International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163
Issue 03, Volume 5 (March 2018) www.ijirae.com

GI-ANFIS APPROACH FOR ENVISAGE HEART ATTACK DISEASE USING DATA MINING TECHNIQUES

D.Richard*
Dept of Computer Science, Research Scholar, Bharathiar University, India
revanrichard@gmail.com;

Dr.R.Mala
Dept of Computer Science, Alagappa University, India

Manuscript History
Number: **IJIRAE/RS/Vol.05/Issue03/MRAE10092**
Received: 22, February 2018
Received: 07, March 2018
Final Correction: 11, March 2018
Final Accepted: 15, March 2018
Published: **March 2018**
Citation: D.Richard & Dr.R.Mala (2018). GI-ANFIS APPROACH FOR ENVISAGE HEART ATTACK DISEASE USING DATA MINING TECHNIQUES. IJIRAE::International Journal of Innovative Research in Advanced Engineering, Volume V, 106-111. **doi://10.26562/IJIRAE.2018.MRAE10092**
Editor: Dr.A.Arul L.S, Chief Editor, IJIRAE, AM Publications, India

Abstract— The process of selecting a subset of relevant features from the feature space for use in model construction and used to carry out the feature selection process is called as pre processing step. The filter approach computationally fast and given accuracy results. The Professional Medical Conduct Board Actions data consist of all public actions taken against physicians, physician assistants, specialist assistants, and medical professional. The Classification and Regression Trees (CART), which described the generation of binary decision trees CART were invented independently of one another at around the same time, yet follow a similar approach for learning decision trees from training tuples. The research used GI-ANFIS is used to data mining technique on heart data sets to provide the diagnosis results.

Keywords— Feature selection; Filter approach; Decision Tree Induction; CART; Gini Index; ANFIS;

I. INTRODUCTION

The data mining knowledge form large of data and cleaning to remove noise and inconsistent data , integration where multiple data sources may be combined to selection where data relevant to the analysis task are retrieved from the database of transformation where data are transformed or consolidated into forms appropriate for mining by performing summary or aggregation operations, for instance Data mining an essential process where intelligent methods are applied in order to extract data patterns Pattern evaluation to identify the truly interesting patterns representing knowledge based on some interestingness measures[1].

The feature selection is known as variable and selection, attribute selection or variable subset selection to process of selecting a subset of relevant features from the feature space for use in model construction and reducing the features from the feature space to a manageable size for processing and analysis [2].

The filer approach to independent of the classifier and Information, correlation, distance between inter and intra class another method wrapper detects to error rate used as a measure Sub set feature selection, performance good for particular type model onlyand another method is hybrid method to takes-all group of techniques which perform feature selection part of the model construction process [2].

The Classification and Regression Trees (CART), [1] which described the generation of binary decision trees CART were invented independently of one another at around the same time, yet follow a similar approach for learning decision trees from training tuples [1]. These two cornerstone algorithms spawned a flurry of work on decision tree induction. CART adopt a greedy (i.e., non-backtracking) approach in which decision trees are constructed in a top-down recursive divide-and-conquer manner. Most algorithms for decision tree induction also follow such a top-down approach

II. METHODOLOGY

The learning and classification steps of decision induction algorithm are simple and fast. In general, induction algorithm classifiers have good accuracy and induction algorithms have been used for classification in many application areas, such as medicine

PRE-PROCESS METHOD DIAGRAM

The below diagram the first steps to be process on input features to set all the second step process on features sub section in this process evaluate data to pass the induction algorithm, this algorithm to adopt CART method, this method classify the data in proper way.The input features means to given data set inputs that is attributes, the attributes is age,sex, occupation, chest pain type (4 values) , food habits, resting blood pressure , fasting blood sugar > 120 mg/dl, resting electrocardiographic results (values 0,1,2) , maximum heart rate achieved , old peak = ST depression induced by exercise relative to rest, the slope of the peak exercise ST segment, thal: 3 = normal; 6 = fixed defect; 7 = reversible defect as follows 14 attributes as to be given input.

Fig 1.Pre-process method

Features sub selection there are three type's wrapper, filter, and hybrid. In this research to analysis to be take filter selection method. Because in this method is also called variable selection or attribute selection. It is the automatic selection of attributes in your data (such as columns in tabular data) that are most relevant to the predictive modeling problem is work. Feature selection is different from dimensionality reduction. Both methods seek to reduce the number of attributes in the dataset, but a dimensionality reduction method do so by creating new combinations of attributes, where as feature selection methods include and exclude attributes present in the data without changing them [5].

ANFIS Data Flow and Process Diagram

An adaptive network is a multi-layer feed forward network in which each node performs a particular function on the incoming signals. The nature and the choice of the node function depend on the overall input-output function. No weights are associated with links and the links just indicate the flow. To achieve desired i/p-o/p mapping the parameters are updated according to training data and gradient-based learning procedure [6].

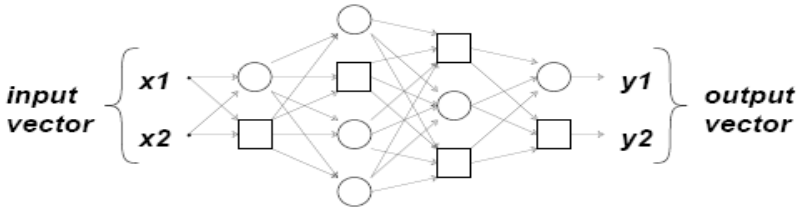

Fig 2. ANFIS Data Flow and Process

ANFIS algorithm Adaptive Neural Fuzzy Inference System (ANFIS) exploit the advantages of NN and FIS by combining the human expert knowledge (FIS rules) and the ability to adapt and learn. Three major components constitute Fuzzy Inference systems (FIS). This includes: a rule base which is made up of a selection of fuzzy rules; a database that defines the membership functions and a reasoning mechanism that is a way of inferring a reasonable output or conclusion. Our approach applies Sugeno fuzzy rules which can be illustrated as follows; for a first-order Sugeno fuzzy inference system with two inputs, a common rule set with two fuzzy if-then rules is the following [2]:

Rule1: If x is A1 and y is B1, then f1 = p1x + q1y + r1
Rule2: If x is A2 and y is B2, then f2 = p2x + q2y + r2

III. GI-ANFIS APPROACH

The Gini index is used in CART. Using the notation described above, the Gini index measures the impurity of D, a data partition or set of training tuples, as

$$\text{Gini (D)} = 1\text{-} \sum_{i-1}^{m} pi^{k2}$$

where pi is the probability that a tuple in D belongs to class Ci and is estimated by

jCi,Dj/jDj. The sum is computed over m classes. The Gini index considers a binary split for each attribute. Let's first consider the case whereA is a discrete-valued attribute having v distinct values, fa1, a2, : : : , avg, occurring inD. To determine the best binary split on A, we examine all of the possible subsets that can be formed using known values of A. Each subset, SA, can be considered as a binary test for attribute A of the form "A 2 SA?". Given a tuple, this test is satisfied if the value ofA for the tuple is among the values listed in SA. If A has v possible values, then there are 2^vpossible subsets. For example, if income has three possible values, namely flow, medium, high, then the possible subsets are flow, medium, high, flow, medium, flow, high, medium, high, flow, medium, high, and fg. We exclude the power set, flow, medium, high, and the empty set from consideration since, conceptually, they do not represent a split. Therefore, there are 2^v-2 possible ways to form two partitions of the data, D, based on a binary split on A.When considering a binary split, we compute a weighted sum of the impurity of each resulting partition. For example, if a binary split on A Partitions D into D1 and D2, the Gini index of D given that partitioning is

$$\text{Gini}_{\text{A}}(\text{D}) = \frac{|\text{D}_1|}{|\text{D}|}\text{Gini }(\text{D}_1) + \frac{|\text{D}_2|}{|\text{D}|}\text{Gini }(\text{D}_2)$$

For each attribute, each of the possible binary splits is considered. For a discrete-valued attribute, the subset that gives the minimum gain index for that attribute is selected as its splitting subset.For continuous-valued attributes, each possible split-point must be considered. The strategy is imilar to that described above for information gain, where the midpoint between each pair of (sorted) adjacent values is taken as a possible split-point. The point giving the minimum Gini index for a given (continuous-valued) attribute is taken as the split-point of that attribute. Recall that for a possible split-point of A, D1 is the set of tuples in D satisfying A _ split point, and D2 is the set of tuples in D satisfying A >split point.The reduction in impurity that would be incurred by a binary split on a discrete- or Continuous-valued attribute A is

$$\Delta\text{DGini(A)} = \text{Gini(D)} - \text{GiniA(D)} \Longrightarrow \text{ANFIS}$$

The selected features were applied to ANFIS to train and test the proposed approach. The structure of the proposed approach where X= {x1 , x2,...., xn}are the original features in dataset, Y={y1 , y2,....,yk } are the features after pplying the Gini Index (features selection), and Z denote the final predication and accuracy after applying Y on ANFIS The Gini Index is then fed to ANFIS.

Fig 1. Proposed Structure

partition, D (training dataset) Attribute list Attribute selection method Output: Decision tree Method: Create a node N; If tuples in D are all of the same class, C then Return N as a leaf node labeled with the class C; If attribute list is empty then Return N as a leaf node labeled with the majority class in D; Apply attribute selection method to find the best splitting rule; Label node N with splitting criterion; Attribute list = attribute list – splitting attribute; For each outcome j of splitting criterion Let Dj be the set of data tuples in D satisfying outcome j;

If Dj is empty then Attach a leaf labeled with the majority class in D to node N; Else attach the node returned by generating decision tree to node N; End for Return N

GI-ANFIS processing is based on the following:

- GI-ANFIS is used to find the first solutions for hard rules (structural identification).
- Precondition part of rules is made fuzzy.
- GI-ANFIS is used to fine tune the parameters (parameter identification).
- Normalization of weights is implicit in the above procedure.

IF x<a AND y <b THEN z =f1
IF x<a AND y >b THEN z =f2
IF x>a AND y >c THEN z =f3
IF x>a AND y >c THEN z =f4

Table 1. Corresponding rule table

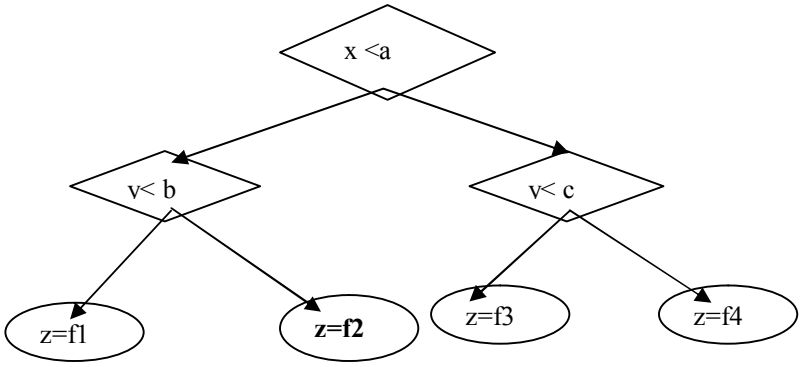

Fig 1. Corresponding Rule Flow Chart

Steps of the proposed algorithm:

Step1: Load data.
Step2: Generate the dataset
Step3: Now generate with given dataset.
Step4: Now using testing dataset with Information, correlation, and distance between inter and intra class and dummy values in it evaluate the file (dataset) using filter search method using in feature selection.
Step5: To remove the dummy values entries all the output values which has <0.5 value is removed i.e. not considered as proper data.
Step6: Now using the evaluated data, again CART tree is generated.
Step7: Find out the GI-ANFIS ratio or perchantage (%).
Step8: Compare this generated to result for (GI-ANFIS) with (IG-ANFIS) using dataset.

IV.RESULTS AND ANALYSIS

The remove the dummy and duplicated data entries all the output values which has <0.5 value is removed

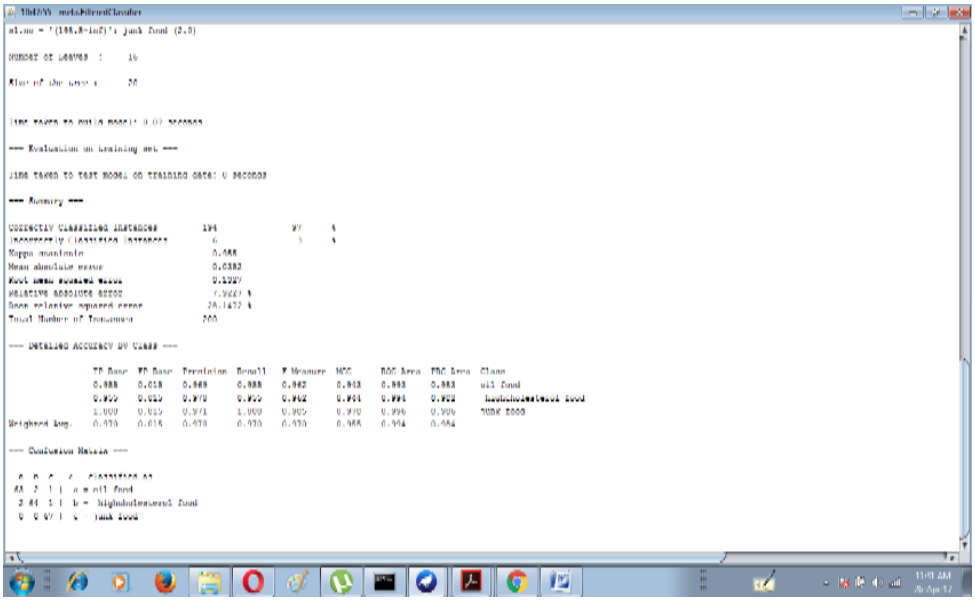

Fig 1.To view of Filter method

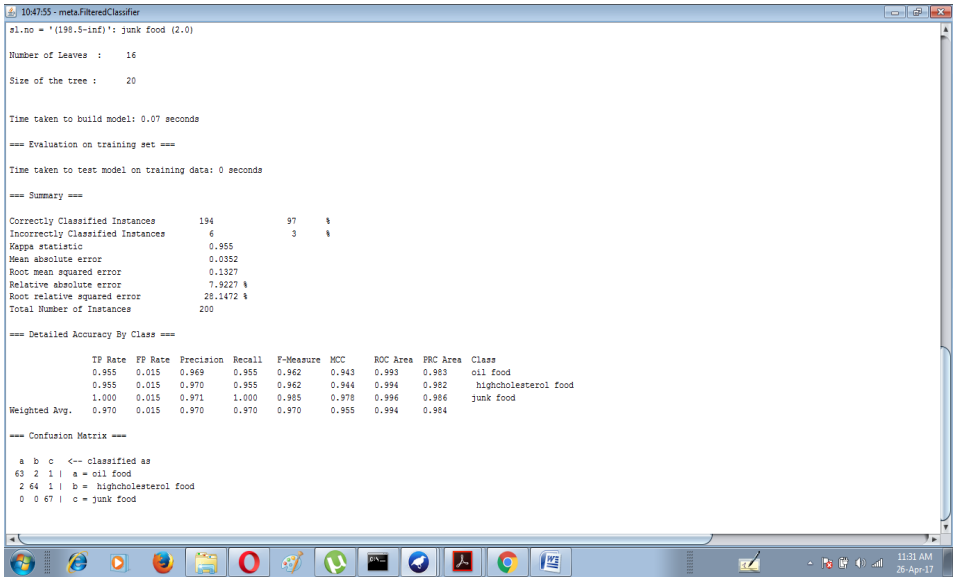

Fig 2.To view of GI-ANFIS Using CART Algorithm

Compare the previous result and accuracy

THE APPROCH	ACCURACY
NNS	95.42 %
ILFN-FUZZY	96.13%
IG-ANFIS	96.56%
GI-ANFIS	97.00%

Table 1.Comparison of classification accuracy

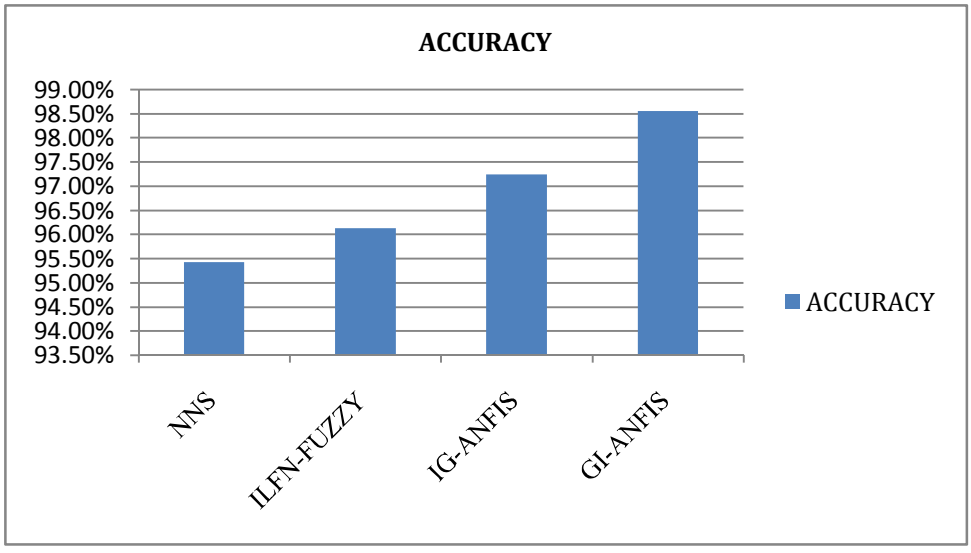

Fig 3 .Comparison of classification accuracy chart

V.CONCLUSION

The research used GI-ANFIS data mining technique on UCI heart data sets to provide the diagnosis results. The results from the approach were so promising. If further attempts are engaged in the application of Information Technology in diagnosing various diseases such as heart then efficient, timely and decent healthcare services will be realized. Large databases that used in the medical sector still have a concern of Missing features values brought about by many factors as discussed early. GI-ANFIS and approach had considerably good results.

REFERENCES

1. Dr.Neeraj Bhargava,et al., "An Approach for Classification using Simple CARTAlgorithm in Weka", Institute of Electrical and Electronics Engineers, 2017.
2. Deepali Chandna,et al., "Diagnosis of Heart Disease Using Data Mining Algorithm, International Journal of Computer Science and Information Technologies", Vol. 5 (2), 2014.
3. A Emam, Ischemia prediction using ANFIS,et al.,"Institute of Electrical and Electronics Engineers", 2010.
4. Benard Nyangena Kiage,et a.,"IG-ANFIS DATA MINING APPROACH FORFORECASTING CANCER THREATS" , International Journal of Engineering Research And Management (IJERM), Volume-01, Issue-06, September 2014.
5. S.S.Baskar,et al ., "Filter Based Feature Selection Approach to Enhance the Classification Accuracy" ,2013
6. W. Altidor, T. M. Khoshgoftaar, J. Van Hulse, A. Napolitano,et al., "Ensemble feature ranking methods for data intensive computing applications", Handbook of data intensive computing, Springer Science + Business media, LLC 2011, pages 349 -376.
7. Y. Saeys, T. Abeel,et al., Y. V. Peer, "Robust feature selection using ensemble feature selection techniques", W. Daelemans et al. (Eds.): ECML PKDD 2008, Part II, LNAI 5212, pages 313–325
8. Micheline Kamber ,et al., "Data Mining:Concepts and TechniquesUniversity of Illinois at Urbana Champaign ",2006
9. L. Yu, H. Liu,et al, "Feature Selection for High-Dimensional Data: A Fast Correlation Based Filter Solution", b Proceedings of the Twentieth International Conference on Machine Leaning, ICML-03, Washington, D.C., August, 2003, pages 856-863.
10. Yen GG, Meesad P,et al.," Constructing a fuzzy rule-based system using the ILFN network and Genetic Algorithm ",2001.
11. D. Gamberger and N. Lavrac, "Conditions for Occam's Razor applicability and noise elimination",In Proccedings of the Ninth European Conference on Machine Learning, 1997.

