**International Journal of Public Health Science (IJPHS)**
Vol. 13, No. 3, September 2024, pp. 1489~1501
ISSN: 2252-8806, DOI: 10.11591/ijphs.v13i3.24213


# Evaluation of cardiovascular disease in diabetic patients using machine learning techniques

**Silpa Nrusimhadri[1], Sangram Keshari Swain[1], Veeranki Venkata Rama Maheswara Rao[2], Shiva Shankar Reddy[3], Mahesh Gadiraju[3]**
[1]Department of Computer Science and Engineering, Centurion University of Technology and Management, Bhubaneswar, India
[2]Shri Vishnu Engineering College for Women (A), Bhimavaram, India
[3]Sagi RamaKrishnam Raju Engineering College (A), Bhimavaram, India

**Article Info**

***Article history:***

Received Nov 14, 2023
Revised Jan 14, 2024
Accepted Jan 25, 2024

***Keywords:***

Cardiovascular disease
Coronary artery disease
Deep learning
Machine learning
Particle swarm optimization

**ABSTRACT**

Machine learning (ML) improves operations in many industries, including medicine. It affects the prognosis of several disorders, including heart disease. If predicted, it may provide doctors with new insights and allow them to treat each patient individually. If anticipated, it may provide medical practitioners with valuable information. Our team uses machine learning algorithms to study heart disease risk. This research will compare decision trees, AdaBoost, support vector machines, artificial neural networks (ANN), and customized ANN. The study will include this analysis. The given model will leverage the dataset of general information and medical test results. Our model uses particle swarm optimization (PSO) and k-nearest neighbors (KNN). Algorithm for feature selection. The model reduces dimensionality using evolutionary algorithms and neural networks. We compared the numerous assessment criteria to the current models, our model, and earlier models. Because of this, the suggested model's suitability was rated with the highest accuracy.



***Corresponding Author:***

Silpa Nrusimhadri
Department of Computer Science and Engineering, Centurion University of Technology and Management
Bhubaneswar, Odisha, India
Email: nrusimhadri.silpa@gmail.com

## 1. INTRODUCTION

Worldwide, heart attacks are the leading cause of death. Causes include stress, genetics, hypertension, and other issues. "Heart disease" is often an umbrella term for various conditions that interfere with the heart's standard structure and function. The 75% or more of the victims were from low- and middle-income countries. Coronary heart disease (CHD), i.e., heart attack, is the most prevalent and lethal of all heart conditions [1]. An estimated 805,000 Americans have a heart attack yearly in the United States, with one occurrence reported every 40 seconds. It can be challenging to identify individuals who are at high risk of health issues due to the range of risk factors involved, including diabetes, hypertension, hyperlipidemia, and others. Doctors and scientists have started using machine learning (ML) techniques to create screening tools as they are better at detecting patterns and classifying data than traditional statistical methods [2]–[4]. Cardiovascular diseases (CVD) include all heart diseases. However, not all heart disorders are cardiovascular ailments. Heart or blood vessel diseases are referred to as CVD.

These methods can estimate outcome probability for individuals aged 30 to 74 throughout 4 to 12 years. Compare the equations with a new CHD prediction equation to see whether a single profile accurately predicts associated endpoints. Individual shapes for diastolic (DBP) and systolic blood pressure (SBP) are presented for all outcomes. An approach to calculating confidence intervals for expected probabilities, hazard ratios, and excess risk estimations is described. Blood is pumped around the body through the circulatory system, also called the cardiovascular system. The circulatory system's primary components are the heart, veins, arteries, and blood capillaries. Heart disorders result from their inability to carry out their duties adequately. Instantaneous death may result from cardiovascular system failure [3]. Heart disease is the leading killer on a global scale. The increased incidence of heart disease makes accurate early disease prediction urgent and unsettling. By combining data mining and neural network approaches, people's heart illness severity was precisely diagnosed [5]. The severity of the disease is categorized using decision trees (DT), genetic algorithms (GA), naive bayes (NB), and K-nearest neighbor (k-NN) algorithms. The diseases must be handled with attention due to the complexity of their problem characteristics. Failure to do so may result in premature mortality or a reduction in the efficacy of organs. Integrating data mining and neural network methodologies accurately identified the degree of cardiac ailments in individuals [6].

The categorization component of data mining techniques is crucial for data study and cardiac disease prediction. The random forest (RF) algorithm is an ensemble ML algorithm. It performs well across various classification and regression predictive modeling tasks and is possibly the most well-known and commonly used ML method. The random forest approach also provides a brute-force parameter adjustment method that facilitates more accessible feature selection. The linear support vector machine (SVM) technique is used for linearly separable data, which means that a dataset is considered linearly separable if it can be separated into two groups with only a single straight line [7]–[9]. After that, the linear SVM classifier is applied to the data to classify it. Linear regression (LR) is a popular ML technique in supervised learning. This approach predicts a dependent variable that falls into categories based on independent factors. LR helps to analyze data and understand the relationship between a dependent binary variable and one or more independent variables [10]. It is a method to determine the connection between a reliant binary variable and one or more autonomous variables.

The result must thus be a value that can be classified as either discrete or categorical, such as yes or no, zero or one, true or false, and so on. Instead of presenting exact values like 0 and 1, it gives probabilistic values ranging from 0 to 1. Ensemble learning is a method that incorporates the model methods of five different classifiers to predict and diagnose the recurrence of cardiovascular illness [11]. These classifiers are the SVM, the artificial neural network (ANN), the NB, the regression analysis, and the RF. The symptoms of cardiovascular disease are shown in Figure 1. It is crucial to remain vigilant for symptoms of heart disease and communicate with your doctor if you have any concerns. Regular medical check-ups can sometimes detect cardiovascular disease at an early stage [12]–[15].

Some cardiac conditions are congenital, meaning they can occur at birth. Throughout your lifespan, various types develop. As shown in Figure 2, coronary artery disease develops gradually over a person's lifetime due to the slow accumulation of a gummy material known as plaque in the arteries that carry blood to the heart muscle. Plaque has been linked to various cardiac conditions due to its ability to limit or block blood flow to the heart muscle [16]. Some of the heart diseases are shown in Figure 3. Heart disease may lead to heart attacks, strokes, and death. Early detection of heart disease signs ensures proper therapy. This paper proposes a hybrid data mining-based solution to detect heart illness. Numerous academics worldwide began analyzing large databases to predict heart-related disorders [17]. Various ML approaches may analyze large datasets and make helpful findings. Different methods in ML models are crucial for effectively indicating the presence or absence of cardiac disorders.

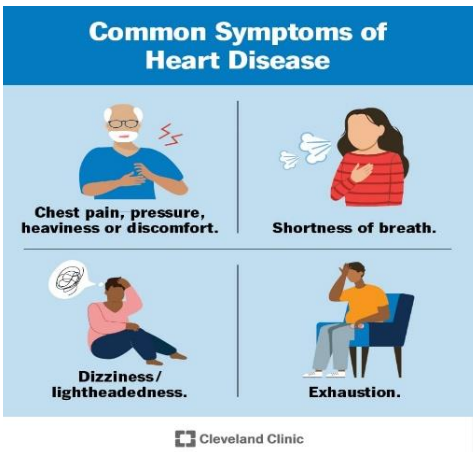


Figure 1. Symptoms of CAD

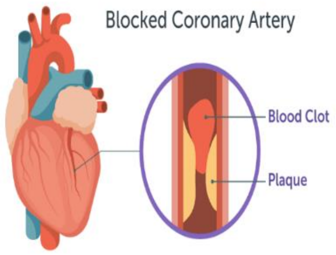


Figure 2. Coronary artery diseases

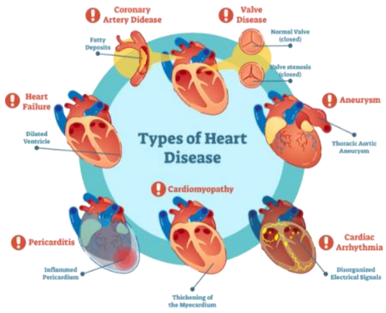


Figure 3. Various types of heart diseases

## 2. RELATED WORK

Li *et al.* [18] created a classification system using SVM, K-NN, artificial neural networks (ANN), LR, DT, and NB. Cleveland heart disease dataset tests the system. They presented a quick conditional mutual information feature selection technique. SVM with fast conditional mutual information (FCMIM) has 92.37% accuracy. Mohan *et al.* [19] concentrated on enhancing the functionality and accuracy of heart disease prediction. Several different approaches, including KNN, DT, genetic algorithm (GA), and NB, are used to categorize the severity of the condition. The University of California Irvine (UCI) dataset was utilized to test a method that predicted heart disease with an accuracy of 87.4%. Yadav *et al.* [20], multiple algorithms, including NB, KNN, and LR, were examined to predict heart illness and analyze the best treatment. The accuracy of the different models was as: NB was 82.91% accurate, Fuzzy KNN was 94.19% accurate, and BP-Neural Network was 98% accurate. The Statistical Update is essential for anyone seeking the best data on these variables and situations [21], including the public, lawmakers, journalists, doctors, healthcare facility administrators, researchers, and health advocates.

Dwivedi [22] analyzed the predictive capabilities of six different ML techniques in forecasting heart disease. The StatLog heart disease dataset, acquired from the ML laboratory at the UCI, was used in this study. The results showed that the logistic regression model had the highest level of classification accuracy at an impressive 85%. In 2020, Yahaya *et al.* [23] published a detailed research paper on using machine learning (ML) to predict heart disease. The findings will be helpful to medical professionals in identifying potential cardiac risks. They proposed a practical approach for determining the existence of heart illness by using the back propagation (BP) feature extractor of ANN to the various online heart disease database categorizations. Allow for more accurate results. The discovery in the study discloses the usage of the WEKA tool; the prediction of heart disease resulted in a rapid turnaround and obtained an accuracy of 97.5%. TR *et al.* [24] developed four methods for comparative assessment and positive performance. This research found that statistical techniques performed worse than ML. The UCI online heart disease data set was used to assess ML's classification algorithms' precision, F1 score, accuracy, and recall. The KNN classification method fared better than any of the other fourteen parameters that were provided.

Nashif *et al.* [25] used two prominent open-access databases to confirm this. The system was tested for heart disease detection in several datasets using 10-fold cross-validation. The SVM approach has 97.53% accuracy, 97.50% sensitivity, and 94.94% specificity. The patient's doctor or caretaker monitors them 24/7 while treating heart disease. The new technology sends data to a centralized server that updates every 10 seconds. The notion of inflame-aging acknowledges that older low-grade inflammatory pathways increase CV risk. Understanding the paths that connect inflammation and aging may uncover novel treatment targets and help address the global aging population [26].

Deep learning neural networks (DNNs) were successfully implemented by Sharma and Parmar [27] utilizing Talos optimization. In DNN, one of the optimization strategies available is called Talos optimization. Talos offers a higher level of accuracy (90.76%) compared to other optimizations. It is used for heart disease datasets to discover a good forecast. They developed a Keras model using the Talos optimization and then deployed it. According to Krishnaiah *et al.* [28], many methods from the field of data mining have been used to forecast the patients who would develop cardiac disease. However, the ambiguity that existed in the data was not eliminated. An effort was made to stop the uncertainty caused by unstructured data by injecting fuzziness into the measured data. It was accomplished by creating and combining a membership function with the estimated value. In addition, an effort was made to categorize the patients by basing the classification on the characteristics obtained from the medical profession. While testing and organizing the system, it was discovered that those in the age range of 50 to 60 years had the same symptoms of heart disease as those in the age range of 40 to 45 years. It was discovered when comparing the two age ranges.

Sharma *et al.* [29] utilized the Cleveland dataset, which contains information regarding cardiac problems and has 1,025 instances. They divided the data into training and testing datasets based on a percentage-based division. The Cleveland dataset mainly comprises studies on coronary heart diseases. They examined the accuracy using 14 different characteristics and used four distinct methods. After the implementation stage, the level of accuracy achieved by random forest is a maximum of 99%, while the level achieved by decision tree is a minimum of 85%. Although it's unclear how childhood CVD risk factors relate to clinical occurrences, they predict subclinical adult VD. The number of risk factors at target can improve cardiovascular-free survival in patients with type 2 diabetes who have a high risk of developing CVD [30]. CVD risk factor assessment's predictive value in cardiotoxic cancer patients is unclear. Prospective multicenter study of moderate/high cardiotoxic anticancer treatment patients [31].

Using the RF classifier and the basic k-means algorithm as ML approaches, Dhar *et al*. [32] created a hybrid method for predicting cardiac conditions. It would seem that a random forest classifier that is applied to specified variables and has a classification accuracy of 100% is the most successful model for predicting individuals who will develop heart disease. Archana Singh and Rakesh Kumar gave research on the accuracy of ML methods for predicting CVD [33]. Following the implementation of the ML strategy for testing and training the accuracy of the KNN, it was shown to be much more effective than other algorithms. The accuracy of the findings was determined to be SVM 83%, DT 79%, LR 78%, and KNN 87%. Bhatt *et al.* [34] created a technique to forecast cardiovascular illnesses accurately to decrease fatalities. This work presents Huang-started k-mode clustering to increase classification accuracy. Nikam *et al*. [35] suggested a model that would use methods from ML to predict cardiovascular illness based on characteristics. One of the aspects that stands out most is the body mass index (BMI). The decision classifier technique was shown to be quite effective in predicting sick individuals by utilizing characteristics such as age, BMI, cholesterol, and many more. The accuracy of the forecast was significantly enhanced with the addition of feature BMI. Not just conventional MS risk factors cause CVD [36]. Multiple factors induce and cause atherosclerosis. But its medium- and long-term effects always induce morbidity and death. Most environmental risk factors cause stress hormone signaling, oxidative stress, and inflammation [37].

## 3. METHOD

In this model, the prediction of heart diseases will be done using ML techniques. The model will use a genetic algorithm with neural networks to reduce dimensionality. The model will also predict the sub-classifications like mild, moderate, and severe. The dataset contains general information along with medical reports in attribute form.

### 3.1. Problem definition

Heart disease is characterized by a high degree of complexity, making it essential to manage it cautiously. If this is not done, it might cause damage to the heart or perhaps lead to an early death. It may be challenging to identify heart disease since there are so many risk factors that might contribute to the condition, such as diabetes, high blood pressure, high cholesterol, an irregular pulse rate, and a significant number of other risk factors. The mortality rate may decrease significantly if the illness is identified when it is still in its early stages and preventive measures are implemented as soon as possible.

### 3.2. Goals

The primary goal is to create an ML-based model that combines particle swarm optimisation (PSO) and K-nearest neighbors (K-NN) algorithms to forecast a person's chance of developing heart disease. A neural network that has undergone cross-validation is used to predict cardiac disease. It also assesses how well the model performs on the selected dataset and compares the outcomes with other machine-learning techniques used for the same problem.

### 3.3. Objectives

Using dataset and ML models may be used to predict CVD, including determining whether people are more likely to acquire heart-related illnesses. By using a dataset, several algorithms will have varying degrees of precision in their results. Objectives of the project are: i) To develop a PSO-integrated KNN to perform feature selection; ii) To design a customized cross-validated neural network that can predict heart stroke; iii) To demonstrate the state of the art, compare ML measures with the current methods.

### 3.4. System architecture

The dataset was taken into consideration as an input. Labeled examples for use in training and assessing models should be included in this dataset. Images may need to be resized, pixel values may need to be normalized, and the dataset may need to be divided into training and validation sets. Create a plan for how your CNN model will work. The hyperparameters that must be determined are the convolutional layers' depth, the filters' width, the number of fully connected neurons, the activation function, and the dropout rate. The entire flow of the proposed work is shown in Figure 4.

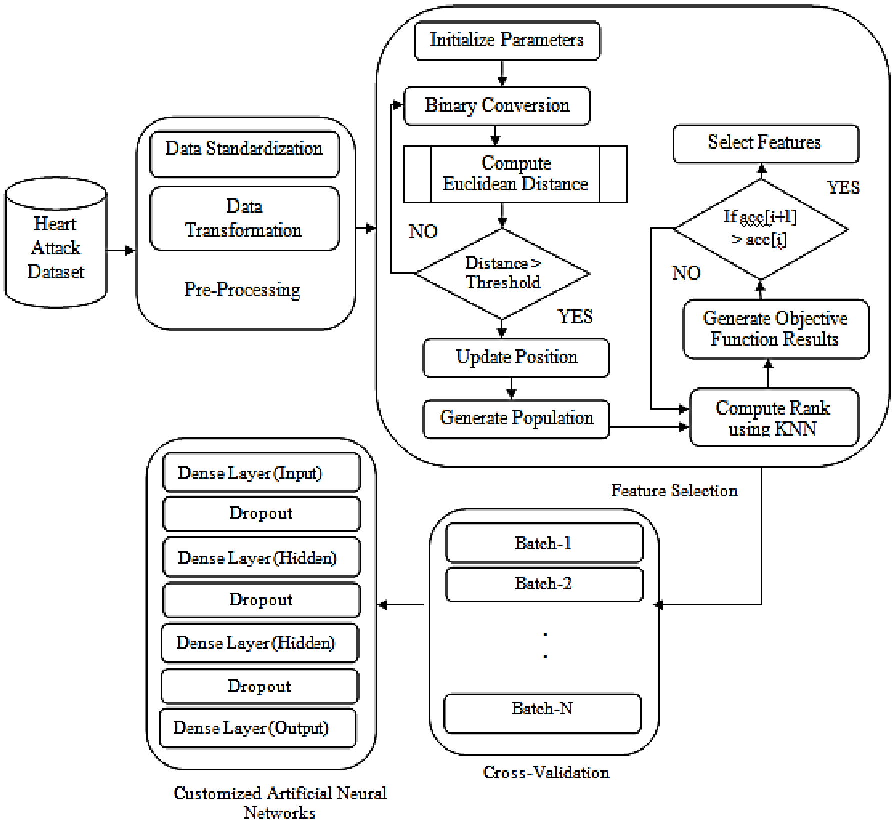


Figure 4. Workflow of the proposed work

To create a CNN model, you must first compile it by deciding on a loss function, optimization technique, and metrics to utilize during training. Set your cross-validation's fold size (K). Five or ten is a popular option. K-Fold By dividing the dataset into K equal parts (folds), cross-validation allows you to evaluate your model's ability to generalize. Separate the dataset into training and validation sets for each fold. PSO automatically tunes KNN hyper parameters, finding the best k and distance metric values. The hyperparameters k and the distance metric substantially affect KNN performance. Therefore, PSO may improve them. PSO efficiently finds optimum hyper parameters compared to brute-force grid search or random search.

### 3.5. Datasets

The data was obtained from the UCI ML repository. Dr Z-Alizadeh Sani collects it for CAD diagnosis. The dataset contains 54 attributes and 304 records. Patients might be CAD or normal. Patients with diameter narrowing of less than 50% have normal diameter narrowing; those with more than 50% have CAD. Demographics, symptoms and evaluation, electrocardiogram (ECG), and laboratory and echo results are the features. Patient demographics, risk factors, and medical history are gathered to study the condition. It includes 304 records and 56 traits. Each patient might be CAD or normal. The attributes of the dataset are shown in Table 1.

Table 1. Various features with various types

| Feature type | Feature name | Range | Feature type | Feature name | Range |
|---|---|---|---|---|---|
| Demographic | Age | 30-36 | Clinical | Blood pressure (BP): mmHg | 90-190 |
| | Weight | 48-120 | | Pluse rate (PR) (ppm) | 50-110 |
| | Length | 140-188 | | Edema | Yes, No |
| | Sex | Male, Female | | Weak peripheral pulse | Yes, No |
| | Body mass index (BMI) | 18.1-40.9 | | Lung rales | Yes, No |
| | Diabetes Mellitus (DM) | Yes, No | | Systolic murmur | Yes, No |
| | Hyper tension (HTN) | Yes, No | | Diasstolic murmur | Yes, No |
| | Current smoker | Yes, No | | Typical chest pain | Yes, No |
| | Current smoker | Yes, No | | Dsypnea | Yes, No |
| | Family history (FH) | Yes, No | | Function class | 1,2,3,4 |
| | Obesity | Yes, No | | Atypical | Yes, No |
| | Cronic renal failure (CRF) | Yes, No | | Nonanginal | Yes, No |
| | Cerebrovasculare accident | Yes, No | | Exertional chest pain (CP) | Yes, No |
| | Thyroid disease congestive heart failure (CHF) | Yes, No | | LowTH Ang (low Thershold angina | Yes, No |
| | Dyslipidemia | Yes, No | | Rhythm | Yes, No |

### 3.6. Attribute selection

During the feature selection process, you will choose the most relevant properties to the prediction model. It is implemented into the system to make it work more effectively. The forecast is based on the patient's gender, the kind of chest pain they are experiencing, their blood pressure while they are fasting, their serum cholesterol, and other patient factors. From algorithm 1, we can easily select the exact attribute to retrieve the precise target class. In developing predictive models, attribute selection is vital, mainly when working with high-dimensional datasets. It includes picking out the most important characteristics (attributes) from the whole collection of features to boost model performance, simplify computation, and improve interpretability. Combining neural networks with the PSO algorithm is one strategy that may be used when selecting attributes [38].

PSO is a natural-inspired optimization technique that mimics the coordinated actions of a group of particles. The optimization issue has several possible solutions, each represented by a particle. Using their collective knowledge and that of their neighbors, PSO's population of particles navigates a solution space in pursuit of the optimal combination of qualities. Particles, locations, velocities, and fitness functions constitute the foundation of PSO. Particles optimize their locations and velocities using information about their previous best states and the best states of their neighbors. They are composed of layered networks of artificial neurons. Because of their adaptability and learning ability, neural networks are increasingly used for classification and regression tasks. Using all available information may occasionally lead to overfitting or higher computational costs, which can significantly impact the performance of neural networks. A population of particles (possible solutions) searches a search space for the best solution in PSO. Each particle in KNN is a possible k-distance metric configuration. The position is a KNN hyper parameter solution, while velocity reflects the particle's search space movement [39].

Start with a pool of particles, each of which stands in for a different set of attributes. To measure how effectively a neural network, trained on the chosen characteristics, performs on a particular job, you must first define a fitness function (such as classification accuracy or mean squared error). After each cycle, particles realign themselves and slow down or speed up depending on where and how fast their neighbors are going. Particle locations stand in for the input variables in a neural network. Use the properties that each particle's location represents to train and assess a neural network on the dataset. Adjust the fitness values of each particle according to the neural network results. Iterate the PSO until a convergence threshold is reached or until a certain number of generations have passed. Once PSO has converged, the optimal attribute subset may be chosen based on the particles' best past placements. Create the final model by training a neural network on only the selected set of attributes. By quickly searching the space of potential attribute combinations, PSO can help neural networks pick a subset of attributes to maximize the model's performance. This method enhances the model's interpretability by zeroing down on the most important features while decreasing the likelihood of overfitting.

Algorithm 1: Particle swarm optimization (PSO) Algorithm+ Neural Networks
Input Parameters: $V_{id} = Velocity$; $X_{id} = position$; $\omega = interia\ Weight$;
$C^1, C^2 = random\ acceleration\ constants$; $P^t_{id} = personal\ best\ t^{th} position\ for\ the\ i^{th}\ particle$;
$P^t_{nd} = local\ best\ position$

BEGIN
Step 1: Set the position and speed of each particle to a random value
Step 2: While not completing all iterations or coming up with a satisfactory solution
do

a. Determine the function value for each solution;
b. The present position and the best locations in history are compared for function value. Update as the current position for each particle if the current location has a higher function value than;
c. Choose a particle from the immediate vicinity of the present particle with the best fitness value; this particle is referred to as the neighborhood best ();
d. for every particle, do

revise the particle's speed following the equation;

Step 3:
$$V_{id}^{t+1} = \omega_i V_{id}{}^{t} + C^1\, rand()(P^t{}_{id} - X^t{}_{id}) + C^2 rand()((P^t{}_{id} - X^t{}_{id})$$

i. Reposition the particle following the equation;

ii. $X^{t+1}{}_{id} = X^t{}_{id} + V_{id}^{t+1}$

END

## 3.7. Models used

### 3.7.1. Artificial neural network (ANN)

It is a popular ML technique for predicting CVDs. The first step in utilizing ANNs to predict cardiovascular illness is to gather a dataset with appropriate patient data. Pretreatment is essential to guarantee the data is clean and acceptable for ANN training. Data cleaning, normalization, scaling, and encoding of categories are all part of this process. Layers of artificial neurons form an ANN for making CVD predictions [40], [41]. The architecture for ANN is shown in Figure 5.

Neurons in the input layer represent data set qualities used to train the model. Each neuron stands for a distinct characteristic. Neurons that process and alter the incoming data are located in one or more hidden levels. A model's number of hidden layers and neurons can significantly affect its performance. In the output layer, one neuron is used to predict the presence or absence of a specific CVD, but multiple neurons are needed to indicate various CVDs. A labeled dataset is fed into an ANN to train it. It is then adjusted to minimize a loss function by changing the synaptic weights and biases. The importance and preferences are typically updated repeatedly using back propagation and the gradient of the loss function.

The ANN is trained to identify intricate data patterns and correlations that predict the danger of developing cardiovascular disease [42]. To check on the model's progress during training and avoid overfitting, a portion of the dataset is often put aside for validation. Adjusting hyper parameters, such as the learning rate, number of hidden layers, neurons per layer, and regularization procedures, can improve model accuracy and generalization.

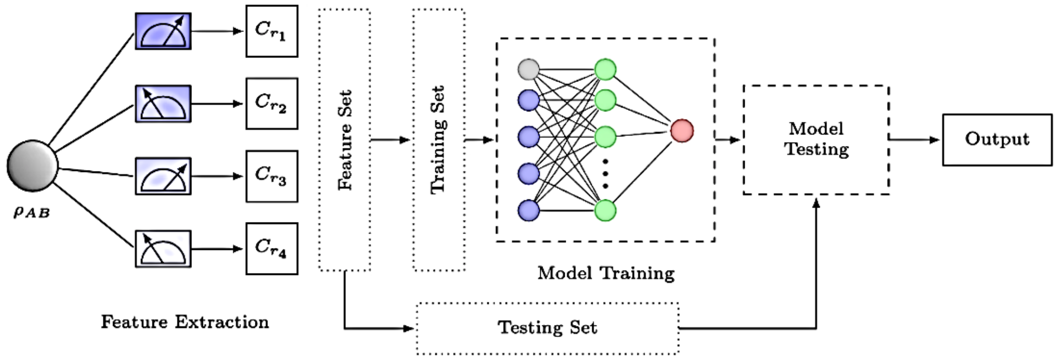


Figure 5. ANN architecture

Learning complicated patterns from patient data, ANNs are a flexible tool for predicting cardiovascular illnesses. They may help doctors spot patients in danger and intervene with them before it's too late [43]. However, before implementing the model in a clinical scenario, it is crucial to guarantee data quality, tweak hyper parameters, and thoroughly evaluate the model.

### 3.7.2. Proposed model (customized artificial neural network)

The proposed model initially pre-processes the data in two ways; first, it converts the categorical data into numerical data using label encoding and normalizes all the values. Then, these values are converted to binary for fast computations using the genetic approaches. The task of the genetic process is to select the features that can maximize accuracy. The proposed model uses the firefly algorithm because it can attract more files based on the brightness. Suppose the space is better than the assumed threshold value. The records whose values are satisfied are passed as input to the KNN; based on the distance between them, it computes the accuracy and prints the attributes. The selected features and randomly cross-validated records are passed as input to the neural network. The proposed model has added dropout layers between the hidden and information since it is a high dimensional data to reduce the computational power. In customized artificial neural networks, the model designs the layers so that for every layer, it adds a dropout layer to filter the unnecessary values and pass only the essential information to the successive layers. The filtration of values helps the networks perform less computation and speeds up the process. The model uses the following estimators in the layers.

– ReLu:

ReLu is a synchronous linear function that outputs one from an input of positivity and zero from a negative value. It has developed into the typical activation function for various neural networks due to a model that uses it. It is the simplest way to train and often yields better results. A neural network's activation function transforms a node's weighted input sum into network activation or a consequence specific to that input. The ReLU model takes much less time because it converges soon during training. This activation function substitutes negative numbers with 0 and leaves positive ones alone. The slopes of small and big values are no longer "killed," and this is computationally much faster.

$$z=\max(0, z)$$

– Adam optimizer

Adaptive moment estimation optimizes gradient descent. The method works effectively for complex challenges with numerous variables or data. It works great and uses little memory. Intuitively, it combines the momentum algorithm with the Root Mean Squared Propagation (RMSprop) approach and gradient descent. The Adam optimizer draws on the two approaches' good qualities and strengths to produce a more optimal gradient descent. It measures the learning rate, including squared gradients, as RMSprop, and, like stochastic gradient descent (SGD) with momentum, it employs momentum with the help of the gradient's arithmetic mean like SGD.Adam is the name of the optimizer, which adjusts the proportional gain for each neural network weight by estimating the first and foremost moments of the gradient. This approach is simple, straightforward, and uses less memory.

## 4. RESULT

SVM, Decision Tree, Ada-boost, ANN, and customized artificial neural network (CANN) are used to predict cardiac illness. There are 56 variables in the dataset. However, only 14 are used for predicting cardiovascular disease. This project considers a wide range of patient characteristics, including gender, chest pain kind, fasting blood pressure, serum cholesterol, etc. Predicting cardiac disease requires evaluating the accuracy of many algorithms and then using the one with the highest accuracy. Accuracy, confusion matrix, precision, recall, and f1-score are the only assessment metrics used to analyze the experiment results. The correlation matrix is shown in Table 2.

During the process of cross-validation for a customized artificial neural network (ANN) using K-fold cross-validation (where K is usually set to 10), you will carry out a series of training and assessment iterations, which are more often known as "folds." The dataset will be partitioned at each fold into various subsets, which will then be used for training and validation. An explanation of what occurs in each of the 10 folds is shown in Figure 6 and Figure 7. The dataset will be partitioned at each fold into various subsets, which will then be used for training and validation. The average score of all folds is shown in Figure 8. The accuracy and loss of the proposed model are shown in Figure 9. The ROC curve is shown in Figure 10, and the AUC-ROC curve in Figure 11.

Table 2. Correlation matrix

| | Age | Weight | Length | Sex | BMI | DM | HTN | Current smoker | Ex-smoker | FH |
|---|---|---|---|---|---|---|---|---|---|---|
| Age | 1.000000 | -0.264585 | -0.163753 | -0.045769 | -0.161414 | 0.072543 | 0.246690 | -0.143879 | 0.076608 | -0.183900 |
| Weight | -0.264585 | 1.000000 | 0.460631 | 0.234529 | 0.725005 | -0.003531 | -0.028532 | 0.157385 | 0.068977 | 0.021963 |
| Length | -0.163753 | 0.460631 | 1.000000 | 0.700279 | -0.269356 | -0.052316 | -0.153668 | 0.335248 | 0.079034 | 0.004488 |
| Sex | -0.045769 | 0.234529 | 0.700279 | 1.000000 | -0.284088 | -0.194348 | -0.149278 | 0.336330 | 0.156932 | -0.071098 |
| BMI | -0.161414 | 0.725005 | -0.269356 | -0.284088 | 1.000000 | 0.045360 | 0.091652 | -0.089398 | 0.005016 | 0.014045 |
| DM | 0.072543 | -0.003531 | -0.052318 | -0.194348 | 0.045360 | 1.000000 | 0.217864 | -0.208458 | -0.120087 | -0.064434 |
| HTN | 0.246690 | -0.028532 | -0.153668 | 0.149278 | 0.091652 | 0.217864 | 1.000000 | -0.169000 | 0.041045 | -0.098467 |
| Current smoker | -0.143879 | 0.157385 | 0.335248 | 0.336330 | -0.089398 | -0.208458 | -0.169000 | 1.000000 | -0.094652 | 0.089532 |
| EX-smoker | 0.076608 | 0.068977 | 0.079034 | 0.156932 | 0.005016 | -0.120087 | 0.041045 | -0.094652 | 1.000000 | -0.080152 |

```
Training for fold 1 ...
Epoch 1/10
8/8 [==============================] - 1s 12ms/step - loss: 0.6127 - accuracy: 0.6653
Epoch 2/10
8/8 [==============================] - 0s 12ms/step - loss: 0.5075 - accuracy: 0.7603
Epoch 3/10
8/8 [==============================] - 0s 11ms/step - loss: 0.4516 - accuracy: 0.7727
Epoch 4/10
8/8 [==============================] - 0s 11ms/step - loss: 0.4373 - accuracy: 0.7810
Epoch 5/10
8/8 [==============================] - 0s 11ms/step - loss: 0.4094 - accuracy: 0.8347
Epoch 6/10
8/8 [==============================] - 0s 11ms/step - loss: 0.4036 - accuracy: 0.8223
Epoch 7/10
8/8 [==============================] - 0s 9ms/step - loss: 0.3809 - accuracy: 0.8306
Epoch 8/10
8/8 [==============================] - 0s 9ms/step - loss: 0.3546 - accuracy: 0.8595
Epoch 9/10
8/8 [==============================] - 0s 8ms/step - loss: 0.3324 - accuracy: 0.8347
Epoch 10/10
8/8 [==============================] - 0s 8ms/step - loss: 0.2879 - accuracy: 0.8678
Score for fold 1: loss of 0.329963356256485; accuracy of 90.32257795333862%
------------------------------------------------------------------------
```

Figure 6. Fold 1 output for customized ANN

```
Training for fold 10 ...
Epoch 1/10
8/8 [==============================] - 1s 13ms/step - loss: 0.5757 - accuracy: 0.6818
Epoch 2/10
8/8 [==============================] - 0s 10ms/step - loss: 0.4920 - accuracy: 0.7769
Epoch 3/10
8/8 [==============================] - 0s 11ms/step - loss: 0.4530 - accuracy: 0.7893
Epoch 4/10
8/8 [==============================] - 0s 11ms/step - loss: 0.4382 - accuracy: 0.7851
Epoch 5/10
8/8 [==============================] - 0s 12ms/step - loss: 0.4302 - accuracy: 0.8264
Epoch 6/10
8/8 [==============================] - 0s 11ms/step - loss: 0.3842 - accuracy: 0.8306
Epoch 7/10
8/8 [==============================] - 0s 11ms/step - loss: 0.3479 - accuracy: 0.8471
Epoch 8/10
8/8 [==============================] - 0s 12ms/step - loss: 0.3395 - accuracy: 0.8512
Epoch 9/10
8/8 [==============================] - 0s 11ms/step - loss: 0.3202 - accuracy: 0.8636
Epoch 10/10
8/8 [==============================] - 0s 12ms/step - loss: 0.2769 - accuracy: 0.8595
Score for fold 10: loss of 0.2417239099740982; accuracy of 93.33333373069763%
```

Figure 7. Fold 10 output for customized ANN

```
Score per fold
------------------------------------------------------------------------
> Fold 1 - Loss: 0.329963356256485 - Accuracy: 90.32257795333862%
------------------------------------------------------------------------
> Fold 2 - Loss: 0.2196495682001114 - Accuracy: 93.54838728904724%
------------------------------------------------------------------------
> Fold 3 - Loss: 0.37064090371131897 - Accuracy: 87.09677457809448%
------------------------------------------------------------------------
> Fold 4 - Loss: 0.3165503740310669 - Accuracy: 89.99997615814221%
------------------------------------------------------------------------
> Fold 5 - Loss: 0.24318501353263855 - Accuracy: 89.99997615814221%
------------------------------------------------------------------------
> Fold 6 - Loss: 0.25850361585617065 - Accuracy: 89.99997615814221%
------------------------------------------------------------------------
> Fold 7 - Loss: 0.3335193395614624 - Accuracy: 86.66666746139526%
------------------------------------------------------------------------
> Fold 8 - Loss: 0.18101267516613007 - Accuracy: 96.66666388511658%
------------------------------------------------------------------------
> Fold 9 - Loss: 0.39297378063201904 - Accuracy: 80.0000011920929%
------------------------------------------------------------------------
> Fold 10 - Loss: 0.2417239099740982 - Accuracy: 93.33333373069763%
------------------------------------------------------------------------
Average scores for all folds:
> Accuracy: 89.76343989372253 (+- 4.3270177141887975)
> Loss: 0.2887722536921501
```

Figure 8. Average score of all folds

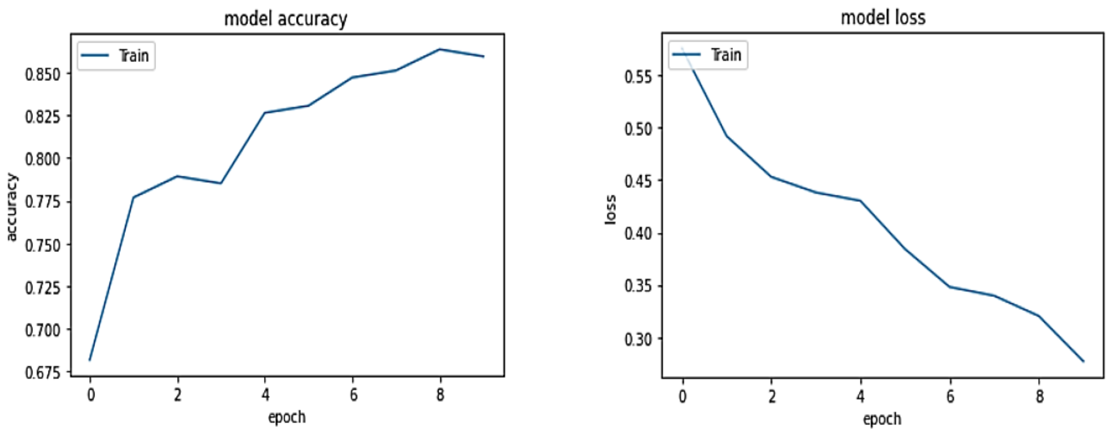


Figure 9. Accuracy & loss of proposed model

The performance analysis of various algorithms is shown in Table 3. We used algorithms like AdaBoost, DT, SVM, ANN, and CANN and compared these algorithms with multiple evaluation metrics. Among them, CANN got the highest accuracy of 94% compared with the remaining algorithms. The 91% in recall, 92% in precision, and 91% for F1 score were obtained for the proposed work. These metrics also got the highest values compared with various models.

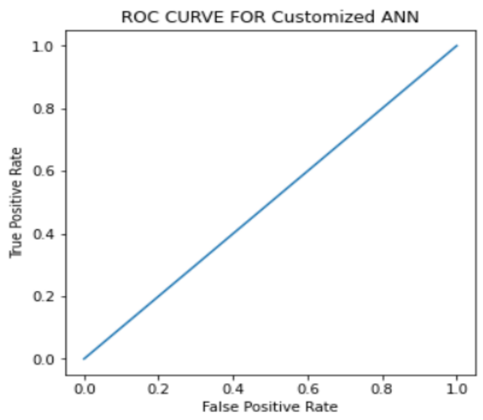


Figure 10. ROC curve for proposed model

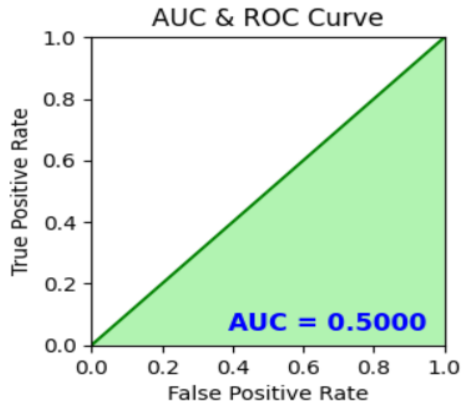


Figure 11. AUC & ROC curve for proposed model

Table 3. Values obtained for various metrics with various models

| Models | Accuracy | Recall | Precision | F1-Score |
|---|---|---|---|---|
| Adaboost | 78% | 78% | 87% | 72% |
| Decision Tree | 73% | 73% | 69% | 69% |
| SVM | 72% | 73% | 89% | 83% |
| ANN | 81% | 82% | 82% | 81% |
| CANN (Proposed Work) | 94% | 91% | 92% | 91% |

Considering Table 3, a graph was obtained to the accuracy shown in Figure 12. In Figure 12, it was proven that CANN got the highest value, and another graph was obtained for various metrics with various models shown in Figure 13. Figure 13 shows that the proposed work had the highest values compared with other models.

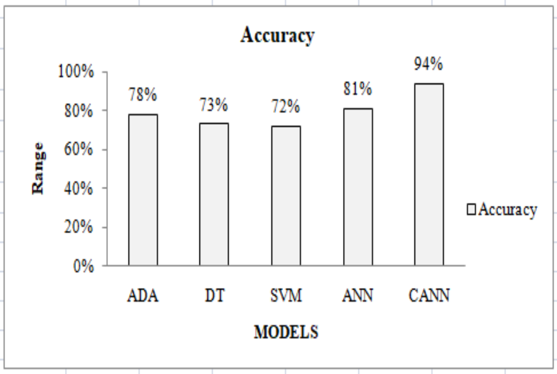


Figure 12. Graph for accuracy with various models

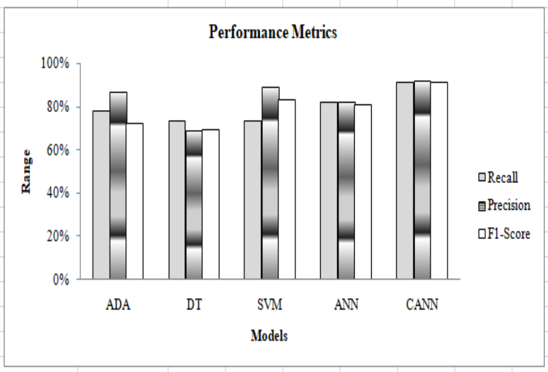


Figure 13. Graph for various performance metrics with various models

## 5. DISCUSSION

Many researchers have worked on prediction, detection, and classification using ML models with various datasets. In this work, we selected a diabetic patient dataset from the UCI repository. We initialized the parameters for that dataset, converted them into binary conversion, and computed the distance using Euclidean distance. The threshold value and the parameter's position will be updated if satisfied. We generated the objective function results and developed the rank using KNN with particles of PSO. From that, the feature for cross-validations with n-batches was selected and forwarded to the customized ANN. The entire process of the attribute selection was explained in Algorithm 1. After the feature selection, the models were applied and compared with the existing models. Hence it was proved that our proposed work had the best accuracy of 94% compared with 92% [7], 91% [10], 92.37% [11], 87.4% [12], 85% [14], 91.5% [15], 90.76% [18], 89.54% [19], 87% [22] and 91% [23]. The proposed work was done with a rare dataset, and the features were selected using PSP with KNN to predict the exact target value so that the patients would not suffer with these correct results.

## 6. CONCLUSION

Based on PSO and KNN, this model predicts heart disease. Our method uses KNN as a classifier to lower the misclassification rate. It uses a PSO-based feature selection measure to choose a few characteristics and enhance the classification performance. The findings imply that the suggested strategy can considerably raise learning accuracy. According to simulation studies, PSO-based feature selection is crucial for categorizing heart disease. Doctors can more accurately predict conditions with common traits thanks to this algorithm. Plans call for the integration of ensemble classifiers with PSO to create a decision support system for the early detection of heart illness. They also include a comparison of GA and PSO for diagnosing heart disease.

## REFERENCES


[1] N. A. ElSayed *et al.,* “10. cardiovascular disease and risk management: standards of care in diabetes—2023,” *Diabetes Care*, vol. 46, no. Supplement_1, pp. S158–S190, Jan. 2023, doi: 10.2337/dc23-S010.

[2] H. Han *et al.*, “Sleep duration and risks of incident cardiovascular disease and mortality among people with type 2 diabetes,” *Diabetes Care*, vol. 46, no. 1, pp. 101–110, Jan. 2023, doi: 10.2337/dc22-1127.

[3] X. Yao *et al.*, “Age at diagnosis, diabetes duration and the risk of cardiovascular disease in patients with diabetes mellitus: a cross-sectional study,” *Frontiers in Endocrinology*, vol. 14, May 2023, doi: 10.3389/fendo.2023.1131395.

[4] S. Reddy, N. Sethi, and R. Rajender, “Risk assessment of myocardial infarction for diabetics through multi-aspects computing,” *EAI Endorsed Transactions on Pervasive Health and Technology*, vol. 6, no. 24, p. 167655, Dec. 2020, doi: 10.4108/eai.17-12-2020.167655.

[5] J. D. Berry *et al.*, “Lifetime risks of cardiovascular disease,” *New England Journal of Medicine*, vol. 366, no. 4, pp. 321–329, Jan. 2012, doi: 10.1056/NEJMoa1012848.

[6] A. Knutsson and H. Bøggild, “Shiftwork and cardiovascular disease: review of disease mechanisms,” *Reviews on Environmental Health*, vol. 15, no. 4, Jan. 2000, doi: 10.1515/REVEH.2000.15.4.359.

[7] S. V. Appaji, R. Shiva Shankar, K. V. S. Murthy, and C. S. Rao, “Cardiotocography class status prediction using machine learning techniques,” *Indian Journal of Public Health Research and Development*, vol. 10, no. 8, pp. 651–657, 2019, doi: 10.5958/0976-5506.2019.01961.2.

[8] N. Townsend, L. Wilson, P. Bhatnagar, K. Wickramasinghe, M. Rayner, and M. Nichols, “Cardiovascular disease in europe: epidemiological update 2016,” *European Heart Journal*, vol. 37, no. 42, pp. 3232–3245, Nov. 2016, doi: 10.1093/eurheartj/ehw334.

[9] R. Shiva Shankar, V. Mnssvkr Gupta, K. V. S. Murthy, and C. Someswararao, “Object oriented fuzzy filter for noise reduction of pgm images,” *Proceedings - ICIDT 2012, 8th International Conference on Information Science and Digital Content Technology*, vol. 3, pp. 776–782, 2012.

[10] S. S. Reddy, N. Sethi, and R. Rajender, “Safe prediction of diabetes mellitus using weighted conglomeration of mining schemes,” *Proceedings of the 4th International Conference on Electronics, Communication and Aerospace Technology, ICECA 2020*, pp. 1213–1220, 2020, doi: 10.1109/ICECA49313.2020.9297390.

[11] J. Delgado-Lista *et al.*, “Long-term secondary prevention of cardiovascular disease with a mediterranean diet and a low-fat diet (cordioprev): a randomised controlled trial,” *The Lancet*, vol. 399, no. 10338, pp. 1876–1885, 2022, doi: 10.1016/S0140-6736(22)00122-2.

[12] A. Groenewegen, F. H. Rutten, A. Mosterd, and A. W. Hoes, “Epidemiology of heart failure,” *European Journal of Heart Failure*, vol. 22, no. 8, pp. 1342–1356, Aug. 2020, doi: 10.1002/ejhf.1858.

[13] C. Berry, D. R. Murdoch, and J. J. V. McMurray, “Economics of chronic heart failure,” *European Journal of Heart Failure*, vol. 3, no. 3, pp. 283–291, Jun. 2001, doi: 10.1016/S1388-9842(01)00123-4.

[14] M. C. Verhaar, E. Stroes, and T. J. Rabelink, “Folates and cardiovascular disease,” *Arteriosclerosis, Thrombosis, and Vascular Biology*, vol. 22, no. 1, pp. 6–13, Jan. 2002, doi: 10.1161/hq0102.102190.

[15] S. Eilat-Adar, T. Sinai, C. Yosefy, and Y. Henkin, “Nutritional recommendations for cardiovascular disease prevention,” *Nutrients*, vol. 5, no. 9, pp. 3646–3683, Sep. 2013, doi: 10.3390/nu5093646.

[16] M. Franco, R. S. Cooper, U. Bilal, and V. Fuster, “Challenges and opportunities for cardiovascular disease prevention,” *The American Journal of Medicine*, vol. 124, no. 2, pp. 95–102, Feb. 2011, doi: 10.1016/j.amjmed.2010.08.015.

[17] M. S. Beeri, R. Ravona-Springer, J. M. Silverman, and V. Haroutunian, “The effects of cardiovascular risk factors on cognitive compromise,” *Dialogues in Clinical Neuroscience*, vol. 11, no. 2, pp. 201–212, 2009, doi: 10.31887/dcns.2009.11.2/msbeeri.

[18] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan, and A. Saboor, “Heart disease identification method using machine learning classification in e-healthcare,” *IEEE Access*, vol. 8, pp. 107562–107582, 2020, doi: 10.1109/ACCESS.2020.3001149.

[19] S. Mohan, C. Thirumalai, and G. Srivastava, “Effective heart disease prediction using hybrid machine learning techniques,” *IEEE Access*, vol. 7, no. 04, pp. 81542–81554, 2019, doi: 10.1109/ACCESS.2019.2923707.

[20] S. S. Yadav, S. M. Jadhav, S. Nagrale, and N. Patil, “Application of machine learning for the detection of heart disease,” *2nd International Conference on Innovative Mechanisms for Industry Applications, ICIMIA 2020 - Conference Proceedings*, pp. 165–172, 2020, doi: 10.1109/ICIMIA48430.2020.9074954.

[21] C. W. Tsao *et al.*, “Heart disease and stroke statistics-2022 update: a report from the american heart association,” *Circulation*, vol. 145, no. 8, pp. E153–E639, 2022, doi: 10.1161/CIR.0000000000001052.

[22] A. K. Dwivedi, “Performance evaluation of different machine learning techniques for prediction of heart disease,” *Neural Computing and Applications*, vol. 29, no. 10, pp. 685–693, 2018, doi: 10.1007/s00521-016-2604-1.

[23] L. Yahaya, N. David Oye, and E. Joshua Garba, “A comprehensive review on heart disease prediction using data mining and machine learning techniques,” *American Journal of Artificial Intelligence*, vol. 4, no. 1, p. 20, 2020, doi: 10.11648/j.ajai.20200401.12.

[24] R. TR, U. K. Lilhore, P. M, S. Simaiya, A. Kaur, and M. Hamdi, “Predictive analysis of heart diseases with machine learning approaches,” *Malaysian Journal of Computer Science*, no. 1, pp. 132–148, Mar. 2022, doi: 10.22452/mjcs.sp2022no1.10.

[25] S. Nashif, M. R. Raihan, M. R. Islam, and M. H. Imam, “Heart disease detection by using machine learning algorithms and a real-time cardiovascular health monitoring system,” *World Journal of Engineering and Technology*, vol. 06, no. 04, pp. 854–873, 2018, doi: 10.4236/wjet.2018.64057.

[26] L. Liberale, L. Badimon, F. Montecucco, T. F. Lüscher, P. Libby, and G. G. Camici, “Inflammation, aging, and cardiovascular disease: jacc review topic of the week,” *Journal of the American College of Cardiology*, vol. 79, no. 8, pp. 837–847, 2022, doi: 10.1016/j.jacc.2021.12.017.

[27] S. Sharma and M. Parmar, “Heart diseases prediction using deep learning neural network model,” *International Journal of*

*Innovative Technology and Exploring Engineering*, vol. 9, no. 3, pp. 2244–2248, 2020, doi: 10.35940/ijitee.c9009.019320.
[28] V. Krishnaiah, M. Srinivas, G. Narsimha, and N. S. Chandra, "Diagnosis of heart disease patients using fuzzy classification technique," in *International Conference on Computing and Communication Technologies, ICCCT 2014*, 2014. doi: 10.1109/ICCCT2.2014.7066746.
[29] V. Sharma, S. Yadav, and M. Gupta, "Heart disease prediction using machine learning techniques," in *2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, IEEE, Dec. 2020, pp. 177–181. doi: 10.1109/ICACCCN51052.2020.9362842.
[30] F. C. Sasso *et al.*, "The number of risk factors not at target is associated with cardiovascular risk in a type 2 diabetic population with albuminuria in primary cardiovascular prevention. post-hoc analysis of the nid-2 trial," *Cardiovascular Diabetology*, vol. 21, no. 1, 2022, doi: 10.1186/s12933-022-01674-7.
[31] J. Caro-Codón *et al.*, "Cardiovascular risk factors during cancer treatment. prevalence and prognostic relevance: insights from the cardiotox registry," *European Journal of Preventive Cardiology*, vol. 29, no. 6, pp. 859–868, 2022, doi: 10.1093/eurjpc/zwaa034.
[32] S. Dhar, K. Roy, T. Dey, P. Datta, and A. Biswas, "A hybrid machine learning approach for prediction of heart diseases," in *2018 4th International Conference on Computing Communication and Automation, ICCCA 2018*, 2018. doi: 10.1109/CCAA.2018.8777531.
[33] A. Singh and R. Kumar, "Heart disease prediction using machine learning algorithms," in *2020 International Conference on Electrical and Electronics Engineering (ICE3)*, IEEE, Feb. 2020, pp. 452–457. doi: 10.1109/ICE348803.2020.9122958.
[34] C. M. Bhatt, P. Patel, T. Ghetia, and P. L. Mazzeo, "Effective heart disease prediction using machine learning techniques," *Algorithms*, vol. 16, no. 2, 2023, doi: 10.3390/a16020088.
[35] A. Nikam, S. Bhandari, A. Mhaske, and S. Mantri, "Cardiovascular disease prediction using machine learning models," in *2020 IEEE Pune Section International Conference, PuneCon 2020*, 2020, pp. 22–27. doi: 10.1109/PuneCon50868.2020.9362367.
[36] J. L. Silveira Rossi, S. M. Barbalho, R. Reverete de Araujo, M. D. Bechara, K. P. Sloan, and L. A. Sloan, "Metabolic syndrome and cardiovascular diseases: going beyond traditional risk factors," *Diabetes/Metabolism Research and Reviews*, vol. 38, no. 3, 2022, doi: 10.1002/dmrr.3502.
[37] T. Münzel *et al.*, "Environmental risk factors and cardiovascular diseases: a comprehensive expert review," *Cardiovascular Research*, vol. 118, no. 14, pp. 2880–2902, 2022, doi: 10.1093/cvr/cvab316.
[38] I. Yekkala, S. Dixit, and M. A. Jabbar, "Prediction of heart disease using ensemble learning and particle swarm optimization," in *Proceedings of the 2017 International Conference On Smart Technology for Smart Nation, SmartTechCon 2017*, 2018, pp. 691–698. doi: 10.1109/SmartTechCon.2017.8358460.
[39] C. Gupta, A. Saha, N. V. S. Reddy, and U. D. Acharya, "Cardiac disease prediction using supervised machine learning techniques," *Journal of Physics: Conference Series*, vol. 2161, no. 1, 2022, doi: 10.1088/1742-6596/2161/1/012013.
[40] O. Y. Atkov *et al.*, "Coronary heart disease diagnosis by artificial neural networks including genetic polymorphisms and clinical parameters," *Journal of Cardiology*, vol. 59, no. 2, pp. 190–194, 2012, doi: 10.1016/j.jjcc.2011.11.005.
[41] K. W. Davidson *et al.*, "Aspirin use to prevent cardiovascular disease: us preventive services task force recommendation statement," *JAMA - Journal of the American Medical Association*, vol. 327, no. 16, pp. 1577–1584, 2022, doi: 10.1001/jama.2022.4983.
[42] J. Azmi, M. Arif, M. T. Nafis, M. A. Alam, S. Tanweer, and G. Wang, "A systematic review on machine learning approaches for cardiovascular disease prediction using medical big data," *Medical Engineering and Physics*, vol. 105, 2022, doi: 10.1016/j.medengphy.2022.103825.
[43] R. Zeleznik *et al.*, "Deep convolutional neural networks to predict cardiovascular risk from computed tomography," *Nature Communications*, vol. 12, no. 1, 2021, doi: 10.1038/s41467-021-20966-2.

## BIOGRAPHIES OF AUTHORS

**Silpa Nrusimhadri** is working as an Assistant Professor in the Department of Computer Science & Engineering at Shri Vishnu Engineering College for Women (A), Andhra Pradesh, India. She is pursuing her Ph.D. in Computer Science and Engineering at Centurion University of Technology and Management (CUTM), Odisha, India. She has 13 years of teaching experience and 8 years of Research Experience. Her research interests include Data Mining, Web Mining, Big Data Analytics, Text Mining, Data Science, Artificial Intelligence and Machine Learning. She is actively involved and successfully implemented two projects funded by DST. She has 14 research Scopus-indexed papers. She can be contacted at email: nrusimhadri.silpa@gmail.com.

**Sangram Keshari Swain** is a Ph.D holder in Computer Science & Engineering having 16 years of teaching experience. He is a Professor in the Department of Computer Science and Engineering along with holding the responsibility of Dean, Examinations and Dean, Students' Welfare in Centurion University of Technology and Management, Odisha. He holds B Tech (IT), M Tech (IT), MBA (HR), LLB & PGDRD degrees. Social Responsibility activity is his Passion always. His research interests and activities include Data Mining, Data Analytics, Big Data, Artificial Intelligence, Machine Learning, Cloud Computing, Internet of Things, Waste Management and Social Development. He has published 02 Patents, 02 Copyrights, 08 Book Chapters, more than 50 Technical Research Papers in different SCI, WoS, Scopus indexed and UGC approved journals. He has presented more than 15 technical papers in different National & International conferences. He has translated and published seven Odia language books on Chemotherapy, Breast Cancer, Lung Cancer, Ovary Cancer, Brain Cancer, Blood Cancer, and Solid Waste Management. He can be contacted at email: sangram@cutm.ac.in.

**Veeranki Venkata Rama Maheswara Rao** is a leading Researcher & Academician in Computer Science & Engineering and holds a Ph.D. degree. He works as a Professor in the Dept. of Computer Science & Engineering at Shri Vishnu Engineering College for Women (A), Andhra Pradesh, India. He is actively involved and successfully implemented three projects funded by DST. He has 54 research papers, 24 of which are Scopus-indexed and 9 of which are Web of Science-indexed. He has 23 years of experience, including 6 years of Industry experience, 19 years of Teaching experience, and 15 years of Research experience. His research interests include Data Mining, Web Mining, Cloud Computing, Big Data Analytics, Data Science, Artificial Intelligence, and Machine Learning. He can be contacted at email: mahesh_vvr@yahoo.com.

**Shiva Shankar Reddy** is an Assistant Professor at the Department of ComputerScience and Engineering in Sagi Rama Krishnam Raju Engineering College, Bhimavaram, Andhra Pradesh, India. He was awarded Ph.D. degree in Computer Science and Engineering with a specialization in medical mining and machine learning from Biju Patnaik University of Technology (BPUT), Rourkela, Odisha, India. His research areas are image processing, medical mining, machine learning, deep learning, and pattern recognition. He published 100+ papers in International Journals and Conferences. S.S. Reddy has filed 05 patents. He can be contacted at email: shiva.csesrkr@gmail.com.

**Mahesh Gadiraju** is a Professor in Department of Computer Science and Engineering and holds PhD Degree. He has received his M. Tech. (CSE) from J.N.T.U. College of Engineering, AP, India in the year 2005 and Ph.D (CSE) from Acharya Nagarjuna University in 2016. His research areas include Content delivery networks, Data Mining, Machine learning. He has 30 research publications including Scopus indexed conference papers and journal papers. Dr. Mahesh acted as coordinator for many International Conferences. He got All India51 rank in GATE exam. He is acting as Research Committee member for S. R. K. R. Engineering College. He can be contacted at email: gmahesh@srkrec.ac.in.