iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** V **Month of publication:** May 2017

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

Volume 5 Issue V, May 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Categorization of Liver Disease Using Classification Techniques

Ashwani Kumar[1], Neelam Sahu[2]
[1,2] Dept of IT, Dr. C. V. Raman University, Kota, Bilaspur, Chhattisgarh, India

Abstract: Health is an important issue in human life and facing various diseases by most of the people. Liver Diseases is very serious liver facing by most of the human being. Machine learning techniques are very effective tool through we classify the liver and non-liver patients with better accuracy. Classification is one of the easy and effective technique through which we get better accuracy to judge liver data. In this paper we use C4.5, Random Forest, CART, Random Tree and REP tree classification method and get better accuracy to detect liver disease. We achieved better accuracy 79.22% in Random Forest using 80-20% training-testing data partition.
Keywords: Classification, liver, Feature Selection.

I. INTRODUCTION

In medical science, healthcare is very challenging task for every doctors and medical institutions. Machine learning techniques can greatly benefit all parties involved in the health care industry. In health care, there is massive data, and this data has no organizational value until converted into information and knowledge, which can help control costs, increase profits, and maintain high quality of patient care.The main aim of this work is to detection of liver Mellitus and classify the liver and non liver data. Classification is one of the most important decision making techniques in many real world problem. In this paper, the main objective is to classify the data as liver or non liver and improve the classification accuracy. P. Mazaheri et al. (2015) [1]have used different classification algorithms like tree Decision algorithm, Artificial Nerve Network algorithm, Back up arrow machine algorithm, K- closest neighbour algorithm, Simple BIZ algorithm and Regression algorithm. ILPD dataset and BUPA dataset are the two datasets that are used for this experimental work. H. Jin et.al,(2014)[2] have suggested classification algorithms such as Naïve Bayes, Decision Tree, Multilayer Perceptron, k-NN, Random Forest and Logistic fr classification of liver patient disease. They have compared the performance of algorithms in different measures like Kappa statistics, Area under ROC curve, RMSE, precision, recall, sensitivity and specificity. A. Gulia et.al,(2014)[3] have used various classifiers like J-48, Multilayer Perceptron, Random Forest ,Support Vector Machine and Bayesian Network classifier for classification of liver patient data. Reetu et al. (2015)[4] have used decision tree induction J48 algorithm to classify liver patient dataset. This experiment implemented the use of 10-fold cross validation method with classification algorithms to generate the classifier. A. K. Tiwari .et al. (2013) [5] have used various classifiers like Artificial Neural Network based predictive models such as Back Propagation (BP), Radial Basis Function (RBF), Self Organizing Map (SOM) and Support Vector Machine (SVM). This work also implemented correlation based feature subset selection method for evaluation of predictor attributes. S. Vijayarani et al., (2015)[8] have used Support Vector Machine (SVM) and Naïve Bayes algorithms for classification of liver patients. Indian Liver Dataset (ILPD) available on UCI repository is used for performance evaluation. These classification algorithms are evaluated on the basis of accuracy and execution time. Experimental result shows that Support Vector Machine (SVM) is better than Naïve Bayes algorithm in classifying liver patient dataset

II. DECISION TREE

Decision tree is data mining based classification techniques to generate the rules and classification of data based on this rule.Decision tree (Han J.et al.., 2006) [6] is most popular and powerful classification techniques in which in the training stage a treelike structure is formed where each non-leaf node is decision node which splits according to the features of training data while leafnode represent class node, Once the decision tree is formed, unknown samples can be presented to the root node of decision tree andultimately reaches to the class node to classify the sample as one of the target class. In this research work, we have used C4.5,SimpleCart , Random tree ,Random Forest and REP Tree as decision tree.

III. DATA SET

This research work used liver data set collected from UCI repository [7]. The data set consist 8 features, 768 instances and 1 class.

Volume 5 Issue V, May 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

The class level consistsliver and non liver.

IV. EXPERIMENT RESULT

In this experiment, we have used decision tree based classification techniques for classifying liver patient. Partitions of data is also one of the important role in classification accuracy. To achieve the better accuracy, we have partition the data set 70-30% and 80-20% training testing ratio. Theaccuracy of Random tree, CART and REP Tree gives better in case of the 80-20% partition as shown in table I. Random Forest, CART and REP Tree provide the good accuracy, sensitivity, specificity, precision and F-measures as compare to 70-30% training testing data partition. We get 79.22% accuracy in Random Forest decision tree method with 80-20% training-testing partition which is highest among all.We use Info Gain feature selection method to achieve better classification accuracy and rank the feature in descending order as 2,6,8,5,4,1,7,3.Table II shows that accuracy of proposed Info Gain-Random Forest model with 80-20% data partitions in different feature subsets. We have also calculated thevarious performance measures like sensitivity, specificity, precision and F-measures of different feature subsets. We achieved better classification accuracy 79.87% with 6 numbers of features.

Table I
Performance of models of with different data partitions

Models	70-30% Tarring-testing					80-20% Tarring-testing				
	C4.5	Random Forest	CART	Random Tree	REP Tree	C4.5	Random Forest	CART	Random Tree	REP Tree
Accuracy	76.52	**78.69**	76.52	71.30	75.21	75.97	**79.22**	77.27	67.53	78.57
Sensitivity	75.0	61.11	59.72	55.55	51.38	63.26	61.22	61.22	59.18	61.22
Specificity	77.21	86.70	84.17	78.48	86.07	81.90	87.61	84.76	71.42	86.67
Precision	60.0	67.69	63.23	54.05	62.71	62.0	69.76	65.21	49.15	68.18
F-measures	66.66	64.23	61.42	54.79	56.48	62.62	65.21	63.15	53.70	64.51

Table II
Accuracy of proposed Info Gain-Random Forest model with 80-20% data partitions

Number of features	Accuracy	Sensitivity	Specificity	Precision	F-measures
8	79.22	61.22	87.61	69.76	65.21
7	77.92	63.26	84.76	65.95	64.58
6	79.87	67.34	85.71	68.75	68.04
5	78.57	61.22	86.67	68.18	64.51
4	77.27	61.22	84.76	65.21	63.15
3	78.57	59.18	87.61	69.04	63.73
2	74.02	59.57	80.37	57.14	58.34

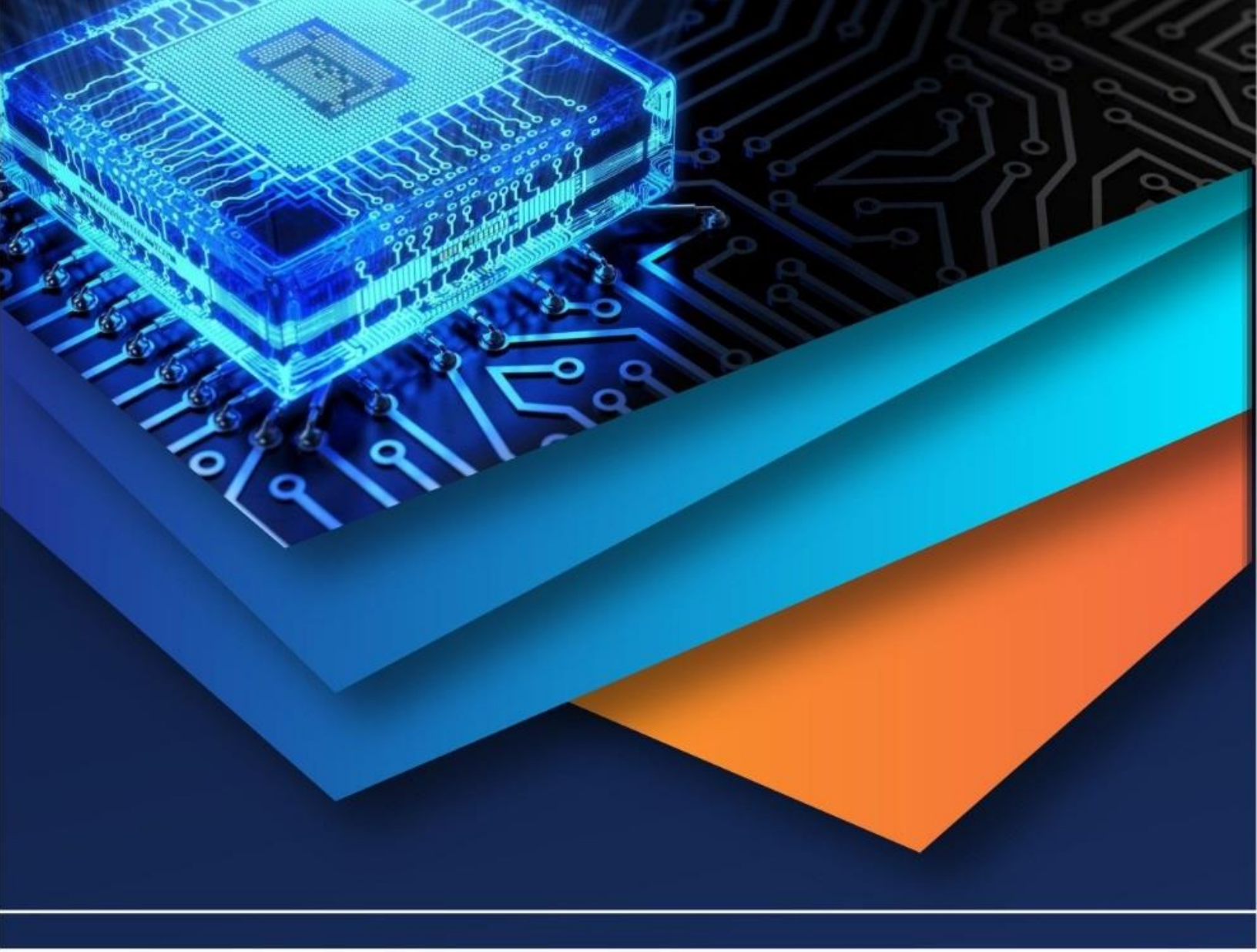
V. CONCLUSION

Healthcare issue is very sensitive and challenging problem in medical science. Classification is one of the good methods to analyze liver and non liver patient in great accuracy. We use theC4.5, Random Forest, CART, Random Tree and REP tree asclassifier for classification of liver data. We achieved 79.22% accuracy in Random Forest using 80-20% data partition with 6 features. We recommended Random forest is better classifier for classification of liver disease among all.

Volume 5 Issue V, May 2017
ISSN: 2321-9653

REFERENCES

[1]. P. Mazaheri, A. Narouzi and A. Karimi (2015), Using Algorithms to Predict Liver Disease Classification, Electronics Information and Planning. 3 :255-259.
[2]. H. Jin , S. Kim and J. Kim (2014), Decision Factors on Effective Liver Patient Data Prediction, International Journal of Bio-Science and Bio-Technology. 6(4): 167-178.
[3]. A. Gulia ,R. Vohra and P. Rani . (2014), Liver Patient Classification Using Intelligent Techniques, International Journal of Computer Science and Information Technologies. 5 (4) : 5110-5115.
[4]. Reetu andN. Kumar (2015), Medical Diagnosis for Liver Cancer Using Classification Techniques, International Journal of Recent Scientific Research. 6(6): 4809-4813.
[5]. A. K.Tiwari , L. K.Sharma and G. Ramakrishna (2013), Comparative Study of Artificial Neural Network based Classification of Liver Patient, Journal of Information Engineering and Application. 3(4).
[6]. J. Han and M. Kamber , Data Mining Concepts and Techniques, Morgan Kaufmann, San Francisco. 2nd ed.,2006.
[7]. Available: http://www.ics.uci.edu/~mlearn/databases/
[8]. S.Vijayarani . and S.Dhayanand (2015), Liver Disease Prediction using SVM and Naïve Bayes Algorithms, International Journal of Science, Engineering and Technology Research (IJSETR). 4 (4): 816-820.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY