

# Thrice Filtered Information Energy Based Particle Swarm Feature Selection (TFIE-PSFS) Method Based Artificial Neural Network Classification for Improving Heart Disease Diagnosis

P. Umasankar<sup>1</sup> and V. Thiagarasu<sup>2</sup>

<sup>1</sup>Research Scholar, Department of Computer Science, Manonmaniam Sundaranar University, Tamil Nadu

<sup>2</sup>Associate Professor, Department of Computer Science, Gobi Arts and Science College, Erode, Tamil Nadu

E-Mail: pus.msu2013@gmail.com

(Received 3 November 2018; Revised 26 November 2018; Accepted 23 December 2018; Available online 2 January 2019)

**Abstract** - Diagnosing the existence of heart disease is really tedious process, as it entails deep knowledge and opulent experience. As a whole, the forecast of heart disease lies upon the conventional method of analysing medical report such as ECG (The Electrocardiogram), MRI (Magnetic Resonance Imaging), Blood Pressure, Stress tests by a Medicinal expert. Nowadays, a large volume of medical statistics is obtainable in medical industry and turns as a excessive source of forecasting valuable and concealed facts in almost all medical complications. Thus, these facts would really aid the doctors to create exact predictions. The innovative methods of Artificial Neural Network models have also been contributing themselves in yielding the main prediction accuracy over medical statistics. This paper targets to predict the presence of heart disease utilizing Back Propagation MLP (Multilayer Perceptron) of Artificial Neural Network. The proposed ANN design targeted to generate the three outputs Yes (Patient having heart disease), No (Patient not having heart disease), and Hesitant (Patient those who are in between yes and no category).

**Keywords:** Artificial Neural Network, Multi-Layer Perceptron, Heart Disease

## I. INTRODUCTION

Heart is an essential organ of all living creature, which assumes a key role of propelling blood to the rest of the organs through the blood vessels of the circulatory system. Every practical issue in the heart has a straight effect on the existence of apprehensive human being; meanwhile it disturbs the further parts of the body such as lungs, brain, liver, kidney etc. Heart Diseases pronounce a range of conditions that influence the heart and remain as a major cause of death all over the globe. The clinical indications of the Heart Disease obfuscate the prediction, as it is inclined by numerous components like functional and pathologic appearance. Subsequently, this could postpone the prognosis of the illness. Hence, there is a requirement for the creation of innovative models to progress the prediction accuracy in a limited time. Disease prognosis through various elements or indications is a multi-layer problem, even that could prompt a false presumption. Like so, an endeavour is made to connect the learning and experiences of the experts and to fabricate a framework that fairly aids the diagnosing process.

Data Mining is the study of substantial datasets to extricate the hidden and formerly unidentified patterns, relationships and knowledge that are hard to investigate with conventional measurements [1]. Data mining methods are the consequence of a long method of research and product improvement. Data Mining is alienated into two assignments such as Predictive Tasks and Descriptive Tasks. Predictive Tasks forecast the estimation of an explicit attribute based on other attribute. Categorization, Regression and Deviation Deduction come under Predictive Tasks. Descriptive Tasks design that outline the connection between the data. Clustering, Association Rule Mining and Sequential Pattern Discovery are future under Descriptive Tasks. Data Mining comprises of few steps from raw data collection to some form of new knowledge. The iterative process comprises of following stages like Data Integration, Data cleaning, Data transformation, Data Selection, Data Mining, Knowledge Representation, and Pattern Evaluation.

Medical Data Mining is a field of experiment which encompasses lot of inaccuracy and vagueness. Establishment of eminence amenities at reasonable cost is the main task confronted in the health care organization. Poor clinical decision may prompt to appalling results. Health care data is huge. Clinical decisions are frequently prepared made dependent on doctor's experience instead of on the knowledge rich data covered up in the data base. In some cases now and again will result in errors, disproportionate medical cost which disturbs the excellence of service to the patients [2]. Medical history data involves various tests to detect the specific disease. It is conceivable to advance the benefit of Data mining in health care by retaining it as an intellectual analytical tool. The researchers in the field of medicine recognize and forecast the disease with the assistance of Data mining techniques [3].

## II. RELATED WORKS

Nahar, Jesmin, *et al.*, [1], examines the sick and healthy aspects which contribute to heart disease for males and females. Association rule mining, a computational knowledge method, is employed to detect these components and the UCI Cleveland dataset, an organic database, is

considered alongside the three rule generation algorithms – Predictive Apriori, Apriori and Tertius.

Vijayarani, S., and S. Sudha [2] this paper investigates the classification tree methods in data mining. The objective of this paper is to explore the investigational results of the performance of various categorization techniques for a heart disease dataset. The classification tree algorithms utilized and experimented in this work are Random Forest, Decision Stump, and LMT Tree algorithm.

Gayathri, P., and N. Jaisankar [3] the review of the research associated to heart disease and furthermore the overview of numerous classifications of heart disease such as coronary artery disease, coronary heart disease, ischemic heart disease, heart failure, congenital heart disease, cardiovascular disease, hypoplastic left heart syndrome and valvular heart disease are presented in the research

Shouman, Mai, Tim Turner, and Rob Stocker [4] the analysis obtainable here is part of work to create devices to help the healthcare practitioners to detect heart disease prior in the desire for prior mediations in this preventable killer. The general precision of common data mining methods in heart disease findings is hard to measure from the literature. This examination explores Naïve Bayes, Decision Tree, and K-nearest Neighbour performance in the analysis of heart disease patients.

Amato, Filippo, *et al.*, [5] this paper ANNs characterize a dominant device to aid physicians to perform analysis and different implementations. In this connection, ANNs have numerous benefits including: (i) Reduced likelihood of overlooking relevant Information (ii) The ability to process large amount of data (iii) Reduction of diagnosis time.

Persi Pamela, I., and P. Gayathri [6] A fuzzy structure is one of the delicate calculating approach is proposed in this paper along with a data mining technique for competent analysis of coronary heart disease. However the database has 76 attributes, only 14 attributes are initiated to be competent for CHD diagnosis as per all the distributed tests and doctors' view. Therefore, the vital attributes are taken from the heart disease database. From these attributes hard guidelines are found by retaining CART decision tree algorithm, which are then realistic to the fuzzy system. A Particle Swarm Optimization (PSO) technique is smeared for the optimization of the fuzzy membership functions where the constraints of the membership tasks are transformed to novel positions.

Chaurasia, Vikas, and Saurabh Pal [7] This research paper proposes to deliver a survey of existing methods of knowledge discovery in databases using data mining techniques which will be valuable for medical practitioners to yield successful decision. The major goal of this work is to forecast more precisely the existence of heart disease with decreased number of attributes.

Thenmozhi, K., and P. Deepika [8] the authors discovered the several decision tree algorithms like J48, ID3, C4.5 and C5.0 and in the classification and prediction of heart disease.

Kim, Jae-Kwon, *et al.*, [9] This paper proposes the Fuzzy Rule-based Adaptive Coronary Heart Disease Prediction Support Model (FbACHD\_PSM), which stretches comfortable reference to coronary heart disease patients. The projected model uses a mining method approved by medical experts to deliver approvals.

Seera, Manjeevan, and Chee Peng Lim [10] in this paper, a hybrid intelligent system that comprises of the Fuzzy Min-Max neural network, the Organization and Random Forest model, and Regression Tree is proposed, and its value as a decision support device for medical data classification is observed. The hybrid intelligent system targets to abuse the benefits of the constituent models and, at the same time, ease their impediments.

Bashir, Saba, Usman Qamar, and M. Younus Javed [11] The goal of the proposed research is to envisage the heart disease in a patient more precisely. The proposed structure customs common vote based novel classifier collaborated to amalgamate different data mining classifiers. UCI heart disease dataset is practiced for assessment and results.

Shabana, ASMI P., and S. Justin Samuel [12] Different data mining techniques such as Decision Tree, Naive Bayes, Association Rule and Linear Regression are practiced to envisage the heart disease. Data mining techniques in overall diagnosis realistic over all disease treatment dataset explore if hybrid data mining techniques can attain comparable (or better) results in classifying appropriate actions as that attained in the diagnosis. In this paper, the proposed work is to more precisely predict the existence of heart disease with new attributes of the disease and using association rules.

Aljaaf, A. J., *et al.*, [13] In this study, a multi-level risk assessment of developing heart failure has been proposed, in which a five levels of risks in heart failure can be predicted using C4.5 decision tree classifier. In contrast, we are enhancing the primary prediction of heart failure over concerning three core risk factors with the heart failure data set.

Bashir, Saba, Usman Qamar, and Farhan Hassan Khan [14] this research paper presents a novel classifier collaborative structure based on improved bagging approach with multi-objective weighted voting structure for analysis and prediction of heart disease. The proposed structure overwhelms the boundaries of orthodox performance by exploiting a collective of five heterogeneous classifiers: linear regression, Naive Bayes, instance-based learner, quadratic discriminant analysis and support vector machines.

Kim, Jaekwon, Jongsik Lee, and Youngho Lee [15] established model for CHD prediction must be aimed bestowing to rule-based procedures. In this study, a fuzzy logic and decision tree (classification and regression tree [CART])-driven CHD prediction model was advanced for Koreans. Datasets derived from the Nutrition Examination Survey VI (KNHANES-VI) and Korean National Health was exploited to produce the proposed method.

Joshi, Sujata, and Mydhili K. Nair [16] In this research, the classification-based data mining techniques are practiced to healthcare data. This research emphasizes on the forecast of heart disease using three classification techniques namely Naïve Bayes, Decision Trees, and K Nearest Neighbour.

Chadha, Ritika, *et al.*, [17] The objective of the research is to accumulate, organize and analyse the numerous data mining techniques that have been implicit and instigated in the latest years for Heart Disease Prediction. This paper attempts to highlight blatant assessments and focus to the boons and banes of each technique.

Choi, Edward, *et al.*, [18] explored whether the practice of deep learning to typical temporal relations among proceedings in electronic health records (EHRs) would advance model performance in forecasting preliminary diagnosis of heart failure (HF) associated to conservative techniques that snub momentarily

Saxena, Kanak, and Richa Sharma [19] In this study, we have structured a framework that can competently discover the doctrines to forecast the risk level of patients considering the given constraint about their health. The main influence of this research is to assist a non-specialized doctor to brand accurate decision about the heart disease risk level. The guidelines engendered by the projected system are arranged as Pruned Rules, Original Rules, rules without duplicates, Sorted Rules, Classified Rules and Polish. The implementation of the structure is evaluated as far as arrangement exactness and the results reveal that the structure has astonishing prospective in expecting the coronary illness risk level all more accurately.

Goldstein, Benjamin A., Ann Marie Navar, and Rickey E. Carter [20] The purpose of this review is to demonstrate the practice of machine-learning techniques for the growth of risk prediction models. Normally offered as black box methods, most machine-learning methods are intended at unravelling specific tasks that rise in data examination that are not well addressed by distinctive regression approaches.

Miranda, Eka, *et al.*, [21] The process of designing the method initiated by classifying the knowledge connected to the cardiovascular disease outline and the level of cardiovascular disease risk factors for adults dependent on the medical record, and scheming a mining technique model consuming a naïve Bayes classifier. Assessment of this research employed two techniques: sensitivity, accuracy, and specificity calculation as well as an assessment term

with internists and cardiologists. The features of cardiovascular disease are recognized by its principal risk factors. Those aspects are the level of lipids in the blood, diabetes mellitus, kidney function, and coronary artery function. Class markers were dispensed in relation to the standards of these factors: risk level 1, risk level 2 and risk level 3.

Chadha, Ritika, and Shubhankar Mayank [22] This paper endeavours to elicit the organization and execution of these techniques Decision Tree, Artificial Neural Networks, and Naive Bayes and strain upon the effects and inference persuaded on the base of exactness and time intricacy. By far, the opinions expose that Artificial Neural Networks outstripped Decision Tree and Naive Bayes.

Singh, Garima, *et al.*, [23] The central objective of this investigation is to cultivate a archetype Health Care Prediction System using, Naive Bayes. The System will determine and abstract hidden data associated to diseases (heart attack, cancer and diabetes) from a chronological heart disease database. It will response for the problematical queries for analysing sickness and consequently assist healthcare practitioners to practice intellectual clinical selections which prehistoric call support systems cannot do with.

Pouriyeh, Seyedamin, *et al.*, [24] This paper aims to probe and link the exactness of dissimilar data mining classification structures and their amalgamations over Ensemble Machine Learning Techniques for predicting heart disease.

Samuel, Oluwarotimi Williams, *et al.*, [25] In this study, the frequently used 13 HF attributes were measured and their contributions were firm by an knowledgeable cardiac clinician. And Fuzzy analytic hierarchy process (Fuzzy\_AHP) technique was cast-off to calculate the universal weights for the attributes constructed on their distinct contribution. At that time, the global weights that epitomize the contributions of the attributes were instigated to train an ANN classifier for the prediction of HF risks in patients.

Tripoliti, Evanthia E., *et al.*, [26] The goal of this paper is to show the state-of-the-art of the machine learning methodologies practiced for the evaluation of heart failure. More precisely, models forecasting the existence, approximating the subtype, evaluating the severity of heart failure and forecasting the existence of confrontational actions, such as re-hospitalizations, destabilizations, , and mortality are presented.

Acharya, U. Rajendra, *et al.*, [27] Electrocardiogram is broadly employed to diagnose the congestive heart failure (CHF). It is the principal non-invasive diagnostic device that can escort in the administration and follow-up of patients with CHF. Heart rate variability (HRV) signals which are nonlinear in nature hold the hidden signatures of

numerous cardiac diseases. As a result, this paper recommends a nonlinear method, empirical mode decomposition (EMD), for a mechanized identification and classification of standard and CHF using HRV signals. In this research, HRV signals are endangered to EMD to acquire intrinsic mode functions (IMFs).

Seah, Jarrel CY, *et al.*, [28] To examine Generative Visual Rationales (GVRs) as a device for envisaging neural system knowledge of chest radiograph features in congestive heart failure (CHF).

Kumar, Priyan Malarvizhi, and Usha Devi Gandhi [29] The data generation speed of IoT sensor devices is very extraordinary. As a result, the size of data produced from the IoT-based health monitoring system is also very high. So as to overwhelm this problem, this paper suggests mountable three-tier architecture to store and method such massive volume of wearable sensor data. Tier-1 emphasizes on collection of information from IoT wearable sensor devices. Tier-2 customs Apache HBase for loading the large size of wearable IoT sensor data in cloud calculating.

Jain, Divya, and Vijendra Singh [30] This paper grants a survey on the exploitation of feature selection and classification techniques for the analysis and forecast of chronic diseases. Ample selection of features shows a substantial role for augmenting accuracy of classification systems. Dimensionality reduction supports in enlightening global performance of machine learning algorithm.

Paul, Animesh Kumar, *et al.*, [31] In this paper, we have industrialised acomputed fuzzy diagnostic system based on genetic algorithm (GA) and a modified dynamic multi-swarm particle swarm optimization (MDMS-PSO) for predicting the risk level of heart disease. Our planned fuzzy diagnostic system (FS) works as follows: i) Effective attributes are selected through statistical methods such as Correlation coefficient, R-Squared and Weighted Least Squared (WLS) method ii) Pre-process the data sets iii) Weighted fuzzy rules are formed other basis of selected attributes using GA, iv) MDMS-PSO is employed for the optimization of membership functions (MFs) of FS, v) Build the ensemble FS from the generated fuzzy knowledge base by fusing the different local FSs.

Manogaran, Gunasekaran, R. Varatharajan, and M. K. Priyan [32] Multiple Kernel Learning with Adaptive Neuro-Fuzzy Inference System (MKL with ANFIS) based deep learning method is proposed in this research for heart disease diagnosis. The anticipated MKL with ANFIS based deep learning method tails two-fold approach. MKL technique is cast-off to distribute constraints between normal individuals and heart disease patients.

### III. METHODOLOGY

Methodology is a process of achieving the projected work with the proposed algorithm. In this paper the Multilayer

Perceptron Back Propagation Algorithm of Neural Network is used Back Propagation Algorithm is also called as „Back Propagation of Errors Algorithm“, is a standard technique that create an actual ANN (Artificial Neural Network) training in conjunction with certain optimization methods like gradient descent. The process figures the depth of the loss function in the input data in reference to all the weights in the system. The ascent techniques are then applied to the optimization methods to regulate the weights to reduce the loss function in the network. Therefore, the algorithm necessitates a recognized and a anticipated output for all inputs so as to calculate the rise of loss function. Generally, an overview of Multi-layered Feed Forward Networks is completed using delta rule which probably brands a series of iterative rules to compute gradients for each layer. Back Propagation Algorithm demands the stimulation function to be altered between the neurons. The current researches on distributed computing, parallel, and computational neuroscience are presently employed with the conceptions of Multi-Layer Perceptron using a Back-Propagation Algorithm. MLP Back Propagation Algorithm has also expanded attention in pattern recognition domain. Consequently, they are expedient in research, owing to their ability in elucidating multifaceted complications, and similarly for their fitness calculation outcomes even with serious forecasts. MLP is one of the Neural Network models, has the same design of Feed-Forward back Propagation for Supervised training. As a whole, MLP network comprises an input, one or more hidden layers and an output layer.

*Step 0:* Initialize the weight and learning rate

*Step 1:* Perform Steps 2 to 9 when stopping condition is false

*Step 2:* Perform Steps 3-8 for each training pair

Feed-Forward Phase (Phase I):

*Step 3:* Each input unit receives input signal  $x_i$  and sends it to the hidden unit ( $i=1$  to  $n$ )

*Step 4:* Each hidden unit  $z_j$  ( $j=1$  to  $p$ ) sums its weighted input signals to calculate net input:

$$Z_{inj} = V_{oj} + \sum_{i=1}^n x_i v_{ij} \quad (1)$$

Calculate output of the hidden unit by applying its activation function over  $Z_{inj}$

$$Z_j = f(Z_{inj}) \quad (2)$$

$$f(Z_{inj}) = \frac{1}{1+e^{-\lambda Z_{inj}}} \quad (3)$$

*Step 5:* For each output unit  $y_k$  ( $k=1$  to  $m$ ), calculate the net input:

$$Y_{ink} = w_{ok} + \sum_{j=1}^p z_j v_{jk} \quad (4)$$

And apply the activation function to compute output signal

$$Y_k = f(Y_{inj}) \quad (5)$$

$$f(Y_{ink}) = \frac{1}{1+e^{-\lambda Y_{inj}}} \quad (6)$$

Back-propagation of error (Phase II):

Step 6: Each output unit  $y_k$  ( $k=1$  to  $m$ ) receives a target pattern corresponding to the input training pattern and computer the error correction term:

$$\delta_k = (t_k - y_k)f'(y_{ink}) \quad (7)$$

$$f'(y_{ink}) = \lambda f(y_{ink})[-f'(y_{ink})] \quad (8)$$

On the basis of the calculated error correction term, the weight and bias is updated as follows:

$$\Delta W_{jk} = \alpha \delta_k Z_j \quad (9)$$

$$\Delta w_{ok} = \alpha \delta_k \quad (10)$$

Also, send  $\delta_k$  to the hidden layer backwards.

Step 7: Each hidden unit ( $z_j$ ,  $j=1$  to  $p$ ) sums its delta inputs from the output units:

$$\delta_{inj} = \sum_{k=1}^m \delta_k w_{jk} \quad (11)$$

The term  $\delta_{inj}$  gets multiplied with the derivative of ( $Z_{inj}$ ) to calculate the error terms:

$$\delta_j = \delta_{inj} f'(Z_{inj}) \quad (12)$$

$$f'(Z_{inj}) = \lambda f(Z_{inj})[1 - f(Z_{inj})] \quad (13)$$

On the basis of calculated  $\delta_j$ , the weight and bias are calculated as:

$$\Delta V_{ij} = \alpha \delta_j x_i \quad (14)$$

$$\Delta V_{oj} = \alpha \delta_j \quad (15)$$

Weight and bias updation (Phase III):

Step 8: Each output unit ( $y_k$ ,  $k=1$  to  $m$ ) updates the bias and weights:

$$w_{jk}(new) = w_{jk}(old) + \Delta w_{jk} \quad (16)$$

$$w_{ok}(new) = w_{ok}(old) + \Delta w_{ok} \quad (17)$$

Each hidden unit ( $z_j$ ,  $j=1$  to  $p$ ) updates its bias and weights:

$$V_{ij}(new) = V_{ij}(old) + \Delta V_{ij} \quad (18)$$

$$V_{oj}(new) = V_{oj}(old) + \Delta V_{oj} \quad (19)$$

Step 9: Check for the stopping condition. The stopping condition may be certain number of epochs reached or when the actual output equals the target output.

Testing Algorithm:

Step 0: Initialize the weights. The weights are taken from the training algorithm

Step 1: Perform steps 2-4 for each input vector

Step 2: Set the activation of input unit for  $x_i$  ( $i=1$  to  $n$ ).

Step 3: Calculate the net input to hidden unit  $x$  and its output. For  $j=1$  to  $p$ ,

$$z_{inj} = V_{oj} + \sum_{i=1}^n x_i v_{ij} \quad (20)$$

$$z_j = f(z_{inj})$$

Step 4: Now compute the output of the output layer unit.

For  $k=1$  to  $m$ ,

$$y_{ink} = w_{ok} + \sum_{j=1}^p z_j w_{jk} \quad (21)$$

$$y_k = f(y_{ink})$$

Sigmoidal activation function is used for calculating the output.

#### IV. DESCRIPTION OF THE DATASET

Dataset practiced for research is contained the data of Heart Disease which was in use from the UCI Machine Learning Repository. The statistics folder actually comprised of 10 different datasets with the dissimilarity in quantity of attributes and missing values. Among them, Cleveland dataset was appeared to be more appropriate with a minor number of 14 attributes (out of 64) with 303 occurrences.

TABLE I HEART DISEASE DATASET

S. No.	Feature Name	Description	Range
1	Age	Patient's Age	29-77
2	Sex	1=male; 0=female	0-1
3	Cp	Value 1: typical angina Value 2: atypical anginal Value 3: non-anginal pain Value 4: asymptotic	1-4
4	trestbps	Resting blood pressure(in mm Hg)	94-200
5	Chol	Serum cholesterol in mg/dl	126-564
6	Fbs	(Fasting blood sugar .120mg/dl ) (1=true; 0=false)	0-1
7	Restecg	electrocardiography results Value 0: normal Value 1: having ST-T wave abnormality (T wave inversions and/or ST Elevation or depression of >0.05mV) Value 2: showing probable or definite left	0-2
8	Thalach	Maximum heart rate achieved	71-202
9	Exang	Exercise induced angina(1=yes;0=no)	0-1
10	OldPeak	ST depression induced by exercise relative to rest	0-6.2
11	Slope	The slope of the peak exercise ST segment Value 1: up sloping Value 2: flat Value 3:down sloping	0-2
12	Ca	Number of major vessels (0-3) Colored by fluoroscopy	0-3
13	Thal	Normal, fixed defect, reversible defect	3-7

### V. RESULTS AND DISCUSSION

The experiment was assessed in MATLAB R2018B. The dataset was alienated into three classes of data as target, input and sample. MLP with Back Propagation Algorithm is accomplished using input data. All the 13 attributes were taken into the attention for attaining the goal. Input weights are mechanically accustomed by the purelin function for aggregating the precision of the output. The error is premeditated as the transformation between the network output and the target output. The objective is to reduce the average of the amount of these mistakes. The LMSE (Least Mean Square Error) algorithm regulates the weights and biases of the linear network in an attempt to reduce this mean square error. Fortuitously, the mean square error performance file for the linear network is a quadratic task. As a result, the performance index will either have a weak minimum or no minimum, one global minimum, reliant on the features of the input vectors. In detail, the features of the input vectors regulate whether or not a unique solution exists. The subsequent table 2 portrays the pre-processed dataset acquired by using the proposed feature selection methodology [33].

TABLE II PRE-PROCESSED DATASET OBTAINED BY PROPOSED TFIE-PSFS METHOD [33]

S. No.	Proposed TFIE-PSFS
1	cp
2	ca
3	thal
4	slope
5	exang
6	chol

Figure 1 gives the confusion matrix obtained for original dataset with 13 attributes by using Artificial Neural Network classification. From the figure 1, the confusion matrix for three outpus of the original dataset, the accurateness of the arrangement about the three classes has epitomized. Figure 2 gives the ROC for the unique dataset achieved by ANN. Figure 3 gives the regression analysis of the original dataset when consuming ANN classification. Figure 4 depicts the Mean Squared Error for original dataset with ANN.

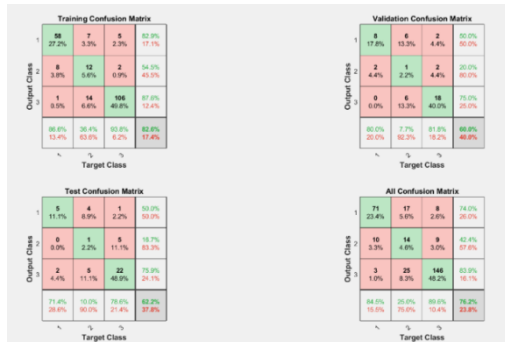


Fig. 1 Confusion Matrix obtained by using MLP for original dataset

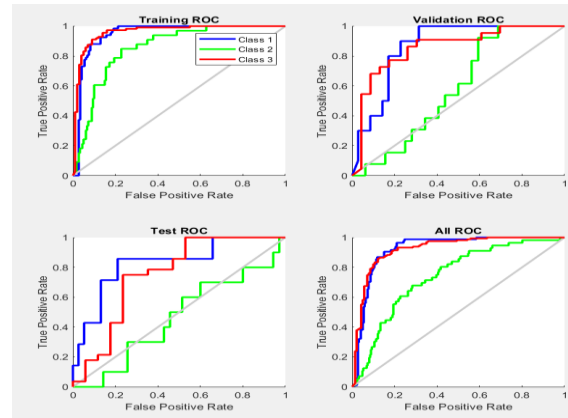


Fig. 2 ROC for original dataset curve obtained by using MLP

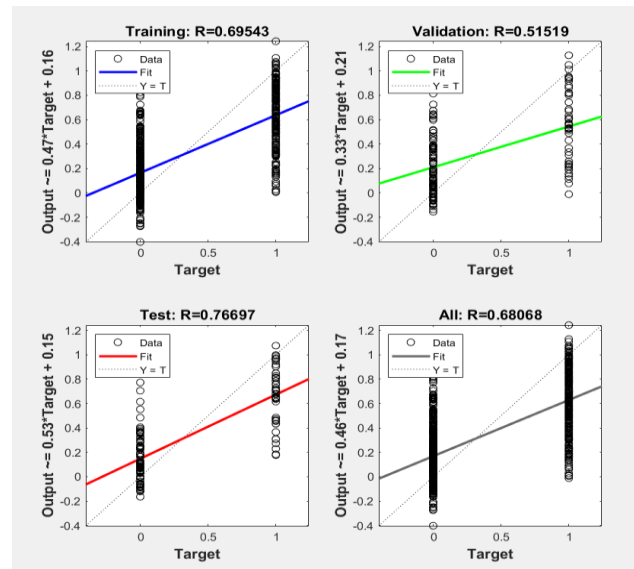


Fig. 3 Regression analysis for original dataset by using ANN

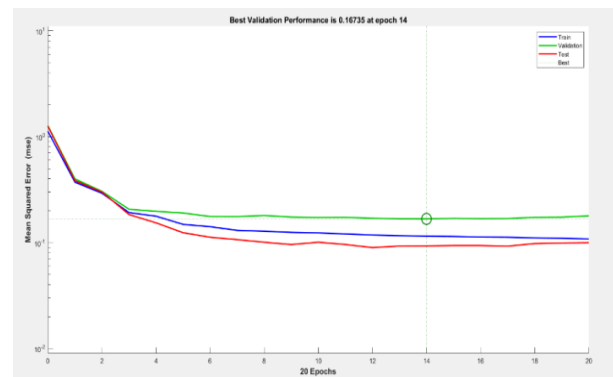


Fig. 4 Mean Squared Error for original dataset with ANN

Figure 5 gives the confusion matrix obtained for pre-processed dataset with 6 attributes by using Artificial Neural Network classification. From the figure 5, the confusion matrix of the pre-processed dataset, the accuracy of the classification about the three class has represented. Figure 6 gives the ROC for the pre-processed dataset obtained by ANN. Figure 7 gives the regression analysis of the pre-processed dataset when using ANN classification. Figure 8 depicts the Mean Squared Error for pre-processed dataset with ANN.



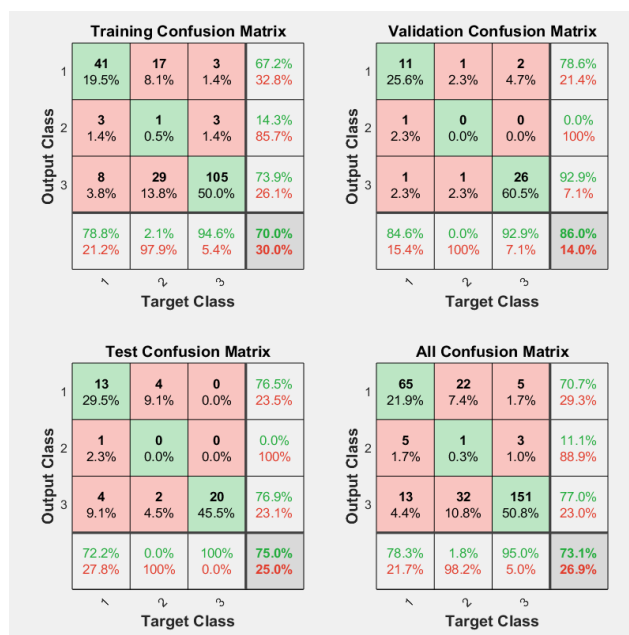


Fig. 5 Confusion Matrix for reduced dataset by using ANN

## VI. CONCLUSION

In this paper, a unique approach based on MLP Back Propagation Neural Network is recommended to predict the Heart Disease by classifying the patient category into three types. The Projected classification employed almost 13 noteworthy medical attributes and pre-processed dataset accomplished by proposed TFIE-PSFS for Heart Disease predictions. The research shows good performance of the proposed algorithm and was associated to related methods above the same dataset. In addition, the precision of the dissimilar training functions of Multilayer Perceptron Algorithm were associated to opt the best training function. The investigational outcomes display the promising results that the MLP with appropriate training algorithm can be an operative device to predict Heart Disease with enhanced precision for the pre-processed dataset by TFIE-PSFS technique. Using the projected three output by the proposed ANN design, the classification accuracy increases due to third category (Hesitation).

## REFERENCES

- Nahar, Jesmin, *et al.*, "Association rule mining to detect factors which contribute to heart disease in males and females", *Expert Systems with Applications*, Vol. 40, No. 4, pp.1086-1093, 2013.
- Vijayarani, S., and S. Sudha, "An efficient classification tree technique for heart disease prediction", *International Conference on Research Trends in Computer Technologies (ICRTCT-2013) Proceedings published in International Journal of Computer Applications (IJCA)(0975-8887)*, Vol. 201, 2013.
- Gayathri, P., and N. Jaisankar, "Comprehensive study of heart disease diagnosis using data mining and soft computing techniques", 2013.
- Shouman, Mai, Tim Turner, and Rob Stocker, "Integrating clustering with different data mining techniques in the diagnosis of heart disease", *J. Comput. Sci. Eng.*, Vol. 20, No. 1, 2013.
- Amato, Filippo, *et al.*, "Artificial neural networks in medical diagnosis", pp. 47-58, 2013.
- Persi Pamela, L., and P. Gayathri, "A fuzzy optimization technique for the prediction of coronary heart disease using decision tree", 2013.

- Chaurasia, Vikas, and Saurabh Pal, "Data mining approach to detect heart diseases", 2014.
- Thenmozhi, K., and P. Deepika, "Heart disease prediction using classification with different decision tree techniques", *International Journal of Engineering Research and General Science*, Vol. 2, No. 6, pp.6-11, 2014.
- Kim, Jae-Kwon, *et al.*, "Adaptive mining prediction model for content recommendation to coronary heart disease patients", *Cluster computing*, Vol. 17, No. 3, pp. 881-891, 2014.
- Seera, Manjeevan, and CheePeng Lim, "A hybrid intelligent system for medical data classification", *Expert Systems with Applications*, Vol. 41, No. 5, pp. 2239-2249, 2014.
- Bashir, Saba, UsmanQamar, and M. YounusJaved, "An ensemble-based decision support framework for intelligent heart disease diagnosis", *Information Society (i-Society), 2014 International Conference on IEEE*, 2014.
- Shabana, ASMI P., and S. Justin Samuel, "An analysis and accuracy prediction of heart disease with association rule and other data mining techniques", *Journal of Theoretical and Applied Information Technology*, Vol. 79, No. 2, pp. 254-60, 2015.
- Aljaaf, A. J., *et al.*, "Predicting the likelihood of heart failure with a multi-level risk assessment using decision tree", *Technological Advances in Electrical, Electronics and Computer Engineering (TAECE), 2015 Third International Conference on.IEEE*, 2015.
- Bashir, Saba, UsmanQamar, and Farhan Hassan Khan, "BagMOOV: A novel ensemble for heart disease prediction bootstrap aggregation with multi-objective optimized voting", *Australasian physical & engineering sciences in medicine*, Vol. 38, No. 2, pp. 305-323, 2015.
- Kim, Jaekwon, Jongsik Lee, and Youngho Lee, "Data-mining-based coronary heart disease risk prediction model using fuzzy logic and decision tree", *Healthcare informatics research*, Vol. 21, No. 3, pp. 167-174, 2015.
- Joshi, Sujata, and Mydhili K. Nair, "Prediction of heart disease using classification-based data mining techniques", *Computational Intelligence in Data Mining, Springer*, New Delhi, Vol. 2, pp. 503-511, 2015.
- Chadha, Ritika, *et al.*, "Application of data mining techniques on heart disease prediction: a survey", *Emerging Research in Computing, Information, Communication and Applications*. Springer, New Delhi, pp. 413-426, 2016.
- Choi, Edward, *et al.*, "Using recurrent neural network models for early detection of heart failure onset", *Journal of the American Medical Informatics Association*, Vol. 24, No. 2, pp.361-370, 2016.
- Saxena, Kanak, and Richa Sharma, "Efficient Heart Disease Prediction System", *Procedia Computer Science*, Vol. 85, pp. 962-969, 2016.
- Goldstein, Benjamin A., Ann Marie Navar, and Rickey E. Carter, "Moving beyond regression techniques in cardiovascular risk prediction: applying machine learning to address analytic challenges", *European heart journal*, Vol. 38, No. 23, pp. 1805-1814, 2016.
- Miranda, Eka, *et al.*, "Detection of cardiovascular disease risk's level for adults using naive Bayes classifier", *Healthcare informatics research*, Vol. 22, No. 3, pp.196-205, 2016.
- Chadha, Ritika, and ShubhankarMayank, "Prediction of heart disease using data mining techniques", *CSI transactions on ICT*, Vol. 4, No. 2-4, pp. 193-198, 2016.
- Singh, Garima, *et al.*, "Heart disease prediction using Naïve Bayes", *International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056*, 2017.
- Pouriyeh, Seyedamin, *et al.*, "A comprehensive investigation and comparison of Machine Learning Techniques in the domain of heart disease", *Computers and Communications (ISCC), 2017 IEEE Symposium on IEEE*, 2017.
- Samuel, Oluwarotimi Williams, *et al.*, "An integrated decision support system based on ANN and Fuzzy\_AHP for heart failure risk prediction", *Expert Systems with Applications*, Vol. 68, pp. 163-172, 2017.
- Tripoliti, Evanthia E., *et al.*, "Heart failure: diagnosis, severity estimation and prediction of adverse events through machine learning techniques", *Computational and structural biotechnology journal*, Vol. 15, pp. 26-47, 2017.
- Acharya, U. Rajendra, *et al.*, "Application of empirical mode decomposition (EMD) for automated identification of congestive

- heart failure using heart rate signals”, *Neural Computing and Applications*, Vol. 28, No. 10, pp. 3073-3094, 2017.
- [28] Seah, Jarrel CY, *et al.*, “Chest radiographs in congestive heart failure: visualizing neural network learning”, *Radiology*: 180887, 2018.
- [29] Kumar, PriyanMalarvizhi, and Usha Devi Gandhi, “A novel three-tier Internet of Things architecture with machine learning algorithm for early detection of heart diseases”, *Computers & Electrical Engineering*, Vol. 65, 222-235, 2018.
- [30] Jain, Divya, and Vijendra Singh, “Feature selection and classification systems for chronic disease prediction: A review”, *Egyptian Informatics Journal*, 2018.
- [31] Paul, Animesh Kumar, *et al.*, “Adaptive weighted fuzzy rule-based system for the risk level assessment of heart disease”, *Applied Intelligence*, Vol. 48, No. 7, pp. 1739-1756, 2018.
- [32] Manogaran, Gunasekaran, R. Varatharajan, and M. K. Priyan, “Hybrid recommendation system for heart disease diagnosis based on multiple kernel learning with adaptive neuro-fuzzy inference system”, *Multimedia tools and applications*, Vol. 77, No. 4, pp. 4379-4399, 2018.
- [33] P.Umasankar, V. Thiagarasu, “A Novel Thrice Filtered Information Energy based Particle Swarm Feature Selection for the Heart Disease Diagnosis”, *International Journal of Pure and Applied Mathematics*, Vol. 119, No. 15, pp. 3485-3499, 2018.