this area. In Phase 3, the proposed methodology is fully explained. In Phase 4, the research using the suggested machine learning models is described in detail. Phase 5 offers suggestions for additional research as its conclusion.

2. Related works

Due to increased forecast precision and effectiveness, machine learning and artificial intelligence algorithms have grown significantly in the past decade. The potential to create and choose algorithms with the maximum accuracy and effectiveness makes this research important. Hybrid approaches are promising disease prediction methods that combine several ML models with information systems.

Khan [21] presented an IoT system allowing a more precise assessment of heart illness utilizing a Modified Deep Convolutional Neural Network (MDCNN). The patient's smartwatch and heart monitor monitor blood pressure and an electrocardiogram (ECG). The method for heart disease prediction is created using a convolution neural network. The MDCNN was employed to categorize the input data from sensors into normal and pathological. The outcomes show that the suggested methodology offers greater accuracy than the other methods.

An efficient cardiovascular disease prediction system for a CDSS was suggested by Fitriyani et al. [22] and includes XGBoost for predicting heart disease, a hybrid SMOTE-ENN to equalize the variance of data used for training, and DBSCAN to identify and remove outliers. Rani et al. [23] developed an integrated system that can help in the earlier identification of heart disease. Employing an integrative selection of features approach that blends the Genetic Approach (GA) and recursive feature removal, the pertinent features from the provided dataset were selected. The pre-processing of the data has also made use of SMOTE and traditional scalar methods. The suggested approach yields far better results than the existing approaches.

LASSO and Relief were created by Ghosh et al. [24] to use rank values to extract the most relevant features from medical sources. This aids in the resolution of machine learning's underfitting and overfitting issues as well. Moreover, several hybrid approaches, like bagging and boosting, speed up testing and decrease execution times.

Ambrish et al. [25] developed the logistic regression (LR) method to classify cardiac illness. Pre-processing the data includes cleaning the dataset, finding the missing values, and selecting features based on correlation with each attribute's intended value. The traits that have the most significant positive association were picked. The presented approach produces superior outcomes when compared to previous methods.

For better heart disease identification, Sarra et al. [26] introduced a novel model of heart disease categorization based on the support vector machine (SVM) method. The χ^2 statistical optimal feature selection method was applied to raise prediction accuracy. The performance of the suggested approach was then verified by contrasting it with conventional models based on several performance metrics. This finding indicates that our approach performed better than other cutting-edge techniques regarding heart disease prediction.

Hasanov et al. [27] have suggested a Sine Cosine Weighted K-Nearest Neighbour (SCA_WKNN) algorithm for forecasting heart disease that trains from the data recorded in the blockchain. The blockchain serves as a haven for patient data storage and a legitimate source of learning data because the data saved there cannot be altered. The suggested strategy performs better when compared to the current methods.

Malibari [28] introduced the EO-LWAMCNet, an EO-optimized lightweight automatic modulation categorization network to forecast an individual's chronic health issues accurately. Each bit of data collected by an implanted sensor in the individual's body was transmitted to the cloud via a gateway. The EO-LWAMCNet model starts the categorization procedure for predicting CVD based on the obtained sensor data. Differentiating from prior approaches, the presented approach yields greater performances.

These techniques, however, require a lot of calculation time to identify CVs and have poor prediction. For more effective and accurate early detection and better recovery and treatment, the CD identification method's prediction accuracy needs to be further refined. Therefore, the primary issues with the current approaches are their lengthy computation durations and poor accuracy,

which may be related to the dataset's use of extraneous information. Additionally, the inability of predictive algorithms for heart disease to accurately determine an individual's risk of developing the disease can lead to a false positive or false negative diagnosis. Using algorithms for prediction, some individuals may be identified as having a significant chance of developing heart disease even without any signs of the condition. Examinations and unnecessary procedures could follow, along with anxiety. For this purpose, we introduced a novel Hybrid Ensemble Learning Machine and LightGBM techniques to predict cardiovascular disease. With the help of the proposed approach, we effectively predict CVD and gain high performances with less computational time.

3. Proposed methodology

The machine learning techniques utilized in this study to create an intelligent cardiovascular disease forecasting system are discussed in this part. The three phases of this hybrid system are data gathering, data pre-processing, and developing a model. Missing values are inputted, features are chosen, and class balance is completed during pre-processing. We hybrid SMOTE and Tomek approaches to generate artificial samples for minor classes, yielding an equal amount of examples for each class. By eliminating extra features from the CVD database, the recommended Gorilla Troops Optimizer (GHO)-based dimensionality reduction strategy finds the most relevant CVD-related attributes. Finally, we ensemble ELM and LightGBM approaches to predict cardiovascular disease. The classifier concludes by predicting whether or not a person has heart disease. Fig. 1 represents the architecture of the proposed methodology. The following subsections give the specific steps and explanations of the dataset modules and performance metrics. Additionally, the suggested model's effectiveness is compared to other cutting-edge models, and the findings are shown in the outputs and discussions section.

Finally, we ensemble ELM and LightGBM approaches to predict cardiovascular disease. The classifier concludes by predicting whether or not a person has heart disease. Fig. 1 represents the architecture of the proposed methodology. The following subsections give specific steps and explanations of the datasets, modules, and performance metrics. Additionally, the suggested model's effectiveness is compared to other cutting-edge models, and the findings are shown in the outputs and discussions section.

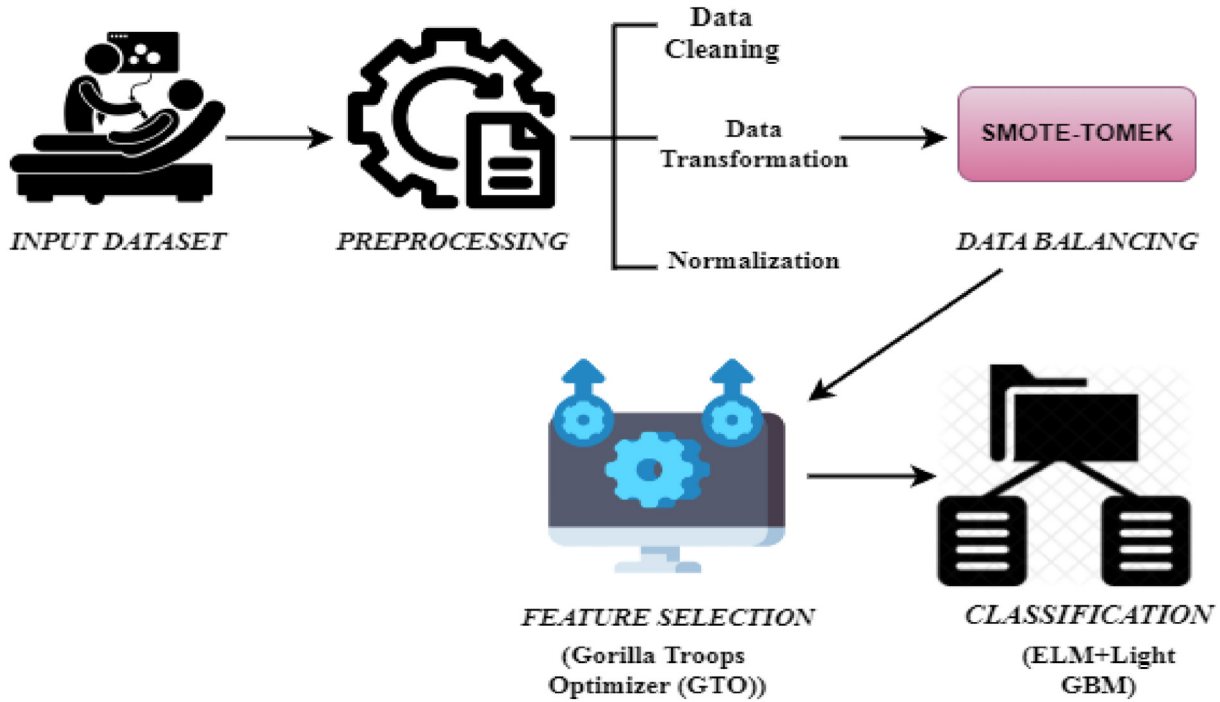


Fig. 1. Overall framework of the presented method.

3.1. Data pre-processing

Data preparation is the most crucial stage before applying Machine Learning algorithms. Since real-world data is frequently noisy, insufficient, and inconsistent, it cannot be used effectively in the prediction task. A pre-processing stage is employed to depict the data for the forecast of heart disease appropriately. Filtering for missing data, normalization, and feature selection are all parts of data preparation.

Noise and missing values are common in the gathered data after cleaning. These data must be cleansed of noise and any missing numbers filled in to produce a precise and efficient output. Transformation is the process of converting data from a single format to another so that it is easier to understand. It entails duties for aggregation, standardization, and smoothing.

3.1.1. Max–min normalization

Min-max normalization is one of the most widely used techniques for data restructuring. Min-max normalization is applied for data transformation. The result of each qualitative characteristic is transformed into a numerical value based on the minimum and maximum values. A scale of 0–1 will be applied to the data. The data are simpler to comprehend because of this homogeneity. Equation (1) is used to recognize data transformation.

$$D_t = \frac{(Z - Z_{min})}{(Z_{min_{max}})} \quad (1)$$

While Z groups the expected values, Z_{min} and Z_{max} indicates Z 's minimal and maximum values.

3.1.2. Data class balancing

Extracting and choosing the best features from a balanced dataset is crucial for a supervised ML algorithm to execute classification accurately and efficiently. It is vital to balance the category distributions during the feature selection procedure since the intelligent meter dataset used in the current study needs to be more balanced, as with most detection data sets. The SMOTETomek method is used in the present research to address this problem. The Intelligence of SMOTE and Tomek connections approaches are combined in SMOTE-Tomek to concurrently over- and under-sample data classes. Removing most class links until both have the same entities completes the indicated objective. Although the SMOTE approach can mitigate the imbalanced data class distribution problem alone, it skews the distributions due to clusters generated by several data classes that are poorly articulated in most real-world smart meter datasets. As a result, it is anticipated that during the oversampling phase of the SMOTE approach, a group of samples from the minority or majority class will predominate. As a result, model

overfitting will result from providing such biased data to the learning predictor.

3.2. Feature selection

Approaches for selecting features are crucial for the ML process since they enable the extraction of the optimal attributes for categorization. The

reducing the area of search by troubling the spaces among the gorillas to strike a balance between both exploration and exploitation and lastly proceeding toward an unknown region to increase the ability of GTO for exploration. These three techniques can all be written numerically as an equation during exploration.

$X(t)$ and $XY(t+1)$ signify the gorilla's actual and

$$XY(t+1) = \begin{cases} (UL - LL) \times r_1 + LL, \text{rand} < p(r_2 - C) \times X_r(t) + L \times H, \text{rand} \geq 0.5 \\ L \times (L \times (X(t) - XY_r(t)) + r_3 \times (X(t) - XY_r(t))), \text{rand} < 0.5 \end{cases} \quad (2)$$

execution time is also shortened as a result of this. We used the Gorilla Troops Optimizer (GTO) algorithm to choose the features.

The GTO algorithm is a desirable option for distinctive task selection due to several advantages that set it against its direct competitors. One of its key benefits is that it can be readily altered to operate in high-dimensional spaces, mainly when traversing a large search area for choosing features. The approach meticulously balances finding and exploiting novel characteristics to identify important ones, and utilizing existing ones to increase selection reliability. Motivated by the Cuckoo behaviour, GTO offers a unique biological perspective for optimization and could be especially helpful in assessing distinctive subsets.

3.2.1. Gorilla Troops Optimizer (GTO)

GTO simulates five distinct strategies by modelling the habits of swarms of gorillas. They adhere to the silverback and migrate toward an urban area in search of other gorillas. Fighting for adult females is one way to do this. Throughout the exploration phase, three strategies are employed: migration to the surviving members, migration to an undiscovered region, and migration to a known area. During the exploitative stage, two methods are used: tracking the silverback and then vying for mature females.

3.2.1.1. Exploration phase. In GTO, every participant can be an outcome, and at every stage of the optimization process, the most effective one can be named the silverback gorilla. Several various strategies were used for the exploration phase: proceeding towards a recognized region to increase the capacity of GTO for examination of the region,

potential positions in the following t iterations. The rand symbols $[0, 1]$ represent a value. The parameter (p) in the range $[0, 1]$ must be specified during any optimization process to determine the possibility that a movement strategy would be chosen for an unidentified location. One gorilla from the gorilla swarms is factor X_r , whereas GX_r is just one gorilla group from the designated region that can be allocated randomly. The factor's lowest and maximum boundaries are indicated by LL and UL , respectively.

$$C = F \times (1 - \text{rand}) \quad (3)$$

$$F = \cos(2 \times r_4) + 1 \quad (4)$$

$$L = C \times l \quad (5)$$

$$H = Z \times X(t) \quad (6)$$

$$Z = [-C, C] \quad (7)$$

Meanwhile, t is the number of iterations. The random value within the range $[0, 1]$ is r_4 . Additionally, the initials l represent a random value between $[-1, 1]$.

3.2.1.2. Exploitation phase. Two phases are employed during the exploitative phase: continuing to pursue and hunting for adult females. One of the two approaches may be chosen based on the outcomes of the analysis of C with W . A swarm of gorillas is said to be led by a silverback, who exercises leadership and directs the troop to food sources. This phase will be used if $C > W$. Numerically describing this phenomenon is possible using equation (8).

$$XY(t+1) = L \times S \times (X(t) - X_{SB}) + X(t) \quad (8)$$

The vector $X(t)$ indicates the location of the gorilla. Furthermore, S is denoted by the following in the equation r_4 :

$$S = \left(\left| \left(\frac{1}{N} \sum_{i=1}^N XY_i(t) \right)^g \right| \right)^{(1/g)} \quad (9)$$

$$g = 2^l \quad (10)$$

$XY_i(t)$ Indicates the location of every potential vector at iteration t , and N represents the total amount of gorillas.

$$XY(i) = X_{SB} - (X_{SB} \times Q - X(t) \times Q) \times A \quad (11)$$

While

$$Q = 2 \times r_5 - 1 \quad (12)$$

$$A = \beta \times E \quad (13)$$

$$E = \{N_1 rand \geq 0.5N_2 rand < 0.5\} \quad (14)$$

Meanwhile, Q represents the impact force determined by equation (10). To assess the parameter A , which represents the combat stage, utilize equation (13). The algorithm starts with a value for B . E is in charge of simulating the impact's forceful aspect on equation (14).

All variables must fall within the range $[0, 1]$, with values greater than 0 showing that the variable is a candidate for categorization selection. As indicated in equation (15), the variable utilized to calculate single fitness must be the threshold to identify the precise traits to be evaluated.

$$BX_{ij} = \{1 \text{ if } X_{ij} > 0.50; \text{ otherwise}\} \quad (15)$$

In this case, X_{ij} it represents the dimension-specific amount for searching agent i at dimension j . We used a straightforward truncation rule to ensure variable bounds while upgrading every search agent's location at specific dimensions while the new value can violate the maximum constraints $[0, 1]$. The system then performs the following calculation for every gorilla in equation (16).

$$Fitness_i = \alpha \times (1 - C_i) + (1 - \alpha) \times \frac{|BX_i|}{D} \quad (16)$$

C_i stands for the accuracy attained by the KNN model, and α falls within the $[0, 1]$ range. The initial training data used as input has a dimension of D . The smallest fitness value will be used to identify the optimal solution. Upgrading continued till the condition reached its end. The GTO technique then generates the best outcome, which comprises the

feature set and is applied as a subsequent step to eliminate pointless characteristics from the test dataset. Our framework is predestined in the final stage by using it on the smaller test set.

The GTO algorithm selects vital characteristics such as age, blood pressure, cholesterol, smoking status, diabetes, BMI, family history of CVD, physical activity level, dietary habits, blood glucose levels, and past CVD events to predict the risk of cardiovascular disease (CVD).

3.2.1.3. Computational complexity. The GTO algorithm's capacity to effectively investigate and navigate complex solution spaces is a hallmark of its computational power. GTO is well-suited for situations where feature selection entails a high number of dimensions because it is made to be computationally efficient, especially in high-dimensional optimization tasks. The algorithm's efficiency is ascribed to its application of bio-inspired dynamics, mainly imitating the natural behaviour of cuckoos.

The three steps essential for figuring out the GTO algorithm's cost of computation are initialization, fitness evaluation, and upgrading of the vultures. The computing complexity of the initialization process is equal to $O(N)$, while the number of gorillas (N) is represented as N . Conversely, the computational complexity of the upgraded mechanism procedure depends on both the exploitation and the exploration stages.

Every phase begins with an upgrading operation on every solution in the optimization space. It ends with the execution of the best solution, which is equivalent to $O(T \times N) + O(T \times N \times D) \times 2$. On the other hand, D denotes the dimensions of the issue, and T represents the maximum number of iterations. Consequently, the GTO approach's computation complexity is $O(N \times (1 + T + TD) \times 2)$.

3.3. Classification

In this portion, the categorization algorithms that are utilized for prediction are covered.

3.3.1. Ensemble Learning Machine (ELM)

The Ensemble Learning Machine categorizer is used for binary categorization of the cardiovascular disease data. One of the best paradigms for learning is the ELM. Compared with more conventional learning methods like Backpropagation, SVM, ANN, Least Square-SVM, etc., it has better generalizing efficiency and an increased learning rate. The model's weights and bias settings are modified

for a single hidden layer. It employs the MP generalized inverse approach rather than gradient descent-based backpropagation learning to identify outcome weights. A schematic representation of the ELM model is shown in Fig. 2.

Equation (17) illustrates the actual outcome of the abovementioned conceptual arrangement.

$$O_j = \sum_{i=1}^L \beta_i g(\omega_i x_j + b_i), i=1,2,3, \dots, L, j=1,2, \dots, n \tag{17}$$

While x and y are the output and input vectors, accordingly, b_i is the bias of the i^{th} hidden neuron; $\omega_i = [\omega_{i1}, \omega_{i2}]$ is the weight of the vector among input and hidden layer; and m and L are the trained samples and hidden node counts, correspondingly. The outcome j^{th} input vector is denoted by O_j . The outcome weight of the matrix (β) connects the neuron's output to its hidden nodes, and $g(x)$ is an activation function. Formula (18) indicates the ELM's initial set of an arbitrary number of trials with zero error.

$$\sum_{j=1}^m \|O_j - y_j\| = 0 \tag{18}$$

While Q and N are the goal quantities and the number of variables, equation (19) illustrates the target training matrix (Y) and outcome weight (β) matrix.

$$H \times \beta = Y \tag{19}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \beta_2^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times Q}, y = \begin{bmatrix} Y_1^T \\ Y_2^T \\ \vdots \\ Y_m^T \end{bmatrix} \tag{20}$$

$$H(\omega_1, \dots, \omega_m, b_1, \dots, b_L, x_1, \dots, x_m) = \begin{bmatrix} g(\omega_1 x_1 + b_1) & \dots & g(\omega_m x_1 + b_L) \\ \vdots & \ddots & \vdots \\ g(\omega_1 x_m + b_1) & \dots & g(\omega_m x_m + b_L) \end{bmatrix}_{m \times L} \tag{21}$$

Equation (21), which provides the output of hidden layer matrix H , may be used to compute utilizing Formula (22), which provides the MP expanded inverse of H .

$$\beta = H^+ T \tag{22}$$

While H^+ stands for the generalized inverse since the input weights and biases of the ELM are chosen randomly, two problems could occur: It performs slowly during testing, and when there are the most hidden neurons present, it performs poorly when generalizing. Therefore, learning systems inspired

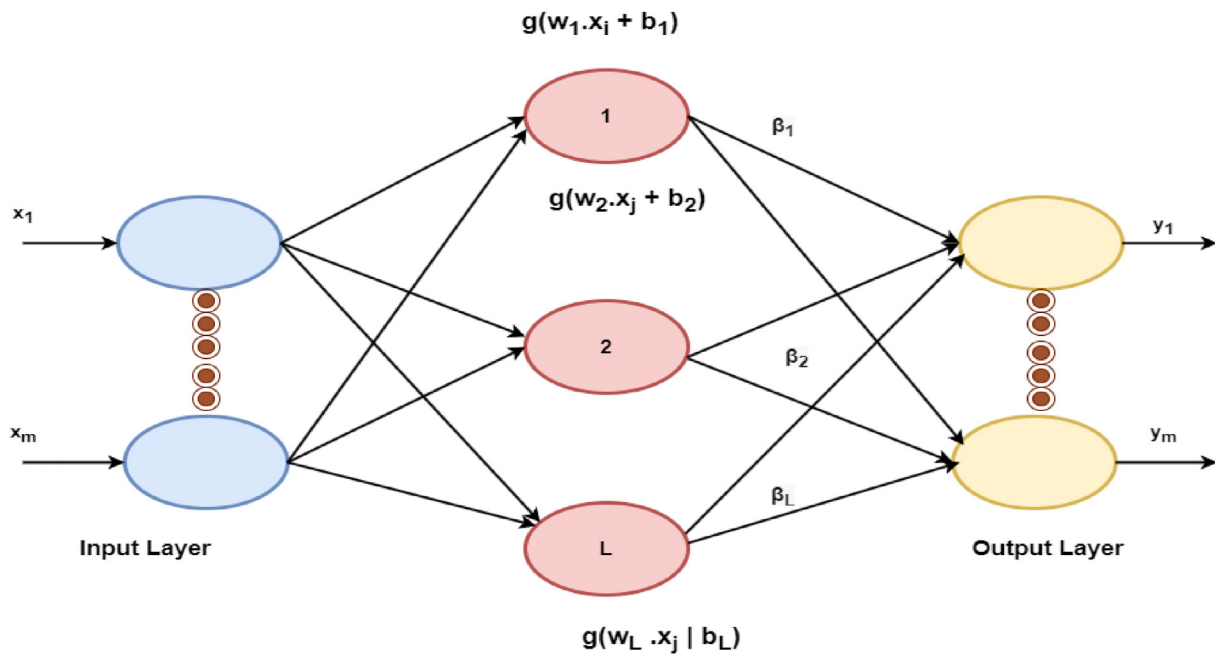


Fig. 2. ELM approach framework.

by nature and biology are more suited to address these problems. To tackle this issue, we ensemble ELM with Light GBM.

3.3.2. LightGBM

This study employs the LightGBM method for precise CVD prediction. A Gradient Boosting Decision Tree (GBDT) framework called Light Gradient Boosting Machine (LightGBM) is based on a decision tree technique that was presented. The GBDT method can discretize continuous features. However, it only employs initial derivative data to optimize the loss function. The decision tree algorithm learns the findings and residuals of every other tree before it.

Additionally, the expansion in data volume poses challenges to GBDT's effectiveness. Despite this, the layer-by-layer training strategy causes the leaf growth mode to expand. LightGBM then used the histogram-based decision tree approach. To speed up colossal data processing and eliminate false alarms and miss detections, LightGBM used the leaf growth approach with depth restriction to work together to minimize unnecessary XGBoost memory usage.

LightGBM was created to reduce the established objective provided by the supervising learning $X = \{(x_i, y_i)\}_{i=1}^N$ data set.

$$Obj = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f_k) \quad (23)$$

This approach uses the logistic function to measure the distinction between the predictions and targets.

$$l(z_i, \hat{z}_i) = z_i \ln(1 + e^{-\hat{y}_i}) + (1 - z_i) \ln(1 + e^{\hat{y}_i}) \quad (24)$$

In LightGBM, a regression model was then applied:

$$F_T(X) = \sum_{t=1}^T f_t(x) \quad (25)$$

The regression tree can also be described in another way $w_{q(x)}, R \in \{1, 2, \dots, N\}$, while N indicates the number of nodes, R represents hierarchies, w shows sample size weight, and the function of the objective is stated as:

$$Obj^{(t)} = \sum_{i=1}^n l(y_i, f_{t-1}(x_i) + f_t(x_i)) + \sum_k \Omega(f_k) \quad (26)$$

The steepest descent approach, which considers the function of loss, is used in conventional GBDT. The goal function is readily approximated in LightGBM using Newton's technique. While $g; h_i$ stands for the 1st function of loss and the 2nd function of loss, respectively.

$$h_i = \partial_{s_{t-1}(x_i)} \psi(z_i, S_{t-1}(x_i)) \quad (27)$$

$$D_i = \partial_{s_{t-1}(x_i)}^2 \psi(z_i, S_{t-1}(x_i)) \quad (28)$$

The formula (11) can be modified as follows to describe the leaf j sample set:

$$Obj^{(t)} = \sum_{j=1}^J \left[\left(\sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) w_j^2 \right] \quad (29)$$

Using quadratic programming, it is possible to determine the limit and the ideal weight for every leaf node given the topology of the tree $q(x)$:

$$b_j^* = - \frac{\sum_{i \in I_j} h_i}{\sum_{i \in I_j} s_i + \lambda} \quad (30)$$

Therefore, the gain computation equation is:

$$G = \frac{1}{2} \left[\frac{\left(\sum_{i \in I_L} h_i \right)^2}{\sum_{i \in I_L} s_i + \lambda} + \frac{\left(\sum_{i \in I_R} g_i h_i \right)^2}{\sum_{i \in I_R} s_i + \lambda} - \frac{\left(\sum_{i \in I} h_i \right)^2}{\sum_{i \in I} s_i + \lambda} \right] \quad (31)$$

LightGBM employs multi-threaded optimization to boost productivity and reduce processing time. It utilizes the optimum tree depth to prune trees and prevent over-fitting.

4. Result and discussions

To assess the efficacy of the suggested approach, we conducted two distinct types of investigations in this section. In the initial study, we employ an ensemble approach to forecast heart disease and assess efficiency using various evaluation criteria. A second investigation is then carried out to evaluate the effectiveness of the suggested model with other machine learning models.

4.1. Experimental setup

Using an Intel i5 2.60 GHz processor and 16 GB of RAM, it runs Windows 10. Python, KERAS, and TensorFlow are used in the examinations, which are carried out in front of the Anaconda3 platform.

4.2. Dataset description

Four well-known heart disease datasets are used in this study to evaluate the proposed methodology. Using the heart disease and heart failure prediction datasets from UCI, we examined and assessed their performance.

1) UCI heart disease dataset

The University of California (UCI, Irvine C.A) repository, publicly accessible on the Kaggle website, provided the dataset of heart patients used for the suggested methodology. The “Cleveland, Switzerland, Long Beach, and Hungary” datasets have been combined to create this dataset. The 1050 patients' records from the Kaggle dataset were included, along with 76 characteristics. 14 of the 76 variables were employed in the heart disease prediction. This is because these attributes significantly impact the disease more than others. The characteristics of both datasets are shown in [Table 1](#).

2) Heart failure prediction dataset

We tested our algorithm with this additional dataset. On the Kaggle website, it is publicly accessible. This dataset consists of 299 cases. Among the 299 individuals, 105 are female and 194 are male. Patients between the ages of 40 and 95 were also taken into account. We also made use of 13 dataset attributes.

4.2.1. IoMT platform

Hardware is an essential part of IoMT. Several gadgets can be utilized to collect real-time data from patients with CVD. A wide range of medical devices can be used for collecting data, such as the ECG, which can be utilized for tracking the heart; the glucometer, which can identify diabetes in the blood; and fitness monitors, which may capture details. These gadgets are connected to the internet and can use connectivity technologies like routers and gateways to deliver data to the cloud. Data are frequently periodically transferred utilizing an API key to establish a secure connection. Utilizing these API credentials, the necessary devices may obtain

and preserve data on the IoT system. Next, machine learning models display software choices that give medical staff members access to data analytics, reporting, and device control opportunities. The report suggests managing CVD data by installing a dependable platform that predicts CVD utilizing and IoMT components.

This part will be developed to provide additional details and support for utilizing the IoMT for CVD. IoMT is essential for revolutionizing the management and prevention of CVD because it makes continuous, real-time vital sign monitoring possible. IoMT allows monitoring of patients remotely through wearable technology and internet-connected sensors, enabling medical practitioners to quickly identify deviations and take appropriate action. With IoMT technology, doctors can keep an eye on their patients from a distance, spot modifications to their health, and provide individualized interventions that encourage patients to take charge of their cardiovascular health.

IoMT also opens up novel possibilities for innovation and research in machine learning and predictive analytics, which contributes to the development of advanced classification of risks models and personalized treatment methods. When all remains to be done, IoMT has a huge opportunity to completely change the way that studies, monitoring, and intervention are provided for CVD patients.

4.3. Evaluation metrics

The following list includes the simulation's evaluation parameters:

Accuracy (ACC): Accuracy mostly depends on how the data is collected. Comparing various

Table 1. Attributes.

Attribute no	UCI dataset	Cardiovascular dataset
	Attributes	Attributes
1	Sex	Age
2	St-depression	Glucose
3	Thal	Weight
4	Fasting blood sugar	Alcohol
5	Chest pain	Systolic blood pressure
6	Age	Physical activity
7	Target	Cholesterol
8	Resting blood pressure	Height
9	Maximum heart rate achieved	Smoking
10	St-slope	Gender
11	No.of major vessels	Diastolic blood pressure
12	Serum cholesterol in mg/dl	Cardiovascular
13	Resting electrographic results	
14	Exercise-induced angina	

measures from the same or different sources allows for evaluation. The evaluation is given by equation (33).

$$\text{accuracy} = \frac{a_p + a_n}{a_p + a_n + b_p + b_n} \quad (32)$$

Precision: This represents the likelihood that a patient with a positive screening test has the illness. Evaluating the PPV is possible using equation (34).

$$\text{PPV} = \frac{a_p}{a_p + b_p} \quad (33)$$

Negative predictive value (NPV): According to equation (35), this represents the likelihood of locating a patient who is not at risk for developing heart disease.

$$\text{NPV} = \frac{a_n}{a_n + b_n} \quad (34)$$

Sensitivity, Recall: This shows the capacity to identify a patient at risk for heart disease and is assessed as stated in equation (36).

$$\text{Sensitivity} = \frac{a_p}{a_p + b_n} \quad (35)$$

Specificity: This can be calculated by dividing the total number of negatives by the true negatives, as shown in (37). Value 1.0 designates the best specificity, while 0.0 designates the poorest.

$$\text{Specificity} = \frac{a_n}{a_n + b_p} \quad (36)$$

F1 Score: F-Measure uses harmonic mean to combine the outcomes of specificity and sensitivity

$$\text{F1} = 2 \times \frac{\text{Precision} \cdot \text{recall}}{\text{Precision} + \text{Recall}} \quad (37)$$

4.4. Evaluation of cardiovascular disease

The evaluation of this proposed model has four stages.

Step 1. Registration Module

In this module, we should register our details to create an account. It has to fill in the user name and password and confirm the password. And all that information is saved in a “datasheet.txt” file. It is shown in Fig. 3.

Step 2. Login-successful

Here, we should provide the details in the proper information we gave in the registration stage. If all the details are correct, it will show “successfully

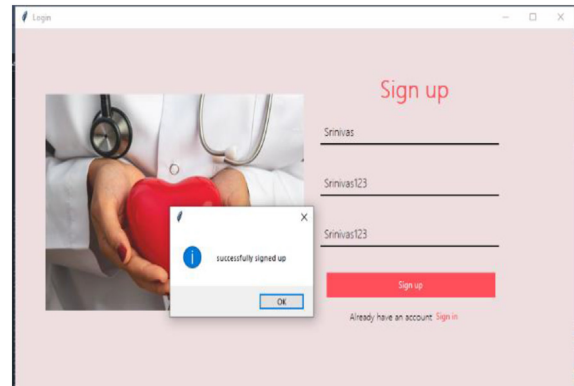


Fig. 3. A registration module.

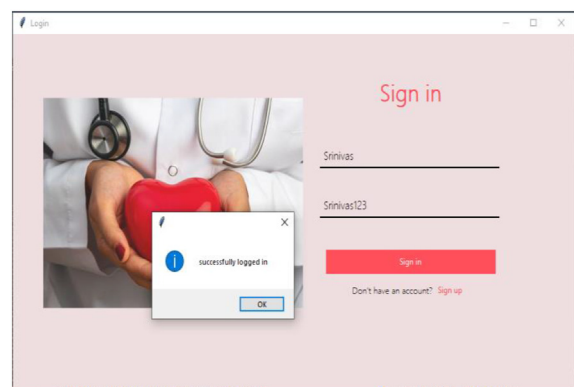


Fig. 4. Login phase.

login” and lead you to access the next page. It is shown in Fig. 4.

Step 3. Login-Unsuccessful

When we don't provide the correct details, it will show as an “invalid username or password.” It is shown in Fig. 5.

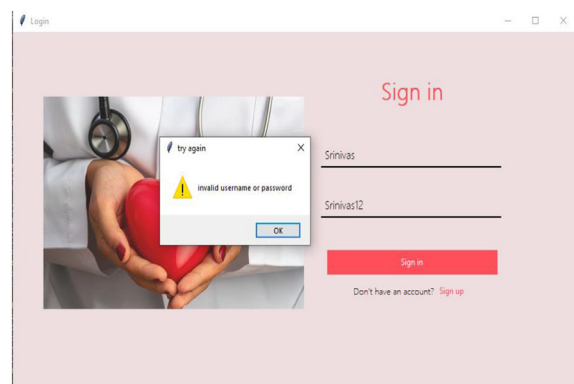


Fig. 5. Login unsuccessful phase.

When the given input details are correct, then it will lead you to access this page. Here, you need to fill in some data, which this page asks in Fig. 6.

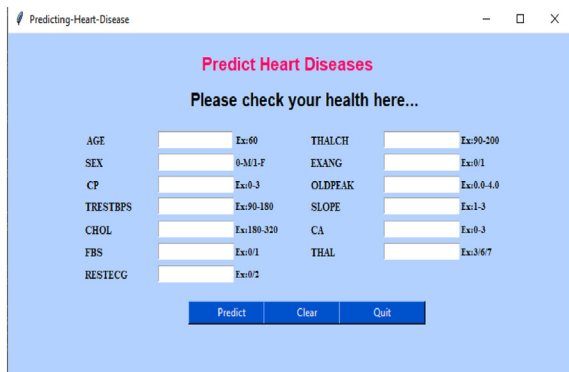


Fig. 6. Data details.

Click Predict after all the data has been entered. Following that, it will display, as in Figs. 7 and 8, whether the individual has cardiac disease or not.

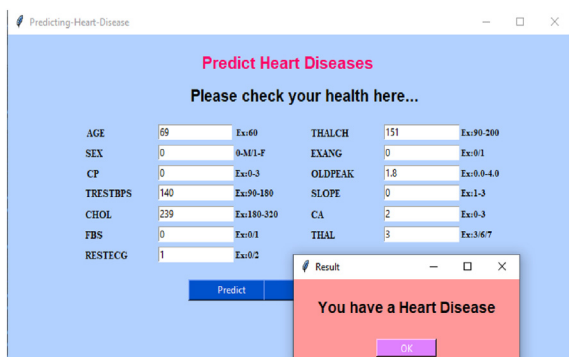


Fig. 7. Prediction outcome for heart patient.

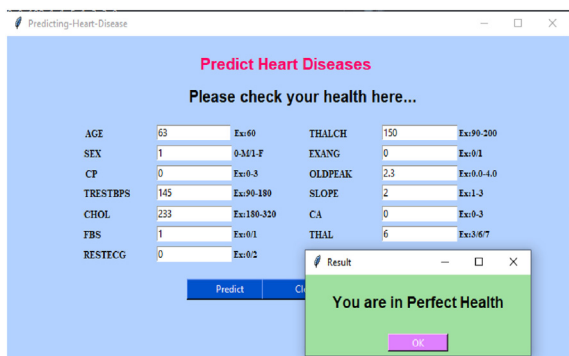


Fig. 8. Prediction outcome for the healthy patient.

4.5. Performance evaluation of the UCI dataset

The evaluation of the suggested method based on several criteria and comparison to relevant past studies on the UCI dataset is shown in this section. The existing approaches like GBM, ANN, XGBoost, SVM, and light GBM are employed to compare with

Table 2. Performance comparison of the UCI dataset.

Approches	Precision	F-Score	Accuracy	Recall	AUC
GBM	89.65	83.87	83.60	78.78	94.04
ANN	85.29	86.56	85.24	87.87	81.01
XGBoost	90.32	87.50	86.88	84.84	87.07
SVM	88.23	89.55	88.52	90.90	88.31
LightGBM	96.66	92.06	91.80	87.87	92.15
Proposed	98.96	99.02	99.18	98.56	99.09

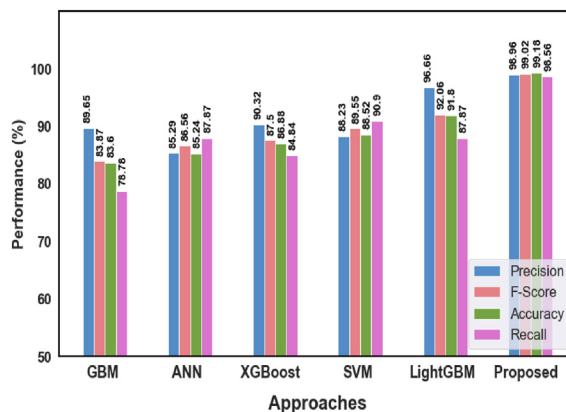


Fig. 9. Performance comparison of proposed vs. existing approach on the UCI dataset.

our proposed approach. Table 2 shows the performance comparison on the UCI dataset.

Fig. 9 demonstrates that, compared to other models, the proposed classifier obtained the best results with 99.18% accuracy, 98.96% precision, 98.56% recall and 99.02% F1-score. When compared to current methods, the suggested strategy performs better on the UCI database.

4.6. Performance evaluation of the cardiovascular disease dataset

In this section, we discuss the performance of using the cardiovascular disease dataset. Table 3 describes the evaluation of the suggested method based on several criteria and comparisons to relevant past studies on cardiovascular disease datasets.

The suggested technique outperforms the current methods when compared to the cardiovascular disease database. The existing approaches like ANN, RFE-GB, XGBoost, MLP, and ML-DL are utilized to compare with the proposed approach. Fig. 10 represents the cardiovascular disease dataset performances. The proposed approach obtained 98.93% of accuracy, 98.47% of precision, 99.03% of recall, and 98.82% F-Score.

Table 3. Proposed vs. existing approach comparison on cardiovascular disease dataset.

Technique	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Machine-deep ensemble classifier	88.70	88.02	88.02	88.01
XGBoost	87.02	89.62	82.11	86.30
Artificial Neural Network	68	69	68	68
Multilayer perceptron (MLP)	87.23	95.23	82.01	88.13
Ensemble Hybrid Voting classifier	84.03	87	84	85
RFE-GB	89.78	86	84	83
Proposed	98.93	98.47	99.03	98.82

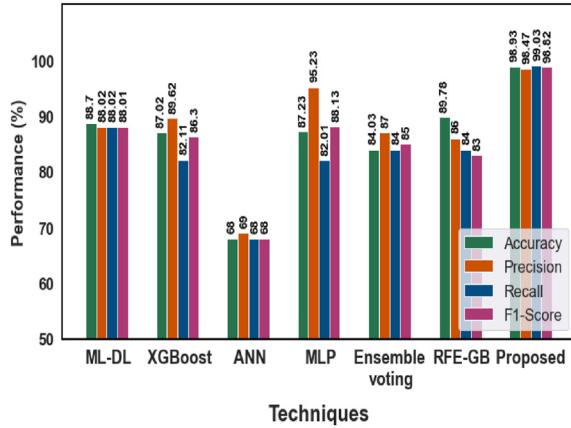


Fig. 10. Differentiation of proposed with existing method on cardiovascular disease dataset.

4.7. Performance evaluation using the class imbalance approach

After normalization, the effectiveness of classifiers was further enhanced by class balance. The SMOTETomek approach was used to balance the classes. Table 4 shows the effectiveness of the proposed approach using data balancing approaches.

The differentiation of using the class balancing approach is shown in Fig. 11. After utilizing the imbalance approach, the proposed approach's performance was improved even more.

Table 4. Utilizing the data balance approach.

Approaches	Sensitivity	Accuracy	Specificity	F-Score	Precision
Adaboost	79.26	82.34	85.36	81.76	84.41
LR	80.48	83.24	85.97	82.75	85.16
NB	82.92	85.07	87.19	84.7	86.62
RF	81.70	83.85	85.97	83.48	85.35
SVM	81.09	84.16	87.19	83.64	86.36
Proposed	98.56	99.18	98.24	99.02	98.96

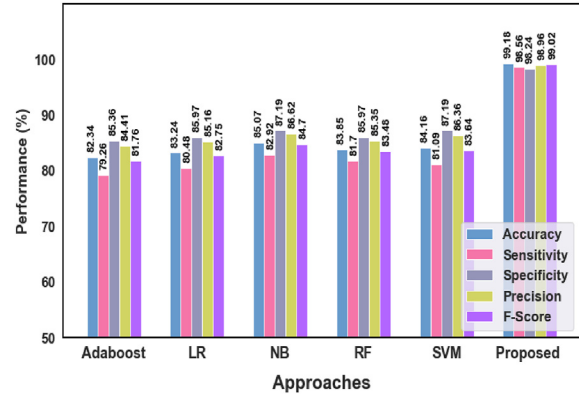


Fig. 11. Differentiation of using class balancing approach.

4.8. Performance evaluation of selecting features

Selecting features helped the efficiency of the categorizer even more. All baseline algorithms' accuracy significantly increased after using the advised feature selection technique.

Table 6 and Fig. 13 represent the performance comparison of the proposed approach vs. the existing approach. Table 5 and Fig. 12 show that while the mean error rate of all the algorithms dropped, the suggested model's accuracy compared to that of other predictors. The suggested ML approach has a 98.14% accuracy rate, the highest of all models. As a result of this research, we can conclude that the suggested feature selection strategy could decrease the complexities and dimensions of the dataset while also extracting valuable details from big medical data for ML algorithms. Additionally, we observe that feature selection-based machine learning models may reliably forecast patient cardiac disease with a constrained set of features.

Table 6 and Fig. 13 represent the performance comparison of the proposed approach vs. the existing approach. Table 5 and Fig. 12 show that while the algorithms' mean error rate dropped, the

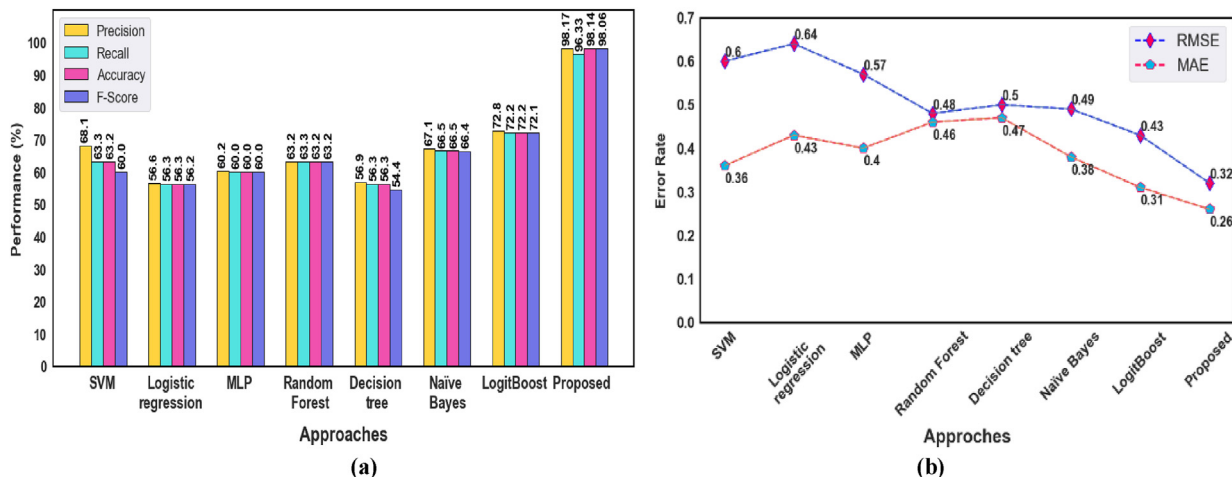


Fig. 12. Graphical representation before selecting features (a) Differentiation of evaluation metrics (b) Error rate comparison.

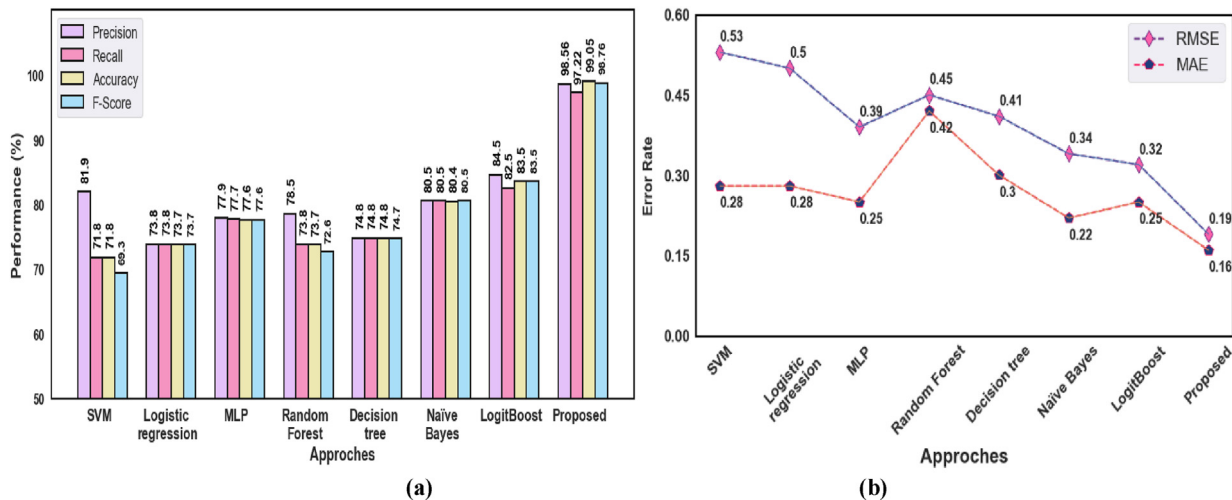


Fig. 13. Graphical representation after using feature selection approach (a) comparison of performance metrics (b) MAE and RMSE comparison.

Table 5. Performance of Proposed vs. existing approach before using feature selection approach.

Approaches	Precision (%)	Accuracy (%)	F-Score (%)	RMSE	Recall (%)	MAE
SVM	68.1	63.2	60.0	0.60	63.3	0.36
Logistic regression	56.6	56.3	56.2	0.64	56.3	0.43
MLP	60.2	60.0	60.0	0.57	60.0	0.40
Random Forest	63.2	63.2	63.2	0.48	63.3	0.46
Decision tree	56.9	56.3	54.4	0.50	56.3	0.47
Naive Bayes	67.1	66.5	66.4	0.49	66.5	0.38
LogitBoost	72.8	72.2	72.1	0.43	72.2	0.31
Proposed	98.17	98.14	98.06	0.32	96.33	0.26

Table 6. Performance of the proposed method after using the feature selection approach.

Approaches	Precision (%)	Accuracy (%)	F-Score (%)	RMSE	Recall (%)	MAE
SVM	81.9	71.8	69.3	0.53	71.8	0.28
Logistic regression	73.8	73.7	73.7	0.50	73.8	0.28
MLP	77.9	77.6	77.6	0.39	77.7	0.25
Random Forest	78.5	73.7	72.6	0.45	73.8	0.42
Decision tree	74.8	74.8	74.7	0.41	74.8	0.30
Naive Bayes	80.5	80.4	80.5	0.34	80.5	0.22
LogitBoost	84.5	83.5	83.5	0.32	82.5	0.25
Proposed	98.56	99.05	98.76	0.19	97.22	0.16

suggested model's accuracy compared to that of other predictors. The suggested ML approach has a 98.14% accuracy rate, the highest of all models. As a result of this research, the suggested feature selection strategy could decrease the complexities and dimensions of the dataset while also extracting valuable details from big medical data for ML algorithms. Additionally, feature selection-based machine learning models may reliably forecast patient cardiac disease with constrained features.

The performance evaluation of the various techniques used to predict and categorize CVD is shown in the provided Table 7. Three primary metrics are used to evaluate each method: accuracy, precision, and recall. Precision quantifies the percentage of true positive predictions among all positive predictions, whereas accuracy indicates how accurate the model is overall at making predictions. Conversely, recall shows the percentage of accurate positive predictions among all occurrences of a specific class in the dataset. Remarkably, the suggested approach performs exceptionally well, attaining 99.05% accuracy, 99.08% precision, and 99.10% recall. This shows that the suggested strategy beats or performs comparably with current techniques, indicating that it is a strong

Table 7. Overall classification approach comparison.

Methods	Accuracy	Precision	Recall
Deep CNN	97.70	–	–
EDCNN	–	99.01%	–
CardioXNet	99.06%	99.5%	99.5%
CWT and CNN	98.64%	–	–
ECNN	98.79%	–	–
Proposed	99.05%	99.08%	99.10%

contender for accurately identifying and categorizing chronic illnesses in scenarios involving healthcare monitoring.

4.9. Evaluation of computational time

The selected classifier's error and computation time are performance indicators. We will consider sensitivity, specificity, and categorization accuracy. Every algorithm considers the computation time. The proposed method offers strong sensitivity, specificity, and accuracy while requiring minimum computation time. Fig. 14 and Table 8 show the computation time for the suggested method.

Table 8. Overall classification approach comparison.

Approaches	Time (Sec)
Naïve Bayes	0.21
Bayes Net	0.13
MLP	14.36
PART	0.99
Proposed	0.08

The model correctly predicted the result as “abnormal.” The features' correlation is analyzed by the heat map in Fig. 15. The light blue and white hues show that the correlation among the variables is close to +1 and 1, accordingly, whereas the dark blue color shows that the correlation is close to 0. The correlation between the qualities chol and FBS and the attribute class is nearly zero, indicating that there may be no association at all. Consequently, we might do away with them to enhance the functionality of our suggested model.

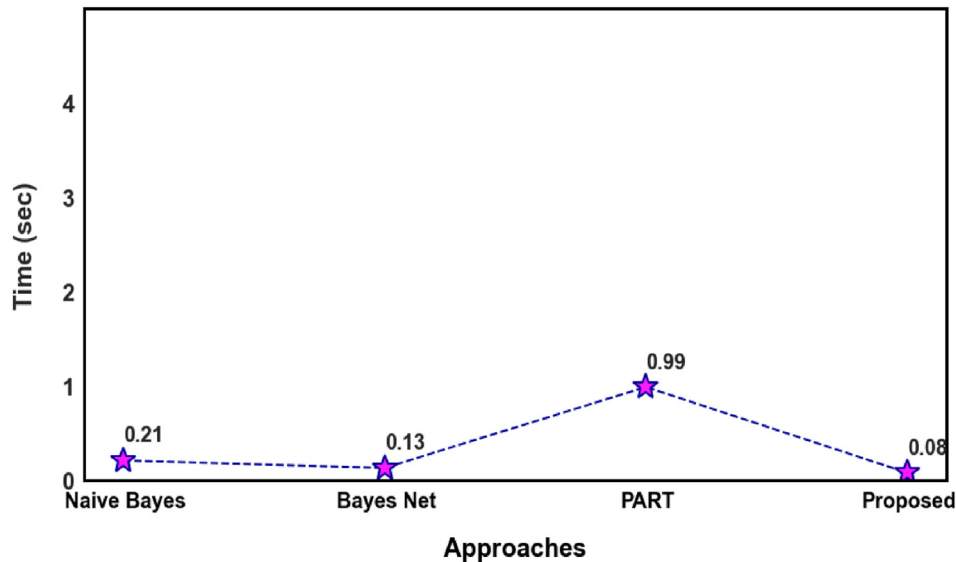


Fig. 14. Execution time comparison over existing approaches.

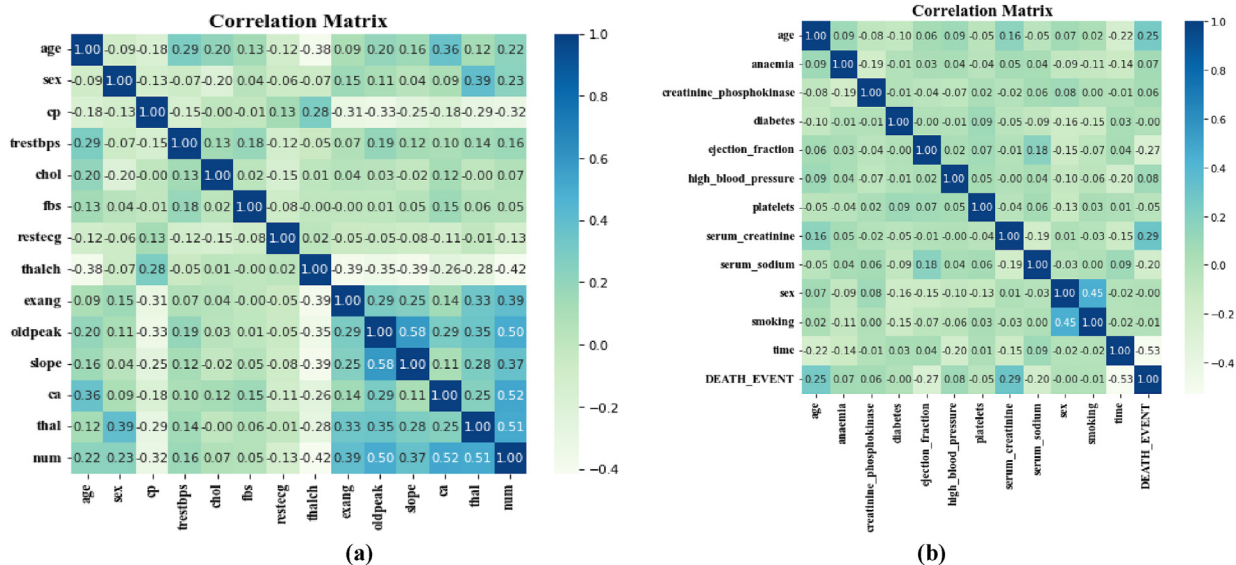


Fig. 15. Heatmap correlation (a) UCI dataset and (b) heart disease prediction dataset.

4.10. Evaluation of training and testing

The graph highlights the advantages of implementing the method suggested in the paper. Training and testing loss functions, as well as training and testing accuracy, are shown in Figs. 16 and 17. Using the prepared training data, the proposed algorithm is trained for 100 epochs during the training phase. A learning rate of 0.01 has been observed.

4.11. Limitations

Overall, research has shown that the performance of various classifiers was satisfactory compared to past studies. However, there are certain limitations, such as the reliance on a single feature selection technique, which can lead to exact results. The dataset's having a lot of missing values could be damaging. Other datasets utilized with this approach must likewise address

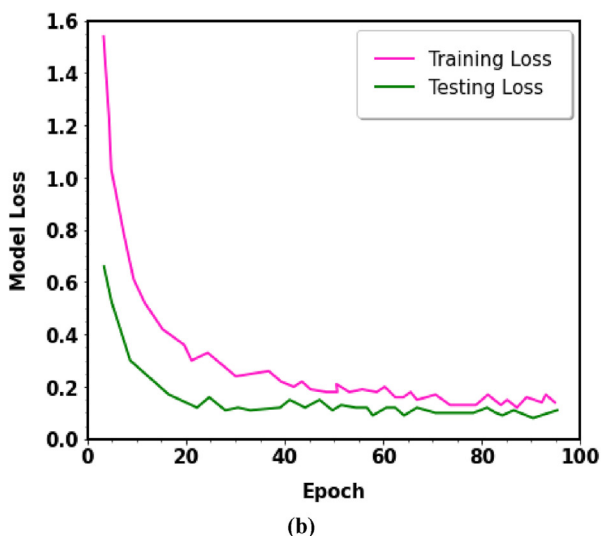
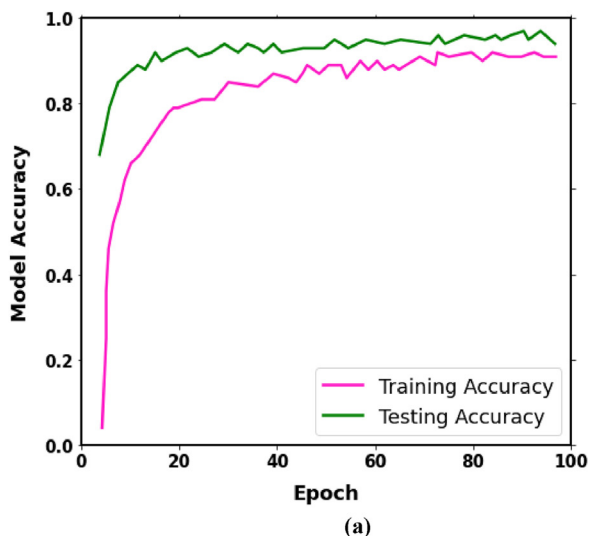


Fig. 16. Comparison of UCI dataset (a) Training vs. testing accuracy (a) Training vs. testing loss.

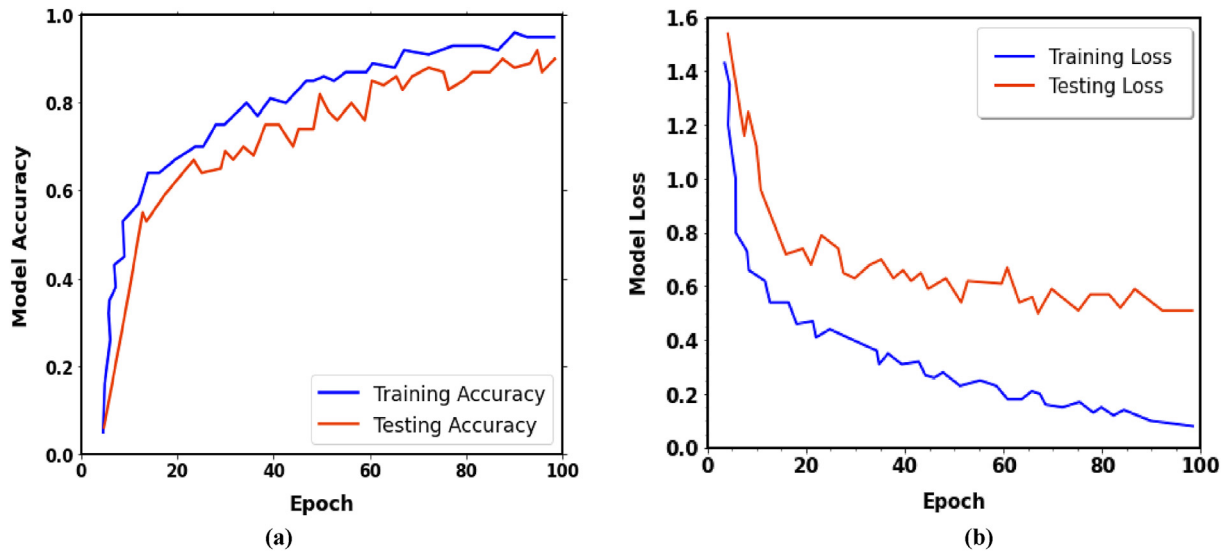


Fig. 17. Comparison of cardiovascular disease dataset (a) Training vs. testing accuracy (a) Training vs. testing loss.

this problem if the missing value is sufficiently significant, as we have demonstrated how to do. Despite the size of our training dataset, more data would improve the model's accuracy.

5. Conclusion & future scope

Regardless of one's ethnic or social origins, accurately predicting the risk of heart disease might significantly impact the overall mortality rate of people. A crucial element in reaching this objective is early diagnosis. With the aid of machine learning, several researchers have previously tried to forecast cardiac disease. This study follows a similar path but uses an enhanced and distinctive methodology and a larger dataset to train the algorithm. Our proposed approach comprises three phases: Pre-processing, feature selection, and classification. In pre-processing, data cleaning, transformation, and normalization are performed. After that, we increased the minority samples with the help of a hybrid SMOTETomek approach. To reduce the dimension and computational time, we selected the essential features by using the Gorilla Troops Optimizer (GTO) approach. Finally, the ensemble of ELM and LightGBM is employed to predict cardiovascular disease. We utilized the UCI and Cardiovascular disease dataset to evaluate and contrast the suggested method. Although the suggested system cannot replace a doctor, it can be employed in isolated and rural locations where there are no contemporary medical facilities or cardiac specialists. Additionally, it can help the doctor make quick decisions. There are also certain restrictions on the suggested system. It can only determine whether or not a person has cardiac disease.

By utilizing different feature selection algorithms and optimization methodologies, we will conduct additional research in the future to improve the efficacy of these predictive models for the identification of heart disease. Additionally, wearable gadgets on the market will be used for training and testing the planned work.

Funding

No funding was received to assist with the preparation of this manuscript.

Ethics approval

This material is the authors' own original work, which has not been previously published Elsewhere. The paper reflects the authors' own research and analysis in a truthful and complete manner.

Acknowledgements

We declare that this manuscript is original, has not been published before and is not currently being considered for publication elsewhere.

References

- [1] R. Valarmathi, T. Sheela, Heart disease prediction using hyperparameter optimization (HPO) tuning, *Biomed Signal Process Control* 70 (2021) 12–25, <https://doi.org/10.1016/j.bspc.2021.103033>, 103033.
- [2] L. Revathi, C. Anjuaravind, Artificial intelligence tool for heart disease prediction using deep, *Learning CNN 1* (2023) 1–38, <https://doi.org/10.1007/s10462-023-10493-5>.
- [3] R.A. Vasudev, B. Anitha, G. Manikandan, B. Karthikeyan, L. Ravi, V. Subramaniaswamy, Heart disease prediction

- using stacked ensemble technique, *J Intell Fuzzy Syst* 39 (2020) 8249–8257, <https://doi.org/10.3233/JIFS-189145>.
- [4] S. Mahmoud, M. Gaber, G. Farouk, A. Keshk, Heart disease prediction using modified version of LeNet-5 model, *Int J Intell Syst Appl* 12 (2022) 15, <https://doi.org/10.5815/ijisa.2022.06.01>.
- [5] M. Roseiro, J. Henriques, S. Paredes, T. Rocha, J. Sousa, An interpretable machine learning approach to estimate the influence of inflammation biomarkers on cardiovascular risk assessment, *Comput Methods Progr Biomed* 230 (2023) 107347, <https://doi.org/10.1016/j.cmpb.2023.107347>.
- [6] J. Xie, R. Wu, H. Wang, H. Chen, X. Xu, Y. Kong, W. Zhang, Prediction of cardiovascular diseases using weight learning based on density information, *Neurocomputing* 452 (2021) 566–575, <https://doi.org/10.1016/j.neucom.2020.10.114>.
- [7] S.K. Arunachalam, R. Rekha, A novel approach for cardiovascular disease prediction using machine learning algorithms, *Concurrency Comput Pract Ex* 34 (2022) 7027, <https://doi.org/10.1002/cpe.7027>.
- [8] J. Sekar, P. Aruchamy, H. Sulaiman Lebbe Abdul, A.S. Mohammed, S. Khamuruddeen, An efficient clinical support system for heart disease prediction using TANFIS classifier, *Comput Intell* 38 (2022) 610–640, <https://doi.org/10.1111/coin.12487>.
- [9] K. Kadam Malav, P. Kamat, Prediction of heart disease using k-means and artificial neural network as a hybrid approach to improve accuracy, *Int J Eng Technol* 9 (2017) 3081–3085, <https://doi.org/10.21817/ijet/2017/v9i4/170904101>.
- [10] K. Dalakleidi, K. Zarkogianni, A. Thanopoulou, K. Nikita, Comparative assessment of statistical and machine learning techniques towards estimating the risk of developing type 2 diabetes and cardiovascular complications, *Exp Syst* 34 (6) (2017) e12214, <https://doi.org/10.1111/exsy.12214>.
- [11] W. Sun, P. Zhang, Z. Wang, D. Li, Prediction of cardiovascular diseases based on machine learning, *ASP Trans Internet Things* 10 (2021) 30–35, <https://doi.org/10.5281/OT.2021.100035>.
- [12] S.P. Patro, G.S. Nayak, N. Padhy, Heart disease prediction by using novel optimization algorithm: a supervised learning prospective, *Inform Med Unlocked* 26 (2021) 100696, <https://doi.org/10.1016/j.imu.2021.100696>.
- [13] M.A. Khan, S. Abbas, A. Atta, A. Ditta, H. Alquhayz, M.F. Khan, R.A. Naqvi, Intelligent cloud based heart disease prediction system empowered with supervised machine learning, *Comput Mater Continua (CMC)* 65 (2020) 1–39, <https://doi.org/10.32604/cmc.2020.011416>.
- [14] S. Mezzatesta, C. Torino, P. De Meo, G. Fiumara, A. Vilasi, A machine learning-based approach for predicting the outbreak of cardiovascular diseases in patients on dialysis, *Comput Methods Progr Biomed* 177 (2019) 9–15, <https://doi.org/10.1016/j.cmpb.2019.05.005>.
- [15] A.A. Nancy, D. Ravindran, P.D. Raj Vincent, K. Srinivasan, D. Gutierrez Reina, Iot-cloud-based smart healthcare monitoring system for heart disease prediction via deep learning, *Electronics* 11 (2022) 2292, <https://doi.org/10.3390/electronics11152292>.
- [16] Y. Chen Zhang, Y. Chen, S. Ye, W. Cai, J. Jiang, Y. Xu, G. Zheng, M. Chen, Heart disease prediction is based on the embedded feature selection method and deep neural network, *J Healthcare Eng* 1 (2021) 1–9, <https://doi.org/10.1155/2021/6260022>.
- [17] L.K. Singh, E. Pooja, H. Garg, M. Khanna, An IoT based predictive modeling for Glaucoma detection in optical coherence tomography images using hybrid genetic algorithm, *Multimed Tool Appl* 81 (26) (2022) 37203–37242, <https://doi.org/10.1007/s11042-022-13540-5>.
- [18] A.K. Faieq, M.M. Mijwil, Prediction of heart diseases utilizing support vector machine and artificial neural network, *Indones J Electr Eng Comput Sci* 26 (2022) 374–380, <https://doi.org/10.11591/ijeecs>.
- [19] L.K. Singh, M. Khanna, R. Singh, An enhanced soft-computing based strategy for efficient feature selection for timely breast cancer prediction: Wisconsin Diagnostic Breast Cancer dataset case, *Multimed Tool Appl* 1 (2024) 1–66, <https://doi.org/10.1007/s11042-024-18473-9>.
- [20] P.E. Rubini, C.A. Subasini, A.V. Katharine, V. Kumaresan, S.G. Kumar, T.M. Nithya, A cardiovascular disease prediction using machine learning algorithms, *Ann Rom Soc Cell Biol* 1 (2021) 904–912, <https://doi.org/10.1155/2023/1406060>.
- [21] M.A. Khan, An IoT framework for heart disease prediction based on MDCNN classifier, *IEEE Access* 8 (2020) 34717–34727, <https://doi.org/10.1109/ACCESS.2020.2974687>.
- [22] N.L. Fitriyani, M. Syafrudin, G. Alfian, J. Rhee, HDPm: an effective heart disease prediction model for a clinical decision support system, *IEEE Access* 8 (2020) 133034–133050, <https://doi.org/10.1109/ACCESS.2020.3010511>.
- [23] P. Rani, R. Kumar, N.M.S. Ahmed, A. Jain, A decision support system for heart disease prediction based upon machine learning, *J Reliab Intell Environ* 7 (2021) 263–275, <https://doi.org/10.1007/s40860-021-00133-6>.
- [24] P. Ghosh, S. Azam, M. Jonkman, A. Karim, F.J.M. Shamrat, E. Ignatius, F. De Boer, Efficient prediction of cardiovascular disease using machine learning algorithms with Relief and LASSO feature selection techniques, *IEEE Access* 9 (2021) 19304–19326, <https://doi.org/10.1109/ACCESS.2021.3053759>.
- [25] G. Ambrish, B. Ganesh, A. Ganesh, C. Srinivas, K. Mensinkal, Logistic regression technique for prediction of cardiovascular disease, *Global Transit Proc* 3 (2022) 127–130, <https://doi.org/10.48550/arXiv.2304.13314>.
- [26] R.R. Sarra, A.M. Dinar, M.A. Mohammed, K.H. Abdulkareem, Enhanced heart disease prediction based on machine learning and χ^2 statistical optimal feature selection model, *Designs* 6 (5) (2022) 87, <https://doi.org/10.3390/designs6050087>.
- [27] H. Hasanova, M. Tufail, U.J. Baek, J.T. Park, M.S. Kim, A novel blockchain-enabled heart disease prediction mechanism using machine learning, *Comput Electr Eng* 101 (2022) 108086, <https://doi.org/10.1016/j.compeleceng.2022.108086>.
- [28] A.A. Malibari, An efficient IoT-Artificial intelligence-based disease prediction using lightweight CNN in healthcare system, *Measurement, Sensors* 26 (2023) 100695, <https://doi.org/10.1016/j.measen.2023.100695>.