Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

MLP-GA-RF:Hybrid Machine Learning with Fog Integration for Advanced Heart Disease Prediction

M. SriRaghavendra[1], S.Md. Shakeer[2],K. Pream[3], B.Suresh Chandra[4], N.Sethu Madhava[5]

[1]sr.meeniga@gmail.com,[2]mshakeer05@gmail.com,

[3]premchintu05@gmail.com,[4]suresh2892003@gmail.com,

[5]nalagarisethu@gmail.com

[(1,2)]*Department of Computer Science and Engineering (Data Science), Rajeev Gandhi Memorial College of Engineering and Technology (Autonomous), Nandyal, Andhra Pradesh, India.*

[(3,4,5)] *students at RGM College of Engineering and Technology (Data Science Department), Nandyal, Andhra Pradesh, India.*

ARTICLE INFO

Received: 17 Dec 2024

Revised: 19 Feb 2025

Accepted: 26 Feb 2025

ABSTRACT

Unawareness of improper detection and diagnosis are related to disease of heart being the main reason of mortality world-wide. This study overcomes the challenge proposed by this by putting forward a hybrid machine learning model, MLP-GA-RF, which is constituted of MLP (Multilayer Perceptron), RF (Random Forest) and Genetic Algorithm (GA). With high GA efficiency in optimizing the MLP parameters, this results in better predictive performance for this MLP, as well as providing a robust classifier for final diagnosis using RF. Finally, the system is integrated with a fog computing framework for further improvement of real-time diagnostic ability. This reason, among others, is why fog computing is so beneficial and differs from other forms of computing: it processes data nearer to the source, thereby lowering latency and decreasing reliance on centralized cloud infrastructure. This is a highly suitable decentralized approach for mobile applications related to healthcare such as monitoring of remote patient, and has the advantage of fast data analysis due to the distributed system. The recall, F1 score and AUC are rigorously evaluated on the model against conventional classifiers: Logistic, Support Vector Machines, Regression, XGBoost, Decision Trees, and Gradient Boosting. It is found that the MLP-GA-RF model outperforms baseline models across all substrates and consistently provides both more accurate and reliable predictions. Fog computing integrated with an optimized hybrid model is a great enhancement for prediction of heart disease.

Keywords: Heart Disease; Neural Networks; Multi-Layer Perceptron; Genetic Algorithm; Random Forest

1.INTRODUCTION

Heart disease is the No. 1 killer in this country, killing more than 1 in 5 people. A heart disease is a disease with a number of illnesses affecting the heart and blood artery. In addition, it's also known as cardiovascular disease. The primary reasons for many deaths world-wide are heart disease and heart disease-related death which often result from high pressure of blood (hypertension), high level of cholesterol, and bad lifestyle. It is very essential to avoid serious health problems to identify the heart disease early.

CVDs (Cardiovascular disease) are the key reason of mortality in the whole world and add more than 70% of all deaths. It states that about 43% of major deaths are due to disease related to cardiovascular [1, 2]. The next factors of risk of heart disease in high income countries [3, 4]. However, chronic illness prevalence is increasing in low income countries [5]. The global burden of disease related to cardiovascular was predicted to reach approximately USD3.7 trillion between 2010 and 2015[6, 7]. The reports of WHO indicate, the total number of deaths from CVDs will increase to 23.6 million by 2030.

Machine learning (ML) has shown promise for improving healthcare outcomes through data-driven decision-making [8, 9]. The primary goal of the technique related to ML is to create code for computers. There are also

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

several best approaches for increasing the model's accuracy. To avoid overfitting during the training process, ML models needs a high amount of samples of data. However, due to the dimensionality curse, the inclusion related to high number of data features is unnecessary. Most medical datasets contain both related and redundant information. Unnecessary characteristics add no relevant information to the prediction task and cause noise in the target's description (output class), resulting in prediction errors.

Fog computing is related to a new model of computing that increase capabilities to the edge of network, where data may be consumed and processed more readily. Its architecture is intended to meet the expanding demands of the IoT, which generates huge amounts of data that requires low latency replies and real-time processing [10].

In this research, I focus on forming strong and scalable system to predict heart disease. With its latency and bandwidth limits, cloud based systems cannot provide analysis on real time and decision making, and are overcoming this barrier using fog computing where processed date reaches its closer to where it's generated; for example space based calculation data, personal data health data. This study aims to use data from the sample in the medical area to improve the diagnostic performance. The study makes use of cutting-edge technologies and methodologies, such as pre-processing approaches for noisy data, feature selection for increased accuracy, and hybrid models to improve system robustness. Modern technologies such as machine learning and fog computing provide heart disease prediction, providing a novel and proactive approach to maintaining cardiovascular health.

Section 2 explains about the different research papers based on the hybrid models. Section 3 explains about the system architecture of the proposed model. Section 4 explains about the dataset. Section 5 explains about the Methodology. Section 6 will do results discussion and Section 7 discuss about the conclusion of the research.

2.LITERATURE REVIEW

In a study conducted by Ali Al Bataineh et al. [11] used both Multilayer Perceptron (MLP) Particle Swarm Optimization (PSO) which helps to enhance heart disease prediction. They tested their model against ten techniques of learning through, including Decision Trees, and XGBoost, using the Disease of Cleveland Heart data set. The MLP-PSO provides best results and outnumbered the other approaches, with the accuracy of 84.61% which is considered maximum.

This study, by Lakshmi et al. [12], improves cardiac prediction. The Framingham dataset is pre-processed, and the EWOA (Enhanced Whale Optimization Algorithm) chooses the most important characteristics. Integrating EWOA with machine learning increases the accuracy to 85.79%.

In the work of [13] a model of machine learning and EDA (Exploratory Data Analysis) of demographic data was used to predict cardiac disease. Logistic Regression, and Decision Tree are examined; Random Forest gives best accuracy (88%). RS Algorithm uses the advantage of Random Forest to achieve 93% accuracy.

In this work [14] Chintan et al. advance to the result of heart disease using advanced models of learning of machine on a Kaggle dataset of 70,000 examples. Models that we use to train are MLP and Xgboost; they use pre-processing procedures such as outlier removal and feature selection before training these models. Improving classification and hyper parameter tuning are done by applying the K mode clustering and GridSearchCV.

Combining GA (Genetic Algorithm) and SVM (Support Vector Machine) in prediction of heart disease, Syeda et al. [15] can improve. Finally, we use the UCI Heart Disease dataset (1025 examples) to allow GA to select features in order to improve SVM classification. It is shown that 86% accuracy is obtained for the Linear SVM, 93% for the Polynomial SVM model and 98% for the GA SVM model. The results attest that GA can help minimize complexity without deteriorating the prediction accuracy.

Optimization techniques and machine learning are applied in this study to improve the diagnosis of heart disease as per Ravichandra et al [16]. A hybrid technique in order to enhance feature selection combines optimization algorithms, like GA, PSO, GWO, with the Random Forest (RF) classifier. Other methods fail and the RF-GWO model tops them for highest accuracy.

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

In this work, models of machine learning are used for forecasting of heart failure in a Kaggle dataset of 303 patient records by Vinod et al. [17]. Cross validation was done to increase accuracy of the models tried, three models XGB (XGBoost), LR (Logistic Regression) and SVC (Support Vector Classifier) tested. An accuracy of Logistic Regression of 88.87%, SVC has 86.27% and XGBoost has 92.51% highest. In this study, the features for Logistic Regression (LR) are selected using PSO (Particle Swarm Optimization) to improve prediction of heart disease (Priyanka at al. [18]).

The preprocessing processes such as removing outlier, feature scaling, are carried out on UCI Heart Disease dataset. PSO selects only the relevant features and decreases the dimensionality while increasing the efficiency of the model. It shows that the feature selection strategy based on the PSO optimized LR model achieves the accuracy of 90.74%, which is superior to traditional feature selection strategies.

In this work, Chethana et al. [19] proposes to use the K Nearest Neighbors classifier to do the same thing for the same dataset from heart disease. Three Optimizable KNNs are considered, and the erroneous model reaches 69% of accuracy with the lowest cost of misclassification. Secondly, the Coarse KNN model could identify heart disease patients with a high True Rate of Positive cases 82.4%. The findings indicate that under a longer training duration, Optimizable KNN proves to be the most effective and balanced model.

In this work, AdaBoost algorithms is used for the improvement of coronary heart disease prediction based on Suhitha et al. [20]. The hybrid model outperforms Naïve Bayes, with 97.43% accuracy, 95.67% True Positive Rate, and 94.65% Precision. It is shown that this method outperforms in heart disease classification.

This study compares Logistic Regression (LR) and Random Forest (RF) model for easily predicting heart disease risk from 304 records of patients. The model using the Random Forest methods was able to achieve 90.16% accuracy after preprocessing and feature analysis and surpassed the Logistic Regression iteration with 85.25%. They [key indications of heart disease] have found chest pain type, ECG readings and thalassemia to be good indicators of heart disease. The betting Random wood has been proven to be the most error and resilient of itinerary bingo than the curtain wire.

This Table 1 illustrates how certain algorithms of machine learning are used to identify heart disease.

Table-1: Comparison of various research papers

Title	**Algorithms**	**Metrics**	**Dataset used**	**Results**
Ali Al Bataineh et al. [11]	MLP, PSO	Accuracy, AUC, F1-score.	Cleveland Heart Disease	MLPPSO achieved 84.61% accuracy.
Lakshmi et al. [12]	SVM, Random Forest (RF), EWOA	Accuracy, Precision, Recall, F1-score.	Framingham Heart Disease	Hybrid SVM-RF achieved 85.79% accuracy.
Vibha et al. [13]	Random Forest (RF), Support Vector Machine (SVM)	Accuracy, AUC, F1-score.	Cleveland Heart Disease	RS Algorithm achieved 93% accuracy.
Chinta et al. [14]	Multilayer Perceptron (MLP),XGBoost (XGB),Decision Tree (DT), RF	Accuracy, AUC, F1-score.	Cleveland Heart Disease	MLP achieved 87.28% accuracy.

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Syeda et al. [15]	Genetic Algorithm (GA), SVM	Accuracy, Sensitivity, Specificity.	Irvine Heart Disease	GA-SVM achieved 98% accuracy.
Ravichandra et al. [16]	Random Forest (RF), GA, PSO, Grey Wolf Optimizer (GWO)	F1-score, Accuracy, Precision, Recall	UCI Heart disease	RF-GWO achieved the highest accuracy among tested models.
Vinod et al. [17]	XGBoost(XGB), Logistic Regression (LR), Support Vector Classifier (SVC).	Accuracy, AUC, F1-score.	Heart Disease Analysis and Prediction Dataset	XGB achieved 92.51% accuracy.
Priyanka et al. [18]	Particle Swarm Optimization (PSO), Logistic Regression (LR).	Accuracy, Precision, Recall, AUC.	Cleveland Heart Disease	PSO-LR achieved 90.74% accuracy
Chethana et al. [19]	Fine KNN, Medium KNN, Coarse KNN, Cosine KNN, Cubic KNN, Weighted KNN, Optimizable KNN	Accuracy, True Positive Rate (TPR), False Negative Rate (FNR).	Cleveland Heart Disease	Optimizable KNN achieved 69% accuracy, with the best prediction speed
Suhitha et al. [20]	Decision Tree (DT), AdaBoost.	Accuracy, Precision, TPR.	Framingham Heart	Hybrid ML achieved 97.43% accuracy.
Soumyalatha et al. [21]	Random Forest (RF), Logistic Regression (LR)	Accuracy, Precision, F1-score.	Cleveland Heart Disease	RF achieved 90.16% accuracy.

3.SYSTEM ARCHITECTURE

In order to aid real time medical diagnosis, this system Figure 1 proposes a Multi-Layer Perceptron (MLP) trained by Genetic Algorithm (GA) and Random Forest (RF) based on fog computing. Data collection is the first part of the process in which various records of a patient's health from hospitals, wearable sensors, healthcare databases are collected. The hybrid approach consists of MLP as the core classifier and GA for MLP hyperparameters' optimization to improve the ML efficiency. Finally, RF is used to rank feature importance and provide other forms of validation, in order to make classification robust.

For the sake of real-time decision making, it was deployed on top of a Fog Computing framework consisting of a centralized Cloud Server and several Fog Node (edge devices) deployed near the healthcare sites. This strategy does not rely on any centralized sources, thus making it possible to reduce latency and impact the cloud infrastructure as less. Iterative Process makes the system a Learner by periodically feeding newly diagnosed instances into it to

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

modify the model. In the final phase, Result Analysis, the prediction results are offered in a simple interface for doctors to use and make reasoned choices.

By the use of such AI driven distributed computing architecture that improves heart disease prediction accuracy, enables real time diagnosis and scalability. By integrating MLP, GA, RF, and Fog computing, the system is formed on the basis of a powerful, efficient, and adaptive medical decision support tool that can lead to considerable benefits of such system to healthcare institutions and remote patient monitoring applications.

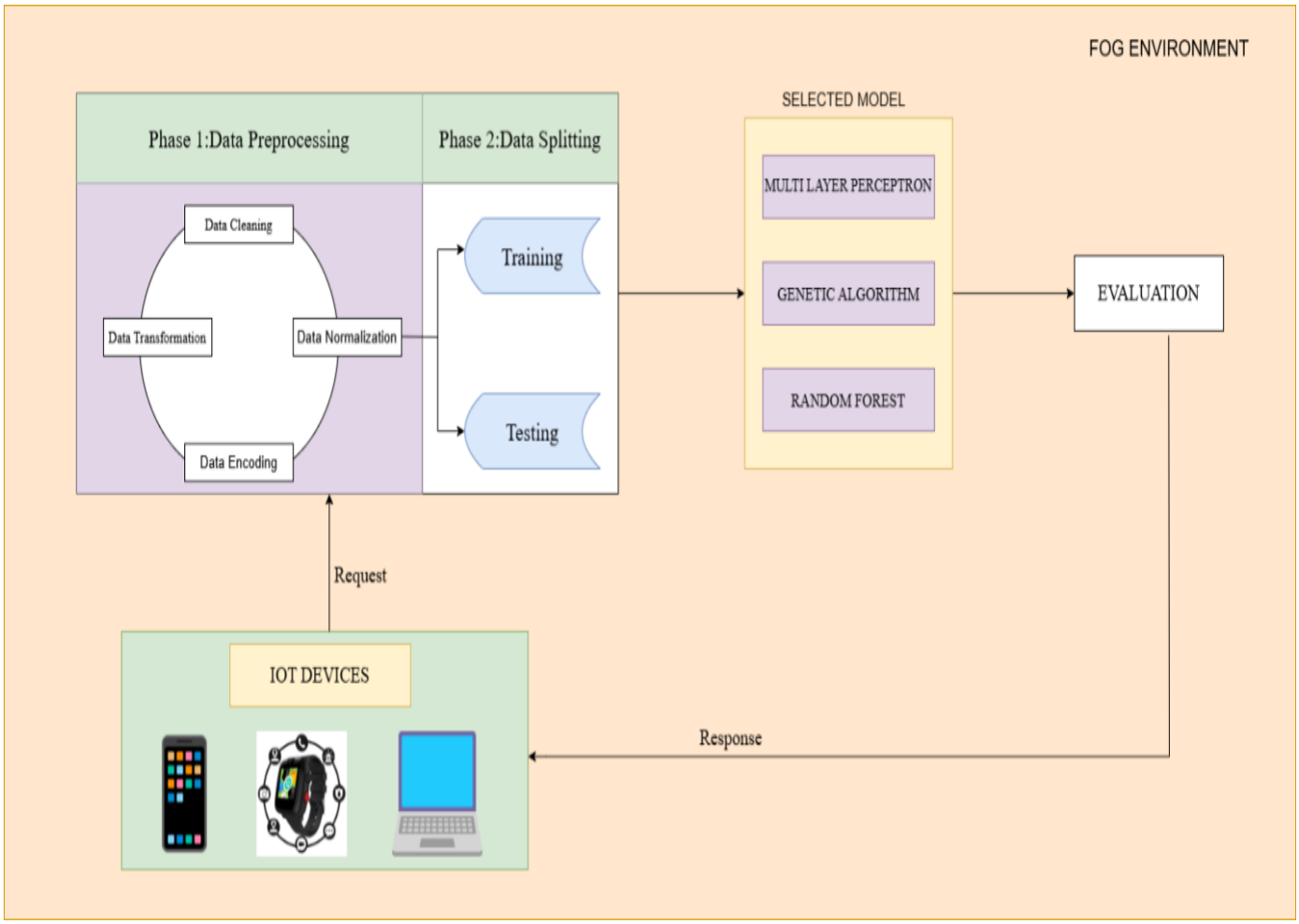


Fig-1: Model Architecture

4.DATA COLLECTION

Dataset source: This is obtained from the public dataset repository of the Kaggle website.

4.1 Dataset Description: For this study the Cleveland Heart is used to identify Disease dataset to predict the heart disease. As information for this dataset, we used a dataset of patient data and prediction of heart disease. The dataset of Disease of Cleveland Heart is popular because it is well structured, clinically relevant characteristics, has a balanced data, and therefore is a good dataset for algorithms of machine learning. Also, it is utilized as a benchmark dataset in research for comparing and evaluating predictive systems. It contains a feature matrix of 303 records and 14 features - demographic data, and diagnostic test results.

Table 2 shows the dataset, which encompasses many features that could be feasible to predict the heart disease. In this case, the result variable 'Target' is used to identify heart disease based on the feature.

Table-2: Dataset Description

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Sl.no	Feature	Description	Data Type	Data Range	Units
1	age	Age	Numeric	29-77	Years.
2	sex	Gender	Binary	1 = male, 0 = female	N/A
3	Cp	Chest pain types	Nomial	0: Typical Angina 1: Atypical Angina 2: Non-anginal Pain 3: Asymptomatic	N/A
4	trestbps	blood pressure	Numeric	94 - 200	mmHg
5	chol	Level of Serum Cholesterol	Numeric	126 - 564	mg/dL
6	fbs	Fasting Blood Sugar > 120	Binary	1 if true, 0 otherwise	mg/dL
7	restecg	Results of Resting Electrocardiographic	Nomial	0: Normal 1: ST-T wave abnormality 2: Left ventricular hypertrophy	N/A
8	thalach	Maximum heart rate	Numeric	71-202	bpm
9	exang	Angina Exercise-Induced	Binary	1 = Yes, 0 = No.	N/A
10	oldpeak	ST depression	Numeric	0.0 – 6.2	mm
11	slope	Slope of Peak Exercise ST Segment	Nomial	0: Upsloping 1: Flat 2: Downsloping	N/Al
12	ca	Number of Major Vessels	Nomial	0 to 3.	count
13	thal	Thalassemia Type	Nomial	1: Fixed Defect 2: Normal 3: Reversible Defect	N/A

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

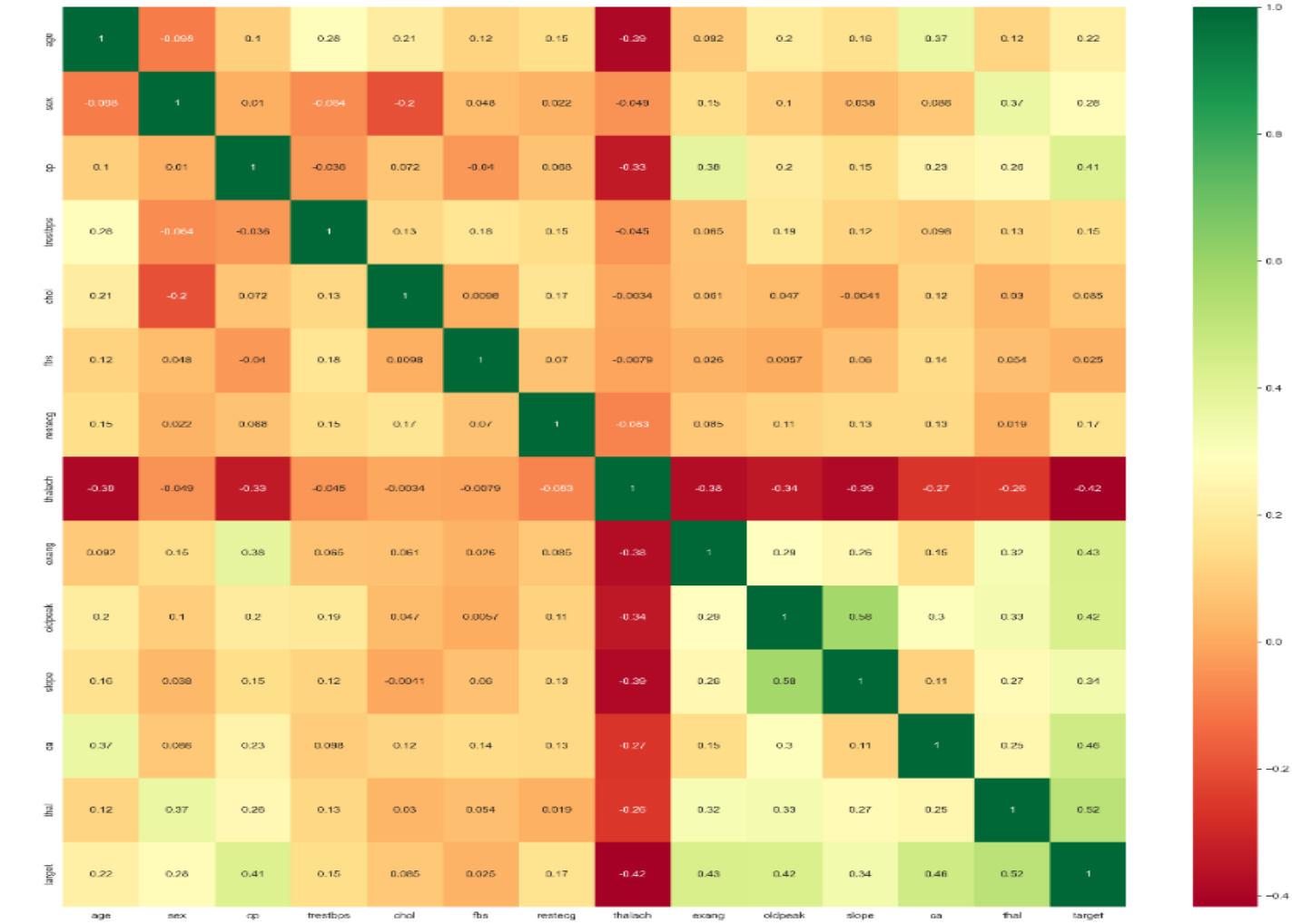
Research Article

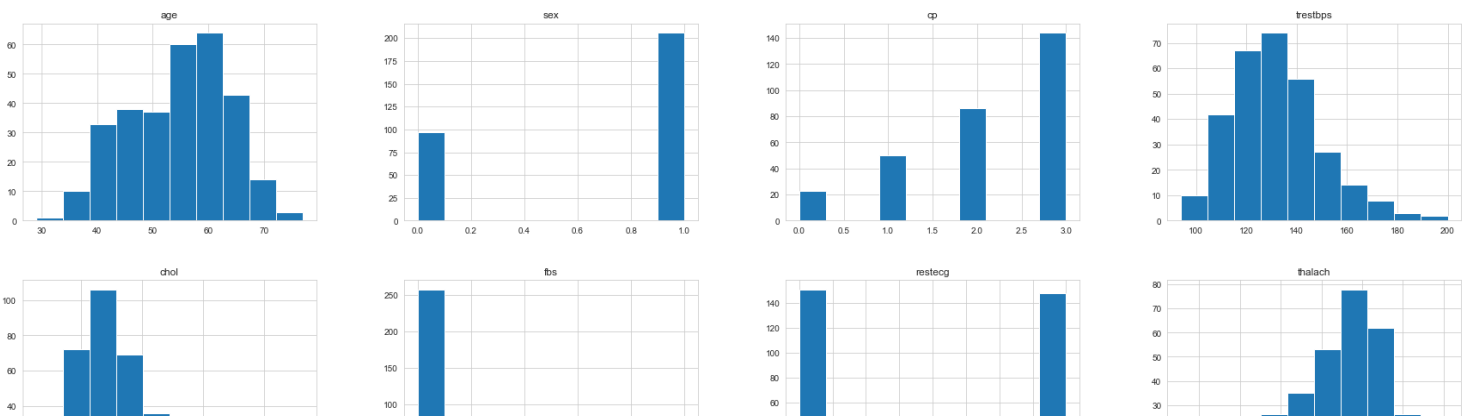
14	target	Indicator of Heart Disease	Binary	1 = Disease, 0 = No Disease	N/A

4.2 Exploratory Data Analysis:

One of the important stages of working with dataset using machine learning models is Exploratory Data Analysis (EDA). In other words, it involves giving a sum total of important statistics, which in pattern show the aberrations if any. EDA aids in the identification of feature correlations, the handling of missing values, and the improvement of feature selection for higher prediction accuracy. Figure 2 depicts the representation of the matrix correlation.

Statistically speaking, the Pearson Coefficient of Correlation is the measure of direction a relationship between two numerical variables. It helps in better grasping how any change in one variable affects another, that this is productive for analysis of data and machine learning. Each attribute calculated to find out Pearson correlation between coefficients. The values of correlation lie from -1 to 1, where a negative correlation is reflected by a value -1, no correlation corresponds to a value 0, and a value 1 signifies a perfect correlation.





Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Fig-3: Visual Representation of each attribute

Fig-3: Visual Representation of each attribute

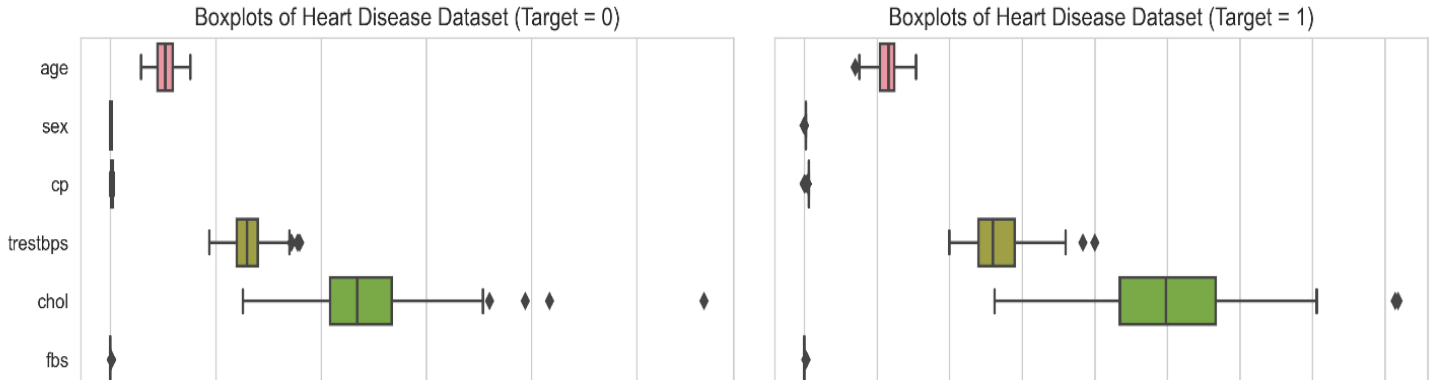
Evaluating data distribution and giving a spike on outlying data are two benefits that boxplots can perform well. Boxplots can apply to a dataset pertaining to heart illness in order to identify, among other things, the distribution of different aspects of heart illness related variables. Boxplot is show in Figure 4.

4.3 Data Preprocessing

Preparation of data is an important first step when making dataset suitable for use with algorithms of machine learning. Preprocessing of data is a method of converting raw data into a required form or structured data. Other such tasks could be data cleansing, normalization, or encoding. The most important goal of preprocessing is to decrease the degree of noise, and outliers. Steps for preprocessing data:

Managing Missing Values: We should be able to manage when any of the values given is missing, we are required to get rid of the non-existing rows, replace them with an obvious value (e.g. mean, median and mode), or bring in more sophisticated imputation methods.

Coding Category Variables: This means, we have to map the category variables to some numerical variables so that the machine learning algorithms take in numeric input. It is possible to encode too, for example by means of one hot (or label) encoding.



Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Fig-4: Visual Representation of Boxplots

Splitting the features: we must split the inputs and the target, that is data split into X (features) and Y (target). To be effective for training and evaluate the models, one needs to split.

Splitting the Dataset: It is partitioned into test and training dataset.

Feature Selection: The crucial phase in pre-processing is feature selection. The feature selection process removes unnecessary features from the data space. Otherwise, introducing undesired and repeated data to the process increases the time and complexity of computation. Important features are selected based on their importance, allowing for easy feature classification without modifying the original subset.

Feature selection is the same as pattern classification, and it is divided into two methods: filter technique and wrapper approach. If the feature selection process is independent, it is referred to as a filter-based technique, and it is determined by the characteristics of the data. If a classifier is used, it is a wrapper approach; the feature obtained from the wrapper method is determined solely by the classification algorithm utilized. Two classifier methods provide two distinct feature subsets. Wrapper method is more effective than filter method, although it is time-consuming.

5.METHODOLOGY OF THE PROPOSED SYSTEM

The Algorithm of Genetics (GA) is a strategy for evolutionary optimization depends on natural selection and genetic inheritance. The process begins with the initialization step, which generates a random population of candidate solutions known as chromosomes. Each chromosome symbolizes one possible solution to the situation at hand. These chromosomes are evaluated for fitness, which is determined by a preset fitness function based on how successfully the problem is solved.

Algorithm 1: Genetic Algorithm (GA)

Input: Population Size P, Mutation Rate μ , Crossover Rate C_r, Maximum Generations G_{max}

Output: Best Solution θ^*

Step 1: Initialize Population

- Generate a population P= { θ_1, θ_2, ..., θ_n }with random MLP parameters.
- Evaluate the fitness for each solution using a predefined fitness function.

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The next step is the selection process, Common selection approaches include Roulette Wheel Selection, which probabilistically Favors fitter individuals, Tournament Selection, in which a subset of the population competes, and Rank-Based Selection, which chooses based on ranked fitness values. In crossover (recombination) the chosen parents engage in two parent chromosomes exchanging genetic material to form offspring. Crossover methods are available in several forms such as single point crossover that exchanges genes at a randomly selected point; two point crossover that exchanges genes from points along the chromosome; and uniform crossover that alternates between two genes located at the same point of the chromosome.

Genetic diversity is maintained in case the children are mutated by a minor random change, for instance, just flipping the bit of a binary chromosome. Also, this prevents the algorithm falling into local optima, and introduces diversity in the search space. When crossing over and mutation takes place, a new generation replaces the old population, ensuring that the best solutions live and evolve. At each iteration of this cycle, the evaluation, selection, crossover, and mutation steps are performed and the process is continued until either a given number of generations are reached or an optimal solution is reached. GA iteratively refines solutions in these steps to come up with near means to an optimization problem or to the best possible solution.

Objective Function:

In particular, the MLPGAN class has a function of fitness which evaluates the quality of a given solution (or a set of MLP parameters) in terms of how well the MLP predicts while minimizing the prediction error, balancing classification accuracy.

Mathematical Formulation:

The objective function $F(\theta)$ for a given set of MLP parameters θ is defined as:

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

$$F(\theta)=\text{Accuracy}(\theta) - \lambda \cdot \text{MSE}(\theta) \quad \text{----------}>\text{Eq(1)}$$

Here, (·) is a function for indicated that returns 1 if the prediction is correct and 0 otherwise.

where:

1. Classification Accuracy:

$$\text{Accuracy}(\theta) = \frac{1}{N}\sum_{1=1}^{N} 1\,(\hat{y}_i \geq 0.5 = y_i) \quad \text{----------}> \text{Eq(2)}$$

Let:

- y_i be the actual target label for the i^{th} sample.
- $\hat{y}_i$ Be the probability which predicted the positive class for the i^{th} sample.
- N is the overall number of samples.

2. Mean Squared Error (MSE):

$$\text{MSE}(\theta) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \quad \text{----------}>\text{Eq(3)}$$

3. Weighting Factor (λ):

✓ However, a coefficient λ=0.4 is used to strike a balance between getting the maximum accuracy and the minimum error for the decision severity.
✓ A lower λ prioritizes accuracy more, while a higher λ penalizes large errors more strictly.

5.1 MLP Trained by Genetic Algorithm (GA)

However, a coefficient λ=0.4 is used to strike a balance between getting the maximum accuracy and the minimum error for the decision severity. A lower λ prioritizes accuracy more, while a higher λ penalizes large errors more strictly. To address these challenges, a Genetic Algorithm (GA) is utilized to improve MLP weights and biases. The training procedure follows these steps:

1.Encoding of Biases and Weights: The structure of MLP is made up of input-hidden and hidden-output weight matrices, layers of hidden biases, and layer of output biases. These are flattened into a one-dimensional vector. Each chromosome showed a real solution in the population.

2.Function of Fitness Evaluation: MLP uses each candidate solution (chromosome) to undertake forward propagation. The calculation of fitness score Eq(4) is done by comparing the output predictions to real labels. The fitness function takes into account accuracy and mean squared error (MSE) to ensure classification performance while minimizing loss.

$$\text{Fitness} = \text{Accuracy} - 0.4 \times \text{MSE} \quad \text{--------}> \text{Eq(4)}$$

3.Selection of Parents: Probabilistic selection methods are used to pick the best-performing individuals (chromosomes) based on the score of their fitness (for example, roulette wheel selection).

4.Crossover Operation: To create children, two selected parent chromosomes are recombined at a random crossover spot.

5.Mutation Operation: Small random alterations are made to child chromosomes to maintain diversity and prevent premature convergence.

6.Evolution Over Generations:

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Steps 2–5 are done for several generations until an optimal solution is identified. The best-performing chromosome (set of MLP weights and biases) is utilized to train the final MLP model.

5.2 MLP-GA with Random Forest:

After training the MLP with GA, the next step is to combine it with a Random Forest (RF) classifier to refine and improve predictions.

1.MLP-GA Predictions as New Features: The predictions provided by the optimized MLP-GA model are included as a new feature in the dataset. This means that the dataset now includes both the original features and the MLP-GA predictions.

2.Training the Random Forest Model: A classifier of Random Forest (with adjustable hyperparameters) is effectively trained on the MLP-GA-augmented dataset. It learns patterns from both the original features and the MLP-GA results.

3.Final Classification & Evaluation: The trained RF model makes the final prediction., F1-Score, Accuracy, Precision, Recall and AUC are all performance measurements.

6.RESULT ANALYSIS

Table 2 compares various machine learning models, including MLP-GA, MLP-GA-RF, Regression of Logistics, SVM, Decision Trees, KNN, MLP-BP, and XGBoost. To improve performance, each model is adjusted with unique hyperparameters, resulting in robust classification and predictive accuracy.

Table-3: Parameters of various ml algorithms

Algorithm	Parameters	Values
MLP-GA	Hidden Layers	1 (64 Neurons)
	Population Size	200
	Generations	100
	Mutation Rate	0.05
	Crossover Rate	0.8
	Activation Function	Leaky ReLU (alpha = 0.01)
	Output Activation	Sigmoid
	Fitness Function	Accuracy - (0.4 * MSE)
MLP-GA-RF	(n_estimators) Number of Estimators	200
	(max_depth)	10
	(min_samples_split)	4
	(min_samples_leaf)	2
Logistic Regression	Regularization (C)	1.5
	Penalty	L2
Support Vector Machine	Type of Kernel	RBF

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

(SVM)	Regularization Parameter (C)	1
	Gamma	0.1
K-Nearest Neighbors (KNN)	Number of Neighbors (n_neighbors)	5
Decision Tree	Criterion	Gini
Random Forest	Number of Estimators (n_estimators)	1000
	Criterion	Gini
Extra Trees	Number of Estimators (n_estimators)	100
Gradient Boosting	Number of Estimators (n_estimators)	100
	Maximum Depth	3
Naïve Bayes (GaussianNB)	Assumption	Gaussian Distribution
XGBoost	Number of Estimators (n_estimators)	300
	Maximum Depth	15
MLP-BP	Hidden Layer Size	(30,)
	Activation Function	ReLU
	Solver	Adam

The report helps to apply key principles of machine learning technology to make a diagnostic for disease related to heart. Our suggested model, MLP-GA-RF, is trained, optimized, evaluated, and compared to existing machine learning methods. Table -4 displays the findings related to experiments for every model of machine learning.

Table-4: comparison of performance using various evaluation metrics

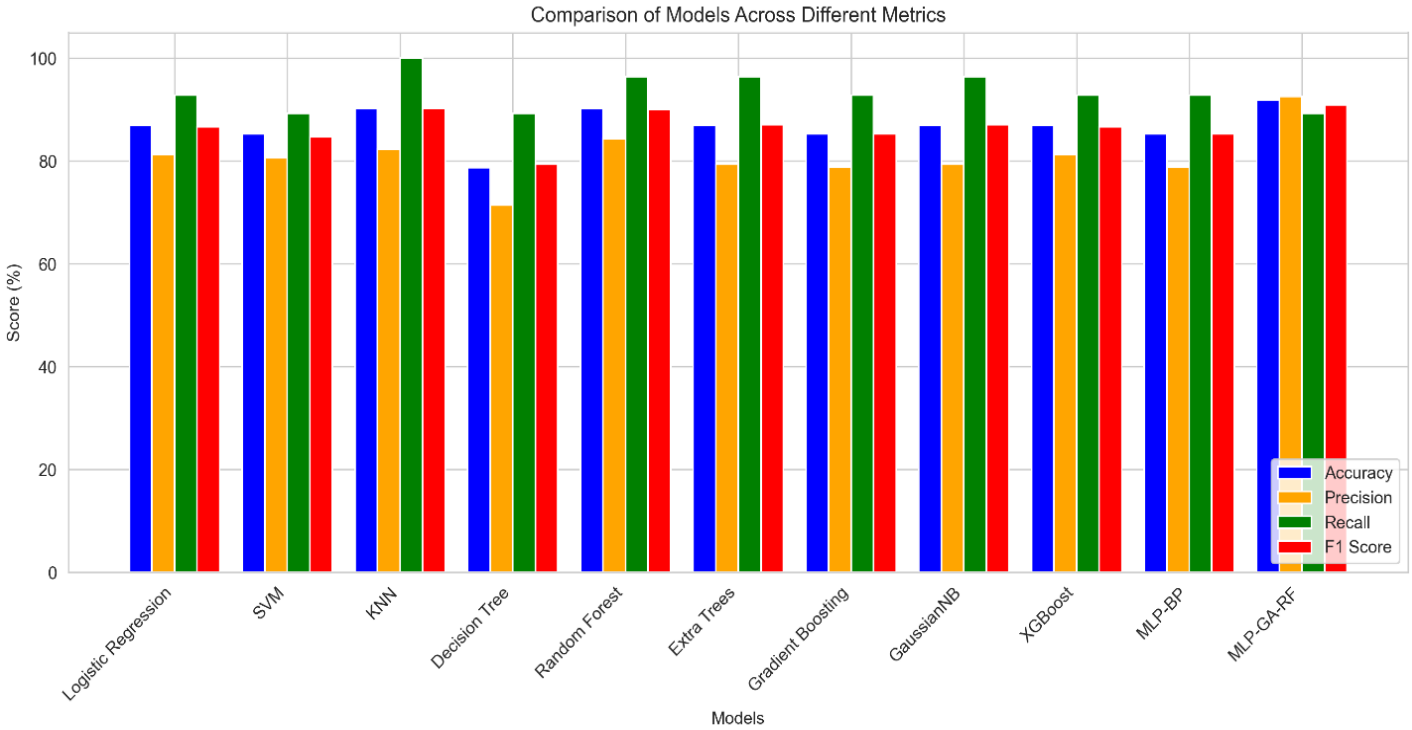
Model	Accuracy	Precision	Recall	F1 Score	AUC
MLP-GA-RF	0.95082	0.962963	0.928571	0.945455	0.949134
KNN	0.901639	0.823529	1	0.903226	0.924242
Random Forest	0.885246	0.818182	0.964286	0.885246	0.955628
Logistic Regression	0.868852	0.8125	0.928571	0.866667	0.952381
GaussianNB	0.868852	0.794118	0.964286	0.870968	0.949134
XGBoost	0.868852	0.8125	0.928571	0.866667	0.906926
MLP-BP	0.868852	0.8125	0.928571	0.866667	0.952381

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

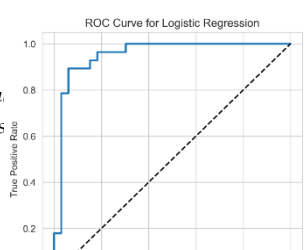
SVM	0.852459	0.806452	0.892857	0.847458	0.944805
Gradient Boosting	0.852459	0.787879	0.928571	0.852459	0.945887
Extra Trees	0.836066	0.764706	0.928571	0.83871	0.945887
Decision Tree	0.721311	0.666667	0.785714	0.721311	0.72619

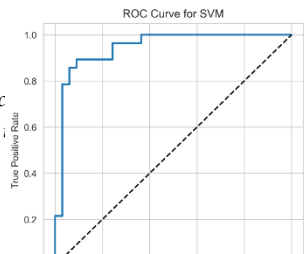
As per the results, the KNN has a score of 90.16, MLP-GA-RF has a score of 95.08 and the Random Forest is at 88.52. Among the models tested, the MLP GA model had the highest value of precision of 96.29, and at the same time a good compromise between recall 92.86 and F1 score 94.55. It also performed well in the Random Forest model with an AUC of 95.56 thus showing it can distinguish between the classes. KNN's performance was interesting in that it had good recall as also 100%, meaning that it detected affirmative case very well but they are not spotted with much precision; Logistic Regression 86.89% and SVM 85.25% made good use of traditional machine learning models, but then ensemble based models performed much better. Since the accuracy of the Decision Tree model was 72.13% and was likely the most vulnerable to overfitting and generalization challenges the model could have, it is most possible. Gradient Boosting (AUC = 85.25%) and Extra Trees (AUC = 84.61%) showed high AUC values (94.5%) on models as to their ability to well deal with the distribution of data. Since this classifier has a high recall of 96.43%, it is appropriate for an application which does not strictly rely on low false negatives. This means that the increased performance of MLP-GA-RF cannot only be attributed to MLPGA, and indeed in results, MLP combined with Genetic Algorithm and Random Forest has an improvement over feature selection and optimization which results to better classification. This puts emphasis on the use of hybrid techniques in the medical diagnosis process where the importance of precision and recall is crucial in the process of correct prediction.



than the others in all criteria. On memory, Logistic Regression and Decision Tree based models like these have very low accuracy and rate more false negs, more than a single them. Appendix A contains the ROC curves for all classifiers.

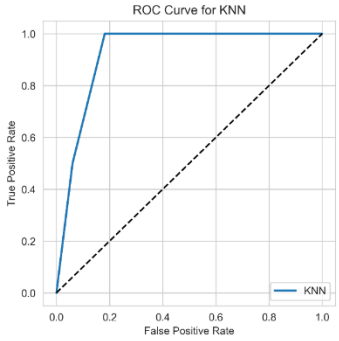
Appendix A: ROC Curves

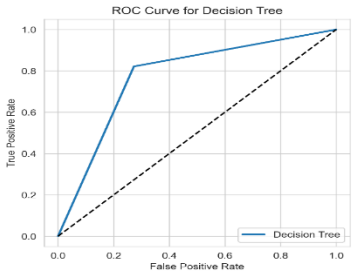


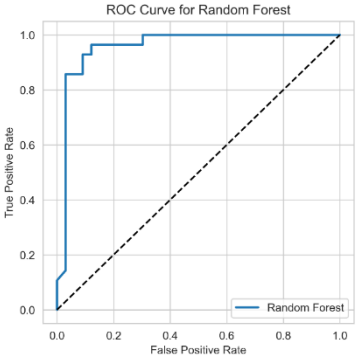


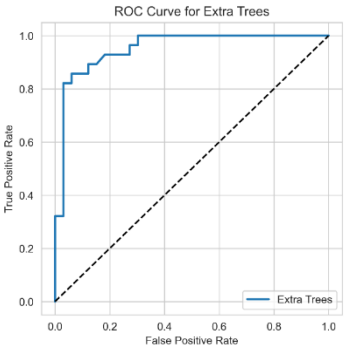
Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

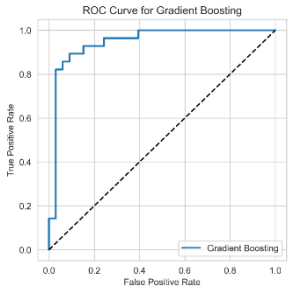
Research Article

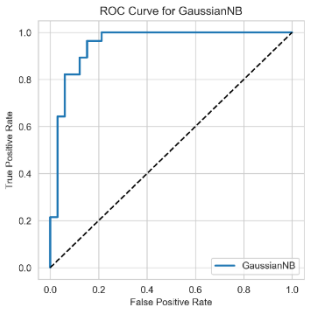


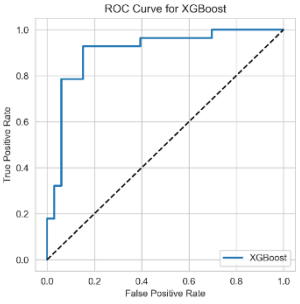


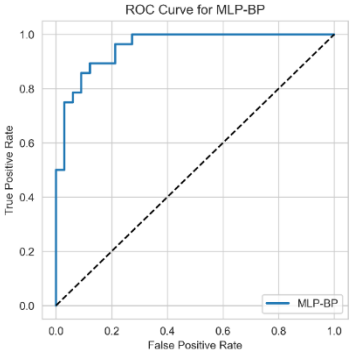






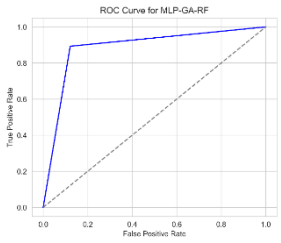






Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article



In this study, the ROC (Receiver Operating Characteristic) curves are shown to depict the level of classification that several key models have for the prediction of heart disease. Therefore, an important metric to evaluate the model's success is AUC, which is the under the Curve Area and has a larger AUC (closely to 1) is a better classification. The more reliable rules are the ones that give more discriminative power i.e the models with high AUC value. The insights in this help in choosing a model that will accurately make medical diagnosis.

7.CONCLUSION

Since more and more people die out from the case of heart disease, we can no longer fail to think of the designing of such a system which can anticipate the case of heart disease precisely and precisely. MLP is predicted to determine heart disease using the genetic algorithm to optimize the weights and bias. It is an optimization technique that has been commonly used in the literature, and it is used to achieve a more accurate prediction by finding the optimization neural network parameters. Results of the experiments indicated that the proposed algorithm MLP-GA-RF had the best performance in all and their accuracies are 95.08%. This model can aid the healthcare providers in making better diagnoses and prescribing much better treatments.

8.REFRENCES

[1] Estes, C. Anstee, Q.M. Arias-Loste, M.T. Bantel, H. Bellentani, S. Caballeria, J.Colombo, M. Craxi, A. Crespo, J. Day, C.P. et al. "Modeling NAFLD disease burden in China, France, Germany, Italy, Japan, Spain, United Kingdom, and United States for the period 2016–2030". J. Hepatol. 2018, 69, 896–904. [CrossRef] [PubMed]

[2] Dro˙ zd˙ z, K. Nabrdalik, K. Kwiendacz, H. Hendel, M. Olejarz, A. Tomasik, A. Bartman, W. Nalepa, J. Gumprecht, J. Lip, G.Y.H. "Risk factors for cardiovascular disease in patients with metabolic-associated fatty liver disease: A machine learning approach. Cardiovasc. Diabetol". 2022, 21, 240. [CrossRef] [PubMed]

[3] Murthy, H.S.N. Meenakshi, "Dimensionality reduction using neuro-genetic approach for early prediction of coronary heart disease. In Proceedings of the International Conference on Circuits, Communication, Control and Computing", Bangalore, India, 21–22 November 2014; pp. 329–332. [CrossRef]

[4] Benjamin, E.J. Muntner, P. Alonso, A. Bittencourt, M.S. Callaway, C.W. Carson, A.P. Chamberlain, A.M. Chang, A.R. Cheng, S. Das, S.R. et al. "Heart disease and stroke statistics—2019 update: A report from the American heart association". Circulation 2019, 139, e56–e528. [CrossRef] [PubMed]

[5] V. Shorewala, "Early detection of coronary heart disease using ensemble techniques". Inform. Med. Unlocked 2021, 26, 100655. [CrossRef]

[6] Mozaffarian, D. Benjamin, E.J. Go, A.S. Arnett, D.K. Blaha, M.J. Cushman, M. de Ferranti, S. Després, J.-P. Fullerton, H.J. Howard, V.J.; et al. "Heart disease and stroke statistics—2015 update: A report from the American Heart Association". Circulation 2015, 131, e29–e322. [CrossRef]

[7] Maiga, J. Hungilo, G.G. Pranowo. “Comparison of Machine Learning Models in Prediction of Cardiovascular Disease Using Health Record Data. In Proceedings of the 2019 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)”, Jakarta, Indonesia, 24–25 October 2019; pp. 45–48. [CrossRef]
[8] Doppala, B.P. Bhattacharyya, D. Chakkravarthy, M. Kim, T.H.” A hybrid machine learning approach to identify coronary diseases using feature selection mechanism on heart disease dataset”. Distrib. Parallel Databases 2021, 1–20. doi: 10.1007/s10619-021-07329-y. [CrossRef]
[9] Mallesh, N. Zhao, M. Meintker, L. Höllein, A. Elsner, F. Lüling, H. Haferlach, T. Kern, W. Westermann, J. Brossart, P. et al. “Knowledge transfer to enhance the performance of deep learning models for automated classification of B cell neoplasms”. Patterns 2021, 2, 100351. [CrossRef] [PubMed]
[10] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, “Fog computing and its role in the internet of things,” in Proc. 1st ed. MCC Workshop Mobile Cloud Comput., 2012, pp. 13–16
[11] Al Bataineh, A. Manacek, S. “MLP-PSO Hybrid Algorithm for Heart Disease Prediction”. J. Pers. Med. **2022**, 12, 1208 [MDPI]
[12] A. Lakshmi and R. Devi, "Heart Disease Prediction Using Enhanced Whale Optimization Algorithm Based Feature Selection With Machine Learning Techniques," 2023 12th International Conference on System Modeling & Advancement in Research Trends (SMART), Moradabad, India, 2023, pp. 644-648, doi: 10.1109/SMART59791.2023.10428617.
[13] V. M B, S. S. R, K. U and K. Y, "Exploratory Data Analysis of Heart Disease Prediction using Machine Learning Techniques-RS Algorithm," 2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI), Coimbatore, India, 2024, pp. 209-216, doi: 10.1109/ICoICI62503.2024.10696414.
[14] Bhatt, C.M. Patel, P. Ghetia, T. Mazzeo, P.L. “Effective Heart Disease Prediction Using Machine Learning Techniques”. Algorithms **2023**, 16, 88 [MDPI]
[15] S. U. Warsi, S. Mohsin, M. Asif, A. Hassan, R. Khan and T. Alyas, "A Hybrid Approach for Heart Disease Prediction using Genetic Algorithm and SVM," 2024 5th International Conference on Advancements in Computational Sciences (ICACS), Lahore, Pakistan, 2024, pp. 1-6, doi: 10.1109/ICACS60934.2024.10473308.
[16] Torthi, Ravichandra, Ajay Dilip Kumar Marapatla, Soumya Mande, Harish Kumar Varma Gadiraju, and Chalapathiraju Kanumuri. "Heart Disease Prediction Using Random Forest Based Hybrid Optimization Algorithms." International Journal of Intelligent Engineering & Systems 17, no. 2 (2024).
[17] V. Jain and M. Agrawal, "Heart Failure Prediction Using XGB Classifier, Logistic Regression and Support Vector Classifier," 2023 International Conference on Advancement in Computation & Computer Technologies (InCACCT), Gharuan, India, 2023, pp. 1-5, doi: 10.1109/InCACCT57535.2023.10141752.
[18] P. Behki and R. Pal, "Prediction of Heart Disease by Feature Selection Technique using Particle Swarm Optimization based on Logistic Regression," 2023 2nd International Conference on Futuristic Technologies (INCOFT), Belagavi, Karnataka, India, 2023, pp. 1-9, doi: 10.1109/INCOFT60753.2023.10425670.
[19] C. C, "Prediction of Heart Disease using Different KNN Classifier," 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2021, pp. 1186-1194, doi: 10.1109/ICICCS51141.2021.9432178.
[20] S. Naveen, S. K. Ravindran, S. G and S. N. Ameen, "Effective Heart disease prediction framework using Random Forest and Logistic regression," 2023 2nd International Conference on Vision Towards Emerging Trends in Communication and Networking Technologies (ViTECoN), Vellore, India, 2023, pp. 1-6, doi: 10.1109/ViTECoN58111.2023.10157078.
[21] S. Katari, T. Likith, M. P. S. Sree and V. Rachapudi, "Heart Disease Prediction using Hybrid ML Algorithms," 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2023, pp. 121-125, doi: 10.1109/ICSCDS56580.2023.10104609.