INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

Corresponding author.

glakshmi.gnc@gmail.com

Early Detection and Classification of Heart Diseases by Employing IFCMML and 2L-C Model with I-GA Machine Learning Methods

G Lakshmi[1]*, **P Sujatha**[2]

1 Research Scholar, Department of Information Technology, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai, Tamil Nadu, India
2 Associate Professor, Department of Information Technology, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai, Tamil Nadu, India

Abstract

Objectives: To identify a new method for early detection and classification of Heart Diseases (HD); to improve the accuracy of the results by employing I-FCMML (Improved Fuzzy C-Means Machine Learning) and 2L-C (Two-Level Classifier) models along with the I-GA (Improved Genetic Algorithm) methods. **Methods:** The I-FCMML algorithm is utilized for feature selection and extraction. Machine learning techniques such as Ensembled Random Forest method (ERFM) and Robust Gradient Boost Method (RGBT) are employed to predict the likelihood of HD and 2L-C, I-GA is used to classify and detect the HD at a premature stage based on features like age, gender, blood pressure, etc. I-FCMML with 2L-C & I-GA extracts all the features from the dataset (Cleveland HD dataset from the Cleveland Clinic Foundation, Ohio, USA) which includes 303 observations, 14 features and selects the suitable function to perform disease classification and detection with high accuracy. To evaluate the performance of the proposed method, MATLAB is used for implementation. The results are compared with existing algorithms such as 3P-ANN, ANN-FAHP, ADWFS, EDSS, and FE-PCA. **Findings:** Early HD detection and classification is achieved with 96.02% accuracy, 95.80% sensitivity, 94.76% specificity, 95% precision, 94% recall, 0.90 True Positive, 0.87 True Negative, and 94.13% F-Score to detect and classify the HD in a robust manner, which is comparatively high than the existing methods. **Novelty:** According to the findings of the comprehensive study, the proposed new method I-FCMML with 2L-C & I-GA has the potential to provide accurate and competent detection and classification of HD at an early stage, which could help for timely treatment and management of HD patients, and it also outperforms the existing methods such as 3P-ANN, ANN-FAHP, ADWFS, EDSS, and FE-PCA.

Keywords: Heart Disease Detection; Classification Algorithm; Genetic Algorithm; Fuzzy C-Means ML; Image Processing

1 Introduction

Heart disease is one of the most significant health problems, affecting millions of individuals worldwide. Chest pain, weakness, shortness of breath, fatigue, dizziness, and swelling are some of the common symptoms of the disease. Researchers are attempting to develop a decision-support system that can predict HD at an early stage, providing for an easy diagnosis. Nevertheless, existing diagnosis systems lack early prediction, classification, and effective feature selection, resulting in low accuracy and a lengthy execution time. The present study focuses on identifying HD at a premature stage and abnormal behaviors of the heart. The study compares the effectiveness of predictions made using I-FCMML genetic algorithms using ERFM and RGBT approaches to the results of existing methods. In today's modern digital era, if there is an effective DSS with medical experts, it is very simple to diagnose heart disease in an effective way. According to the ESC, 4.8 million people were diagnosed, and about 28 million were affected. Because heart disease cannot be accurately diagnosed using standard techniques with the help of the patient's medical history and physical report, Hence, a machine-learning-based DSS based on classifiers and FS approaches is being developed for early-stage heart disease prediction. All previous research works demonstrated that classification and prediction have less accuracy at the premature HD stage, and hence this study works on efficient ML-based genetic method to improve prediction and classification accuracy.

As a part of LR, many ML techniques were proposed for HD diagnosis. CNN based automatic detection on heart beats with and without noise to predict the disease. But no feature selection or extraction was performed. Naïve Bayes and SVM with XG boost method were used and blood flow is measured for the prediction. However, the coronary artery blockage and blood clot is not measured by the system. The accuracy is achieved up to 89% to identify the abnormal movement of the heart[1]. Cardiac classification is done using CWTM, where the ECG signals are converted to 2D and 3D signals. D-CNNM was proposed for dimensionality reduction in signals to classify the type of heart disease. Standard cardiac ARC was predicted by testing the samples in the timeframe window. Activation functions are multiplexed with max-pooling, and prediction is achieved with an F-Score of 87%[2]. Future trends of ML and DL applications related to heart sounds were analyzed using CNN and RNN. All 4 steps in heart sounds are tested, which include de-noising, segmentation, f-extraction, and f-classification. Wavelet mode, empirical mode, cestrum mode, etc. are used for the segmentation process. The analysis shows that ML and DL have a huge advantage in terms of predicting coronary artery disease[3]. The UCI dataset was used to predict the HD by utilizing ML and DL methods, and all the irrelevant data was handled using I-Forest and compared with the predicted and actual tested values. The proven results showed a predicted accuracy level of 89%. The system has minimal drawbacks where it won't work in terms of early prediction[4]. Hyper-tuning of ML classifiers was used along with K-Fold CV to improve the HD detection accuracy experimented in the Z-Alizadeh Sani dataset. RFT and BFT techniques are employed for absolute monitoring of heartbeats and failure prediction. To identify the optimal values, a 10-fold CV was used and attained 86% accuracy in prediction and classification[5]. C-Learning model for heart disease prediction using ML on a new RSR was developed by the authors, where it combines with PCA in terms of prediction accuracy. Lasso, Ridge, and Elastic Net RSR were used to excel modelling data into various learning levels to train and test. The proven results show that the weight matrix SD obtains 89% HD detection accuracy[6]. Logistic Regression with SVM was proposed for heart failure detection by testing the set of features using ML. Receiver optimistic edges and area under the edges for each classifier were performed in this method. KNN, SVM, DT, and NB were used in this system and compared with baseline versions; they attained the accuracy level of 89.9% with minimal drawbacks of early-stage prediction. All the feature sets are measured with Min-Max scalar sets[7]. The authors' objective was to predict the heart rate variability by employing an ML model and to assess the autonomic function with 15 parameters to classify the disease. ECG recordings are measured using a non-invasive method and tested with 300 instances using the I-Neldermead technique. The results are achieved with up to 90% sensitivity and specificity and a 90.8% accuracy rate. The drawbacks of the system are early prediction failure and speed[8]. An ANN-based DSS was introduced to identify the 3 main HDs such as MS, AS, and VSD. The model was trained using the supervised learning method, where the characteristics of patients were given as inputs. Patterns are used here, where the IP-ANN learns the pattern and generates the output. The shortcomings of this model are that it takes much execution time and preprocessing time.[9] Hybrid Linear analysis with modified ant-lion optimization model was designed to predict heart failure where 14 FH attributes were used. Analytic hierarchal model was used to record the variations in heart beats and also fuzzy hybrid method was used to detect the failure rate at middle stage which helps the clinician to reduce the risk of the patients. The results achieved up to 87% accuracy in failure prediction in all iterations[10]. Public UCID was tested by employing DCNN to bifurcate whether the tested instances belonged to infected or not-infected hosts. The sigmoid-M function was used to get the probability score, and the Nadam optimizer was employed to normalize the variables. Due to its effective functionality, the model succeeded up to 91%, where it fails to predict the HD at a premature stage[11]. The system was used to detect the abnormal actions of the heart and to extract the datasets for efficient classification using ML techniques. It was tested in an IoT environment, and the model proves a 78% accuracy rate because of the utilization of LSTM and RSTM methods[12]. A new method was proposed to test the cytokine serum levels in CAD at

a premature level by using genetic algorithms, where the results are achieved with minimal drawbacks. It was proved that the frequencies of risks are minimal after early diagnosis[13]. FL and Hybrid approaches for HD detection study were proposed to compare the performance of ML techniques[14]. A hybrid ML and DL system was proposed to detect heart disease by making use of KNN, where positive instances are measured at the initial stage and evaluation is done after pattern recognition. The method fails at early prediction and efficacy in the classification of HD in real time. The 3P-ANN classifier algorithm is used for disease spotting with minimal accuracy, but it has potential drawbacks like an increase in false positive rates, bias in the data, limited flow data, etc. The proven results show that the accuracy rate is not remarkable. ANN-FAHP model was trained to classify the disease based on the feature sets matching. Biotic and Abiotic search was performed during the prediction of HD and the accuracy results attained up to 86%[14–17]. The ADWFS method was proposed by the researchers by employing ADM and FST, where the imbalance dataset problem is solved, and the coronary syndrome is identified. The model has the disadvantage of not predicting HD at the initial stage. The system removes high effect highlights in new projection by FE-PCA. PCA extracts the projection vectors that contribute the largest covariance and uses these projection vectors to reduce feature dimensionality. IWD-ARP detection method was introduced to implement in networking and embedded environment where the failure prediction novelty is appreciable and the proven results shows 92% accuracy rate. SMOTE was applied for feature extraction and classification to reduce the risk rate and increase the efficacy of the trained model[18–21]. To perform periodic operations by identifying HD in the beginning stages, all of the preceding approaches must meet the following conditions. Some of the steps often used in existing approaches such as,

- **Data Preprocessing:** It incorporates cleaning and processing raw data into an analysis-ready format which includes data integration, reduction, augmentation, sampling and encoding.
- **Standardization of datasets :** Several ML approaches, including I-SVM, PCA, and M-LR, use standardization as a data preprocessing approach.
- **Hyper-Tuning:** The process of selecting the optimal set of hyper parameters for a machine learning (ML) algorithm. All key parameters are determined before the training process begins, including the learning rate, number of hidden layers, and regularization strength.
- **Employing ML approaches :** Choosing an appropriate ML algorithm, preparing the data, and training the algorithm to make predictions or classifications on new data are all part of the process.

The purpose of this research is to develop a novel method to (i) classify the disease; (ii) make early-stage HD predictions; (iii) maximize the accuracy level of prediction by employing the I-FCMML with 2L-C and I-GA techniques; and (iv) identify the abnormal behavior of the heart.

2 Methodology

The model is mainly focused on detecting HD at an early stage as well as identifying abnormal behaviors of the heart. I-FCMML offers a unique characteristic for predicting HD when combined with 2L-C and I-GA. Significant features are identified via improved recursive feature removal. When compared to baseline versions such as 3P-ANN, ANN-FAHP, ADWFS, EDSS, and FE-PCA, the model combines fuzzy clustering, machine learning techniques, and improved genetic algorithm optimization to provide accurate and competent detection, which could aid in the timely treatment and management of HD patients.

2.1 Proposed Methodology

The novel method aims to classify the type of heart disease, and then identify the abnormal behavior of the heart to find the optimal prediction and maximize the true positive rate of the model. ERFM and RGBT, as well as 2-LC and I-GA, are used to predict heart disease quickly. The risk levels of patients are clearly identified by using FCM-ML with the help of trained and tested sample attributes. The positive and negative rates are determined based on the cluster partition and set transformation. The abnormal behavior region is specifically recorded in order to distinguish between deterministic and nondeterministic values. All the abnormal activities are measured at frequent intervals, and the interval time is called latency time. The FCM-ML clustering algorithm was applied to identify the patterns in the data, and the number of clusters was determined using the ERFM method. FCM-ML in real time assigns each data point to a cluster based on the similarity between the data point and the cluster centroid. Learning the fuzzy patterns of HD helps detect or diagnose the symptoms at an early stage with high sensitivity and specificity. Assume that the number of clusters C, M denotes the fuzziness parameter. Initialize the membership F-Matrix U with random values. Initialize the F-Cluster centroids V with random values. Calculate the distance between DP and CC using

any one of the distance metric. Calculate new membership value for each F-Cluster by using the below equation,

$$NMP\ FCluster\ U_{ij} = \left(\frac{d_{ij}}{sum\_k(d_{ik})} \right)^{\hat{}} (2/(M-1)) \tag{1}$$

where, where d_ij is the distance between data point i and cluster centroid j and $sum\_k(d_{ik})$ is the sum of distances between data point i and all the cluster centroids. As a second step update the F-cluster centroids V. Calculate the new F-cluster centroid by using the below equation

$$NMP\ FCentroid\ V_{ij} = \left(\frac{sum_i(U_{ij}\hat{}\ M * x_i)}{sum_i(U_{ij}\ \hat{}\ M)} \right)^{\hat{}} (2/(U-1)) \tag{2}$$

where, x_i is the data point i, U_{ij} is the membership value of data point i to cluster j and M is the fuzziness parameter of the ML method which captures the pattern. During the initial point the HD prediction is carried out using FCM as a test method and IFCMML is employed with the same attributes to identify the disease at early occurrence.

2.2 Data attainment, acquisition and pre-processing

The study uses the raw heart disease dataset from Cleveland Clinic Foundation, USA. The datas are completely extracted for pre-processing and quality representation. Pre processing techniques such as SS and MMS have been applied to the CHD-dataset. The limitations of using this dataset are computational time complex where it takes max time for training and testing.

Table 1. Cleveland Heart Disease available dataset

S.No	F-Name	F-Code	Description
1.	Age	Age	Years
2.	Gander/Sex	SEX	(1 = male; 0 = female)
3.	CP-Chest Pain	CP	0: typical angina 1: atypical angina, 2: non-anginal pain, Value 3: asymptomatic
4.	RBP-Resting Blood Pressure	RBP	In mm Hg on admission to the hospital
5.	SC-Serum Cholesterol	SC	In mg/dl
6.	FBS-Fasting Blood Sugar	FBS	Fasting blood sugar > 120 mg/dl) (1 = true; 0 = false
7.	RE-Resting Electrocardiographic	RE	0: normal, 1: having ST-T wave abnormality, 2: showing probable
8.	Heart Rate Max/Min	HRM	Thalach-Max
9.	Exercise	EX	1 = yes; 0 = no
10.	ST Depression Levels	STD	ST depression induced by TL
11.	Exercise ST Segment Slope	STS	0: upsloping, 1: flat, 2: downsloping
12.	No. of M-vessels	MV	No. of MV (0-3) colored by fluoroscopy
13.	Thallium Scan	THA	0 = normal; 1 = fixed defect; 2 = reversible defect
14.	Label	LB	0 = no disease, 1 = disease

The distribution of the data is somewhat consistent with the training data and tested data that was gathered. Here, the I-FCM method is used to evaluate the exposed data to a given degree which includes,

- Collection of raw CHD-dataset & Preprocessing the datasets to remove the incoherent datas
- Data segmentation process is done to improve the pre-processing accuracy
- The preprocessed data is then fed into the proposed algorithm, which identifies patterns in the data by grouping similar data points into clusters. IFCMML assigns each data point a membership value that indicates how strongly it belongs to each cluster.
- Cluster analysis is performed
- Classification of HD is obtained
- The model will be trained for Prediction accuracy at early stage

2.3 Improvised Fuzzy C-Means Machine Learning Method

To calculate the accuracy and stability of the clustering, the improved FCM-ML technique employs ensemble methods like bagging and boosting. By minimizing the dimensionality of data, Improved Recursive Feature Elimination (RFE) is applied to recognize significant features for HD prediction. The IFCM-ML model automatically calculates the time for the evaluation of classifiers. Apply the IFCM-ML prediction model to cluster the data into groups according to the similarity between the variables. The method computes adaptive weights for each feature depending on its significance in predicting heart disease. Each data point is assigned to one or more clusters with varying degrees of membership during the clustering process. Each cluster reflects a different type of heart disease pattern. Let us assume that $O(I\ and\ J)$ as dataset matrix where I is the instances and J is the output values. The Equation for the same is written as,

$$Dmatrix\ = O(I, J) = \left\{ (I_x, J_y) \mid I_x \text{€} R_n, J_y \text{€} \{0, 1\} \right\} K^{i=1} \tag{3}$$

where, I_x can be written as, $I_x = \{I_1,\ I_2,\ I_3,\ I_n\ \}$. The pre-processing and segmentation techniques are applied and measure the DP and CP value. The feature selection is done in a different way by selecting based on the relevant outputs instead of selecting based on given input values. To maximize the efficiency in FS, it is important to reduce the number of irrelevant DPs and CPs. The steps involved in IFCMML for FS is described as follows,

- Pre process the available dataset and initialize the selected features based on RO
- Compute the CPs and DPs
- Derive the feature subset and Check the redundancy and F-weights
- Randomly choose the given input for FS
- Identify the pattern based on values of output & Select the best optimum level for FSE

The determined values of instances are derived by the following equation to select the subset of features.

$$Subset = J\ (U, V)\ pop^{i+1} = Frequency\left(U^{features\ st}, \eta\right)\ - \left(q^{i+1}.V^{finite+h}\right) \tag{4}$$

where, $(U, V)\ pop^{i+1}$ denotes finite number of instances are selected and pushed up in frequent intervals and the optimum level is reached to classify the HD rate / ECG beats for the next process. The enhanced method is suggested in this research work to pick the most feasible and compact function subset.

I-FCMML Algorithm
1. Input: Initialize the datasets Set the number of clusters as c, Set the fuzziness parameter as m
Initialize the membership matrix as U, Initialize the cluster centroids as V
2. Calculate the DP and CC distance
3. Update U and V
4. Identify the characteristics of each cluster
5. **function I-FCM (X, c, m, max_iter) :**
6. U = initialize_membership_matrix*(X, c)*
7. V = initialize_cluster_centroids*(X, c)*
8. for i in range*(max_iter)*
9. U = update_membership_*matrix(X, V, m)*
10. V = update_cluster_centroids*(X, U, m)*
11. return V
12. **function HDP (X) :**
13. X = preprocess_data*(X)*
14. V = fuzzy_c_means*(X, c, m, max_iter)*
15. ERFM *(X, c, m, max_iter)*
16. RGBT *(X, c, m, max_iter)*
17. cluster_characteristics = identify_cluster_characteristics*(V)*
18. identify_RGBT_ERFM classification
19. model =train_heart_disease_prediction_model(cluster_characteristics)
20. prediction = predict_heart_disease*(model, X_new)*
21. return prediction
22. Repeat the steps to find the maximum accuracy level
23. End

2.4 2-LC and I-GA Learning for Classification and Prediction

The finest classification is done using 2-LC technique to experiment the classification and prediction tasks in terms of separating linear and non-linear, data groups and training and testing datasets. Let's take the training set and testing set as input (X_{Train}, Y_{Train}) and (X_{Test}). The predicted class labels are $(Y_{predict})$. Initialize and train the linear classifier with the available parameters and train the available training set. Then compute the distance from each set to the decision boundary of the 2-LC classifier. Split the (X) into two different groups based on the calculated distance which is available and bifurcate into positive and negative distances. In addition to that train the 2-NLCto obtains the classification accuracy. Compute the predicted class label for each dataset (X_{Test}). If the DP has positive distance use 2-LVC and if it's negative use 2-NLC. Finally save all the predictions as $(Y_{predict})$. Various iterations can be done for the optimal solution with the help of below equation,

$$Optimal\ 2-LC\ Solution\ (Y_{predict}) = (X_{Train}, Y_{Train}) | (X_{Test})\ (It_{1-n}) \tag{5}$$

where, (It_{1-n}) various iteration levels to perform the 2-LC and 2-NLC operations.

2.5 Prediction of HD at premature stage

Following the HD classification and FSE, the premature stage prediction model is finalized with the support of the I-GA model. In this paper, I-GA is an enhanced genetic algorithm that may be utilized for 2-LC model optimization. It use a combination of GA and heuristic search approaches to determine the best settings for optimal performance. The steps taken by the I-GA are as follows.

- Initialize a population of solutions with random feature sets and model parameters.
- Evaluate the fitness of each solution using a performance metric.
- Select the best-performing solutions for reproduction and create offspring solutions using genetic operators such as mutation and crossover.
- Evaluate the fitness of the offspring solutions and select the best-performing solutions for the next generation.
- Repeat steps 3-4 until a stopping criterion is met, such as a maximum number of generations or convergence of the solution set.

2-LC or 2-NLC and I-GA used for classification and prediction tasks. 2-LC combines a linear and nonlinear classifier to improve classification accuracy, while I-GA uses a genetic algorithm to optimize feature selection and model parameters for improved performance is shown in eq.6.

$$I-GA\ FV = (X_{Train}, Y_{Train}) | (X_{Test})\ (It_{1-n})\ *\sum_{i=0}^{n} \left(\frac{i}{j}\right) x^i \tag{6}$$

where, the term FV denotes the fitness value and with the help of available data (X_{Train}, Y_{Train}) the most prediction state is obtained.

2.6 Classification Accuracy and FS effectiveness

It is important to carefully select and preprocess the features. The 2-LC and 2-NLC model architectures are used for classification purposes. The model assessed performance across multiple testing datasets to ensure that the best value was obtained. The four steps carried out for high accuracy are: splitting, training, predicting, and evaluating based on the iteration levels with the available datasets. Some of the limitations of 2-LC and 2-NLC are that they learn only linear boundaries, which may not be sufficient for a large number of instances. But the excessive class datasets are correctly predicted due to this distinctive feature. All two-fold classification problems can be solved with the help of the 2-LC and 2-NLC models. In I-FCMML with 2-LC and I-GA, the accuracy is up to 96%.

2.7 About MATLAB R2020a Implementation Process

The proposed research against existing HD detection methods is evaluated using the MATLAB tool. It gives users access to a wide range of machine learning and nature inspired methods, which include supervised and un-supervised learning (DL and RL), as well as tools for feature selection, extraction, and ML model optimization. Because of its ease of use, MATLAB is used by many IT and engineering applications, hospitals, and researchers. It allows users to interactively explore data, generate visualizations, and compare various machine learning algorithms and parameters.

2.8 Performance Analysis Metrics

The following are the metrics used to compare the performance of the proposed I-FCMML with 2L-C & I-GA algorithm to the existing techniques 3P-ANN[16], ANN-FAHP[17], ADWFS[18], EDSS and FE-PCA[19] which were selected as the baseline schemes in the preceding section. The following are the PE formula to evaluate the performance.

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \times 100 \tag{7}$$

$$Sensitivity = \frac{TP}{(TP+FN)} \times 100 \tag{8}$$

$$Specificity = \frac{TN}{(TN+FP)} \times 100 \tag{9}$$

$$Precision = \frac{TP}{(TP+FP)} \times 100 \tag{10}$$

$$CC = \frac{T_1}{\sqrt{T_2 \times T_3 \times T_{4 \times T_5}}} \times 100 \tag{11}$$

where, CC equation is derived using, $T_1 = (TP \times TN - FP \times FN)$, $T_2 = (TP+FP)$, $T_3 = (TP+FN)$, $T_4 = (TN+FP)$, and $T_5 = (TN+FN)$.

- **Sensitivity and Specificity** - Statistical method to evaluate the proposed prediction model, where sensitivity measures TP and specificity measures TN.
- **Accuracyand F score** - A&F determines the efficacy of a classification model that detects both positive and negative classes, whereas F-score is an accuracy of model that considers both P & R.
- **True Positive Rate and True Negative** - TPR shows the percentage of API detected as positive by a prediction model, whereas TN shows the percentage of ANI identified as negative by a Prediction Model.
- **False Positive Rate and False Negative** - Inaccuracy Classification in the Prediction Model
- **Precision and Recall** - Recall estimates the amount of TP among all real positive cases, whereas precision evaluates the amount of TP among all positive predictions.

3 Results and Discussion

The results of the novel Machine Learning approach I-FCMML with 2L-C & I-GA against the 3P-ANN[16], ANN-FAHP[17], ADWFS[18], EDSS and FE-PCA[19] were shown in this section. I-FCMML with 2L-C & I-GA outperforms the conventional techniques in terms of early HD detection, feature extraction efficiency, and prediction accuracy rating. Figures 1 to 8 specify the X axis plotted with competence metrics and the Y axis plotted with absolute values and proportion of performances. The results show the clarity of pattern and identify the best solution out of given values.

3.1 Sensitivity and Specificity Performance Analysis

The performance analysis of I-FCMML with 2L-C & I-GA against 3P-ANN[16], ANN-FAHP[17], ADWFS[18], EDSS and FE-PCA[19] is evaluated and presented in Figure 1 . It is noted that the proposed ML technique produced remarkable performance with enhanced results. In ANN and ADWFS, the roulette wheel selection method is used, resulting in lower performance than ML approach. The proposed improvised Fuzzy C-Means technique employing the I-FCMML with 2L-C & I-GA algorithm performs well up to 95.80% in terms of forecasting the HD at an early stage due to significant classification and extraction.

Table 2. Performance Values of Sensitivity and Specificity

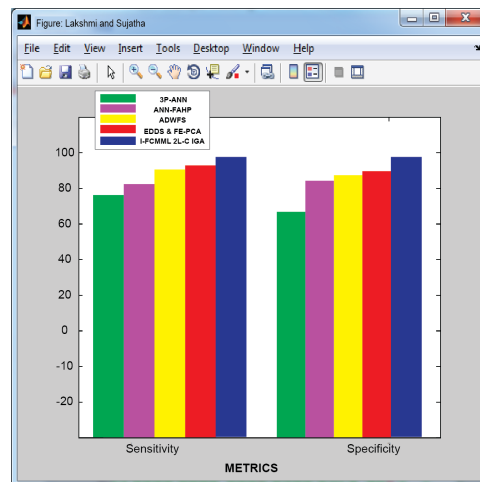| Metrics/ Schemes	3P-ANN[16]	ANN-FAHP[17]	ADWFS[18]	EDSS and FE-PCA[19]	I-FCMML with 2L-C & I-GA
Sensitivity	79.67	81.09	87.01	90.15	95.80
Specificity	61.07	80.76	83.04	86.47	94.76

Fig 1. Sensitivity and Specificity Performance Analysis

3.2 Accuracy and F-Score Performance Analysis

Figure 2 illustrates the accuracy level and F-Score performance evaluation of the proposed approach I-FCMML with 2L-C & I-GA against 3P-ANN [16], ANN-FAHP [17], ADWFS [18], EDSS and FE-PCA [19]. The I-FCMML with 2L-C & I-GA works well with the efficient use of genetic algorithms and ML concepts for the extraction and classification of 14 features. Out of the instances tested with the new method, a high number of predicted instances were found. According to certain fuzzy partitions, the data point is accurately measured. I-FCMML with 2L-C & I-GA outperformed and resulted in 96% accuracy and a 94% F-Score.

Table 3. Performance Values of Accuracy and F-Score Analysis

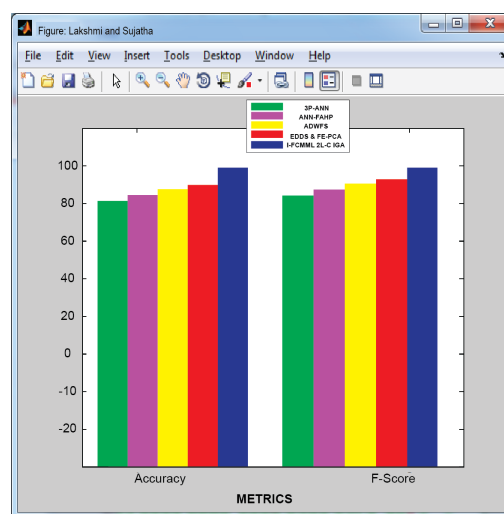| Metrics/ Schemes	3P-ANN [16]	ANN-FAHP [17]	ADWFS [18]	EDSS and FE-PCA [19]	I-FCMML with 2L-C & I-GA
Accuracy	80.06	83.04	82.08	91.45	96.02
F-Score	81.98	82.16	83.09	90.59	94.13

Fig 2. Performance of Accuracy and F-Score

3.3 True Positive and True Negative Rate Performance Analysis

Figure 3 reveals the performance analysis of the true positive and true negative of the proposed I-FCMML with 2L-C and I-GA against 3P-ANN [16], ANN-FAHP [17], ADWFS [18], EDSS and FE-PCA [19]. The I-FCMML with 2L-C & I-GA performs better than the existing algorithms on the selected datasets with 14 features and 303 multiples of instances. Due to its perfect hit ratio and unique extraction and classification, the proposed ML technique has a high accuracy in the early prediction of heart disease. The confusion matrix is used to calculate TPR and TNR based on the dataset values.

Table 4. Performance Values of TP and TN

Metrics/ Schemes	3P-ANN [16]	ANN-FAHP [17]	ADWFS [18]	EDSS and FE-PCA [19]	I-FCMML with 2L-C & I-GA
True Positive	154 (0.76)	166 (0.82)	175(0.84)	185 (0.86)	194 (0.90)
True Negative	125 (0.74)	134 (0.79)	115 (0.81)	109 (0.83)	95 (0.87)

3.4 False Positive and False Negative Rate Performance Analysis

Figure 4 portrays the detailed analysis of False Positive and False Negative of the proposed ML approach I-FCMML with 2L-C and I-GA against 3P-ANN [16], ANN-FAHP [17], ADWFS [18], EDSS and FE-PCA [19]. On the chosen datasets with features and instances, the ML technique outperformed the existing methods due to its high accuracy rate and measured the proportion of negative cases that are incorrectly identified as positive and vice versa. The I-FCMML with 2L-C and I-GA performs well in prediction and classification of heart disease due to RGBT, and the results are given below.

Table 5. Performance Values of FP and FN

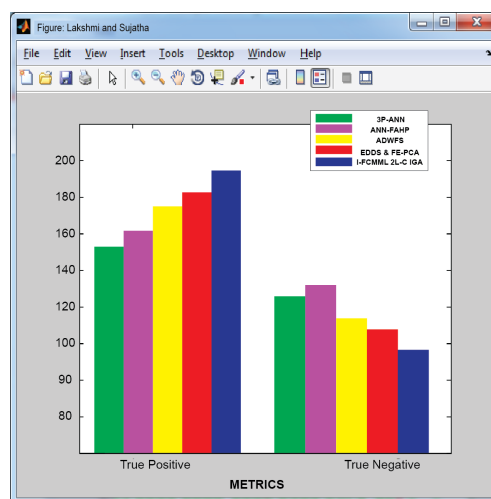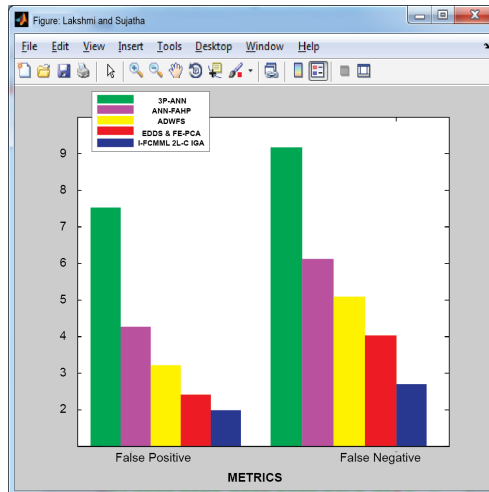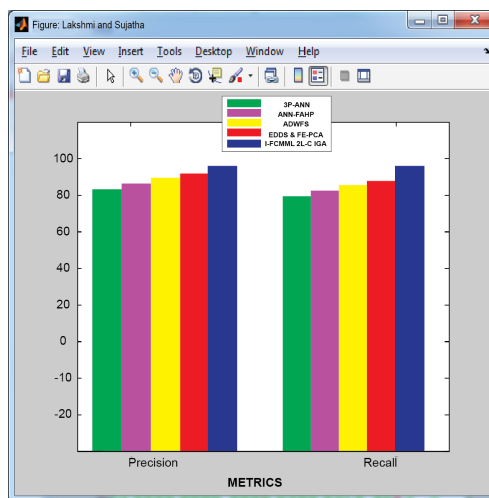| Metrics/ Schemes	3P-ANN [16]	ANN-FAHP [17]	ADWFS [18]	EDSS and FE-PCA [19]	I-FCMML with 2L-C & I-G
False Positive	7.5 (0.75)	4.5 (0.45)	3.5 (0.35)	2.6 (0.26)	2 (0.20)
False Negative	9 (0.90)	6.3 (0.63)	5.4 (0.54)	4.2 (0.42)	1.7 (0.17)

Fig 3. Performance of TP and TN

3.5 Precision and Recall Performance Analysis

Figure 5 shows the performance of precision and recall of the proposed I-FCMML with 2L-C and I-GA against 3P-ANN [16], ANN-FAHP [17], ADWFS [18], EDSS and FE-PCA [19]. The new method achieves more predictable results compared to the existing methods in terms of image scanning and deep pointing. Because of the image segmentation and assigned cluster centroids, the I-FCMML with 2L-C and I-GA performs well where other schemes fail to perform in a robust manner.

Table 6. Performance Values of Precision and Recall

Metrics/ Schemes	3P-ANN [16]	ANN-FAHP [17]	ADWFS [18]	EDSS and FE-PCA [19]	I-FCMML with 2L-C & I-GA
Precision	81	84	86	91	95
Recall	78	80	84	89	94

Fig 4. Performance of FP and FN Performance

Fig 5. Performance of Precision and Recall

4 Conclusion

This research work has presented a novel Prediction and Classification ML approach called I-FCMML with 2L-C and I-GA to increase the system's accuracy level and to identify Heart Disease at an early stage. The Cleveland HD dataset from the Cleveland Clinic Foundation in Ohio, USA, has been used in the proposed technique, and 303 cases with 14 features are used to select the set of feature values using an evolutionary genetic algorithm (GA). The trained and test values are compared to the available datasets to make a prediction. Several iterations are carried out which contains patient information, such as medical history, lab results, and imaging data, to predict the likelihood of a patient having a specific type of heart disease. The results show that the new ML technique prediction is more accurate and quicker than the other prediction model with 96.02% accuracy, 95.80% sensitivity, 94.76% specificity, 95% precision, 94% recall, 0.90 True Positive, 0.87 True Negative, and 94.13% F-Score,

and that it successfully classified the HD. For the cardiologist, it is also a time-saving and user-friendly application. It is noted that ML decision support system is more suitable for HD prediction. The proposed approach is limited in that it is based only on a predefined set of features, and the prediction and accuracy levels may vary from case to case. In the future, the algorithm can be enhanced to predict Cardiomyopathy and heart failure accuracy rates even better and making the system more flexible based on the patients stability.

References

1) Nagavelli U, Samanta D, Chakraborty P. Machine Learning Technology-Based Heart Disease Detection Models. *Journal of Healthcare Engineering*. 2022;2022:1–9. Available from: https://doi.org/10.1155/2022/7351061.

2) Jamil S, Rahman M. A Novel Deep-Learning-Based Framework for the Classification of Cardiac Arrhythmia. *Journal of Imaging*. 2022;8(3):70. Available from: https://doi.org/10.3390/jimaging8030070.

3) Chen W, Sun, Qiang, Chen, Xiaomin, Xie, et al. Deep Learning Methods for Heart Sounds Classification: A Systematic Review. *Entropy*. 2021;23:667–667. Available from: https://doi.org/10.3390/e23060667.

4) Bharti R, Khamparia A, Shabaz M, Dhiman G, Pande S, Singh P. Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning. *Computational Intelligence and Neuroscience*. 2021;2021:1–11. Available from: https://doi.org/10.1155/2021/8387680.

5) Saboor A, Usman M, Ali S, Samad A, Abrar MF, Ullah N. A Method for Improving Prediction of Human Heart Disease Using Machine Learning Algorithms. *Mobile Information Systems*. 2022;2022:1–9. Available from: https://doi.org/10.1155/2022/1410169.

6) Albahr A, Albahar M, Thanoon M, Binsawad M. Computational Learning Model for Prediction of Heart Disease Using Machine Learning Based on a New Regularizer. *Computational Intelligence and Neuroscience*. 2021;2021:1–10. Available from: https://doi.org/10.1155/2021/8628335.

7) Yilmaz R, Yağin FH. Early Detection of Coronary Heart Disease Based on Machine Learning Methods. *Medical Records*. 2022;2022(1):1–6. Available from: https://doi.org/10.37990/medr.1011924.

8) González-Corrales R, García-García C, Casal-Moro R, Berlanga-Llavori R, Martínez-Álvarez RP, García-Sáez G. A machine learning-based approach to predict cardiovascular disease risk using bioelectrical impedance analysis and clinical information. *Journal of Healthcare Engineering*. 2021. Available from: https://doi.org/10.1155/2021/6623843.

9) Ghwanmeh S, Mohammad A, Al-Ibrahim A. Innovative Artificial Neural Networks-Based Decision Support System for Heart Diseases Diagnosis. *Journal of Intelligent Learning Systems and Applications*. 2013;05(03):176–183. Available from: https://doi.org/10.4236/jilsa.2013.53019.

10) Manimurugan S, Almutairi S, Aborokbah MM, Narmatha C, Ganesan S, Chilamkurti N, et al. Two-Stage Classification Model for the Prediction of Heart Disease Using IoMT and Artificial Intelligence. *Sensors*. 2022;22(2):476. Available from: https://doi.org/10.3390/s22020476.

11) Arooj S, Rehman SU, Imran A, Almuhaimeed A, Alzahrani AK, Alzahrani AK. A Deep Convolutional Neural Network for the Early Detection of Heart Disease. *Biomedicines*. 2022;10(11):2796. Available from: https://doi.org/10.3390/biomedicines10112796.

12) Al-Shabi M, Abuhamdah A. Using deep learning to detecting abnormal behavior in internet of things. *International Journal of Electrical and Computer Engineering (IJECE)*. 2022;12(2):2108. Available from: http://doi.org/10.11591/ijece.v12i2.pp2108-2120.

13) Wafa M, Ansari SE, Humphries AK, Naveed OJ, Khan DA, Khan E, et al. Effect of Coronary Artery Disease risk SNPs on serum cytokine levels and cytokine imbalance in Premature Coronary Artery Disease. *Cytokine*. 2019;122. Available from: https://doi.org/10.1016/j.cyto.2017.05.013.

14) Kaur J, Khehra BS. Fuzzy Logic and Hybrid based Approaches for the Risk of Heart Disease Detection: State-of-the-Art Review. *Journal of The Institution of Engineers (India): Series B*. 2022;103(2):681–697. Available from: https://doi.org/10.1007/s40031-021-00644-z.

15) Mohan S, Thirumalai C, Srivastava G. Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques. *IEEE Access*. 2019;7:81542–81554. Available from: https://doi.org/10.1109/ACCESS.2019.2923707.

16) Sharma N, Singla A. Early detection of heart disease using 3P ANN classifier. *IEEE Computing Methodologies and Communication*. 2020;4(6):210–215. Available from: https://doi.org/10.1109/ICCMC51329.2020.9190445.

17) Li JP, Haq AU, Din SU, Khan J, Khan A, Saboor A. Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare. *IEEE Access*. 2020;8(1):107562–107582. Available from: https://doi.org/10.1109/ACCESS.2020.3001149.

18) Kumar PS, Kumar N, Khanna N. Detection of Coronary Heart Disease Using Advanced Data Mining and Feature Selection Techniques. *International Journal of Innovative Technology and Exploring Engineering*. 2019;8(5S2):636–640. Available from: https://doi.org/10.35940/ijitee.K8733.0985S219.

19) Saadon HD, Bakar AM, Sulaiman MY. Early Detection of Heart Diseases Using Feature Extraction and Principal Component Analysis. *Journal of Healthcare Engineering*. 2021;p. 1–8. Available from: https://doi.org/10.1155/2021/6615692.

20) Nithyanandh S, Jaiganesh V. Quality of service enabled intelligent water drop algorithm based routing protocol for dynamic link failure detection in wireless sensor network. *Indian Journal of Science and Technology*. 2020;13(16):1641–1647. Available from: https://doi.org/10.17485/IJST/v13i16.19.

21) Bhatt CM, Patel P, Ghetia T, Mazzeo PL. Effective Heart Disease Prediction Using Machine Learning Techniques. *Algorithms*. 2023;16(2):88. Available from: https://doi.org/10.3390/a16020088.