iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** VII **Month of publication:** July 2018

DOI: http://doi.org/10.22214/ijraset.2018.7134

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: 6.887
Volume 6 Issue VII, July 2018- Available at www.ijraset.com

Proficient Prognosis of Cardiac Disorders using Machine Learning, Low-Density Lipoprotein &Triglycerides

Nahida Nazir[1], Sharifa Rajab[2], Vinod Sharma[3]
[1, 2, 3]*University of Jammu, Jammu, India*

Abstract: ***Abnormalities related to the Lipid profile, Electrocardiogram (ECG), Blood Pressure (BP), chest pain are major causes of cardiovascular diseases. Researchers working on Artificial Intelligence techniques are playing an important role to help medical domain in the prediction of heart disease by examining the various tests on different softwares to prevent the loss in the early stages of heart problem. Since there are various AI techniques present, but in this context a comparative analysis of various techniques have been done and the techniques used are Decision Tree, Neural Networks, Random-forest-means based on the certain characteristics of these techniques. Techniques were trained and tested on the primary dataset having 9 attributes and time taken to build the model. Addition of few parameters like LDL and TG have increased the efficiency of the above mentioned algorithms in the efficient diagnosis of cardiac disease. Here paired t-test that is mean based difference test is used to determine which of the technique is more statistically significant.***

Keywords: Artificial Neural Network, Artificial Intelligent Technique, clustering, Low Density Lipoprotein, Electrocardiogram

I. INTRODUCTION

Problems of heart and its blood vessels give birth to the various heart related challanges called cardio vascular diseases .Death occurs within seconds if heart stops pumping blood[13].Heart disease can be referred with the alternative name that is cardiovascular disease i.e. a disease of heart and the arteries(blood vessels of heart).There are number of factors that have direct impact on the improper functioning of the heart like Family History, Smoking, High LDL, High cholesterol, High Triglycerides, High BP, improper intake of diet,physical inactivity, Hypertension.There are no. of reasons associated with arteries that causes the arteries to narrow and sometimes may block the entire Arteries, one of the reason is deposition of fat , the condition is called Atherosclerosis. Atherosclerosis may affect the left main coronary artery and the right coronary artery that results in a heart attack and sometimes heart stroke .There are different types of heart diseases that affects various organs of the body like congenital Heart disease, this is the defect of the heart present since birth Arrhythmia means abnormal heartbeat sometimes heartbeat may be very fast called tachycardia and when the heartbeat is too slow called Bradycardia generally Arrthemia occurs due to improper pulses generated by the heart. Myocardial infarction also known as coronary thrombosis is a cut in blood flow that destroys the muscles of heart .Hypertrophic cardiomyopathy is a genetic problem associated with the heart so there is more than fifty percent chance of this disease to transmit it to the children .pulmonary stenosis it a condition when heart could not pump blood through the right ventricle to the pulmonary artery. since there are no. of factors present in causing heart disease that do not have the same contribution some factors have greater influence to create the cardiac related issues like 13% of the heart disease occurs due to high blood pressure. Chewing or intake of tobacco causes 6% of death ,5% 0f deaths occur due to obesity[1].In Asian countries, one fifth of lives are lost due to cancer, cardiac problems and diabetes [25].As per statistical records of Australia Bureau 33.7% deaths occur due to cardiac issues [26]. Decisions made by the Doctors or some medical practitioners are based either on experience or some intuitions on analyzing the no. of test reports which may lead to the unnecessary bias [11].

AI plays an integral task in the progression of medical field .Existing AI techniques show different accuracy levels dependingon the attributes ,testing and training methods used to achieve the result [5]. AI techniques not only play an important role in medical domain but in emerging fields like Robotics ,Transportation using Fuzzy Logic .Various techniques like Neural Networks, Decision Tree, Random Forest,k -means clustering play vital role in prediction of heart disease [2]. So keeping in mind the problems faced by the traditional system there is necessity to suggest a system that efficiently diagnose the cardiac disease in lesser time, decreasing traditional errors done by doctors while examining the no. of unrelated tests, and enhancing the patient's satisfaction as livelihood

This paper presents a comparison of Decision tree, Neural Networks, Random Forest, K-means along with the statistical significance test (paired t test).The dataset was collected from the Government Medical College Srinagar and Medicare Diagnostic Centre Srinagar from January to March 2018.Data of 200 patients were collected including both men and women aged above 25.Research was conducted on mainly two tests that is lipid profile test that is the screening of blood including Cholesterol, Triglycerides, Blood Sugar and the ECG that records the electrical impulses of the heart, the rest remaining attributes taken for research do not involve any particular test like Blood Preasure, Age.

The rest remaining part of the paper is organized as Literature Review presented in section 2. Section 3 consists materials and methods .Section 4 presents techniques implemented .Experimental results are presented in section 5 followed by the conclusion and future work in separate sections.

II. LITERATURE REVIEW

Till date most of the research has been done on the Cleveland data that is easily available on the internet and used similar attributes but the only difference was in techniques and the algorithm .In fact accuracy was high in some algorithms but most of the important attributes to heart disease detection were missing that may lead to the inappropriate results and diagnosis when applied in the real environment . Vikas Chaurasia Saurabh Pal (2013) used almost 10 attributes implemented on decision table,CART,ID3 and the accuracy of these techniques were 82.50%,72.93%, 83.49% respectively but the criticism regarding this research is that it lacks three most important attributes like blood pressure ,Triglycerides, LDL and included one of the parameter that is heart rate that may vary depending on the stress of patient ,emotionally disturbed, due to exercise so . A. SankariKarthiga, M. SafishMary(2017) used only two algorithms decision tree and Naïve Bayes, the experiment was conducted on the MATLAB software conclusion of the experiment was that the Decision trees using J48 performed better than Naïve Bayes in the prediction of heart disease.Mai Shouman, Tim Turner, Rob Stocker in (2012) implemented K-Nearest Neighbor on the Cleveland dataand the early investigation shows that it could not perform well but when applied voting improves the efficiency in heart disease prediction.

V. Manikantan , S. Latha (2013) implemented Machine learning techniques that helped in prediction of various diseases like cancer,Heart disease and diabetes using various datasets for the respective diseases not only this most of the death occurs due to cardiac issues in USA and UK. G. Karthiga, C. Preethi, R. Delshi Howsalya Devi (2013) presented Clustering and classification techniques respectively to improve the efficiency of cardiac disease prediction using K means ,MAFIA and the preprocessing of the data was done efficiently in order to provide the better results not only this but accuracy of various techniques also depend on the attributes used .R. Tamilarasi, Dr. R. Porkodi (2015) implemented K nearest ,Naïve Bayes , multilayer Perceptron techniques for the diagnosis of heart disease. Tool used was WEKA the result of this experiment were that the K-Nearest performed better in comparison with the other corresponding techniques. N. Aditya Sundar, P. Pushpa Latha, M. RamaChandra (2012) proposed a CANFIS system that stands for Coactive Neuro Fuzzy Inference system thatinvolved neural networks,fuzzy logic and genetic algorithm in the prediction of cardiac disease .A. Dewan and M. Sharmain (2015)used Back propagation algorithm that is only classification technique that works on nonlinear relationship for the prediction of disease.

Ajinkya Kunjir, HarshalSawant,Nuzhat F Sheikh (2017) worked on multiple datasets for detection of multiple diseases and the techniques to visualize the result were a) Line graph b) Bar graphs c) Pie charts and experimental resultsshow that Naïve Bayes performed better than Decision Tree. Marjia Sultana, Afrin Haider and Mohammad Shorif Uddin (2016) implemented Kstar,MLP,J48,Bayes Net,SMOthe efficiency comparison was done using ROC curve study and the experiment was conducted on weka ; Bayes Net and SMO performed better than other techniques. Salma Banu N.K, Suma Swamy (2016) performed a study of different AI models since 2006 to 2016. In (2006)Hongmei Yan et al used Multi-layer perceptron and the accuracy was 90%. In (2007) HeonGyu Lee et al used Bayesian classification, CMAR, C4.5, SVM and the accuracy were 81%,80%,78%, 85%.In 2013 used Naïve Bayes and the accuracy was 80.7%.Lokanath et al (2015) used GA technique and the efficiency of Genetic algorithm in the prediction of disease was about 90%. Mohammad A. M. Abushariah, Assal et al (2014) implemented ANN and Fuzzy logic experiment was conducted on MATLAB and the result was that ANN performed better than the Fuzzy Logic. Prof. Samiksha H. Zaveri inpaper (2017) highlighted a brief summary of various existing AI Algorithms in the diagnosis of various existing diseases in the medical domain. R. Subha ,K. Ananda kumar, A. Bharathi (2016)provided a description of various AI techniques for the classification of cardiac diseases, classification of such a disease depends on different parameters that are used to conduct the research that are actually the classifiers some act as the strong classifier some act as a weak classifier. Tülay Karay, lanTülay Karay, lanv (2017) implemented Multilayer perceptron in for the presence or absence of cardiac disease, Cleveland data was used and the algorithm was trained with Back propagation algorithm that resulted in the 95% accuracy rate. J. S. Sonawane and D. R. Patil(2014) presented Vector quantization with random incremental training accuracy of the model in the prediction of heart disease was

85.55%.Feature selection approach along with the support vector machine by Hidayet TAKCI (2018) provided the better results as compared to other techniques in the prediction of disease.

So from the above literature it is clear that all the existing AI techniques can perform better based on the data along with the preprocessing. Researchers who work in the medical domain frequently use AI techniques in the prediction of heart disease .Different techniques ended with different accuracy along with different time taken to build the model, but the common point about the previous research was usage of the cleveland data available on UCI repository, problem with the Cleveland data was that preprocessing of data has not been done correctly.so the main drawback of the existing system is in the organization and preprocessing of data .

Like BP consists of two values that is upper value (systolic) and lower value (diastolic) in Cleveland data only systolic value is given and the diastolic value is missing, so how come an algorithm trains the dataset when there is no appropriate format of the data .One of the dataset that was downloaded from UCI repository was “male heart disease data” that too had the same problem mentioned above .

If we talk about the other attribute that is fasting blood sugar it is written there present or absent so it seems somewhat illogical to conclude in the presence or absence of disease .

So by analyzing the previous work and keeping in mind the problems related with that research an efficient system have been suggested that work on 9 attributes in the prediction of cardiac disease .

The different types of cardiac diseases have also been identified based on the defects associated with parameters then on the basis of disease there is a final conclusion whether heart disease is present or not .

If all the parameters are normal excluding high BP then the type of problem is hypertension so on and so.forth.

Data preprocessing has been done very efficiently like the fasting blood sugar has been taken in numeric values rather than specifying in terms of yes or No.

Another parameter that is BP includes here both the corresponding systolic value and the diastolic value .Most importantly LDL parameter and TG have been introduced that has not been introduced before so by including these two parameters we can achieve the higher accuracy and the logical result not only this we can prevent the human loss.

III. MATERIALS AND METHODS

This section consists of the techniques and the tools used to perform the comparison of the techniques for the diagnosis of cardiac problems .The main steps that were followed to achieve the goal is given below.

A. Data Collection

data was collected from the GMC Srinagar having almost 200 patients.

B. Preprocessing

after collecting all the relevant information the data was first sorted into the EXCEL to convert the file into CSV format then from CSV format to the ARFF file that is only the format that wekacan import.so the tools used were EXCEL and WEKA .

C. Selection of The Techniques

this step involves the different AI techniques

D. Testing and Training

percentage split method was used in which 75% of the data was set to train the data and the rest remaining 25% was used to test the model.

E. Analysis of the Summary

finally the result was analyzed with the help of confusion matrix.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue VII, July 2018- Available at www.ijraset.com

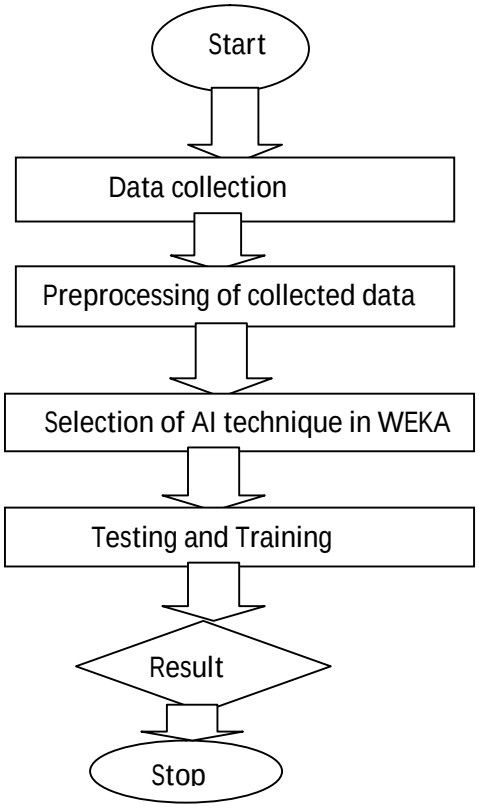


Figure 1.1: Flow chart of the Methodology

IV. TECHNIQUES IMPLEMENTED

A. *Decision Tree*

Decision Tree is a tree like structure which consists ofseveral nodes and each node is attached to the decision tree through the edge that is path connecting two nodes [5].In medical domain the node represents the particular test and the result is represented by the corresponding Edge [4][5]. Root node is also called decision node that has the best predicting capability among all the existing data values then further splitting occurs on various parameters like information gain. An interesting fact about the decision tree is that it works on the numerical data and categorical data, easy to implement and interpret [2].Decision tree follows the top-down approach while constructing tree that means starting from the decision node and reaching to the leaf nodes of the tree. It is a classification technique and with each input data value there is an associated target value that is known as supervised learning [4]. The rules that are followed to build any decision tree is IF-THEN rules and the learning process in decision tree is known as decision treeinduction [2].Algorithms that are used to implement the decision tree are CART, Iterative Dichotomiser 3 usually represented as ID3, C4.5 that is the extension of the ID3, J48 [2].CART stands for classification and regression that works on the high dimensional data. CART involves the use of Gini Index as a portioningparameter [3].ID3 came into existence with the help of Ross Quinlan to represent the decision tree. It is actually the fundamental algorithm in the construction of decision tree, it is based on greedy approach. The two parameters that are involved in the construction of decision tree is entropy and information gain.

1) Entropy: it is used to measure the impurities associated with the data mathematically it is calculated as

$$E= \sum_{i=1}^{c} -p_i \log_2 p_i$$

P_i is the probability of class i

2) *Information Gain:* it is actually a splitting attribute so the attribute with the highest information gain is chosen as the splitting attribute

Major flaw of ID3 was that it does not provide any guarantee to optimalsolution. Another difficulty in implementing ID3 is that it could not perform well with the continuous data values [27]J48 is actually the extension of c4.5 algorithm .Here J48 is used to

implement the decision tree the reason behind the j48 selection was its best performance.J48 has the property of reducing the size of the tree that is called tree pruning [4][5].

B. Artificial Neural Networks

A neural network is a connected set of input/output units where all the connections are associated with weights. Neural networks are associated with the learning phase in which the neuron learns by adjusting its weight in order to provide the better class label this phase is also called as connectionist phase. Structure of artificial neural network resembleswith the structure of the brain. There are no. ofartificial neural network algorithms, most important is back propagationalgorithm that became popular in 1980s.Learning of back propagation algorithm is based on the feed forward neural network. ANN consist of different layers that are categorized as input layerwhich takes the input and the units that constitute the input layer is called input unit, Hidden layer(but not present in all networks) and the output layer which represents the output and the units in both the layers are called neurods.Input range of ANN lies in the range of 0 and 1.Net input weight of neural network is calculated by multiplying each unit with its associated weight then perform the sum operation to calculate the net weight. The output of the node is determined by the activation functions. Neural network can work with continuous value of input output unlike decision trees. The main goal of ANN was to figure out the problems in the same way as humans could do. They find the applications in the real world environment like handwritten recognition, pathology, and medicine .Due to the parallel nature of the neural networks the computational speed is enhanced to a greater extent and find the applications in both the classification as well as in prediction. Neural networks can produce results even on the incomplete information.ANN are highly fault tolerant [11][12] One of the major drawbacks of ANN are difficult interpretability, lack of universal structural representation of the network.

C. Random Forest

Random forests also called as random decision forest are based on ensemble learning for the tasks of classification and regression. Ensemble learning is based on divide and conquer approach used to improve the performance the basic idea behind this learning is that the group of weak learners can come together to form a strong learner As the name suggests this algorithm creates a forest with no. of decision trees In general more no. of trees the more robust is the algorithm and the more it would be accurate. It follows the same approach of constructing the forest as decision trees follow like information gain and Gini index. It is one of the most powerful and efficient algorithm that is based on the supervised learning and is a classification technique. There is a direct relationship between the accuracy and the no. of trees generated so more the number of trees generated higher is the algorithm accuracy. In random forest multiple trees are generated to represent a single tree for the model representation. To classify a new object based on the attributes each tree gives the classification and we say the tree votes for that class. The forest chooses the classification having the most votes of all the other trees in the forest. There are number of advantages of using random forest like it deals with both the classification and regression tasks, it prevents the over fitting of the model, can handle large data sets with high dimensionality. Helps with feature selection based on importance of each variable once we know which variables are important then we can use that information in any other classification or prediction models[9][10]. Random forests are user friendly methodology as it deals with the two parameters only first one is no. of trees represented as ntree, the default value for the no. of trees is 500 and the next parameter is mtry that is variables randomly sampled as candidate s at each split Main drawbacks of Random forests are ;perform good with classification problems but not as good as for regression problems as it does not give the precise continuous nature predictions and it does not predict beyond the range of training data and may over fit dataset that are particularly noisy.

D. K Mean Clustering

The concept of k means came into existence in early 1950s but remain unpopular for a longer period until computing power became available. K means clustering is an exploratory data analysis technique that means it will explore the complete data set, implements nonhierarchical method of grouping objects together that is it will not follow any hierarchical method but it will take the datasets as they are coming and will group them together. K nearest is based on comparing the given tuple with the training tuple and assign the data point on the basis of similarity. Euclidean distance is used to measure the closeness between two data points

E. Steps of k Means

1) Input the no. of clusters that we need to create whether we need to cluster in a size of two or a cluster of 3 so on so forth.
2) After deciding the clusters calculate the centroid
3) After calculating the centroid ,the distance between the two centroids and clusters are created and they are clustered together on the basis of minimum distance

4) The process is repeated until we get the consistency in the cluster size and the consistency in the cluster assignment then we stop or till all the dataset have been completed.

The main disadvantage of using Euclidean distance is that smaller values have lesser influence than the larger ones, so to get the same influence of the values continuous values are normalized first to get the accurate result. All the training data must be present in the memory at the run time so it is also called as memory based classification. It is one of the simplest algorithm that is based on the unsupervised learning .It divides the data into the K clusters. This technique needs k no. of clusters and the training set. The data point is assigned to the cluster that is closer to the particular cluster [5].The main disadvantage of k means here is that it could not perform well as the clusters are of different sizes and densities [14].It deals with continuous as well as discrete values [5].

Dataset

Name	Type	Description
Age	Continous	Age factor varies in men and women ,in men signs become visible in 40 years to 50 years old and in women it is near about 58 to 65 years old
Sex	Nominal	women have higher chances of cardiovascular disease than men one of the basic reason is intake of the birth control tablets that increases the blood pressure
Blood Sugar	Continous	Normal range of blood sugar is between 70-110 and the values that do not lie in this range is considered to be problematic
LDL	Continous	.Normal range of LDL is 25-50.LDL more than 190 is considered to be very high Lower the level of LDL lesser are the chances of getting heart stroke
Triglycerides	Continous	TG less than 150 is considered to be normal,200-499 is considered to be high and TG level more than 500 is considered to be very high
Cholestrol	Continous	Cholesterol level less than 200mg/dl is normal,200-239mg/dl moderate high ,above 240 is considered to be high
BP	Continous	High BP is also referred as hypertension, 120/80 is normal BP.This parameter indicates where the vascular system is functioning properly or not.
Chest Pain	Nominal	An uncomfortable feeling in the chest occurs frequently in the cardiac patients
Resting ECg	Nominal	An electrocardiogram is also known as cardiac imaging test that is used to analyze the defects of heart, also fore-tell about the heart attack.
Diagnosis	Nominal	Present or Absent Heart Disease based on above parameters

Table 1:Primary Datasetof heart disease

IV. EXPEIMETAL RESULT

As already discussed ,the techniques like Decision Tree, Neural Networks ,Random-Forest, k-Means were tested and trained on the below dataset using percentage and split method for first three techniques and classes to cluster evaluation for K-Means. There are no. of parameters associated with each technique some of the important ones are discussed as for Random forest the no. of iterations were taken as 100 and seed for random no. generator equals to 1.In multilayer perceptron the learning rate can range from 0-1 here the value taken was 0.3,momentum rate lie in domain of 0 and 1 here the value for momentum rate was taken as 0.2.Threshold value should be greater than 0 here the value for threshold was taken as 20.the no. of hidden layers were set on 2.For J48 pruning confidence were set on 0.25.No. of folds equals 3.Minimum no. of instances per leaf equals to 3.K-Means was implemented using classes to cluster evaluation .Below are given the experimental performances of each technique.

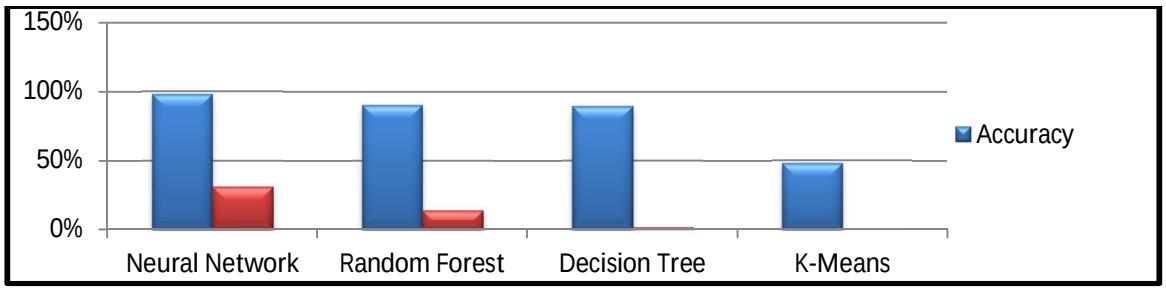


Figure1.7: Bar Graph showing the comparison of techniques

Technique	Testing and Training Method (75% training and 25%)	Accuracy	Time taken to build the model	Statistical Significance
Multilayer perceptron	Percentage split	98%	0.31 sec	Highly significant
Random Forest	Percentage split	96%	0.14 sec	Lesser significant than MLP
Decision Tree	Percentage split	82%	0.02 sec	Lesser Significant than MLP and Random Forest
K Means	Classes to cluster evaluation	54%	0 sec	Least significant

Table 2: Analysis of Techniques

V. CONCLUSION

From the above experiment it is concluded that Multilayer perceptron has the highest accuracy in the prediction of the heart disease that is 98% with time taken by the model 0.31 sec, Random forest provides 96% accuracy, Decision tree has 82% accuracy and K-means accuracy is only 54% not only this the statistical significance test that is paired t-test was conducted on the above four mentioned techniques and the result shows that the Multilayer Perceptron is highly statistical significant than other respective techniques , Random Forest significance is lesser than MLP, Decision Tree is lesser significant than MLP and Random Forest. The statistical significance of K Means is least so K means plays a minor role in the diagnosis of heart disease because of its least significance as well as its least accuracy in comparison with other techniques

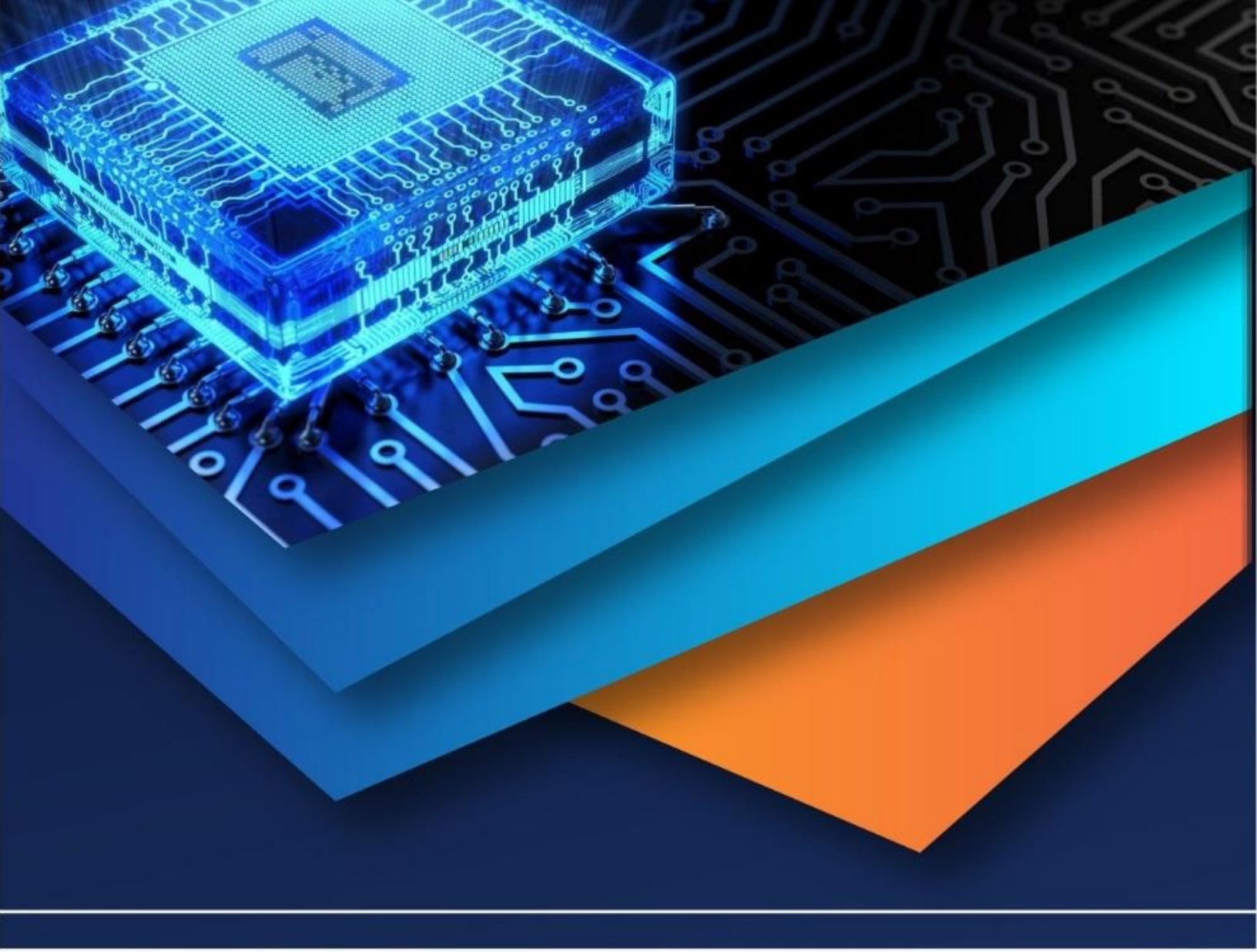
VI. FUTURE WORK

Improvement of K Means clustering in the diagnosis of Cardiac disease.

REFERENCES

[1] E.P.Ephzibah1, Dr. V. Sundarapandian, "A NEURO FUZZY EXPERT SYSTEM FOR HEART DISEASE DIAGNOSIS", an International Journal of Computer Science and Engineering Vol.2, No.1, February 2012.
[2] Ajinkya Kunjir Harshal Sawant, Nuzhat F. Sheikh, "Data Mining and Visualization for prediction of Multiple Disease in Healthcare", International Conference on Big Data Analytics and Computational Intelligence, IEEE 2017.
[3] VikasChaurasiaSaurabh Pal, "Early Prediction of Heart Diseases Using Data Mining Techniques", Caribbean Journal of Science and Technology, 2013, Vol.1, 208-217.
[4] A.Sankari Karthiga, M. Safish Mary,M. Yogasins, "Early Prediction of Heart Disease Using Decision Tree Algorithm", International Journal of Advanced Research in Basic Engineering Sciences and Technology, Vol.3, Issue 3, March 2017
[5] Mai Shouman, TimTurner, Rob Stocker, "Applying K-Nearest Neighbour in Diagnosing Heart Disease patients", International journal of Information and Education Technology,Vol 2,No 3,June 2012.
[6] Prof. Samiksha H. Zaveri, "A Comparative Study of Data Analysis Techniques in the domain of Medicativecare for Disease Predication", International Journal of Advanced Research in Computer Science, Volume 8, No. 3, March – April 2017.
[7] Marjia Sultana, Afrin Haider and Mohammad ShorifUddin, "Analysis of Data Mining Techniques for Heart DiseasePrediction", International Conference on Electrical Engineering and Information and Communication Technology, IEEE 2016.
[8] Salma Banu N.K, Suma Swamy, "Prediction of Heart Disease at early stage using Data Mining and Big Data Analytics: A Survey", 2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques, IEEE 2016.
[9] EeshaGoel, Er.Abhilasha, "Random Forest", International Journal of Advanced Research in Computer science and software Engeneering", Volume 7, Issue 1, January 2017.
[10] Chi Zheng, Jingxin Liu, "Tuberculosis Bacteria Detection based on Random Forest using Fluorescent Images" ,IEEE International Congress on Image and Signal Processing,2016
[11] Mohammad A. M. Abushariah, Assal et al , "Automatic Heart Disease Diagnosis System Based on Artificial Neural Network (ANN) and Adaptive Neuro-Fuzzy Inference Systems (ANFIS) Approaches", Journal of Software Engineering and Applications 2014
[12] Miss.ManjushaB.Wadhonkar, Prof. P. A. Tijare , Prof. S. N. Sawalkar**, "**Classification of Heart Disease Dataset using Multilayer Feed forward backpropogation Algorithm", International Journal of Application or Innovation in Engineering and Management, Volume 2,Issue 4,April 2013
[13] Jairam P. Kelwade, Suresh S. Salankar, "Prediction of heart abnormalities using Particle Swarm optimization in Radial basis function neural network", 2016 International Conference on Automatic Control and Dynamic Optimization Techniques, IEEE 2016.
[14] M.Akhiljabbar, B.L DeekshatuluaPriti Chandra, "Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm" ,International Conference on Computational Intelligence: Modeling Techniques and Applications (CIMTA) 2013

[15] SheenalPatel,Hardik Patel, "SURVEY OF DATA MINING TECHNIQUES USED IN HEALTHCARE DOMAIN", International Journal of Information Sciences and Techniques (IJIST) Vol.6, No.1/2, March 2016.
[16] V. Manikantan, S. Latha, "Predicting the analysis of heart disease symptoms using medicinal data mining methods", International Journal of Advanced Computer Theory and Engineering, vol. 2, pp.46-51, 2013
[17] G.Karthiga1, C.Preethi2, R.Delshi Howsalya Devi3, "Heart Disease Analysis System Using Data Mining Techniques",International Journal of Innovative Research in Science, Engineering and Technology Volume 3, Special Issue 3, March 2014
[18] R. Tamilarasi, Dr. R. Porkodi, "A Study and Analysis of Disease Prediction Techniques in Data Mining for Healthcare",International Journal of Emerging Research in Management &Technology ISSN: 2278-9359 (Volume-4, Issue-3).
[19] N. Aditya Sundar, P. Pushpa Latha, M. Rama Chandra, "performance analysis of classification data mining techniques over heart diseases data base", international journal of engineering science and advanced technology, 2012.
[20] R. Subha, K. Anandakumar,A. Bharathi**,"**Study on Cardiovascular Disease Classification Using Machine Learning Approaches",International Journal of Applied Engineering Research ISSN 0973-4562 Volume 11, Number 6 (2016).
[21] Tülay Karay, lanTülay Karay, lanv, "Prediction of Heart Disease Using Neural Network", 2nd International Conference on Computer Science and Engineering (UBMK, 17), IEEE 2017.
[22] A. Dewan and M. Sharma, "Prediction of heart disease using a hybridtechnique in data mining classification," 2015 2nd International Conferenceon Computing for Sustainable Global Development (INDIA.Com), New Delhi,2015, pp. 704-706.
[23] S. Sonawane and D. R. Patel, "Prediction of heart disease usinglearning vector quantization algorithm," 2014 Conference on IT in Business, Industry and Government (CSIBIG), Indore, 2014, pp. 1-5.doi:10.1109/CSIBIG.2014.7056973.
[24] Hidayet TAKCI, "Improvement of heart attack prediction by the feature selection methods", Turkish Journal of Electrical Engineering & Computer Sciences (2018),doi: 10.3906/elk-1611-235.
[25] ESCAP. (July 2017-Febuary 2011). [Online]. Availablehttp://www.unescap.org/stat/data/syb2009/9.Health-risks-causes-of-death.asp
[26] Australian Bureau of Statistics. (July 2017-Febuary 2011).[Online]. Available:http://www.ausstats.abs.gov.au/Ausstats/subscriber.nsf/0/E8510D1C8DC1AE1CCA2576F600139288/$File/33030_2008.pdf
[27] Quinlan, J.R.1986.Induction of Decision Trees.Mach.Learn.1.1 (Mar.1986), 81-106.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY