iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** I **Month of publication:** January 2018

DOI: http://doi.org/10.22214/ijraset.2018.1202

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 I, January 2018- Available at www.ijraset.com

Support Vector Machine Algorithm Based On Feature Selection For The Heart Disease Diagnosis System

Vinny Sareen[1], Er. Sukhpreet Kaur[2]
[1] *M.Tech (Scholar), Head of Department, Department of Computer Science Engineering, Suscet, Tangori*

Abstract: Current medicine produces a great deal of information deposited in the medical database. Extracting useful knowledge and making scientific decision for diagnosis and treatment of disease from the database increasingly becomes necessary. Data mining in medicine can deal with this problem. It can also recover the management level of hospital information and sponsor the development of tele-medicine and community medicine. The medical field is primarily directed at patient care activity and only secondarily as a research resource. The only justification for collecting medical data is to benefit the individual patient. In data mining there are numerous ways, methods to predict any illness and different researches are still going on. Data mining technique in the history of medical data found with enormous investigations found that the prediction of heart disease is very important in medical science. In the medical past it is observed that the un-structured data as mixed data and it is observed that the data formed with different attributes should be analyzed to predict and provide information for making a diagnosis of a heart patient. Various techniques in Data Mining have been applied to predict the heart disease patients. But, the indecision in the data was not detached with the systems accessible in data mining and implemented by various authors. To remove uncertainty of unstructured data, an attempt was made by introducing selective in the measured data. A membership function was designed and incorporated with the measured value to remove uncertainty and selection data was used to predict the heart disease patients.. Further, an attempt was made to classify the patients based on the attributes collected from the medical field. The maximum distance calculates the based on K-means, genetic and support vector machine classifier was designed to classify the training and testing data belonging to different classes. It was found that genetic and support vector machine classifier suits well as compared with other classifiers of parametric techniques.

Keywords: Data Mining, Medical Dataset, research resources, heart disease detection and classification (SVM).

I. INTRODUCTION

The most of the people will work like a machine to leave a deluxe and comfortable life in the future and to earn more money, but during this type of situation people forgets about their health and even don't take any suitable rest. Because of this they are affected from various types of diseases at a very early age, which cause our health as Diabetes, Cancer, Heart Disease and many more [1]. Now a day's heart disease is a major health problem and cause of death all over the world. The heart is a very valuable part of our body, and plays a very significant role in our life. Our whole life depends on effectual working of the heart. Heart disease is caused due to narrowing or blockage of coronary arteries. This is caused by deposition of fat on the internal walls of arteries & also due to build up cholesterol. There are some of major heart disease factors which include Diabetes, high blood pressure, obesity, smoking, eating habits, alcohol that affects our whole body.

Nowadays, many hospitals keep their present data in electronic form through some hospital database management structure. These systems generate big volumes of data on a daily basis. This data may be in the form of free text, structured as in databases or in the form of images. [3] This data may be used to extract meaning information which may be used for decision making. Data mining (sometimes called data or knowledge discovery) is the process of analyzing and summarizing the data from different perspective & converting these data into valuable information; it plays an important role in the intelligent medical system. The relationships of disorders, real cause of it and the effects of indications that are seen in the patients can be easily and quickly evaluated by the users via well-constructed software with different data mining techniques[2].

In a medical organization like hospitals and medical centers, generates a large amount of data which contains a wealth of unknown information, but these data aren't used correctly. Hence, that unused data can be converted into useful information by using different data mining techniques. In the modern world, cardiovascular diseases are the maximum flying diseases and in every year more than 12 million deaths occur worldwide due to heart problems. Cardiovascular diseases also cause maximum casualties in

©IJRASET (UGC Approved Journal): All Rights are Reserved

India & its diagnosis is a very intricate procedure. Due to the limitation on the potential of the medical experts and their unavailability at certain places, puts the patients at a high risk. Normally, these diseases can be identified using intuition of the medical specialist & it would be highly advantageous if the techniques used for diagnosis will be integrated with the medical information structures. A decision support or computer based information structure which will facilitate accurate diagnosis at reduced cost. This integration of different data mining techniques with existing medical decision support system needs comparison of different data mining methods for extracting the suitable data for the said job.

Classification is one of the most popularly used methods of Data Mining in Healthcare sector. It divides data samples into target classes. The classification method predicts the target class for every data point. With the help of classification approach a risk factor can be associated with patients by analyzing their patterns of diseases. It is a supervised learning method having known class categories. Binary and [4] multilevel is the two methods of classification. In binary classification, only two possible classes such as, "high" or "lower" risk patient may be considered while the multi-class approach has more than two targets, for example, "high", "medium" and "low" risk patient. The Data set is partitioned as training & testing data set. It consists of predicting a sure outcome based on a given input. The training set is the algorithm which consists of a set of attributes in order to predict the consequence. In order to predict the outcome, its efforts to discover the relationship between attributes. Goal or prediction is its outcome. There is another algorithm known as prediction set. It consists of a similar set of attributes as that of the training set. But in forecast set, prediction attribute is yet to be known. In order to process the prediction it mainly analyses the input. The term which describes how "good" the procedure is its accuracy [5].

Table 1: Training And Prediction Sets For Pawti Medical Database

Age	Heart rate	Blood pressure	Heart problem
45	75	140/64	Yes
28	85	101/60	No
38	62	105/55	No

In order to disembosom the knowledge, classification predictes rules. Prediction rules are divulged\\\\ in the form of IF-THEN rules. With the help of the above example, a rule forecasting the first row in the training set may be signified as follows: IF (Age=45 AND Heart rate>75) OR (Age>44 AND Blood pressure>139/60) THEN Heart problem=yes.

II. RELATED WORK

Dbritto, Rovina, et al., 2016 presented Comparative Analysis of Accuracy on Heart Disease Prediction using Classification Methods concern with four algorithms like Naïve Bayes, Decision tree, K nearest neighbor & Support vector machine. They suggested three phases. In that, the size of the data set will be increased by using first two algorithms and predict the heart disease by using logistic regression [6]. Lemke, Frank, et al., 2003 descriptive Medical data analysis using self-organizing data mining technologies‖. In this paper the author constructs the decision tree by analyzing each medical procedure or medically problematic. The problem of review to collect large amounts of data for constructing decision tree [7]. Parthiban, Latha, et al., 2001 [8] presented model on the basis of Coactive Neuro-Fuzzy Inference System (CANFIS) for prediction of heart disease‖. In this model, the author compressing the different methods that include the neural network adaptive abilities, the fuzzy logic qualitative approach and further integrating with genetic algorithm. Li, Wenmin, Jiawei et al., 2001 [9] presented an accurate and efficient classification based on multiple association rules. The proposed a novel method for their classification and named as predominant correlation. The proposed a fast filter technique can identify relevant features devoid of pairwise correlation analysis. Chandra, Priti, et al., 2012 [10] offered Prediction of Risk Score for Heart Disease using Associative Classification & Hybrid Feature Subset Selection. Association

©IJRASET (UGC Approved Journal): All Rights are Reserved

rule process is used to determine relations between attributes of data instances. The classification algorithm is used to forecast the classes in the patient data set.

The feature selection procedure determines attributes which help in predicting heart diseases. Rani, K. Usha [11] proposed Disease Dataset using neural network methods. In this paper the author suggests an artificial neural network that syndicates forward and backward propagation algorithm. The experiment is conducted by single and multilayered neural network models. Parallelism is applied to speed up the learning process at every neuron in all hidden and output layers. Dewan, Ankita et al., 2015 [12] presented the prediction of Heart Disease Using a Hybrid Method in Data Mining Classification .They proposed combination of data mining algorithm for predicting the heart disease. The result of this concludes that the neural network is best amongst all the classification methods for a nonlinear data. They suggested that the Back Propagation algorithm is the best classifier of Artificial Neural Network which is a common method of training. But it has the main drawback of held in a local minimum. Palaniappan, Sellappan et al., [13] presented the results illustrated the uncouth strength of every of the methodologies in comprehending the goal of the quantified mining objectives. IHDPS was capable of responding queries that the traditional decision support systems were not capable of it facilitated the establishing of crucial knowledge, such as patterns, relationships amid medical factors connected with heart disease. IHDPS subsists well -being web-based, user-friendly, reliable, scalable and expandable.

Generally, those huge amounts of messy medical registers have no meaning for operators. Using CBR, a technique which solves a new problem by remembering a previous case and by reusing information and knowledge of that case, CBR goes for those data into useful info that can help to make decision support system for the diagnosis of heart disease. This system can be used to assist doctors and support education for the scholar and postgraduate fresh physicians as a tool to improve the quality of care for the patients. This system can be used as a reference for those students and new doctors. Currently, doctors have problems in determining heart disease in a new patient who does not have existing medical record. Therefore, these data can be used to diagnose heart sickness for new patients who do not have existing medical proceedings. This system is designed to assist doctors and health professionals in determining the diagnosis of patient data. Therefore, this system could assist doctors and health experts to determine the diagnosis and examination of the patient health status. Main Point in this problem as follow [14]:

1) The detection of Heart disease is not more accurate due to the variation of input sample for detection. The system needs to design an accurate detection technique for this.
2) The parameters like FAR and FRR are used to calculate the error rate in the system. Existing systems are having less efficiency to reduce these parameters of the system. It causes an increase in the detection time with high error rate.

III.DESIGN AND IMPLEMENTATION

The study and implement the existing techniques of diagnosis of heart diseases that is genetic algorithm. To propose and implement the K-mean clustering algorithm, feature extraction (PCA) and classification using SVM algorithm.To compare existing and proposed technique using the performance parameters as, false acceptance rate, false rejection rate, accuracy, Specificity and Sensitivity.

1) *Step 1:* To select a dataset in heart disease from the UCI machine learning repository site. In dataset describe the various rows and columns in the heart disease attributes.
2) *Step 2:* First , upload the heart disease dataset shows that the number of attributes in the table form and plotting the histogram in the axes form. The histogram shows that the maximum and minimum intensity of the attribute values.
3) *Step 3:* Applied the K-means clustering algorithm. In k-means clustering approach to divide the heart disease data in clusters form. Like example, we divide the data into the form of clusters with the help of distance measure. We select the index and centroid of the cluster form sets.
4) *Step 4:* Implement the principle component analysis algorithm to extract the algorithm. It identifies the unique properties of the dataset in heart disease attributes. In PCA algorithm to extract the features in the two forms, i.e, Eigen Values and Eigen Vectors.
5) *Step5:* Applied the Genetic algorithm, initialize the population size to search the fit value in the extracted features in the databases.
6) *Step 6 :* Classification **:** Implement the support vector machine for classifying the detected data. The main aim of creating a hyper plane by SVM to separate the data points.

©IJRASET (UGC Approved Journal): All Rights are Reserved

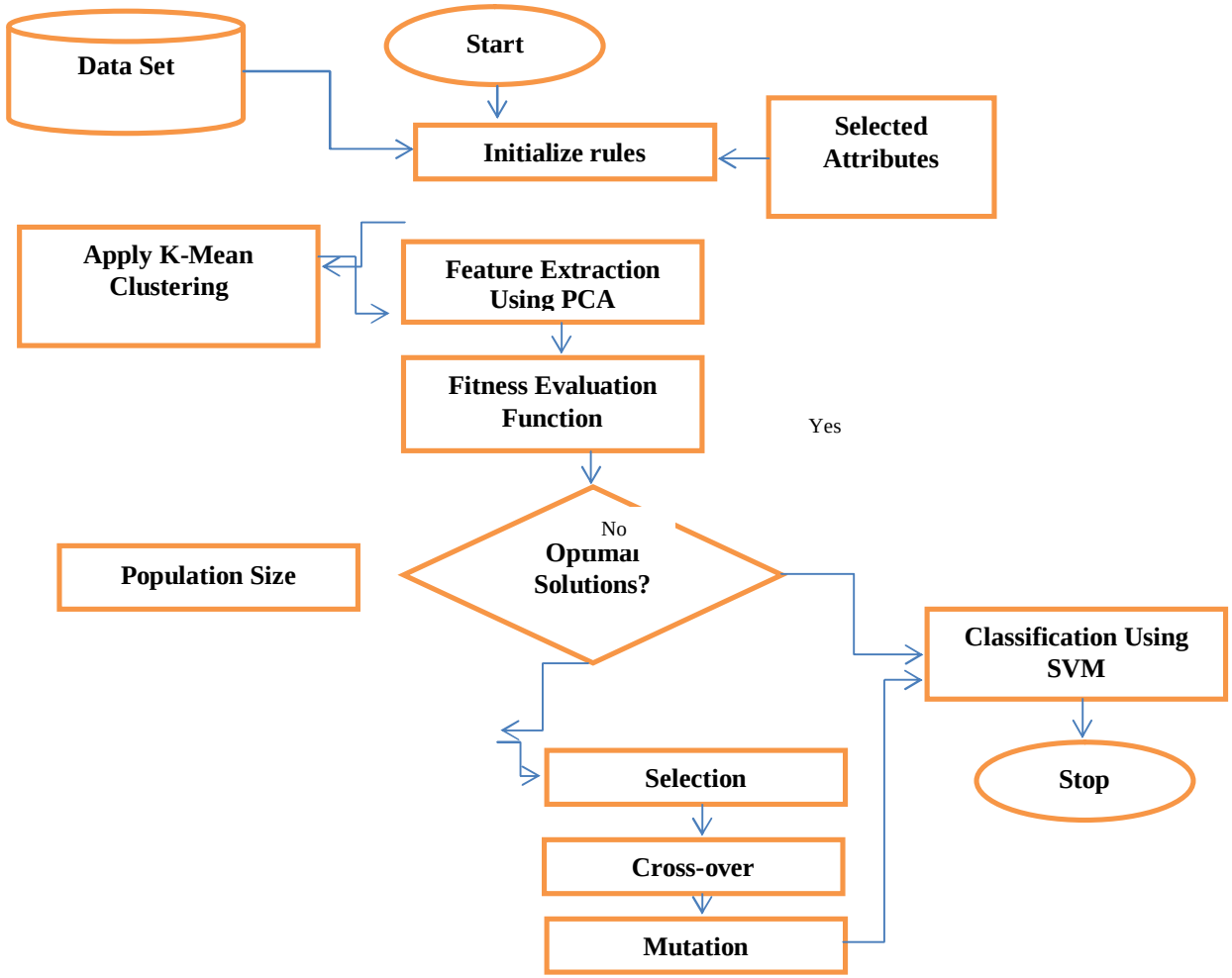


Fig 1. Proposed Flow chart

IV. RESULT AND DISCUSSIONS

The GUI defined that the upload the dataset and FUZZY with GENETIC algorithm implement the rules and detect the disease, detection and calculate the performance parameters in base paper. In the second phase , an implement the proposed model. The performance parameters with genetic algorithm with support vector machine. It evaluates the performance the parameter, i.e accuracy value is 98.65, sensitivity value is 0.99 , specificity value is 0.996 , falser acceptance rate value is 0.00951 and False rejection rate value is 0.003949.

Table 2. Comparison accuracy between proposed and existing approach with (FGA and GSVM).

Base Paper with FGA	Proposed Paper with GSVM
89.5	98.6

©IJRASET (UGC Approved Journal): All Rights are Reserved

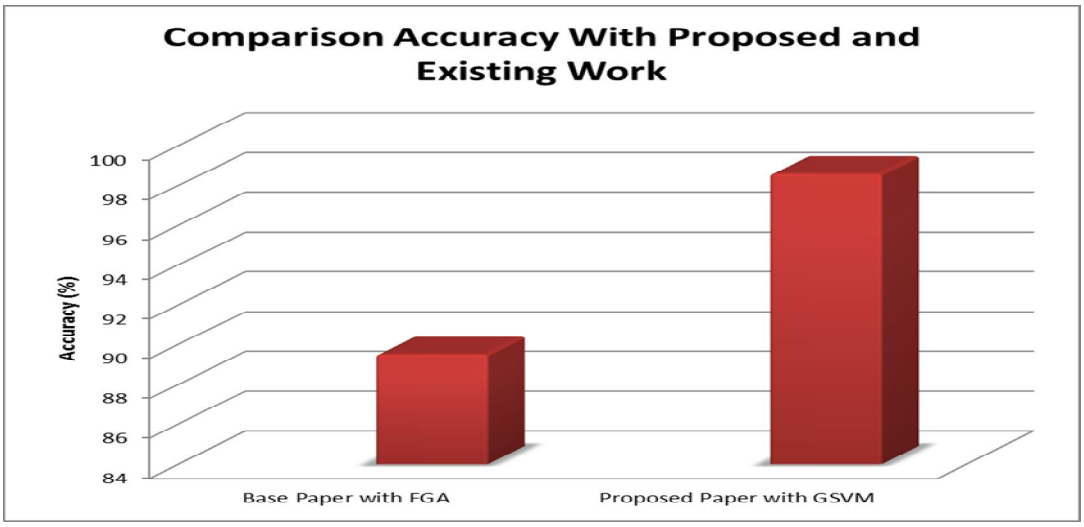


Fig 2. Comparison accuracy with proposed and existing work

The above figure shows that the comparison between accuracy with proposed and existing work.

Table 3. Comparative sensitivity between proposed and existing approach with (FGA and GSVM).

Base Paper with FGA	Proposed Paper with GSVM
0.90	0.97

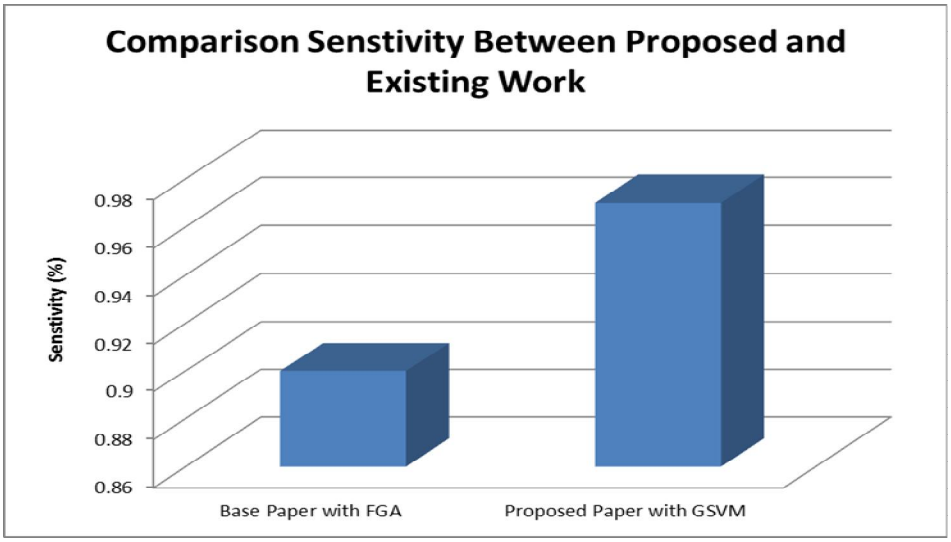


Fig 3. Comparative sensitivity with proposed and existing work

The above figure shows that the comparison between sensitivity with proposed and existing work.

Table 4 . Comparison specificity between proposed and existing approach with (FGA and GSVM).

Base Paper with FGA	Proposed Paper with GSVM
0.996	0.998

©IJRASET (UGC Approved Journal): All Rights are Reserved

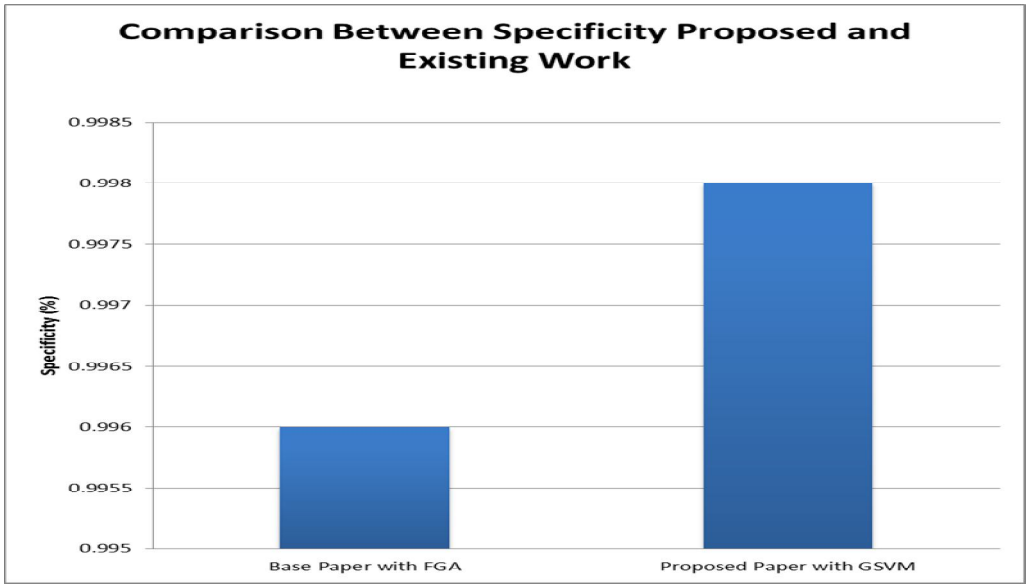


Fig 4. Comparison specificity with proposed and existing work

The above figure shows that the comparison between specificity with proposed and existing work.

V. CONCLUSION AND FUTURE SCOPE

Data mining in medication can deal with this problem. It can also recover the management level of hospital information and encourage the development of tele medicine and community medicine. Medical field is primarily directed at patient care activity and only secondarily as a research resource.

The amount of people feeling sick and receiving admitted into hospitals and clinics are increasing proportionately. The growing number of patients indirectly increases the amount of data that are required to be stored. As the size of data increases, computer storage also increases. Due to the vast amount of data that has been created, humans invented algorithms (FUZZY and GA) algorithm applied in existing work that produce results once a query is supplied. The goals that have been achieved by the developed system are:

1) Simplified and reduced manual work.
2) Large volumes of data can be stored.
3) It provides Smooth workflow.

The overall objective of the research work is to predict the heart disease patients with more accuracy which is useful for health care information systems. To remove the uncertainty of the unstructured data, an attempt has been made with GA+ SVM classifier embedded with Symbolic approach. The present result shows that interval approach in making data as symbolic data found to be successful in providing more accuracy of the system. This work can be enhanced by increasing the number of attributes for the existing system. The symbolic genetic and SVM classifier can be tested with the unstructured data available in health care industry data base by modifying into fuzzified structured data with increased attributes and with a collection of more number of records to provide better accuracy to the system in predicting and diagnosing the patients of heart disease.

Future work, It will use Particle Swarm Optimization using feature subset selection and classification data mining techniques to predict the heart disease.

REFERENCES

[1] Sharma, Mamta, Farheen Khan, and Vishnupriya Ravichandran. "Comparing Data Mining Techniques Used For Heart Disease Prediction." (2017).
[2] Kundra, Divya, and N. Kaur. "Review on prediction system for heart diagnosis using data mining techniques." Int. J. Latest Res. Eng. Technol 1 (2015): 9-14.
[3] Joshi, Sujata, and Mydhili K. Nair. "Prediction of heart disease using classification based data mining techniques." In Computational Intelligence in Data Mining-Volume 2, pp. 503-511. Springer, New Delhi, 2015.
[4] D'Agostino Sr, Ralph B., Scott Grundy, Lisa M. Sullivan, Peter Wilson, and CHD Risk Prediction Group. "Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation." Jama 286, no. 2 (2001): 180-187.
[5] McGregor, Carolyn, Christina Catley, and Andrew James. "A process mining driven framework for clinical guideline improvement in critical care." In Proceedings of the Learning from Medical Data Streams Workshop. Bled, Slovenia (July 2011). 2011.

©IJRASET (UGC Approved Journal): All Rights are Reserved

[6] Dbritto, Rovina, Anuradha Srinivasaraghavan, and Vincy Joseph. "Comparative Analysis of Accuracy on Heart Disease Prediction using Classification Methods." International Journal of Applied Information Systems (IJAIS)–ISSN (2016): 2249-0868.

[7] Lemke, Frank, and Johann-Adolf Mueller. "Medical data analysis using self-organizing data mining technologies." Systems Analysis Modelling Simulation 43, no. 10 (2003): 1399-1408.

[8] Parthiban, Latha, and R. Subramanian. "Intelligent heart disease prediction system using CANFIS and genetic algorithm." International Journal of Biological, Biomedical and Medical Sciences 3, no. 3 (2008).

[9] Li, Wenmin, Jiawei Han, and Jian Pei. "CMAR: Accurate and efficient classification based on multiple class-association rules." In Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on, pp. 369-376. IEEE, 2001.

[10] Chandra, Priti, and B. L. Deekshatulu. "Prediction of risk score for heart disease using associative classification and hybrid feature subset selection." In Intelligent Systems Design and Applications (ISDA), 2012 12th International Conference on, pp. 628-634. IEEE, 2012.

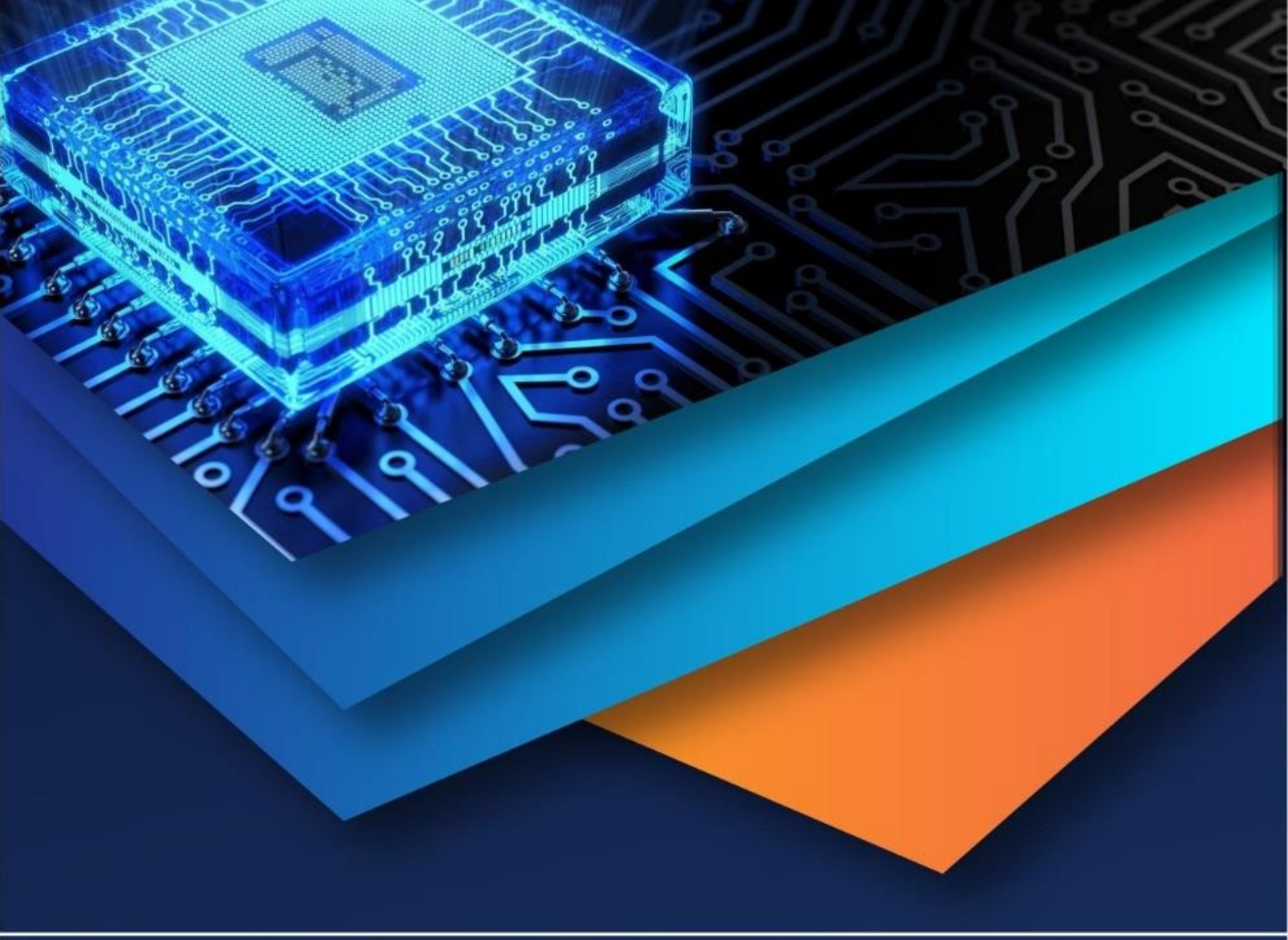
[11] Rani, K. Usha. "Analysis of heart diseases dataset using neural network approach." arXiv preprint arXiv:1110.2626(2011).

[12] Dewan, Ankita, and Meghna Sharma. "Prediction of heart disease using a hybrid technique in data mining classification." In Computing for Sustainable Global Development (INDIACom), 2015 2nd International Conference on, pp. 704-706. IEEE, 2015.

[13] Palaniappan, Sellappan, and Rafiah Awang. "Intelligent heart disease prediction system using data mining techniques." In Computer Systems and Applications, 2008. AICCSA 2008. IEEE/ACS International Conference on, pp. 108-115. IEEE, 2008.

[14] Guru, Niti, Anil Dahiya, and Navin Rajpal. "Decision support system for heart disease diagnosis using neural network." Delhi Business Review 8, no. 1 (2007): 99-101.

©IJRASET (UGC Approved Journal): All Rights are Reserved



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)