iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** XII **Month of publication:** December 2023

DOI: https://doi.org/10.22214/ijraset.2023.57508

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue XII Dec 2023- Available at www.ijraset.com

Heart Disease Detection using Machine Learning

K. Jashveen[1], M. Jayaprakash[2], T. Jayaprakash[3], D. Jeevani[4], P. Joel Sastry[5], K. Jayasree[6], K. Divya Bharathi[7]
[1, 2, 3, 4, 5, 6]*B.Tech, School of Engineering, Hyderabad, India*
[7]*Guide, Assistant Professor, School of Engineering, Mallareddy University*

Abstract: ***Day by day the cases of heart diseases are increasing at a rapid rate and it's very Important and concerning to predict any such diseases before hand. This diagnosis is a difficult task i.e. it should be performed precisely and efficiently. The research paper mainly focuses on which patient is more likely to have a heart disease based on various medical attributes. We prepared a heart disease prediction system to predict whether the patient is likely to be diagnosed with a heart disease or not using the medical history of the patient. We used different algorithms of machine learning such as logistic regression and KNN to predict and classify the patient with heart disease. A quite Helpful approach was used to regulate how the model can be used to improve the accuracy of prediction of Heart Attack in any individual. The strength of the proposed model was quiet satisfying and was able to predict evidence of having a heart disease in a particular individual by using KNN and Logistic Regression which showed a good accuracy in comparison to the previously used classifier such as naïve bayes etc. So a quiet significant amount of pressure has been lifted by using the given model in finding the probability of the classifier to correctly and accurately identify the heart disease. The Given heart disease prediction system enhances medical care and reduces the cost. This project gives us significant knowledge that can help us predict the Patient's Heart disease.***

I. INTRODUCTION

Heart disease describes a range of conditions that affect your heart. Today, cardiovascular diseases are the leading cause of death worldwide with 17.9 million deaths annually, as per the World Health Organization reports. Various unhealthy activities are the reason for the increase in the risk of heart disease like high cholesterol, obesity, increase in triglycerides levels, hypertension, etc… There are certain signs which the American Heart Association lists like the persons having sleep issues, a certain increase and decrease in heart rate (ir regular heart beat), swollen legs, and in some cases weight gain occurring quite fast; it can be 1-2 kg daily. All these symptoms resemble different diseases also like it occurs in the aging persons, so it becomes a difficult task to get a correct diagnosis, which results in fatality in near future. But as time is passing, a lot of research data and patients records of hospitals are available. There are many open sources for accessing the patient's records and researches can be conducted so that various computer technologies could be used for doing the correct diagnosis of the patients and detect this disease to stop it from becoming fatal. Now a days it is well known that machine learning and artificial intelligence are playing a huge role in the medical industry. We can use different machine learning and deep learning models to diagnose the disease and classify or predict the results. A complete genomic data analysis can easily be done using machine learning models. Models can be trained for knowledge pandemic predictions and also medical records can be transformed and analyzed more deeply for better predictions. Many studies have been performed and various machine learning models are used for doing the classification and prediction for the diagnosis of heart disease.

II. PROBLEM STATEMENT

The major challenge in heart disease is its detection. There are instruments available which can predict heart disease but either they are expensive or are not efficient to calculate chance of heart disease in human. Early detection of cardiac diseases can decrease the mortality rate and overall complications. However, it is not possible to monitor patients every day in all cases accurately and consultation of a patient for 24 hours by a doctor is not available since it requires more sapience, time and expertise. Since we have a good amount of data in today's world, we can use various machine learning algorithms to analyze the data for hidden patterns. The hidden patterns can be used for health diagnosis in medicinal data.

III. LITERATURE REVIEW

In literature various machine learning based diagnosis techniques have been proposed by researchers to diagnosis HD. This research study present some existing machine learning based diagnosis techniques in order to explain the important of the proposed work. Detrano et al. [11] developed HD classification system by using machine learning classification techniques and the performance of the system was 77% in terms of accuracy.

©IJRASET: All Rights are Reserved

Cleveland dataset was utilized with the method of global evolutionary and with features selection method. In another study Gudadhe et al. [22] developed a diagnosis system using multi-layer Perceptron and support vector machine (SVM) algorithms for HD classification and achieved accuracy 80.41%. Humar et al. [23] designed HD classification system by utilizing a neural network with the integration of Fuzzy logic. The classification system achieved 87.4% accuracy. Resul et al. [19] developed an ANN ensemble based diagnosis system for HD along with statistical measuring system enterprise miner (5.2) and obtained the accuracy of 89.01%, sensitivity 80.09%, and specificity 95.91%. Akil et al. [24] designed a ML based HD diagnosis system. ANN-DBP algorithm along with FS algorithm and performance was good. Palaniappan et al. [17] proposed an expert medical diagnosis system for HD identification. In development of the system the predictive model of machine learning, such as navies bays (NB), Decision Tree (DT), and Artificial Neural Network were used. The 86.12% accuracy was achieved by NB, ANN accuracy 88.12% and DT classifier achieved 80.4% accuracy. Olaniyi et al. [18] developed a three-phase technique based on the artificial neural network technique for HD prediction in angina and achieved 88.89% accuracy. Samuel et al. [20] developed an integrated medical decision support system based on artificial neural network and Fuzzy AHP for diagnosis of HD. The performance of the proposed method in terms of accuracy was 91.10%. Liu et al. [25] proposed a HD classification system using relief and rough set techniques. The proposed method achieved 92.32% classification accuracy. In [26] proposed a HD identification method using feature selection and classification algorithms. Sequential Backward Selection Algorithm (SBS FS) for Features Selection. The classifier K-Nearest Neighbor (K-NN) performance has been checked on full and on selected features set. The proposed method obtained high accuracy. In another study MOHAN et al. [27] designed a HD prediction method by using hybrid machine learning techniques. He also proposed a new method for significant feature selection from the data for effective training and testing of machine learning classifier. They have been recorded 88.07% classification accuracy. Geweid et al. [28] designed HD identification techniques by using improved SVM based duality optimization technique. In the above literature the proposed HD diagnosis methods limitation and advantages have been summarized in Table 1 for better understanding the important of our proposed approach. All these existing techniques used numerous methods to identify the HD at early stages. However, all these techniques have lack of prediction accuracy and high computation time for prediction of HD. According to Table 1 the prediction accuracy of HD detection method need further improvement for efficient and accurate detection at early stages for better treatment and recovery. Thus, the major issues in these previous approaches are low accuracy and high computation time and these might be due the use of irrelevant features in dataset. In order to tackle these problems new methods are needed to detect HD correctly. The improvement in prediction accuracy is a big challenge and research gap.

IV. METHODOLOGY

All the research materials and techniques background are discussed in the following subsections.

A. Data Set

Cleveland Heart Disease [29] dataset is considered for testing purpose in this study. During the designing of this data set there were 303 instances and 75 attributes, however all published experiments refer to using a subset of 14 of them. In this work, we performed pre-processing on the data set, and 6 samples have been eliminated due to missing values. The remaining samples of 297 and 13 features dataset is left and with 1 output label. The output label has two classes to describe the absence of HD and the presence of HD. Hence features matrix 297*13 of extracted features is formed. The dataset matrix information's are given in Table 2.

B. Pre-Processing Of Data Set

The pre-processing of dataset required for good representation. Techniques of pre-processing such as removing attribute missing values, Standard Scalar (SS), Min-Max Scalar have been applied to the dataset.

C. Standard State Of The Art Features Selection Algorithms

After data pre-processing, the selection of feature is required for the process. In general, FS is a significant step in constructing a classification model. It works by reducing the number of input features in a classifier, to have good predictive and short computationally complex models [30]. We have been used four standard state of the art FS algorithms and one our proposed FS algorithm in this study.

1) *Relief:* Relief [31] algorithm assigns weights to each data set features and updated weights automatically. The features having high weight values should be selected and low weight will be discarded. Relief and K-NN algorithm process to determine the weights of features are the same [32]. The algorithm relief repeated through m random training samples (R_k), without selection substitution, and m is the parameter. Each k, R_k is the 'target' sample and weight W of the is updated [33]. The algorithm 1 is the Pseudo-code for Relief FS algorithm.

2) *Minimal-Redundancy-Maximal–Relevance:* MRMR algorithm chooses features that are suitable for the prediction and selected features that are non redundant. It does not take care of the combination of features [32]. The MRMR pseudo code is given in algorithm 2 [34].
3) *Least-Absolute-Shrinkage-Selection-Operator Algorithm:* LASSO choose feature based on modifying the absolute coefficient value of the features. Then these features coefficient values set to zero and finally zero coefficient features are eliminated from the features set. In the selected features set those features to include who coefficient have a high value. Sometime LASSO selects irrelevant features and includes in the subset of feature [35].

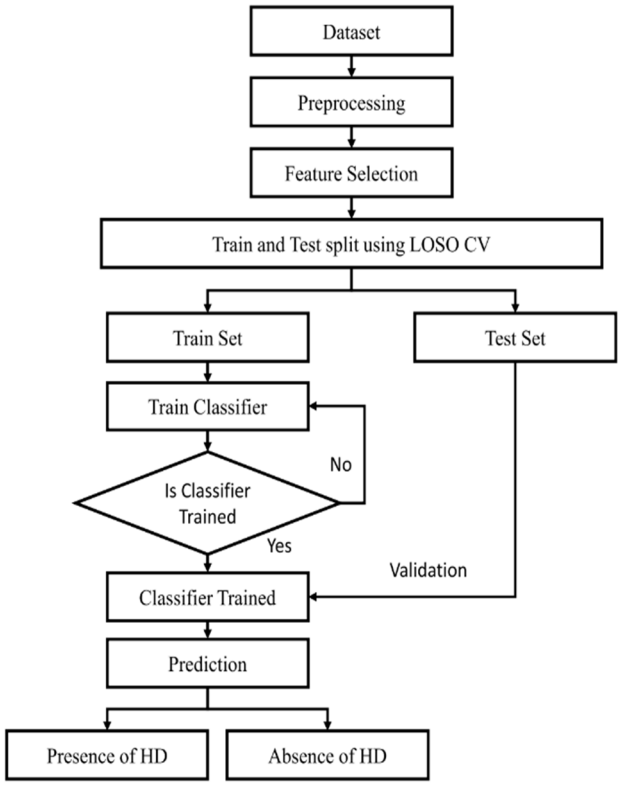


Fig 4.1 Architecture

.

V. EXPERIMENT RESULTS



Fig 5.1 Classes Distribution

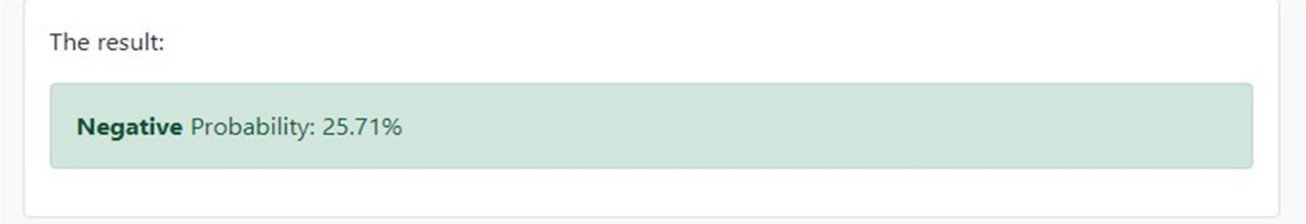


Fig 5.2 Features visualization

	precision	recall	f1-score	support
Benign	0.99	0.96	0.97	1004
Malware	0.99	1.00	0.99	2919
accuracy			0.99	3923
macro avg	0.99	0.98	0.98	3923
weighted avg	0.99	0.99	0.99	3923

Fig 5.3 Classification Report

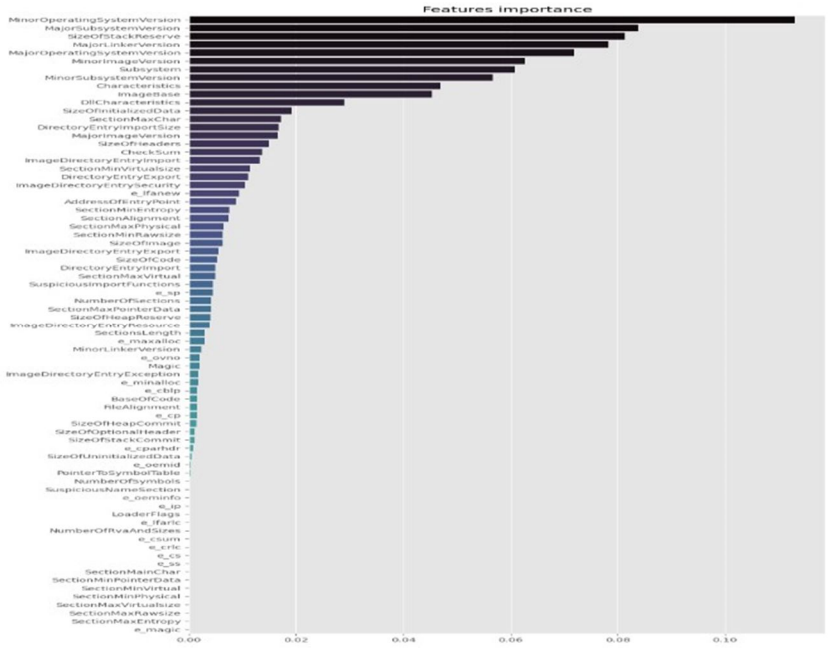


Fig 5.5 Features Importance

VI. CONCLUSION

In this study, an efficient machine learning based diagnosis system has been developed for the diagnosis of heart disease. Machine learning classifiers include LR, K-NN, ANN, SVM, NB, and DT are used in the designing of the system. Four standard feature selection algorithms including Relief, MRMR, LASSO, LLBFS, and proposed a novel feature selection algorithm FCMIM used to solve feature selection problem. LOSO cross-validation method is used in the system for the best hyperparameters selection. The system is tested on Cleveland heart disease dataset. Furthermore, performance evaluation metrics are used to check the performance of the identification system. According to Table 15 the specificity of ANN classifier is best on Relief FS algorithm as compared to the specificity of MRMR, LASSO, LLBFS, and FCMIM feature selection algorithms. Therefore for ANN with relief is the best predictive system for detection of healthy people. The sensitivity of classifier NB on selected features set by LASSO FS algorithm also gives the best result as compared to the sensitivity values of Relief FS algorithm with classifier SVM (linear). The classifier Logistic Regression MCC is 91% on selected features selected by FCMIM FS algorithm. The processing time of Logistic Regression with Relief, LASSO, FCMIM and LLBFS FS algorithm best as compared to MRMR FS algorithms, and others classifiers.

Thus the experimental results show that the proposed features selection algorithm select features that are more effective and obtains high classification accuracy than the standard feature selection algorithms. According to feature selection algorithms, the most important and suitable features are Thallium Scan type chest pain and Exercise-induced Angina. All

FS algorithms results show that the feature Fasting blood sugar (FBS) is not a suitable heart disease diagnosis. The accuracy of SVM with the proposed feature selection algorithm (FCMIM) is 92.37% which is very good as compared previously proposed methods as shown in Table 17. Further, the performance of machine learning based method FCMIM-

SVM is high then Deep neural network for detection of HD. A little improvement in prediction accuracy have great influence in diagnosis of critical diseases. The novelty of the study is developing a diagnosis system for identification of heart disease. In this study, four standard feature selection algorithms along with one proposed feature selection algorithm is used for features selection. LOSO CV method and performance measuring metrics are used. The Cleveland heart disease dataset is used for testing purpose. As we think that developing a decision support system through machine learning algorithms it will be more suitable for the diagnosis of heart disease. Furthermore, we know that irrelevant features also degrade the performance of the diagnosis system and increased computation time. Thus another innovative touch of our study to used features selection algorithms to selects the appropriate features that improve the classification accuracy as well as reduce the processing time of the diagnosis system. In the future, we will use other features selection algorithms, optimization methods to further increase the performance of a predictive system for HD diagnosis. The controlling and treatment of disease is significance after diagnosis, therefore, i will work on treatment and recovery of diseases in future also for critical disease such as heart, breast, Parkinson, diabetes.

VII. ACKNOWLEDGEMENT

I would like to express my heartfelt gratitude to my guide K.Divya Bharathi and head of department, Dr . Thayyaba Khatoon, for their invaluable guidance and unwavering support throughout the development of this project. Their insightful feedback helped me to refine my ideas and develop a comprehensive understanding of the subject matter.

Their mentorship was instrumental in shaping my approach towards the project, and I am grateful for the knowledge and experience they shared with me. Without their encouragement and support, this project would not have been possible. Once again, I extend my sincere thanks to my guides and head of department for their unwavering support and guidance.

Our sincere thanks to all the teaching and non-teaching staff of Department of Computer Science and Engineering (AI&ML) for their support throughout our project work.

REFERENCES

[1] L. Bui, T. B. Horwich, and G. C. Fonarow, ‘‘Epidemiology and risk profile of heart failure,’’ Nature Rev. Cardiol., vol. 8, no. 1, p. 30, 2011.

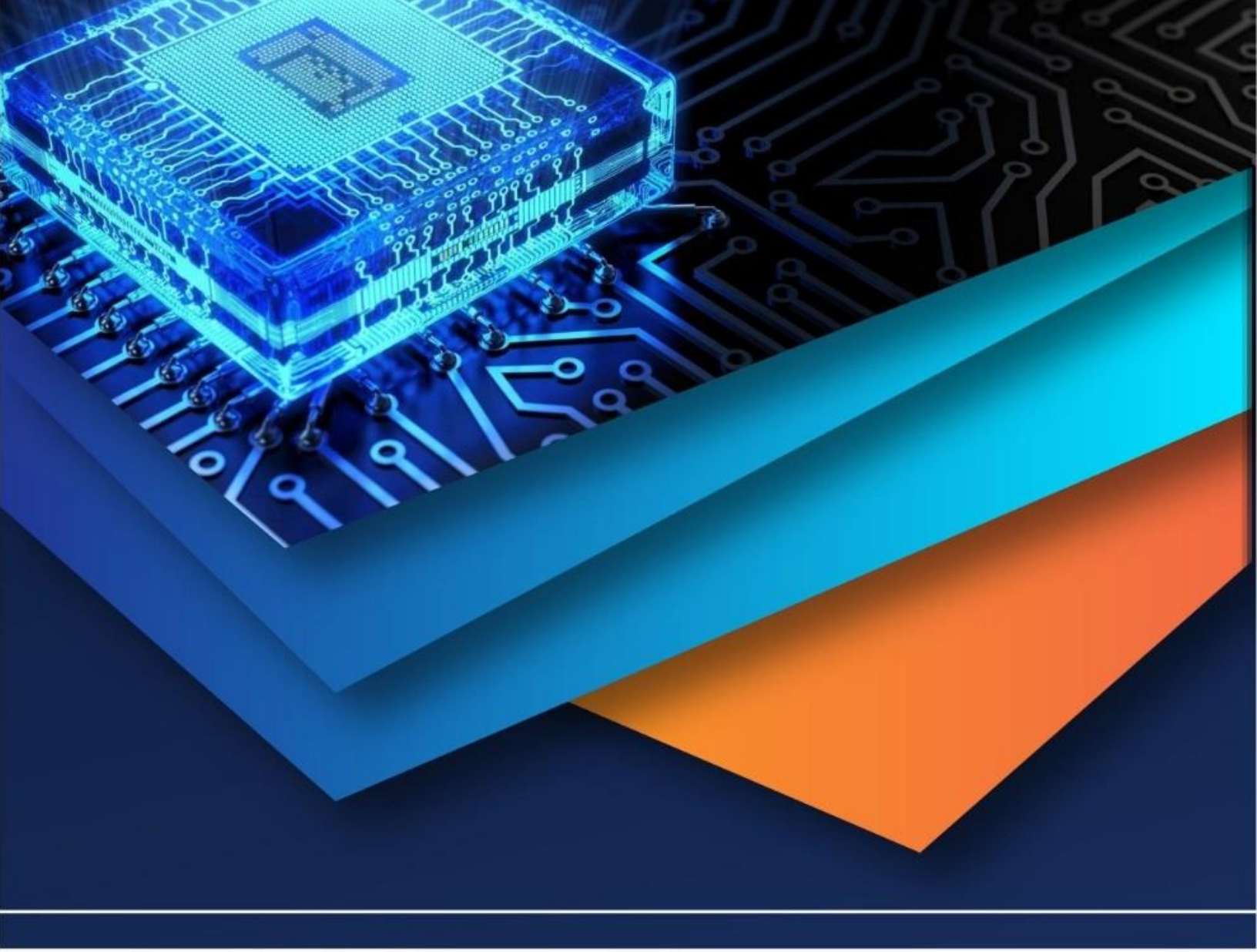
[2] M. Durairaj and N. Ramasamy, ‘‘A comparison of the perceptive approaches for preprocessing the data set for predicting fertility success rate,’’ Int. J. Control Theory Appl., vol. 9, no. 27, pp. 255–260, 2016.

[3] L. A. Allen, L. W. Stevenson, K. L. Grady, N. E. Goldstein, D. D. Matlock, R. M. Arnold, N. R. Cook, G. M. Felker, G. S. Francis, P. J. Hauptman, E. P. Havranek, H. M. Krumholz, D. Mancini, B. Riegel, and J. A. Spertus, ‘‘Decision making in advanced heart failure: A scientific statement from the American heart association,’’ Circulation, vol. 125, no. 15, pp. 1928–1952, 2012.

[4] S. Ghwanmeh, A. Mohammad, and A. Al-Ibrahim, ‘‘Innovative artificial neural networks-based decision support system for heart diseases diagnosis,’’ J. Intell. Learn. Syst. Appl., vol. 5, no. 3, 2013, Art. no. 35396.

[5] Q. K. Al-Shayea, ‘‘Artificial neural networks in medical diagnosis,’’ Int. J. Comput. Sci. Issues, vol. 8, no. 2, pp. 150–154, 2011.

[6] J. Lopez-Sendon, ‘‘The heart failure epidemic,’’ Medicographia, vol. 33, no. 4, pp. 363–369, 2011.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)