Advances in Modelling and Analysis B
Vol. 61, No. 2, June, 2018, pp. 99-105
Journal homepage:http://iieta.org/Journals/AMA/AMA_B

Identification of heart disease using fuzzy neural genetic algorithm with data mining techniques

Srikanth Meda[1*], Raveendra Babu Bhogapathi[2]

[1] Research scholar, Acharya Nagarjuna University, Guntur & Associate Professor in the Department of Computer Science and Engineering at R.V.R. & J.C. College of Engineering, Guntur 522019, India
[2] Professor, Department of Computer Science and Engineering, R.V.R & J.C College of Engineering, Guntur 552019, India

Corresponding Author Email: medasrikanth@gmail.com

Received: April 12 2018
Accepted: June 20 2018

Keywords:
data mining, heart information, fuzzy neural systems, bunching, risk prediction

ABSTRACT

Continuous coordinated efforts between Heart specialists and researchers are taking a step ahead for utilization of information mining strategies to the field of individual patient health determination based on clinical records. In this unique situation, this proposed work introduces the utilization of instance acknowledgment and information mining procedures into identification of heart diseases. The information is demonstrated and ordered by utilizing information mining procedures with some constraints in both supervised and unsupervised learning strategies. These constraints are overcome in this examination by utilization of Fuzzy neural system models which can certainly recognize complex nonlinear connections amongst needy and open factors and the capacity to distinguish every single conceivable cooperation between indicator factors. The proposed work recommends that a selection sensitively supports system for Heart solution that can be assembled using the proposed disease prediction models and characterization methods and can be stretched out for other medicinal spaces. A specific concentration is the utilization of Fuzzy Neural Genetic Algorithm (FNGA) and the utilization of mining techniques to create a novel prescient model for use in the Heart Disease Prediction area.

1. INTRODUCTION

Heart disease is the main cause of death now a days. The World Health Organization has announced that, the primary driving reason for death is high or low blood pressure. The major cause of Heart disease is because of hypertension, over weight, diabetes and rheumatics. Heart disease occurrences are 4.2 -5.3 million, with a yearly frequency of 647,300 to 2.7 million as indicated by the review.

Heart disease discovery is a technique where disease determination should be possible with the historical backdrop of past information, and the consequences of test traits required for the analysis of Heart disease prediction. Heart disease is the main enemy of people in India. Keeping in mind the end goal to build up an effective arrangement for heart disease prediction, distinctive sorts of information mining procedures are included.Additionally extraordinary information mining procedures are utilized with various classifiers to enhance the precision of the proposed framework. This work is intended to help clinicians to experience the conclusion procedure with the expanded rate of exactness.

The most vital point of this paper is to ponder different information mining systems utilized as a part of expectation of Heart disease, by utilizing distinctive information methods. Life is reliant on effective working of heart, since heart is fundamental part of our body. In the event that functionality of heart isn't appropriate, it will influence the other body parts of human, for example, mind, kidney and so on. Heart disease is an infection that effects on the functionality of heart. There are number of elements which expands danger to Heart .The World Health Organization has expected that 18 million deaths happen because of heart disease in the world. Prediction by utilizing information mining strategies gives us precise consequences of disease. Fuzzy Neural Genetic Algorithm(FNGA) can discover and extract enclosed information related with Heart disease from a Heart disease database. It can answer complex inquiries for diagnosing Heart disease and subsequently help medicinal services experts and specialists to settle on clinical choices.

Heart disease types are Heart disorder, Heart Stroke. Stroke is a sort of Heart disease; it is caused by narrowing, blocking, or solidifying of the veins that go to the cerebrum or by hypertension [41-42]. In proposed methods of different authors for the effective prediction of heart disease, performance analysis should be performed on those methods. A specific concentration is the utilization of fuzzy neural systems and the utilization of mining techniques to create a novel perceptive model for use in the Heart Disease prediction Area.

2. RELATED WORK

Heart disease is a stage that has numerous conditions related to heart. These clinical conditions clarify the abnormal health conditions that straightly impact the heart and all its parts. Some of papers specified after Numerous works in writing with related Heart disease finding information mining procedures have inspired our work. A concise writing overview is introduced here.

Beant Kaur, et. al. [13] projected out a work, "Utilization of Modular Neural Network for Heart Disease".This paper for the most part concentrated about the analysis of the Heart disease.

For the most part two sorts of the analysis strategy are utilized, one is manual and other is programmed finding which comprises of conclusion of problem with the assistance of shrewd master framework and measured neural system is utilized to determine the Heart disease. The properties are isolated and given to two neural system models Back propagation Neural Network (BPNN) and Radial Basis Function Neural Network (RBFNN) for preparing and testing.

Ben-Gal I, et. al. [14] suggested, "Enhanced Study of Heart Disease Prediction System utilizing Data Mining Classification Techniques". This paper has predicted frameworks for Heart disease utilizing more number of info traits. The framework utilizes medicinal terms, for example, sex, circulatory strain, cholesterol like 13 ascribes to foresee the likelihood of patient getting a Heart disease. In this examination work included two more properties i.e. heftiness and smoking. The information mining characterization methods, in particular Decision Trees, Naïve Bayes, and Neural Networks are examined on Heart disease database. The execution of these strategies is analyzed in view of exactness.

Chih Fong Tsai, et. al. [15] proposed, "An Efficient Classification Tree Technique for Heart Disease Prediction". This paper breaks down the strategies in information mining. The point of this paper is to look at the test after effects of the execution of various mining methods for a Heart disease dataset. The order tree calculations utilized and tried in this work are Decision tree, Random Forest, and LMT Tree calculation. Similar examination is finished by utilizing Waikato Environment for Knowledge Analysis (WEKA). It is open source programming which comprises of a gathering of machine learning calculations for information mining methods.

With the appearance of Association Rule, the Heart disease is anticipated by Deekshatulu.B.L et al [19]. In the information distribution center, the clinical records of heart patients are kept up.The information from information distribution center is gathered and preprocessed keeping in mind the end goal to productively mine the procedure. The part of preprocessing is to clean the information. At that point, the Equal-Interval binning approach is connected to get the precise incentive over the Pima Indian Heart assault records. Each record is dealt with as Item. The help and certainty esteem for each record is assessed. A client characterized limit is set over the continuous examples, in particular, the certainty esteem is experienced.

Dhivya. S, et al. [23] surrounded k-mean grouping with the decision based tree framework to foresee the heart issues. They suggested various centroid choice methodologies for K-way grouping to build the effectivity. The 13 input qualities have been assembled from Cleveland medicinal foundation of coronary heart issue learning set. The affectability, specificity, and exactness are computed with particular starting centroids assurance strategies and unmistakable quantities of clusters. The 10 cycles were directed, and at each level, the normal and best esteem are assessed. Moreover, coordinating k-mean grouping and choice tree may accomplish more prominent exactness than the paging calculation in the examination of Heart disease patients. The precision completed was once 83.9% through the empowering influence process with two groups.

Georgeena.S et al. [27] proposed method to arrange the Cerebrovascular issue by method for using fake neural system with Back spread blunder approach. The Multi-layer perceptrons engineered neural systems with Back-proliferation blunder process were bolster forward nets with at least one layer of hubs between the entrance and engress hubs. These further layers contain shrouded models or hubs that were not quickly associated with every entrance and engress hubs. Prescient models had been used in kind of areas for the conclusion. Dataset are gathered with 100 records from government medicinal fields. The information esteems purchased from the records of the assortments, the info factors in the entrance layer with 16 hubs. The neural system weights have been instated arbitrarily. This work scope of the weights used to be between [-0.5 and 0.5] and the preparation cost used to be set in the vicinity of 0.1 and 0.9.

Guijun Chen et al. Confined an enhanced Prediction of Heart disease with work Subset determination utilizing Genetic Algorithm [28]. At initial 13 attributes, engaged with forecast of coronary heart issue, proposed prevalent expectation of heart affliction with include subset determination using Genetic Algorithm utilizing 10 characteristics for anticipating and data mining techniques subsequent to consolidating the Feature subset choice in higher time utilization. Order strategies are Naïve Bayes, Decision Tree and Classification by methods for grouping.

3. PROPOSED METHOD

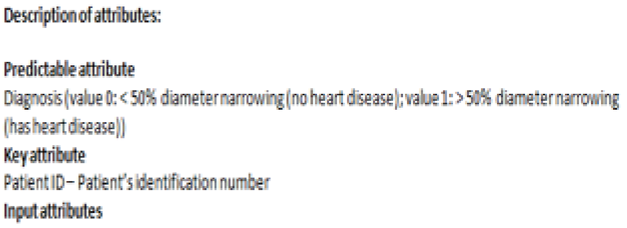
The Fuzzy [30, 33] validation of parameters of anomalous information smoothly decide the anomalies in the extensive size database. The presentation of fuzzy set idea effectively manages the vulnerability issues. The participation work in fuzzy sets achieves the qualities inside the threshold limit [0, 1].

The restorative information gathered from various instruments is indeterminate in nature. The information vulnerability causes the learning flaws to be specific, ambiguous and incoherent. Harsh set hypothesis is utilized to characterize the information. It is proposed to actualize a classifier utilizing fuzzy and unpleasant set for a powerful anomaly arrangement with less computational intricacy. The fuzzy validation use strange information effectively to decide the anomalies in the expansive size database [15].

4. ROUGH SET APPROACH USING FUZZY TECHNIQUE

The distribution of qualities makes the conveyance calculations more delicate. The exceptions show in the information may misinform the preparation stage in the consequence of information mining calculations. In this manner, powerful exception recognition is finished by joining the fuzzy and unpleasant set idea in building a classifier. The fuzzy arrangement is utilized as the preprocessing advancement to group the information into lower, medium and higher classes.

The attributes used for identification of heart disease are



Description of attributes:

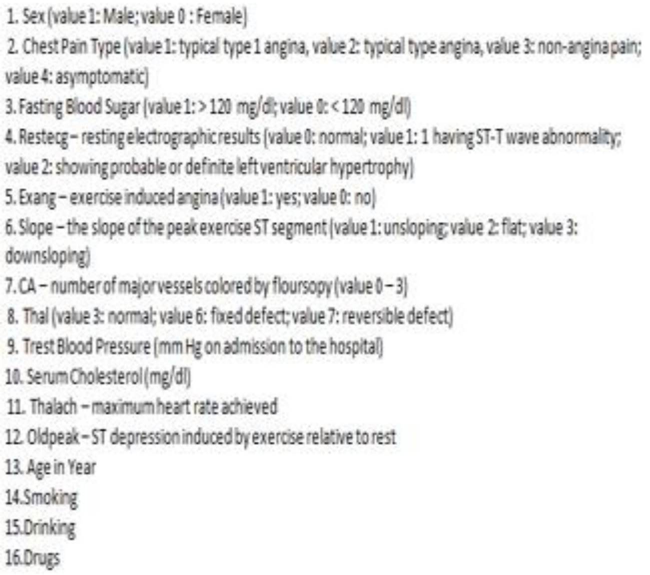
Predictable attribute

Diagnosis (value 0: < 50% diameter narrowing (no heart disease); value 1: > 50% diameter narrowing (has heart disease))

Key attribute

Patient ID – Patient's identification number

Input attributes



1. Sex (value 1: Male; value 0 : Female)
2. Chest Pain Type (value 1: typical type 1 angina, value 2: typical type angina, value 3: non-angina pain; value 4: asymptomatic)
3. Fasting Blood Sugar (value 1: > 120 mg/dl; value 0: < 120 mg/dl)
4. Restecg – resting electrographic results (value 0: normal; value 1: 1 having ST-T wave abnormality; value 2: showing probable or definite left ventricular hypertrophy)
5. Exang – exercise induced angina (value 1: yes; value 0: no)
6. Slope – the slope of the peak exercise ST segment (value 1: unsloping; value 2: flat; value 3: downsloping)
7. CA – number of major vessels colored by floursopy (value 0 – 3)
8. Thal (value 3: normal; value 6: fixed defect; value 7: reversible defect)
9. Trest Blood Pressure (mm Hg on admission to the hospital)
10. Serum Cholesterol (mg/dl)
11. Thalach – maximum heart rate achieved
12. Oldpeak – ST depression induced by exercise relative to rest
13. Age in Year
14. Smoking
15. Drinking
16. Drugs

The half breed fuzzy and harsh set hypothesis based information collection diminishes the computational complexities and enhances the exception location capacity. The novel assertion of set by utilizing their components inside the limit [0, 1] determines a component either inside or outside is thought about rather than vulnerability. This work investigates the utilization of fuzzy validation in the characterization of information and recommends a technique that can decide anomalies in the heart informational index and foresee them by settling the unusual information.

In the testing stage the test dataset is gathered and fuzzy characterization is connected to sort the information into low, medium and higher. At that point the individual probabilities of every attribute are figured to locate the general likelihood of each property. In light of the general likelihood, the records are characterized. It is then checked with the limit.

On the off chance that the value is not as much as the limit the qualities are proclaimed as the generally safe records. If not they are recorded as high risk records. The proposed FNGA is contrasted and the current characterization techniques with respect to the accompanying execution parameters: Average Running time, Average execution time, Execution Time, False Negative, False Positive, True Negative, True Positive, Precision, Recall, and Accuracy.

5. PREPROCESSING USING FUZZY METHOD

The fuzzy technique is utilized as the preprocessing advance to order the information into lower, medium and higher classes. The information is changed over into semantic factors utilizing the enrollment capacities characterized. The UCI Heart disease informational index is considered.

The dataset comprises of 18 test properties. The 18 test qualities are considered as info factors and unique class mark is considered as output variable. From the 18 properties the 8 qualities are allocated with the fuzzy variable utilizing the participation capacities characterized underneath. Utilizing the enrollment capacities the preparation information are changed over into fuzzy phonetic factors.

5.1 AGE member functions

In the age characteristic the fuzzy phonetic factors like child, adult, old and exceptionally old are utilized. The child and exceptionally old are set as trapezoidal and the adult and old are set as triangular capacity as demonstrated as follows.

Attribute	**Range**
Age	<35
	33-45
	40-59
	55>

$$\mu_{young}(a) = \begin{cases} 1 & a < 30 \\ \dfrac{(35-a)}{5} & 30 \le a \le 35 \end{cases}$$

$$\mu_{middle}(a) = \begin{cases} \dfrac{(a-33)}{7} & 33 \le a < 40 \\ 1 & a = 36 \\ \dfrac{(45-a)}{9} & 36 \le a \le 45 \end{cases}$$

$$\mu_{old}(a) = \begin{cases} \dfrac{(a-40)}{5} & 40 \le a < 45 \\ 1 & a = 45 \\ \dfrac{(59-a)}{6} & 45 \le a \le 59 \end{cases}$$

$$\mu_{veryold}(a) = \begin{cases} \dfrac{(a-55)}{20} & 55 \le a < 65 \\ 1 & a \ge 65 \end{cases}$$

5.2 Proposed algorithm

The preparation dataset is given as the info. After this progression, the likelihood is figured for the individual traits, which is conveyed to discover the likelihood for every one of the properties. At that point apply the harsh set order to arrange the information into lower and upper approximations. In the wake of finding the records in the limit area the touchy or basic characteristics are removed. With the after effects of the delicate characteristics the constraints are separated which are connected to the testing information for the expectation of the heart disease data repository.

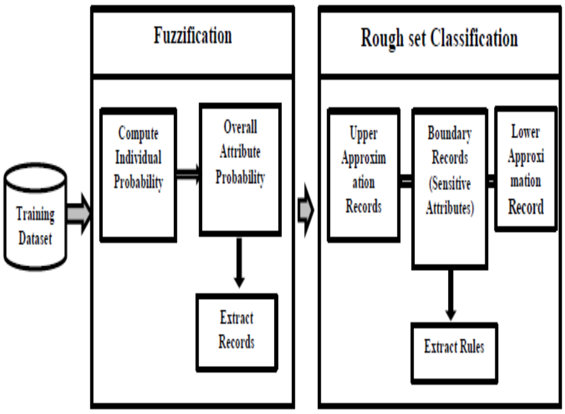


Figure 1. Training phase

```
Algorithm
Input: $D_{tr}$ – Training dataset
Output: R- Rules
Step 1: Extract training dataset and load into database
Step 2: Separate the numerical values as records (x) and attributes (a) in the dataset
        $D_{tr}(x)=\{D_{tr}(x1),D_{tr}(x2),\dots\dots D_{tr}(x_n)\}$
        $D_{tr}(a)=\{pid,age(a),gender(g)heart\ rate(hr)\}$
Step 3: Select the threshold value ($th_{hr}$) for different ages
```

Age a	Heart Rate Threshold Limit th_{hr}	Age a	Heart Rate Threshold Limit th_{hr}
$a\leq 20$	200	$40<a\leq 45$	175
$20<a\leq 25$	195	$45<a\leq 50$	170
$25<a\leq 30$	190	$50<a\leq 55$	165
$30<a\leq 35$	185	$55<a\leq 60$	160
$35<a\leq 40$	180	$60<a\leq 65$	155

```
Step 4: Split the stage based on the membership functions of age, heart rate, blood pressure and cholesterol.
Step 5: Compute the individual probability ($P(x_i)$)
$H(a_i) = \text{maxvalue } [D_{tr}(a_i)]$
$V(a_i) = \text{Currentvalue}[D_{tr}(a_i)]$
$P_{ind}(a_i) = H(a_i) / V(a_i)$
$P(x_i) = \{P_{ind}(a1), P_{ind}(a2)\dots\dots P_{ind}(a_n)\}$
Step 6: Compute the overall probability ($P_{ov}(x_i)$)
$P_{ov}(x_i) = 1/n \sum_{i=1}^{n} P(x_i)$
Step 7: Getting lower limit and upper limit of overall probability $P_{ov}(x_i)$ for each status
$L_1 \rightarrow$ Lower Limit for Status 1
$U_1 \rightarrow$ Upper Limit for Status 1
$L_2 \rightarrow$ Lower Limit for Status 2
$U_2 \rightarrow$ Upper Limit for Status 2
Step 8: Perform Rough set classification
Lower Approximate Region $R_l(x) \rightarrow L_1 \leq R(x) < L_2$
Upper Approximate Region $R_u(x) \rightarrow U_1 < R(x) \leq U2$
Boundary Region $R_b(x) \rightarrow R_l(x) \cap R_u(x) \rightarrow L_2 \leq R(x) \leq U_1$
Step 9: Getting sensitive Attributes from $R_l$ and $R_u$ records
$aL_l \rightarrow$ frequency Count $[R_l(a)]$
$aL_u \rightarrow$ frequency Count $[R_u(a)]$
$S_{attri} \rightarrow aL_l \cap aL_u$
Step 10: Frame rules based on sensitive attributes $S_{attri}$
$S_{attri}(a) = S_{t1}, S_{t2}\dots\dots S_{tN}$ and value of each attribute is $\{1,2,3\}$
Assigned set of rules are $r=\{r_1,r_2,\dots.r_Z\}$
```

$S_{attri}(a)$	S_{t1}	S_{t2}	…………………	S_{tN}
r_1	1	1	…………………	1
r_2	1	1	…………………	2
r_3	1	1	…………………	3
….				
….				
…..				
r_Z	3	3	…………………	3

```
Step 11: Calculate index for the result of fuzzy formulation
Set the value -1 for the attribute status $a_{status}$ =2
For each $r_i$
    $a_s \rightarrow$ match (sum of $a_{status}$ from $R_b(x),r_i$)
    if $(a_s > 0)$ $a_s \rightarrow 1$
    if $(a_s = 0)$ $a_s \rightarrow 2$
    else $a_s \rightarrow 3$
    $RU_i \rightarrow r_i, a_s$
Step 12: Output the fuzzy rule set
$R = \{RU_1, RU_2, RU_3, \dots \dots RU_i\}$
```

The proposed algorithm effectively identifies heart disease with the parameters considered in the work.

5.3 Testing method

The fuzzy order is connected to classify the information into low, medium and higher. At that point the individual probabilities of every parameter ascribe is processed to locate the general likelihood of each property threshold limit. In light of the general likelihood the records are grouped. It is then checked with the limit. In the event that the value is not as much as the edge the qualities are proclaimed as the generally safe records. If not they are recorded as high risk records. These records are thusly checked with the limit locales. It is being checked with the standards created. Lastly, the principles are classified as low, direct and high risk records.

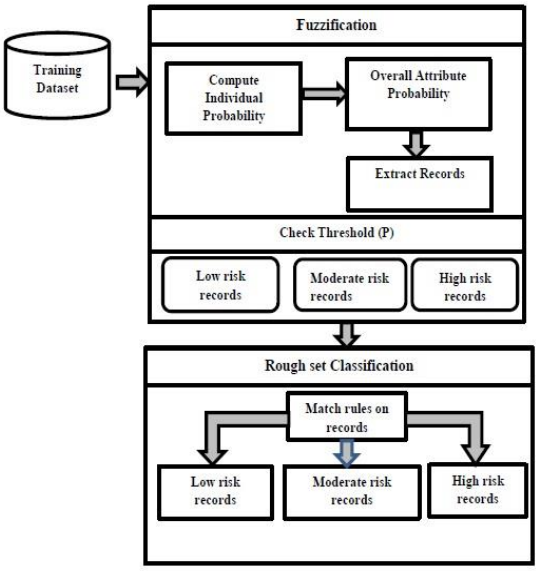


Figure 2. Testing phase

6. EXPERIMENTAL RESULTS

6.1 Average under curve

The AUC is a metric utilized for assessing a double classifier. The likelihood arbitrarily chose positive examples which will be positioned higher than a arbitrarily chosen negative example. This measure is likewise superior to the exactness of disease prediction. The normal under curve esteems contrasted and the current techniques is appeared in Fig 3 and the qualities are classified in Table 1.

$$AUC = \frac{S_0 - n_0(n_0+1)/2}{n_0 n_1}$$

where, S_0 - The sum of rank of abnormal instances prediction

n_0 - The number of test instances belongs to normal

n_1 - The number of instances belongs to the outliers

Table 1. Average AUC

Average AUC							
% of Noise	NSVDD	SSLM	R-SVDD	N-SVDD	DW-SVDD	SVDD	FNGA
0	0.812	0.745	0.704	0.696	0.691	0.691	0.844
10	0.81	0.736	0.695	0.695	0.69	0.688	0.823
20	0.793	0.728	0.694	0.694	0.675	0.675	0.8
30	0.776	0.713	0.69	0.687	0.674	0.674	0.729
40	0.767	0.705	0.68	0.675	0.673	0.67	0.76

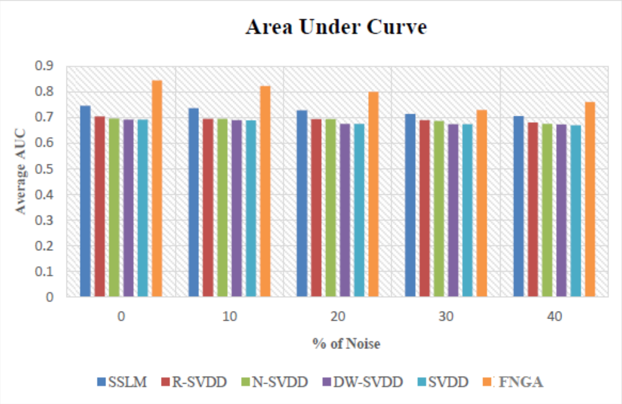


Figure 3. Average area under curve

6.2 ROC (Receiver Operating Characteristics)

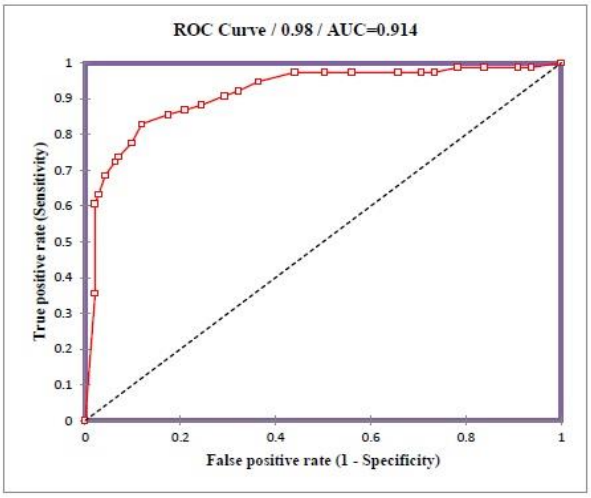


Figure 4. ROC curve

The classifier execution can be measure with the ROC curve with its whole working reach. This curve can be utilized to choose the edge in contrasting the execution of the classifier and the other existing classifiers. It limits the false positive and limits the false positive rates. It is the plot between the affectability and the false positive rate. As per the investigation directed with the Heart disease informational index the AUC ROC is delivered as appeared in Fig. 4.4.

6.3 Accuracy

The precision of the classifier depends on the quantity of example that is accurately ordered. The precision of the technique is appeared in the Fig. 5 and values are classified in the Table 2.

Table 2. Accuracy

Accuracy	
Technique	Percentage
Support Vector Machine	84
Naïve Bayes	86.5
Decision Tree	87.2
Fuzzy C Means	93
FNGA	98.6

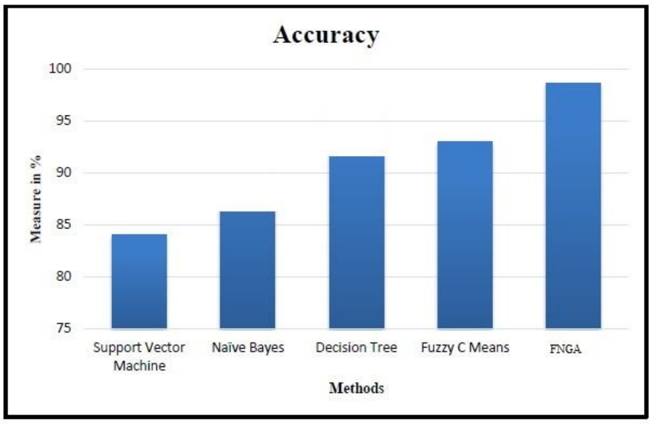


Figure 5. Accuracy

The restorative database contains extensive measure of information identified with the patient's medicinal conditions.This information contains the anomaly called as information commotion or quality clamor. These anomalies are distinguished and are evacuated utilizing the fuzzy and unpleasant set hypothesis. They are the numerical instrument to manage choice framework in giving the questionable and uncertain choices. The Efficient Fuzzy Neural Genetic Algorithm for a compelling anomaly characterization with less computational intricacy is examined. The proposed FNGA is contrasted with current characterization techniques with respect to the execution parameters like Average Running time, Average execution time, Execution Time, False Negative, False Positive, True Negative, True Positive, Precision, Recall, and Accuracy. Further to enhance the adequacy of the classifier, the trait decrease procedure is utilized to expel the unessential properties.

7. CONCLUSION

In this paper we proposed an algorithm for effective Heart disease prediction utilizing FNGA and different element choice strategies for Andhra Pradesh Population. We connected distinctive component choice techniques to rank the properties which contribute more towards order of Heart disease, which in a round about way diminishes the no. of determination tests to be taken. Our exploratory outcomes demonstrate that FNGA and highlight subsets choice gives better prediction of heart disease. Our proposed technique eradicates pointless and distortive information. This exploration will contribute solid and speedier programmed Heart disease identification framework, where simple determination of Heart disease will save lives. Heart disease prediction can be taken care effectively if more research is empowered here. The proposed method identifies heart diseases accurately than the traditional methods.

REFERENCES

[1] Adidela Daveedu R, Navya Sri M, Lavanya Devi G. (2016). Attribute subset selection by mixed weighting mean classification method. Proceedings of the International Conference on Electrical, Electronics, and Optimization Techniques 3523-3531.

[2] Kibler AD, Albert MK. (1991). Instance-based learning algorithms. Machine Learning 6(1): 37–66.

[3] Ahmad TA, Bouaynaya N, Polikar R. (2015). Inductive learning based on rough set theory for medical decision making. Fuzzy Systems (FUZZ-IEEE).

[4] Akhil JM, Chandra P, Deekshatulu BL. (2013). Classification of heart disease using artificial neural network and feature subset selection. Global Journal of Computer Science and Technology 13(3): 5-14.

[5] Akhil Jabbar M, Chandra P, Deekshatulu BL. (2012). Heart disease prediction system using associative classification and genetic algorithm. Proceedings of the International Conference on Emerging trends in Electrical, Electronics and Communication Technologies- ICECIT.

[6] Adeli A, Neshat M. (2010). A fuzzy expert system for heart disease diagnosis. Proceedings of the International Multiconference of Engineers and Computer Scientists I.

[7] Mahdavi-Shahri A, Mahboobeh H, Yaghoobi M, Jalali M (2016). Applying an ensemble learning method for improving multi-label classification performance. ICSPIS.

[8] Anbarasi M. (2010). Enhanced prediction of heart disease with feature subset selection using genetic algorithm. International Journal of Engineering Science and Technology 2(10): 5370-5376.

[9] Asha Gowda Karegowda ASM, Jayaram MA. (2010). Comparative study of attribute selection using gain ratio and correlation based feature selection. International Journal of Information Technology and Knowledge Management 2(2): 271-277.

[10] Jabar AL, Rashid TA. (2015). Combining fuzzy rough set with salient features for hrm classification. Proceedings of the International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing 244-251.

[11] Fawzy A, Mokhtar HMO, Hegazy O. (2013). Outliers detection and classification in wireless sensor networks. Egyptian Informatics Journal 14(2): 157-164.

[12] Baldwin. JF. (1978). Fuzzy logic and fuzzy reasoning. International Journal of Man-Machine Studies 11: 465-480.

[13] Beant K, Williamjeet S. (2014). Review on heart disease prediction system using data mining techniques. International Journal on Recent and Innovation Trends in Computing and Communication 2(10): 3003-3008.

[14] Ben-Gal I, Maimon O, Rockach L. (2005). Data mining and knowledge discovery handbook: a complete guide for practitioners and researchers. Kluwer Academic Publishers, ISBN 0-387-24435-2.

[15] Fong Tsai C, Lin CY. (2010). A triangle based nearest neighbor approach to intrusion detection. Pattern Recognition 43: 222-229.

[16] Chitra R, Seenivasagam V. (2013). Review of heart disease prediction system using data mining and hybrid intelligent techniques. Journal on Soft Computing 3(4): 605-609.

[17] Peña D, Prieto FJ. (2001). Multivariate outlier detection and robust covariance matrix estimation. Technometrics 43(3): 288-310.

[18] Sathiaraj D, Triantaphyllou E. (2013). On identifying critical nuggets of information during classification tasks. IEEE Transactions on Knowledge and Data Engineering 25(6): 1354-1367.

[19] Deekshatulu.BL, Akhil Jabbar M, Chandra P. (2013). Classification of heart disease using artificial neural network. Global Journal of Computer Science and Technology Neural & Artificial Intelligence 13(3): 5-10.

[20] Dhanwani D, Wadhe A. (2013). Study of hybrid genetic algorithm using artificial neural networks in data mining for the diagnosis of stroke disease. International Journal of Computational Engineering Research 3(4): 95-100.

[21] Chandna D. (2014). Diagnosis of heart disease using data mining algorithm. International Journal of Computer Science and Information Technologies 5(2): 1678-1680.

[22] Moorthi DKM. (2014). A novel approach for identifying critical nuggets using PCA with ANOVA. International Journal of Computer Science and Information Technology Research 2(4): 271-278.

[23] Dhivya S, Merlin Mercy E (2016). Heart disease classification and its co-morbid condition detection using WPCA genetic algorithm. International Journal of Multidisciplinary Research and Development 3(4): 287-291.

[24] Chao DF, Peng T, Snooke N, Shen Q. (2014). Feature selection inspired classifier ensemble reduction. IEEE Transactions on Cybernetics 44(8): 1259–1268.

[25] Armento E, Vluymans S, Verbiest N, Caballero Y, Bello R, Cornelis C, Herrera F. (2014). IFROWANN: Imbalanced fuzzy- rough ordered weighted average nearest neighbor classification. IEEE Transactions on Fuzzy Systems 23(5): 1622–1637.

[26] Filzmoser P. (2005). Identification of multivariate outliers: A performance study. Austrian Journal of Statistics 34(2): 127–138.

[27] Thomas GS, Budhkar SS, Cheulkar SK, Choudhary AB, Rohan S. (2015). Heart disease diagnosis system using apriori algorithm. International Journal of Advanced Research in Computer Science and Software Engineering

5(2): 430– 432.

[28] Chen GJ, Zhang XY, Wang ZZJ, Li FL. (2015). Robust support vector data description for outlier detection with noise or uncertain data. Knowledge-Based Systems, Elsevier 90: 129-137.