Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 10, October-2019

A Comparative Study on Predicting the Probability of Liver Disease

R. Kalaviselvi [1]
[1]Assistant Professor, Department of Computer Science and Engineering, Kumaraguru college of technology, Coimbatore, Tamilnadu, India.

G. Santhoshni[2]
[2] M.E, Department of Computer Science and Engineering, Kumaraguru college of technology, Coimbatore, Tamilnadu, India.

Abstract- **In recent times, many people suffer from liver disease due to their food habits, alcohol consumption, stress, and many more unusual practices. Early diagnosis of liver disease may help in more possibility of getting cured, if it is not treated properly at an early stage it leads to a serious health condition. Though the existing techniques are good in prediction, they are inefficient when data grows[15]. As the clinical test report consists of high volume of data, it is very difficult to predict any respective disease. To overcome this kind of issues, generally the medical domain collaborates with automation technologies. Many computing techniques like machine learning, classification algorithms, data analytics and many more are used. To sort out the issues in liver disease prediction, a detailed study on prediction algorithms was made and a comparative analysis was carried out to find the best algorithm with high accuracy. Though the existing solutions are good, their accuracy, execution time, specificity and sensitivity have to be focused to bring up an effective system[3,4]. A comparative result among various algorithms are tabulated and efficiency of the existing methods is discussed.**

keyword:- Machine learning, Data Mining, Accuracy, Specificity, Sensitivity.

I. INTRODUCTION

In recent years, data mining provides a best interpretative features in automating the prediction of diseases. The process of extracting the required data from the large database is called data mining and the extracted data is used for predicting the hidden facts for further analysis in healthcare[2]. Data mining has a major part in prediction process. The liver has a vital role in the human body functions from protein production to removing toxins from the body and it is essential for the survival. The failure of liver functioning leads to serious health conditions. The functioning of the liver is examined by two types of tests such as imaging test and liver function tests which help to diagnose liver diseases. Liver diseases are caused by many factors such as stress, food habits, consumption of alcohol drug intake, etc. In recent days, it could be found that it is very difficult to detect at an early stageas symptoms are very hard toidentify. The physician often slips to detect the liver disease which leads to improper medical treatment. Various data mining algorithms can be used to predict the various disease stages including early stage so that it could be help the physician to give the proper treatment.

II. RELATED WORK

Nazmun Nahar and Ferdous Ara et al., [1] implemented decision tree algorithms: J48, LMT, Random Tree, Random Forest, REPTree, Decision Stump and Hoeffding Tree to predict the liver disease. A comparative study also has been carried out among these algorithms. The system analyzes the performance of all the algorithms by measuring their accuracy, precision, recall, mean absolute error, F-measure, kappa statistic[14] and run time. From the analysis, it was found that the Decision Stump algorithm works effectively when compared to other algorithms and its accuracy rate is 70.67%.

Classification is an yet another technique used to distinguish the various types of data and predicting the accuracy [2,3,5]. Clustering is the process of making a group of abstract objects into classes of similarobjects. Association rule mining is process of findingrules that may govern associations and causal objects between set of items.A survey on various classification techniques to predict the liver diseases was discussed by Sindhujaetal., [6].Algorithms such as C4.5, Naive Bayes, Decision Tree, SVM, Back Propagation Neural Network and Classification and Regression Tree algorithms were compared and evaluated based on the criteria like speed, accuracy, performance and cost. It was concluded that C4.5 algorithm is best when compared to other algorithms.

Vijaranietal.,[5] implemented classification algorithms such as Navies Bayes and Support Vector Machine(SVM) using MATLAB 2013 tool to identify the disease and the algorithms' performance. Based on the comparison done by considering accuracy and execution time it was observed that SVM algorithm works better when compared to Navies Bayes.

Chieh-Chen Wua et al., [8] used the data from New Taipei city hospital and implemented machine learning algorithms such as random forest, Naive Bayes, Artificial Neural Networks (ANN), and logistic regression to predict Fatty liver disease(FLD). Based on the comparison, the performances ROC curve and accuracy. It is concluded that Random forest model showed higher performance RN(87.48) when compared to other classification models could help doctors to classify fatty liver patients for early treatments.

Sadiyah Noor Novita Alfisahrin ei al.,[7] designed a model in WEKA tool where the liver function test attributes such as age, gender, total bilirubin , direct bilirubin , alkaline phosphotase,total proteins, albumin Aspartateamino transferase,ratio albumin and globulinare considered with

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 10, October-2019

classification algorithms such as Decision Tree, Navies Bayes, NBTree to predict the liver disease. Also Chai-Squared ranking method was used to measure the impact of different attributes. The performance of each algorithmare evaluated by measuring execution time. Confusion matrix wsas used to measure the accuracy. As an experimental result, it shows that NBTree has highest accuracy and Navies Bayes algorithm gives the fastest computation time. Alice Auxilia etal.,[9] designed model to analyze various liver disease disorders using R tool. The datasets are well trained and tested using machine learning techniques such as decision tree, support vectormachine and Naive Bayes algorithm. Pearson correlation is applied to measure the accuracy, specificity and sensitivity of each algorithm and as a result the decision tree gives better accuracy than other classification algorithms.

Automated prediction and diagnosis of disease is the one of the most challenging aspect of medical data mining. Sina Bahramirad, etal.,[10] constructed a classification model based on two real liver patient dataset and used eleven algorithms like Logistic,Linear Logistic Regression, Gaussian Processes,Logistic Model Trees ,Multilayer Perceptron,K-STAR ,RIPPER, Neural Net, Rule Induction ,Support Vector Machine, Classification and Regression Trees. By using these algorithm, a comparative study is done for two types of datasets namely Andhra Pradesh state of India (AP dataset), California state of USA (BUPA dataset) and their performance were evaluated to measure accuracy, precision and recall. As a result, it is observed that AP dataset is better than BUPA dataset in terms of accuracy but in terms of precision andrecall BUPA dataset is more accurate than AP dataset.

Ashwani Kumar etal.,[11] implemented the info-gain feature selection method in classification algorithms like C4.5, Random forest, CART, Random tree and REP to get best algorithm. The datasets were divided into two set(70-30% and 80-20%) of training testing ratio to achieve better accuracy. Based on comparison,the performance are evaluated.As a result, it is concluded that an accuracy of 79.22% is achieved in Random forest using 80-20% training-testing data partition with 6 features.

Anju Guliaetal.,[12] designed a hybrid model with different algorithms like J48,MLP,SVM,RandomForest,BayesNet and a comparison between algorithmswas done to improve accuracy. The model was divided into three phases. In 1st phase, classification algorithm is applied on original dataset in 2nd that features influencing the liver disease are selected and in 3rd phase the result of original dataset with and without features are compared each other. Based on the experiments, the performance was evaluated by measuring the algorithms' accuracy. As a result, SVM algorithm is considered as the best before applying feature selection. After feature selection Random Forest algorithm is considered as the better performance algorithm than other algorithms.

Sanjay Kumar etal.,[13] used real liver diseases patient data for developing models using various classification algorithms to detect the liver disorders where the liver function test attributes contains age, gender ,DB, Alkphos, total bilirubin, SgptTP, ALB, A/G Ratio, Sgot,Selector field where considered with classification algorithm such as Naive-Bayes, Random forest, K-means, C5.0 and K-Nearest Neighbors(KNN). As a result, the Random Forest algorithm gives high accuracy before the implementation of adaptive boosting algorithm. But after the implementing the C5.0 algorithm gives better accuracy.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 10, October-2019

III. DIAGRAM

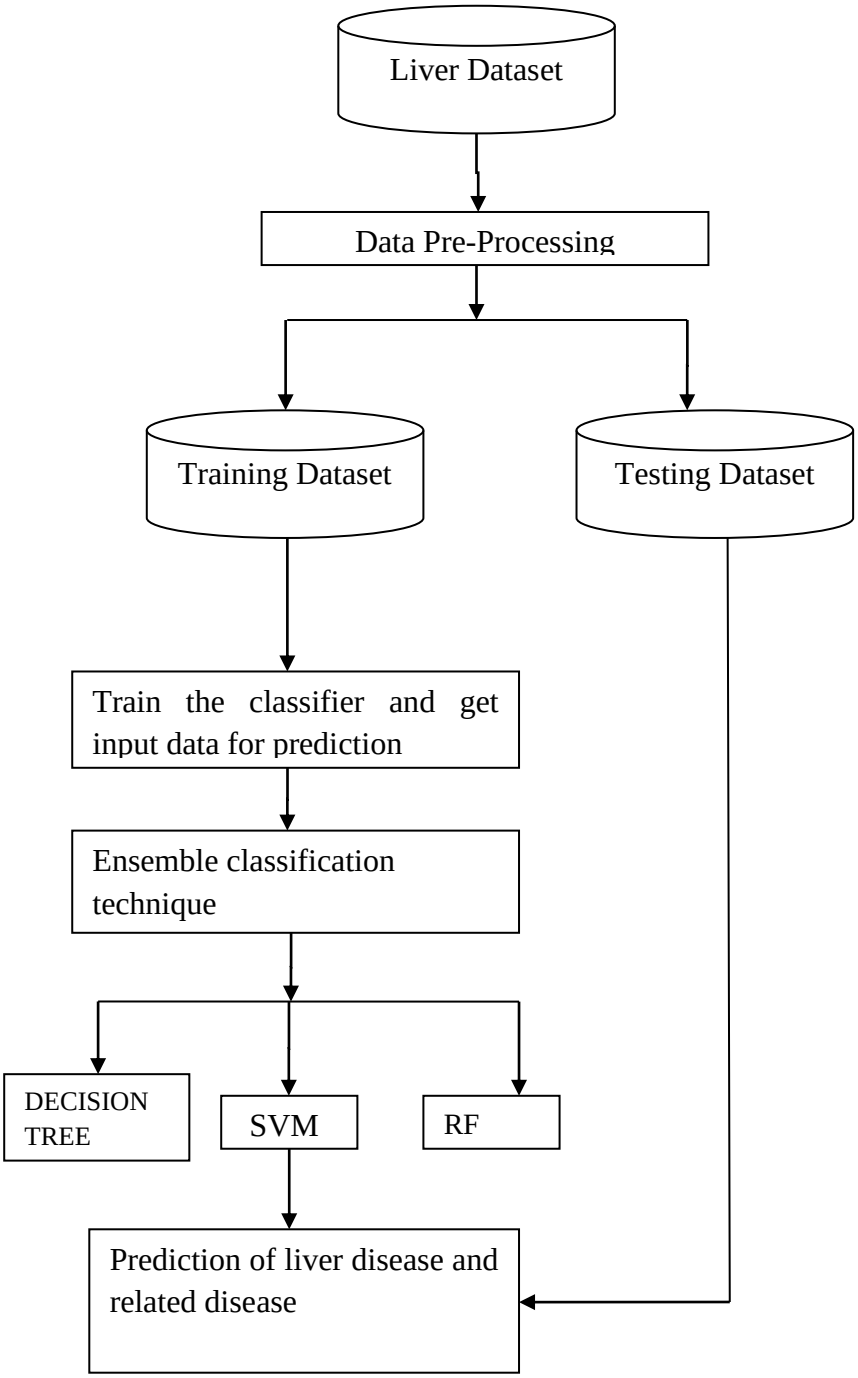

Fig 1 : Functional blocks of liver disease prediction

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 10, October-2019

IV. COMPARSION TABLE

S.No	**Title**	**Algorithm**	**Environment**	**Comparing Performance of Algorithm**	**Accuracy**
1	Liver disease prediction by using different Decision tree techniques.	J48, LMT, Random Tree, Random Forest, REPtree, DecisionStump and Hoeffding Tree	WEKA	Accuracy, Precision, Recall, MeanAbsolute Error, F-Measure, Kappa Statistic and Run time.	Decision Stump algorithm- 70.67%.
2	Liver disease prediction using SVM and Navies Bayes	Navies Bayes, Support Vector Machine	MATLAB 2013 tool	Accuracy and Execution time	SVM-79.66%
3	A Survey on Classification Techniques in Data Mining for Analyzing Liver Disease Disorder	C4.5, Naive Bayes, Decision tree, Support Vector Machine, Back propagation , Neural network, Classification and regression tree algorithm	WEKA	Speed, Accuracy, Performance and Cost	C4.5 algorithm has good accuracy when compared to the mentioned.
4	Data Mining Techniques For Optimatization of Liver Disease Classification	1.Decision Tree, Navies Bayes, NB tree algorithm. 2.Chai-Squared Ranking method	WEKA tool	Time complexity, confusion matrix to measure the accuracy	NB Tree has highest accuracy -67.01%.Navies Bayes algorithm gives the fastest computation time-0.04 Seconds
5	Prediction of fatty liver disease using machine learning algorithms	random forest , Naïve Bayes, artificial neural networks and logistic regression	WEKA tool	Receiver operating characteristic curve and its accuracy	Random forest model - 87.48,%
6	Accuracy Prediction using Machine Learning Techniques for Indian Patient Liver Disease	1.Decision Tree, Naïve Bayes, Random Forest, SVM and Artificial Neural Network. 2. Pearson correlation.	WEKA tool	accuracy, specificity and sensitivity	Decision tree-81%
7	Classification of Liver Disease Diagnosis: A Comparative Study	Logistic,Linear Logistic Regressio n,GaussianProcesses,Logistic Model Trees,Multilayer Perceptron,K-STAR	WEKA tool	Accuracy, Precision and Recall	AP dataset was slightly better than the BUPA dataset in terms of Accuracy.
8	Categorization of Liver Disease Using Classification Techniques	C4.5,Random forest, CART, Random Tree and REP	WEKA tool	Accuracy, Sensitivity, Specificity, Precision and F-measures	Random forest -79.22%
9	Effective Analysis and Diagnosis of Liver Disorder by Data Mining	Naive-Bayes, Random forest, K-means, C5.0 and K-Nearest Neighbors.	WEKA tool	Accuracy, Precision and Recall	Random Forest gives high accuracy before implementation adaptive boosting algorithm. But after the implementing the C5.0 algorithm gives better accuracy-75.19%
10	Liver Patient Classification Using Intelligent Techniques	J48,MLP,SVM,RandomForest ,BayesNet. FeatureSelection	WEKA tool	Accuracy	Before Feature selection: SVM-71.3551 After Feature Selection:Random Forest71.8696

V. CONCLUSION

This study represents the various machine learning and classification algorithms to detect liver diseases at early stage. The algorithms are analyzed and compared based on various factors like accuracy, execution time, precision and so on to achieve the best solution. The future work could include comparison of Decision tree, Adaptive neuro fuzzy inference system and K-nearest neighbor with respect to Accuracy, sensitivity, specificity and precision. The proposed method results could infer the best algorithm with highest accuracy.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 10, October-2019

REFERENCES

[1] Nazmun Nahar and Ferdous Ara, "Liver disease prediction by using different decision tree techniques", International Journal of Data Mining & Knowledge Management Process (IJDKP),Vol.8, No.2, March 2018.

[2] Divya, B, Kalaiselvi, R, "Review on Confidentiality of the Outsourced Data", Research Journal of Science and Engineering Systems, vol.1, pp.1-7, 2017.

[3] Hassoon, M, Kouhi, M S, Zomorodi-Moghadam, M, & Abdar, M, "Rule optimization of boosted c5. 0 classification using genetic algorithm for liver diseaseprediction". In 2017 International Conference on Computer and Applications (ICCA), pp. 299-305, IEEE, 2017.

[4] S. Dinesh1 , Metin KOK, "A Review on Different Parameters Effecting the VehicleEmission Gases of Different Fuel Mode Operations", Research Journal of Science and Engineering Systems, Vol, 3, 2018.

[5] Dr.S.Vijayarani, Mr.S.Dhayanand, "Liver disease prediction using SVM and Navies Bayes", International Journal of Science Engineering and Technology Research,Vol.4,Issue 4,April 2015.

[6] D.Sindhuja ,R. Jemina Priyadarsini ,"A Survey on Classification Techniques in Data Mining for Analyzing Liver Disease Disorder",International Journal of Computer Science and Mobile Computing, Vol.5 Issue.5, May- 2016.

[7] Sadiyah Noor Novita Alfisahrin,Teddy Mantoro,"Data Mining Techniques For Optimatization of Liver Disease",International Conference on Advanced Computer Science Applications and Technologies , 2013.

[8] Chieh-Chen Wu , Wen-Chun Yeh , Wen-Ding Hsu , Md. Mohaimenul Islam , Phung Anh (Alex) Nguyen , Tahmin Nasrin Posly , Yao-Chin Wang, "Prediction of fatty liver disease using machine learning algorithms ", Computer Methods and Programs in Biomedicine, Vol.170, Pages 23-29, March 2019.

[9] L. Alice Auxilia,"Accuracy Prediction using Machine Learning Techniques for Indian PatientLiver Disease", 2nd International Conference on Trends in Electronics and Informatics (ICOEI 2018)IEEE Conference,IEEE Xplore, ISBN:978-1-5386-3570-4.

[10] Sina Bahramirad, Aida Mustapha, Maryam Eshraghi, "Classification of Liver Disease Diagnosis: A Comparative Study", 2013 Second International Conference on Informatics & Applications (ICIA),Lodz, Poland,IEEE Explore,2013.

[11] Ashwani Kumar, Neelam Sahu,"Categorization of liver disease using classification techniques", International Journal for Research in Applied Science & EngineeringTechnology (IJRASET),Vol. 5, No.5, 2017.

[12] Anju Gulia, Rajan Vohra, Praveen Rani, "Liver Patient Classification Using IntelligentTechniques", International Journal of Computer Science and Information Technologies, Vol. 5, No. 4, 2014.

[13] Sanjay Kumar,Sarthak Katyal,"Effective Analysis and Diagnosis of Liver Disorder byData Mining", Proceedings of the International Conference on Inventive Research in Computing Applications (ICIRCA 2018).

[14] Sasikala B S, Vinai George Biju, C. M. Prashanth,"Kappa andAccuracy Evaluations of MachineLearning Classifiers", 2017 2nd IEEE International Conference On Recent Trends In Electronics Information & Communication Technology, pp. 19-20, 2017.

[15] Shambel Kefelegn, Pooja Kamat, "Prediction and Analysis of Liver Disorder Diseasesby using Data Mining Technique: Survey", International Journal of Pure and Applied Mathematics, Vol.118,No. 9, pp.765-770,2018.