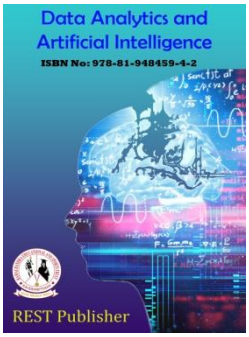
Data Analytics and Artificial Intelligence

Vol: 3(1), 2023

REST Publisher; ISBN: 978-81-948459-4-2

Website: http://restpublisher.com/book-series/daai/

DOI:https://doi.org/10.46632/daai/3/1/18



Analysis Machine Learning Methods For Forecasting Liver Disease

***P. Brinda, A. Keerthana, K. Kusuma Lalitha Sree, S. Losheni**

veltechhightechdr.rangarajandr,sakunthala engineering college,Chennai,TamilNadu, India.

*Corresponding Author Email: brinda13it@gmail.com

Abstract: *The majority of people worldwide are affected by chronic liver disease, which is the leading cause of death worldwide. It is now extremely challenging for researchers in the healthcare industry to predict diseases from the extensive databases. We use classification algorithms from machine learning to resolve this issue. Predicting liver disease is the primary objective of this project. We have implemented five machine learning algorithms: Naive Bayes, SVM, K-Nearest Neighbor, Logistic Regression, and Random Forest. The comparison of these classifier algorithms is entirely based on performance, classification accuracy and execution time. As a result, the objective of our project is to compare and contrast the overall performance of various machine learning algorithms in order to lessen the exorbitant cost of liver disease prediction*

Keywords: *Machine Learning, Liver Disease, Classification, Support Vector Machine, K- Nearest Neighbors, Naīve Bayes.*

1. INTRODUCTION

Consumption of alcoholic beverages, inhalation of potentially hazardous gases, drug use, and contaminated food all contribute to an ever-increasing number of liver disease patients. In the healthcare industry, it is difficult for researchers to predict diseases from the numerous clinical databases. Medical issues can be resolved quickly and easily using machine learning techniques, and the cost of the analysis can be reduced.We use machine learning algorithms to determine whether a person has liver disease or not in this study, which aims to predict results more accurately and reduce the cost of medical diagnosis. We have used Naive Bayes, SVM, Logistic Regression, Random Forest, and KNN to predict liver disease in this study. Our project's objective is to use machine learning algorithms to predict liver diseases from the dataset.

2.LITERATURE REVIEW

Machine learning has attracted a huge amount of researches and has been applied in various fields in the world. In medicine, machine learning has proved its power in which it has been employed to solve many emergency problems such as cancer treatment, heart disease, dengue fever diagnosis and so on. A. Aneeshkumar used a method to powerful type of liver and non-liver sickness dataset. Pre-processing technique is used to cleaning the records for powerful type, after cleaning the records. 15 attributes of actual clinical records are accrued from dataset. C4.five and Naive Bayes are the 2 algorithms utilized in his take a look at. He divided datasets into 3 one of a kind styles of ratio primarily based totally on common and popular deviation of every issue of each elegance and evaluated the accuracy. The bring about his take a look at after compare the accuracy, he stated C4.five is offers higher accuracy than Naive Bayes, as it offers extra accuracy with the minimal time taken.B. Bendi , proposed a Modified Rotation Forest set of rules to calculate the accuracy of the liver type strategies in UCI liver dataset the usage of the combination of function choice method and decided on type method set of rules.C. Onwodi Gregory [3] has proposed actual liver affected person datasets have been investigated for constructing type fashions so that it will expect liver analysis. Eleven records mining type algorithms have been implemented to the datasets and the overall performance of all classifiers are as compared in opposition to every different in phrases of accuracy, precision, and recall. Based at the experimental outcomes the type accuracy is discovered to be higher the usage of FT Tree set of rules evaluate to different algorithms., it additionally suggests the improved overall performance in step with the attributes and it offers 78.0% of Accuracy, 77.five% of Precision, 86.4%

of Sensitivity and 38.2% of Specificity outcomes respectively[8].D. Dr.S.Vijayarani, Mr.S.Dhayanand et al., [4] has proposed description of this studies paintings is to expect liver sicknesses the usage of type algorithms. The algorithms used on this paintings are Naïve Bayes and guide vector system (SVM Comparisons of those algorithms are completed and it's miles primarily based totally at the overall performance elements type accuracy and execution time. From the outcomes, this paintings concludes the SVM classifier is taken into consideration as a exceptional type set of rules due to its maximum type accuracy values. On the alternative hand, even as evaluating the execution time, the Naïve Bayes classifier wishes minimal execution time from the implementation outcomes it's miles found that the SVM is a higher Classifier for expect the liver sicknesses and evaluating the execution time, the Naïve Bayes classifier wishes minimal execution time.E. P.Rajeswari, G.SophiaReena et al.,[5]has proposed the records type is primarily based totally on liver disorder. The schooling dataset is advanced with the aid of using amassing records from UCI repository includes 345 times with 7 one of a kind attributes. This paper offers with outcomes withinside the area of records type received with Naïve Bayes algorithms .FT tree algorithms, and KStar algorithms and at the complete overall performance made recognize FT Tree set of rules while examined on liver sickness datasets, time taken to run the records for end result is speedy while evaluate to different set of rules with accuracy of 97.10 sed at the experimental outcomes the type accuracy is discovered to be higher the usage of FT Tree set of rules evaluate to different algorithms.F. Sa'diyah Noor NovitaAlfisahrin, Teddy Mantoro et al., [′□ □ ,] have proposed to become aware of if the sufferers have the liver sickness primarily based totally on the ten crucial attributes of liver sickness the usage of a Decision Tree, Naive Bayes , and NB Tree algorithms. The end result suggests NB Tree set of rules has the very best accuracy; but the Naïve Bayes set of rules offers the quickest computation time. For destiny take a look at, the overall performance of NB Tree set of rules may be the goal of development of the accuracy with the aid of using locating the maximum large issue in figuring out liver sickness sufferers. For destiny take a look at, the overall performance of NB Tree set of rules may be the goal of development of the accuracy with the aid of using locating the maximum large issue in figuring out liver sickness sufferers.

3.METHODOLOGY

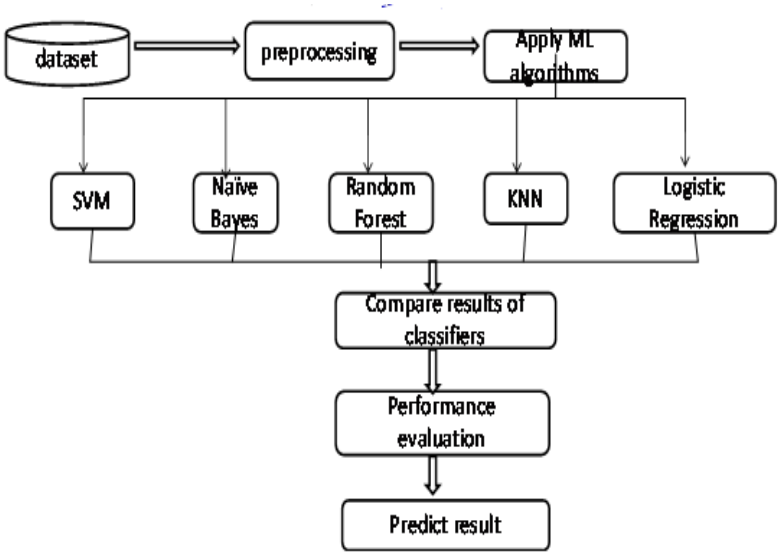


FIGURE 1. Block diagram

Data set collection

Indian Liver Patient Dataset (ILPD) has been taken from Kaggle. This dataset has 576 instances and 10 attributes. Attributes are Age, Gender, TB, DB, ALP, Sgpt, Sgot, TP, ALB and A/G Ratio. This dataset consists of Liver Function Test details (LFT).

Data preprocessing

We analyzed 583 records from patients with liver disease in this study, with 416 samples from liver disease patients and 167 samples from non-liver patients. The dataset is found to have four missing values and is cleaned using various data preprocessing techniques.

FIGURE 2. Data preprocessing

Implementing algorithm

In this module, we are using five algorithms SVM, KNN, Naïve Bayes , Random Forest, Logistic Regression.

Support Vector Machine

Support Vector Machine (SVM) is a supervised classification and regression machine learning algorithm. Even though we refer to them as regression problems, classification is more appropriate.

K-Nearest Neighbour

By calculating the distance between all of the training points and the test data, KNN attempts to determine the appropriate class for the data.

Naïve Bayes

A simple probabilistic classifier with a strong independent assumption known as a Naive Bayes classifier is based on applying Bayes' theorem.

Random Forest

Random forests are a machine learning regression technique for classifying data. At the time of training, liver data is transformed into a large number of decision trees, and the output of each tree is the class, which represents the mode of the classes

Logistic Regression

Based on previous observations of a data set, the statistical analysis technique known as logistic regression can predict a binary outcome, such as yes or no. Analyzing the relationship between one or more existing independent variables allows a logistic regression model to predict a dependent data variable.

Predicting the result

The accuracy, sensitivity, specificity, precision, and f1 measure, among other assessment methods, are used to predict the classification techniques' overall performance. The confusion matrix is used to select the exhibition assessment variables as a result. True Positive (TP) here: A person with a liver disorder is effectively identified by the prediction result. Positive False (FP): A person with a liver disorder is incorrectly identified as the end result of the prediction. Authentic Negative (TN): The prediction's final result effectively rejects the diagnosis of a liver disorder. Negative False (FN): The prediction's final result incorrectly rejects a person's diagnosis of liver disease.

$$Accuracy = \frac{(\text{True Positive} + \text{True Negative})}{(\text{True Positive} + \text{False Positive} + \text{True Negative} + \text{False Negative})}$$

$$Sensitivity = \frac{\text{True Positive}}{(\text{True Positive} + \text{False Negative})}$$

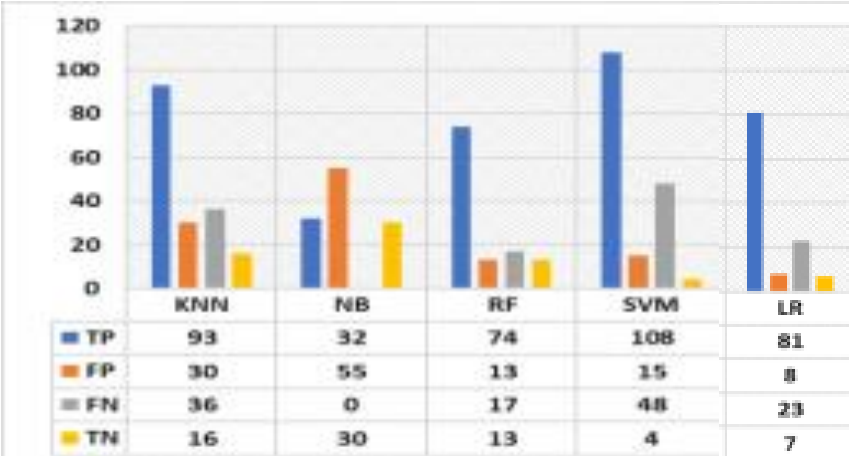


FIGURE 3. Predicting The Result

4.PROPOSED SYSTEM

The concept of machine learning is used in this system, and the models are first trained and then tested. In the end, the outcome will be predicted by the most accurate model.The system initially prompts you to enter your age, gender, total bilirubin, direct bilirubin, total proteins, albumin, A/G ratio, SGPT, SGOT, and alkphos, among other information. The user's blood test report can reveal the values of the final eight parameters listed here. The system takes these inputs from the user, compares them to the most accurate model's training dataset, and then predicts whether the person has a liver disease or not.

The system has following advantages:

No medical training is necessary: To use this application, you don't need to know anything about liver diseases or medical science to predict them. You will receive the prediction results by simply entering the requested information, some of which, like age and gender, are already known in the blood test report.

High precision: For the dataset we used to create this application, the system predicts the results with 100% accuracy. Even though there may be variations in some instances, the accuracy will still be high enough to be reliable on a large scale.

Immediate outcomes: After entering the information, the outcomes are predicted in a matter of seconds. In contrast to the conventional approach, you won't have to wait for a doctor to arrive.

5. FUTURE SCOPE

For the purpose of predicting liver disease, we will soon be able to collect the most recent data from a variety of locations worldwide. Our research has some implications for subsequent research in this area. Only a few well-known supervised ML algorithms have been applied here. In order to make better predictions about liver diseases and boost overall performance, more algorithms can be used. Additionally, the coronary heart dataset and disease category can be used to identify coronary heart diseases using the proposed method in this thesis.

6.CONCLUSION

Using six distinct supervised machine learning classifiers, the primary objective of this work is to develop a robust diagnostic tool for patients with chorionic liver disease. When we looked at how each classifier worked with the patient's data parameters, we found that the LR classifier has the highest extended order accuracy (75 percent) for predicting liver disease on the F1 degree and the lowest precision (53 percent). The application may offer the option to anticipate liver

disease beforehand and recommend the health condition. In low-income countries with a lack of medical foundations and only fewer specialists, this application may be extremely lucrative.

REFERENCES

[1].BendiVenkataRamana, Surendra. Prasad Babu. M, Venkateswarlu. N.B, A Critical Study of Selected Classification Algorithms for Liver Disease Diagnosis, International Journal of Database Management Systems (IJDMS), Vol.3, No.2, May 2011 page no 101-114

[2].Dhamodharan. S, Liver Disease Prediction Using Bayesian Classification, Special Issue, 4th National Conference on Advanced Computing, Applications & Technologies, May 2014, page no 1-3.

[3].Han, J.; Kamber, M., "Data Mining Concepts and Techniques". 2nd Edition, Morgan Kaufmann, San Francisco.

[4].Karthik. S, Priyadarishini. A, Anuradha. J and Tripathi. B. K, Classification and Rule Extraction using Rough Set for Diagnosis of Liver Disease and its Types, Advances in Applied Science Research, 2011, 2 (3): page no 334-345

[5].Rosalina. A.H, Noraziah. A. Prediction of Hepatitis Prognosis Using Support Vector Machine and Wrapper Method, IEEE, (2010), 2209- 22

[6].B. V. Ramana, M. R. P. Babu and N.B. Venkaeswarlu, "A Critical Study of Selected Classification Algorithms for Liver Disease Diagnosis", International Journal of Database Management Systems (IJDMS), Vol.3, no.2, (2011) , pp. 101-114.

[7].S.Dhamodharan, "Liver Disease Prediction Using Bayesian Classification", 4th National Conference on Advanced Computing, Applications & Technologies, Special Issue, May 2014. International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.8, No.2, March 2018 9

[8].P.Rajeswari and G.S. Reena, "Analysis of Liver Disorder Using data mining Algorithms", Global Journal of Computer Science and Technology, Vol.10, no. 14 (2010), PP. 48- 52.

[9].H. Sug, " Improving the Prediction Accuracy of Liver Disorder Disease with Oversampling", Applied Mathematics in Electrical and Computer Engineering, American-MATH 12/CEA12 proceedings of the 6th Applications and proceedings on the 2012 American Conference on Appied Mathematics (2012), PP. 331-335.

[10]. Hashem, M. Esraa, S. Mai, A study of support vector machine algorithm for liver disease diagnosis. American Journal of Intelligent Systems 4.1 (2014); 9-14.

[11]. P. Sajda, "Machine learning for detection and diagnosis of disease." Annu. Rev. Biomed. Eng. 8 (2006); 537-565.

[12]. UCI Machine Learning Repository. ILPD (Indian Liver Patient Dataset) Data Set. https://archive.ics.uci.edu/ml/datasets/ILPD+(Indian+Liver +Patient+Dataset)

[13]. P. Rajeswari ,G. Sophia Reena , Analysis of Liver Disorder Using Data Mining Algorithm,Global Journal of Computer Science and Technology,2010.

[14]. Sa'sdiyah Noor NovitaAlfisahrin,TeddyMantoro, Data mining Techniques For Optimatization of Liver Disease Classification,International conference on advanced Computer Science Application and Technologies,2013.

[15]. S. Dhamodharan , Liver Disease Prediction Using Bayesian Classification , National Confrence on Advanced Computing,Application&Technologies,2014

[16]. A.S.Aneeshkumar,Dr.C.JothiVenkateswaran , A novel approach for Liver disorder Classification using Data Mining Techniques ,Engineering and Scientific International Journal ,ISSN 2394- 7179,ISSN 2394-7187,2015.

[17]. Onwodi Gregory, Prediction of Liver Disease (Biliary Cirrhosis) Using Data Mining Technique, International Journal of EmergingTechnology&Research, ISSN (E):2347-5900, ISSN (P):2347-6079, 2015.

[18]. Dr.S.Vijayarani,Mr.S.Dhayanand,Liver Disease Prediction using SVM and Naïve Bayes Algorithms, International Journal of Science, Engineering and Technology Research(IJSETR), 2015.

[19]. Han, J.; Kamber, M., "Data Mining Concepts and Techniques". 2nd Edition, Morgan Kaufmann, San Francisco.

[20]. P. Rajeswari ,G. Sophia Reena , Analysis of Liver Disorder Using Data Mining Algorithm,Global Journal of Computer Science and Technology,2010.

