A Novel Decision Tree Based Classifier For Accurate Multi Disease Prediction

Sagar Manel[1], Dhaval Patel[2]

[1]PG Scholar, [2]Assistant Professor

Department of Computer Science, Hasmukh Goswami Collage Of Engineering, Vahelal, Ahmedabad

Abstract - The objective of our work is to predict more accurately the presence of heart disease with reduced number of attributes. Originally, fourteen attributes were involved in predicting the heart disease. In our work, Genetic algorithm is used to determine the attributes which contribute more towards the diagnosis of heart ailments which indirectly reduces the number of tests which are needed to be taken by a patient. fourteen attributes are reduced to seven attributes using genetic search. Subsequently, Decision Tree are used to predict the diagnosis of patients with the same accuracy as obtained before the reduction of number of attributes. Also, the observations exhibit that the Decision Tree data mining technique After applying J48 algorithm in weka tool perfect decision tree generate. After that applying predication system it's display output from main three disease. Based on that prediction system apply fuzzy rules for find perfect and higher intensity for disease out of three disease. So as compare to previous algorithm this proposed method gives increased and improved performance and accuracy.

Index Terms- DataMining,GeneticAlgorithm, j48 Algorithm,Decision Tree,Fuzzy Logic

I.INTRODUCTION

Decision tree learning, used in data mining, statistics, and machine learning. A decision tree also uses a predictive model which found the observations about an item and conclusions about the target value. The descriptive model names for such tree are **classification trees** and **regression trees**. In these decision trees the structures of leaves display the class labels and branches shows the features that to lead those class labels. In the decision analysis, a decision tree is used to visually and explicitly represent the decisions and decision making. In the data mining process the decision tree describes the data but not decisions. The classification tree can be display an input result for decision on making. This page deals with the decision trees

data mining.

The decision tree learning method is commonly use in data mining. The goal is creating a model of predicts value of the target variable base on several input variables. The example is shows that the each node is corresponds to the one input variables; there are edges of children for each possible values of the input variable. The represents value of the target variable gives the values of the input variables are explained by the path from the root to the leaf.

Based on the attribute value test the decision tree can be explained by splitting the source set into subsets. This decision tree process is frequently repeated on derived subset in a recursive manner. The recursive mode is call a recursive partition. The recursion complete when the subset node has the same value of the target variable. The splitting has no longer value to the predictions.

II.LITERATURE REVIEW

In paper [1] Different approaches have different aspects in diagnosing diseases. By using the but the hidden layers description cannot be evaluated. In fuzzy logic approach the weighted rules are generated initially and then the fuzzy rule decision is provided and the accuracy obtained id around 79.05%. In naive bayes classification approach helps in predicting

whether the patient is prone to heart disease or not and depicting the risk factor for heart attack. The accuracy observed for naive bayes approach was around 90%. Similarly by using Support vector machines concept the accuracy was achieved around 84.12%. While as by using decision tree approach the accuracy increased up to 96%.

Table 1 Experimental result

Parameters	Neural Network	Fuzzy logic	SVM	Naïve Bayes	Decision Tree
Example Algorithms	Back propagation	Thresholds and weight applied on IF-THEN rules	Maximun & optimal margins by Gaussian theorem	Posterior probability-Bayes Theorem	C4.5,CART J48 using splitting attribute entropy
0Formula	Input layer $wij = wij + \Delta wij$ hidden layer $wjk = wjk + \Delta wjk$	Fuzzy Set $\mu: x \rightarrow [0,1]$	Margin equations w . x − b =1 w . x − b = -1	$p(A/B)$ $= \frac{p(B/A)p(A)}{P(B)}$	Information gain
Advantages	Minimized error in each level	Specification is obtained	Large data set is analyzed	Minimum error occur	No domain knowledge is required
Disadvantage	Very slow working	Comparison Increases	Range should be precise else outliner are observed	Multiple symptoms cannot dependency in attributes	Selection of splitting attribute & over fitting
Approximate Accuracy	80-90%	78-85%	85-90%	90-95%	94-96%

In paper [2] Existing literature shows that Classification task in Data Mining plays a vital role in heart disease prediction when compared with Clustering, Association Rule and Regression. In Classification Decision Tree outperforms in some cases, where Neural Network and Naïve Bayes outperforms in some other cases. Each technique has its own merits and demerits. When combined with each other or with Fuzzy logic, Heart disease prediction with Data Mining techniques will become most successful with less number of attributes. Text mining the medical data is another extension found in predicting the health care data.

Table 2 Experimental result

Algorithm used	Accuracy	Time taken
Naïve bayes	52.33%	609 ms
Decision List	52%	719 ms
KNN	45.67%	1000 ms

In paper [3] Data set of 294records with 13 attributes is used and the outcome reveals that the Naïve Bayes outperforms and sometime Decision Tree. In Future Genetic algorithm will be used in order to

reduce the actual data size to get the optimal subset of attribute sufficient for heart disease prediction. Prediction of the heart disease will be evaluated according to the result produced from it. Improvement is done to increase its consistency and efficiency. Benefit of using genetic algorithm is the prediction of heart disease can be done in a short time with the help of reduced dataset. Genetic algorithm will be implemented with the MATLAB.

Table 3 Experimental result

Techniques	Accuracy
Naïve Bayes	86.53%
Decision Tree	89%
ANN	85.53%

In paper [6] They are using the Data mining techniques like K Means and Weighted Association rule for the destruction of standard work and simple extraction of data precisely from electronic record, assign into protected electronic system of medical records which will salvage lives and decrease the cost

of the healthcare services. K-means clustering is a usually used data clustering for unsupervised learning tasks. Decision tree is used to prediction process.

The weighted association classifier is a idea of weighted association rule for classification. Weighted Association Rule Mining uses Weighted Support and Confidence Framework to extract Association rule from data warehouse. Classification rule mining takes a training data set and generates a small set of rules to organize future data. The k-means clustering is the technique to cluster the attributes from the patient record. The decision tree with K-means clustering can enhance the classifier's performance in diagnosing heart disease. The initial centroid selection technique among the five K-means clustering techniques is used here because it can provide better performance over heart disease prediction. The experiments shows that K-means with decision tree technique make the system more accurate and efficient.

In paper [4] the prediction of heart disease based on feature selection by using multilayer perceptron with back-propagation algorithm and k-nearest neighbor algorithm based on an explicit similarity measure with biomedical test values to diagnose heart disease is presented. The main motivation for this paper is to classify the heart disease with reduced number of attributes. They use the weight information by a multilayer perceptron to determine the attributes which reduces the number of attributes which is needed to be taken from original datasets (13 attribute is reduced to 8 attributes). Afterward, they used k-nearest neighbor algorithm to predict the diagnosis of heart disease after the reduction of a number of attributes. The accuracy differs between 13 attributes and 8 attributes in testing data set is 93% and 90%, respectively. The experimental results show that our propose classification help in the best prediction of heart disease which even helps doctors in their diagnosis decisions.

In this paper, the k-nearest neighbor algorithm based on an explicit similarity measure in data mining are compared for predicting heart disease with reduced number of attributes based on multilayer perceptron with back-propagation algorithm. Original dataset of 14 attributes is reduced to 8 attributes using back-propagation algorithm.

Also, the observations exhibit that the k-nearest neighbor algorithm based on an explicit similarity measure is successful after incorporating feature subset selection with the back - propagation algorithm. Our results show that the proposed algorithm achieves highest SE, SP, and AC values are 0.80, 0.80, and 81.90%, respectively.

In paper [5] the heart disease details to be the main source of death worldwide. It is tough for medical practitioners to conclude the heart attack as it is a complicated work that requires practice and knowledge. The health sector today consists of mysterious information that can be crucial in decisions. Data mining algorithms such as J48, Naïve Bayes, CART, and Bayes Net are tested in this analysis for anticipate heart attacks. The research decision display prediction efficiency of 99%. Data mining permit the health sector to predict arrangement in the dataset. The research undertook an experiment on application of various data mining algorithms to predict the heart attacks and to compare the best method of prediction. The research results do not presents a dramatic difference in the prediction when using different classification algorithms in data mining. The experiment can serve as an important tool for physicians to predict risky cases in the practice and advise accordingly. The model from the classification will be able to answer more complex queries in the prediction of heart attack diseases. The predictive accuracy determined by J48, REPTREE and SIMPLE CART algorithms suggests that parameters used are reliable indicators to predict.

III. PROBLEM STATEMENT

In above literature review we study about different kind of algorithm using of supervise learning and unsupervised learning like decision tree,svm(support vector machine),naïve Bayes, nearest neighbor, k-mean, back propagation neural network.
The problem of learning this all algorithms in neural network working very slow and fuzzy logic create comparison increases, the svm contain range should be precise else outliner are observed and naïve bayes contain multiple symptoms cannot handle and dependency in attributes, at last in decision tree (c4.5,cart and j48) selection of spitting attribute and over fitting is the problem.

In C 4.5 algorithm is complex structure, neural network require –training data set and also contain computational overhead its require more system development time and also define training data set. Decision-tree learners can generate over-complex trees that do not generalize the data well. This is called over fitting Mechanisms such as pruning is necessary to avoid this problem.

IV. PROPOSED METHOD

4.1 Flow chart of the proposed algorithm
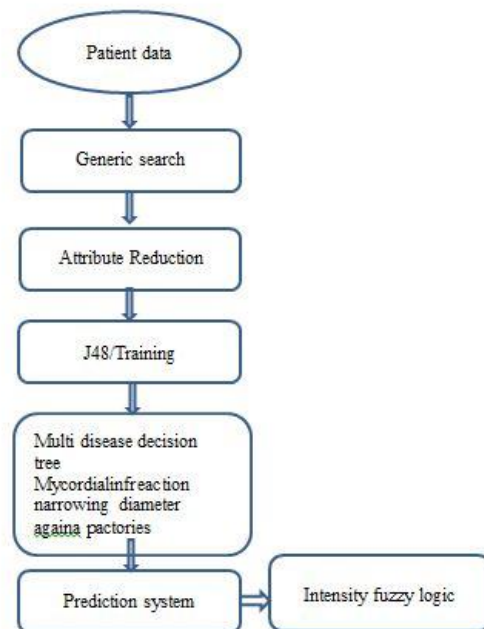
Fig 1 Flow chart for proposed method

Flow chart presented above shows the process of the suggested algorithm. To find efficient pattern is main idea of this algorithm.

4.2 proposed method

Step 1: Collect the patient information.
Step 2: Using cf subset evolution in weka tool for attribute selection.
Step 3: Apply genetic algorithm for attribute reduction.
Step 4: Used J-48 algorithm for training data set.
Step 5: Used three decision tree for multi disease analysis.
Step 6: Apply prediction system on three disease.
Step-7: Find intensity of disease based on fuzzy logic.

4.3 Attribute selection

In this chapter we display only the 50 attributes out
of the 300 attributes in this practical the 14 attribute
is selected

Table 4 Attribute selection

Sr.no	age	sex	cp	trestbps	chol	fbs	Reste cg	Thalac h	exan g	oldpe ak	slop e	ca	tha l	num
1	63	1	1	145	233	1	2	150	0	2.3	3	0	6	0
2	67	1	4	160	286	0	2	108	1	1.5	2	3	3	2
3	67	1	4	120	229	0	2	129	1	2.6	2	2	7	1
4	37	1	3	130	250	0	0	187	0	3.5	3	0	3	0
5	41	0	2	130	204	0	2	172	0	1.4	1	0	3	0
6	56	1	2	120	236	0	0	178	0	0.8	1	0	3	0
7	62	0	4	140	268	0	2	160	0	3.6	3	2	3	3
8	57	0	4	120	354	0	0	163	1	0.6	1	0	3	0
9	63	1	4	130	254	0	2	147	0	1.4	2	1	7	2
10	53	1	4	140	203	1	2	155	1	3.1	3	0	7	1
11	57	1	4	140	192	0	0	148	0	0.4	2	0	6	0
12	56	0	2	140	294	0	2	153	0	1.3	2	0	3	0
13	56	1	3	130	256	1	2	142	1	0.6	2	1	6	2
14	44	1	2	120	263	0	0	173	0	0	1	0	7	0
15	52	1	3	172	199	1	0	162	0	0.5	1	0	7	0
16	57	1	3	150	168	0	0	174	0	1.6	1	0	3	0
17	48	1	2	110	229	0	0	168	0	1	3	0	7	1
18	54	1	4	140	239	0	0	160	0	1.2	1	0	3	0
19	48	0	3	130	275	0	0	139	0	0.2	1	0	3	0
20	49	1	2	130	266	0	0	171	0	0.6	1	0	3	0
21	64	1	1	110	211	0	2	144	1	1.8	2	0	3	0
22	58	0	1	150	283	1	2	162	0	1	1	0	3	0
23	58	1	2	120	284	0	2	160	0	1.8	2	0	3	1
24	58	1	3	132	224	0	2	173	0	3.2	1	2	7	3

25	58	1	3	132	224	0	2	173	0	3.2	1	2	7	3
26	60	1	4	130	206	0	2	132	1	2.4	2	2	7	4
27	50	0	3	120	219	0	0	158	0	1.6	2	0	3	0
28	58	0	3	120	340	0	0	172	0	0	1	0	3	0
29	66	0	1	150	226	0	0	114	0	2.6	3	0	3	0
30	43	1	4	150	247	0	0	171	0	1.5	1	0	3	0
31	40	1	4	110	167	0	2	114	1	2	2	0	7	3
32	69	0	1	140	239	0	0	151	0	1.8	1	2	3	0
33	60	1	4	117	230	1	0	160	1	1.4	1	2	7	2
34	64	1	3	140	335	0	0	158	0	0	1	0	3	1
35	59	1	4	135	234	0	0	161	0	0.5	2	0	7	0
36	44	1	3	130	233	0	0	179	1	0.4	1	0	3	0
37	42	1	4	140	226	0	0	178	0	0	1	0	3	0
38	43	1	4	120	177	0	2	120	1	2.5	2	0	7	3
39	57	1	4	150	276	0	2	112	1	0.6	2	1	6	1
40	55	1	4	132	353	0	0	132	1	1.2	2	1	7	3
41	61	1	3	150	243	1	0	137	1	1	2	0	3	0
42	65	0	4	150	225	0	2	114	0	1	2	3	7	4
43	40	1	1	140	199	0	0	178	1	1.4	1	0	7	0
44	71	0	2	160	302	0	0	162	0	0.4	1	2	3	0
45	59	1	3	150	212	1	0	157	0	1.6	1	0	3	0
46	61	0	4	130	330	0	2	169	0	0	1	0	3	1
47	58	1	3	112	230	0	2	165	0	2.5	2	1	7	4
48	51	1	3	110	175	0	0	123	0	0.6	1	0	3	0
49	50	1	4	150	243	0	2	128	0	2.6	2	0	7	4
50	65	0	3	140	417	1	2	157	0	0.8	1	1	3	0

4.2 Snap Shot

4.2.1   Attributes of weka tool

```
=== Run information ===

Evaluator:    weka.attributeSelection.CfsSubsetEval
Search:weka.attributeSelection.GeneticSearch -Z 20 -G 20 -C 0.6 -M 0.033 -R :
Relation:     whatever-weka.filters.unsupervised.attribute.Normalize-S1.0-T0
Instances:    297
Attributes:   14
              age
              sex
              cp
              trestbps
              chol
              fbs
              restecg
              thalach
              exang
              oldpeak
              slope
              ca
              thal
              num
Evaluation mode:evaluate on all training data
```

Fig.2 14 attributes are used in weka tool

4.2.2 Results of genetic search

```
Search Method:
        Genetic search.
        Start set: no attributes
        Population size: 20
        Number of generations: 20
        Probability of crossover:  0.6
        Probability of mutation:  0.033
        Report frequency: 20
        Random number seed: 1

Initial population
merit        scaled       subset
0.23885      0.24581      1 2 4 7 9 11 12 13
0.2241       0.22832      1 4 5 6 9 11 12
0.25674      0.26702      4 5 7 9 10 12
0.10414      0.08608      9
0.23719      0.24384      4 7 8 10 13
0.23313      0.23903      1 9 10 11 12
0.22641      0.23106      1 2 4 6 7 8 9 12 13
0.12842      0.11487      3 6 7 11
0.1268       0.11294      1 4 6 7
0.14149      0.13037      12
0.20252      0.20273      1 2 3 4 6 13
0.19344      0.19196      6 10 13
0.21467      0.21714      3 6 7 8 9 11 12
0.2245       0.22879      3 4 8 11
0.27512      0.28882      4 5 6 8 10 11 12 13
0.25265      0.26217      2 3 4 5 7 9 12
0.24641      0.25478      1 2 4 6 7 8 10 11 12 13
```

Fig.3 display the result of genetic search

4.2.3 Attribute subset evaluator

```
Generation: 20
merit        scaled       subset
0.30858      0.36035      3 5 8 9 10 12 13
0.30858      0.36035      3 5 8 9 10 12 13
0.30467      0.32557      3 5 9 10 11 12 13
0.30153      0.29763      5 8 9 10 11 12 13
0.30264      0.30753      2 3 5 8 9 10 11 12 13
0.30858      0.36035      3 5 8 9 10 12 13
0.30294      0.31017      5 8 9 10 12 13
0.26804      0            2 3 4 8 9 10 11 13
0.30858      0.36035      3 5 8 9 10 12 13
0.29932      0.27806      3 4 5 8 9 10 11 12 13
0.30858      0.36035      3 5 8 9 10 12 13
0.29404      0.23112      3 5 8 9 10 11 13
0.30127      0.2954       3 5 10 12 13
0.30467      0.32557      3 5 9 10 11 12 13
0.30802      0.35539      3 5 8 9 10 11 12 13
0.30802      0.35539      3 5 8 9 10 11 12 13
0.30166      0.29885      2 3 5 8 9 10 12 13
0.28562      0.15628      1 5 8 9 10 12 13
0.30858      0.36035      3 5 8 9 10 12 13
0.30644      0.34134      3 5 8 9 11 12 13

Attribute Subset Evaluator (supervised, Class (nominal): 14 num):
        CFS Subset Evaluator
        Including locally predictive attributes
```

Fig.4 14 attributes subset evaluator in weka

4.2.4 Attribute reduction on weka tool

```
Attribute Subset Evaluator (supervised, Class (nominal): 14 num):
    CFS Subset Evaluator

    Including locally predictive attributes


Selected attributes: 3,5,8,9,10,12,13 : 7
                      cp

                      chol

                      thalach

                      exang

                      oldpeak

                      ca

                      thal
```

Fig.5 display the result of reduce attribute

Table 4 Attribute selection

Sr.no.	cp	chol	thalach	Exang	oldpeak	ca	thal
1	1	233	150	0	2.3	0	6
2	4	286	108	1	1.5	3	3
3	4	229	129	1	2.6	2	7
4	3	250	187	0	3.5	0	3
5	2	204	172	0	1.4	0	3
6	2	236	178	0	0.8	0	3
7	4	268	160	0	3.6	2	3
8	4	354	163	1	0.6	0	3
9	4	254	147	0	1.4	1	7
10	4	203	155	1	3.1	0	7
11	4	192	148	0	0.4	0	6
12	2	294	153	0	1.3	0	3
13	3	256	142	1	0.6	1	6
14	2	263	173	0	0	0	7
15	3	199	162	0	0.5	0	7
16	3	168	174	0	1.6	0	3
17	2	229	168	0	1	0	7
18	4	239	160	0	1.2	0	3
19	3	275	139	0	0.2	0	3
20	2	266	171	0	0.6	0	3
21	1	211	144	1	1.8	0	3
22	1	283	162	0	1	0	3
23	2	284	160	0	1.8	0	3
24	3	224	173	0	3.2	2	7
25	4	206	132	1	2.4	2	7
26	3	219	158	0	1.6	0	3
27	3	340	172	0	0	0	3
28	1	226	114	0	2.6	0	3

29	4	247	171	0	1.5	0	3
30	4	167	114	1	2	0	7
31	1	239	151	0	1.8	2	3
32	4	230	160	1	1.4	2	7
33	3	335	158	0	0	0	3
34	4	234	161	0	0.5	0	7
35	3	233	179	1	0.4	0	3
36	4	226	178	0	0	0	3
37	4	177	120	1	2.5	0	7
38	4	276	112	1	0.6	1	6
39	4	353	132	1	1.2	1	7
40	3	243	137	1	1	0	3
41	4	225	114	0	1	3	7
42	1	199	178	1	1.4	0	7
43	2	302	162	0	0.4	2	3
44	3	212	157	0	1.6	0	3
45	4	330	169	0	0	0	3
46	3	230	165	0	2.5	1	7
47	3	175	123	0	0.6	0	3
48	4	243	128	0	2.6	0	7
49	3	417	157	0	0.8	1	3
50	3	197	152	0	1.2	0	3

4.2.6 J48 decesion tree

```
J48 pruned tree
------------------

thal = 6
|   ca = 0
|   |   exang = 0: 0 (5.0)
|   |   exang = 1: 1 (3.0/1.0)
|   ca = 3: 1 (2.0)
|   ca = 2: 1 (4.0)
|   ca = 1: 1 (4.0)
thal = 3
|   ca = 0: 0 (115.0/13.0)
|   ca = 3: 1 (6.0/1.0)
|   ca = 2: 0 (14.0/7.0)
|   ca = 1
|   |   cp = 1: 0 (3.0/1.0)
|   |   cp = 4: 1 (10.0/1.0)
|   |   cp = 3: 0 (10.0)
|   |   cp = 2: 0 (6.0/2.0)
thal = 7
|   cp = 1: 0 (8.0/3.0)
|   cp = 4: 1 (77.0/7.0)
|   cp = 3
|   |   slope = 3: 0 (1.0)
|   |   slope = 2: 1 (14.0/4.0)
|   |   slope = 1: 0 (7.0/1.0)
|   cp = 2
|   |   ca = 0: 0 (6.0/2.0)
|   |   ca = 3: 0 (0.0)
|   |   ca = 2: 0 (0.0)
|   |   ca = 1: 1 (2.0)
```

Fig.5 display j48 prune tree

4.2.7 J48 decesion tree

```
Number of Leaves  :    21

Size of the tree :    29


Time taken to build model: 0 seconds

=== Evaluation on test set ===
=== Summary ===

Correctly Classified Instances      254         85.5219 %
Incorrectly Classified Instances     43         14.4781 %
Kappa statistic                       0.7064
Mean absolute error                   0.2203
Root mean squared error               0.3319
Relative absolute error              44.3269 %
Root relative squared error          66.5798 %
Coverage of cases (0.95 level)      100      %
Mean rel. region size (0.95 level)   95.2862 %
Total Number of Instances           297

=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.913	0.212	0.834	0.913	0.872	0.905	0
	0.788	0.088	0.885	0.788	0.834	0.905	1
Weighted Avg.	0.855	0.154	0.858	0.855	0.854	0.905	

```
=== Confusion Matrix ===

  a   b   <-- classified as
146  14 |   a = 0
 29 108 |   b = 1
```

Fig.6 display tree and size of tree

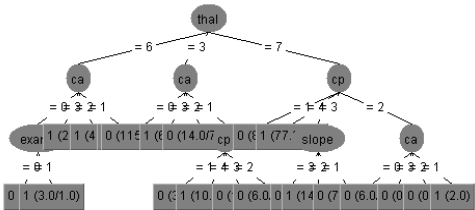### 4.2.8 Decision tree generate in weka

Fig.7 display tree and size of tree

4.2.9 Decision tree

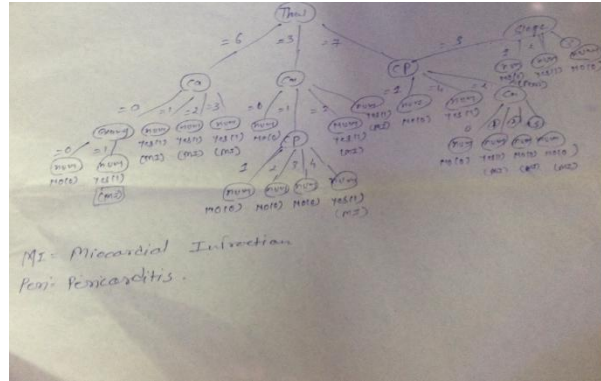MI = Miocardial Infraction
Peri= Pericarditis.

Fig.8 display j48 prune tree
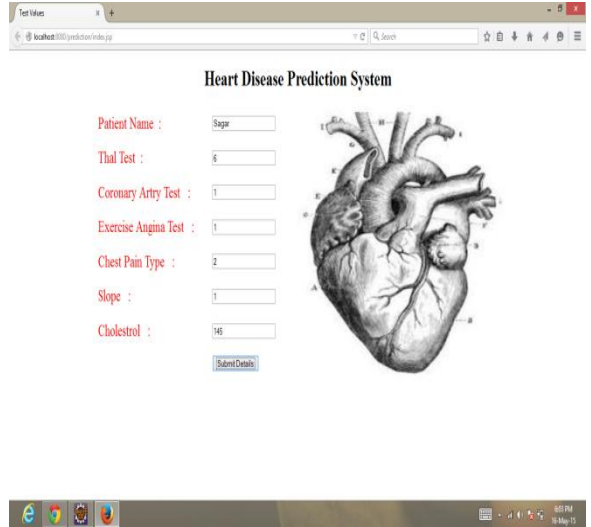
4.2.10

Fig.9 prediction system and its attributes

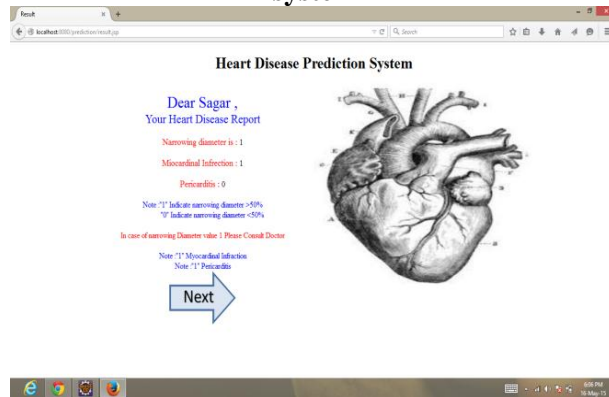### 4.4.1 Result analysis of multi disease prediction system

Fig.10 Result analysis of multi disease

4.5 Fuzzy rules

Table 5 fuzzy rule

| Sr. no.		Fuzzy data	Fuzzy rules
1	Thai	1405/298 = 4.72 = 5 D min =2 D max =3	If Thai < 2 than risk level less than 50% If Thai level >3 than risk level more than 50%
2	Ca	201/298 = 0.67 = 1 D min =1 D max = 2/3	If Ca < 1 than risk level less than 50% If Ca > 2 than risk level more than 50%
3	Slop	476/298= 1.59 = 2 D min = 1 D max = 3	If slope <1 than risk level less than 50% If slope > 3 than risk level more than 50%
4	CP	938/298 = 3.14= 3 D min =2 D max =1	If CP < 2 than risk level less than 50% If CP >1 than risk level more than 50%
5	Chol	73463/298 = 246.52= 247 D min = 121 D max = 317	If Chol < 121 than risk level less than 50% If Chol > 317 than risk level more than 50% If Chol <121 and Chol > 317 than lines in boundery
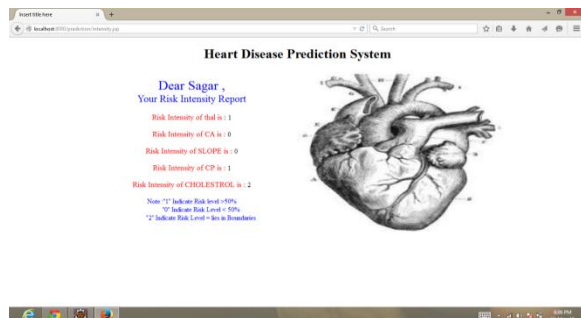
4.6 Intensity

Figure 6 Intensity Result

Table 6 Comparative Table of the three classifiers

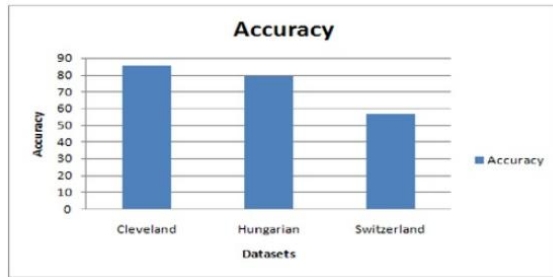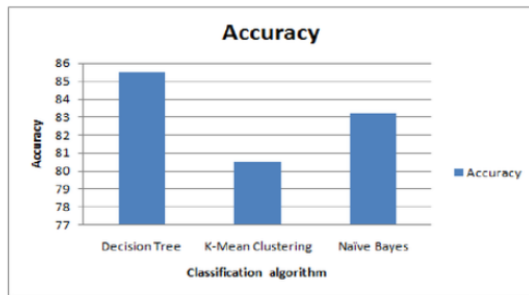| Data Set	Total instance	Correctly Classified	Accuracy
Cleaveland	297	254	85.5%
Hungarian	294	233	79.50%
Switzerland	123	69	56.47%

Fig 4.13 result Discussion

Fig 4.14 result Discussion

V.CONCLUSION

Research focus heart disease prediction with three modules which are attribute reduction, classification and fuzzy approach ,attribute reduction module contain attribute selector and genetic algorithm, the output of attribute reduction model is reduce more precious attribute as attribute selector. classification module take reduce attribute dataset as input and generate decision tree as a output Classification done by j48 algorithm based on that prediction system develop .Research gives intensity of heartdisease using fuzzy approach intensity are classified in class format.

VI. FUTURE WORK

6.1 Future Work
Future work for Dissertation phase-II will be implementation of proposed algorithm using java Research gives intensity of heart disease using fuzzy approach intensity is classified in class format. After that try to find the accuracy of intensity value.

REFERENCES

[1] Kulkarni Rashmi Ravindranath, Survey of Different Approaches for Diagnosing Heart Diseases for Clinical Decision Support System ,International Journal of Scientific and Research Publications, Volume 4, Issue 9, September 2014

[2] Sivagowry , Dr. Durairaj. Persia.An Empirical Study on applying Data Mining Techniques for the Analysis and Prediction of Heart DiseaseInternational conference on information communication and embedded systems icices 2013.

[3] V. Manikantan & S. Latha , Predicting the Analysis of Heart Disease Symptoms Using Medicinal Data Mining Methods ,International Journal on Advanced Computer Theory and Engineering (IJACTE).

[4] Early Heart Disease Predictionusing Data Mining Techniques,Aditya Methaila, Prince Kansal, Himanshu Arya, Pankaj Kumar.S

[5] Kittipol Wisaeng , Predict the Diagnosis of Heart Disease Using Feature Selection and k-Nearest Neighbor Algorithm ,Applied Mathematical Sciences, Vol. 8, 2014, no. 83, 4103 – 4113

[6] Hlaudi Daniel Masethe, Mosima Anna Masethe Prediction of Heart Disease using Classification Algorithm Proceedings of the World Congress on Engineering and Computer Science 2014 Vol II.