International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Special Issue: | Apr 2023 www.irjet.net p-ISSN: 2395-0072

International Conference on Recent Trends in Engineering & Technology- 2023 (ICRTET-3)
Organised by: VSM College of Engineering, Ramachandrapuram

A Comparative Analysis of Liver Diseases by Using Machine Learning Techniques

B.Prasanna Rani [1], K.Manasa [2], D.Bhargav Gowd [3], K.V.Padma Sai [4], V.Likitha Sri [5], N.S.Sri Durga [6]

[1]Assistant Professor, [2,-6]Students B.Tech. Computer Science Engineering, V. S. M. College of Engineering, Ramachandrapuram, A.P, India

---

Abstract - *In a human body function of the liver is important. Many persons are suffering from liver disease, but they don't know it. The identification of liver diseases in the early stage helps a patient get better treatment. If it is not diagnosed earlier, it may lead to various health issues.To solve these issues, physicians need to examine whetherthe patient has been affected by liver disease or not, basedon the multiple parameters. In this paper, we classify the patients who have liver disease or not by using different machine learning algorithms by comparing the performance factors and predicting the better result.*

Key Words: Machine Learning, Liver Disease, Classification, Supervised learning, Regression, Random Forest, Decision Tree, Support Vector Machine, K-Nearest Neighbors.

1. INTRODUCTION

Patients with liver problems that are difficult to detect in the early stage will help to continue their function normally even if they are partially damaged. There are chances for a patient surviving a liver disease to be better if they are diagnosed early.The liver is an important organ that performs many functions energy storage, linked to metabolism, and waste cleansing. It also aids in the digestion of food, the change of food into energy, and the storing of energy until needed. It also helps in the removal of potentially dangerous compounds from our bloodstream. The disease may be a general term that refers to any condition affecting the liver. The liver is the largest organ of the body and it is essential for digesting food and releasing the toxic element of the body. The viruses and alcohol use lead the liver towards liver damage and lead a human to a life-threatening condition. There are many types of liver diseases whereas hepatitis, cirrhosis, liver tumors, liver cancer, and many more. Among them liver diseases and cirrhosis as the main cause of death [1]. Therefore, liver disease is one of the major health problems in the world. Every year, around 2 million people died worldwide because of liver disease [2]. According to the Global Burden of Disease (GBD) project, published in BMC Medicine, one million peoples are died in 2010 because of cirrhosis and million are suffering from liver cancer [3].

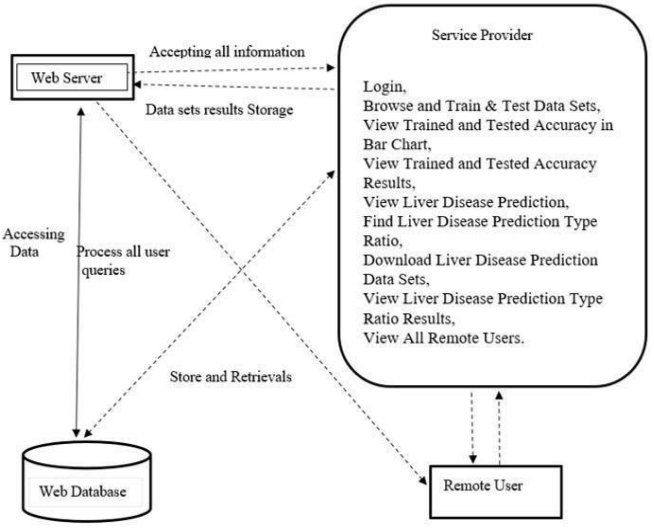

Fig -1: System Architecture

2. IDENTIFY, RESEARCH AND COLLECT IDEA

In [1], Nazmun Nahar,implemented by using various decision trees techniques like LMT, J48, Hoeffding Tree, Decision Stump, and Random tree. for calculating expected time predication of disease affected to liver finally, the Decision Stump gives the highest accuracy results among other techniques.

In [2], A Saranya,explained the applications in data mining techniques and also used Medical Data Mining (MDM) to diagnose liver diseases. This technique includes prediction in the early stage, the existence and also complexity of the disease which helps partial assistance to the physicians.

In [3], S Dhamodhar considers three major liver diseases like cirrhosis, hepatitis, and liver cancer. The fundamental purpose of this forecast is to find the type of disease by using classifications techniques such as cirrhosis, hepatitis, liver cancer, and "no disorders."Then compare the accuracy of the FT and Naive Bayes tree algorithms and shows that the Naive Bayes algorithm accuracy is significantly higher than that of the other methods.

In [4], Kemal Akyol, Yasemin Gultepe by using the dataset which has shown a balanced result by using sampling technique for getting accuracy. the Stability Selection technique is used for selection based on attributes. For improving the performance, a blend of Stability Selection and Random Forest methods is used.

In [5], Shambel Kefelegn, Pooja Kamat for getting better results different data mining classification techniques are compared with the earlier liver prediction methods. The accuracy is measured with the help of confusion matrices for getting the better performance of the accuracy.

In [6], Fadl Mutaher Ba-Alwi using various machine learning algorithms compared the Hepatitis prognostic data among them. In that Naive Bayes, technique gave good accuracy and also takes less time to build a model.

In[7],K.Thirunavukkarasu, Used different classification techniques for predicting liver diseases. They compare the results of accuracy score and confusion matrix with Logistic Regression, SVM, and K-Nearest Neighbour. BendiVenkata,used different classifications algorithms, they checked the accuracy, precision, sensitivity, and specificity on liver datasets.

In [8], Tapas Ranjan Baitharua has proposed an Intelligent medical decision support system to help physicians diagnose liver disorders through a learning pattern technique. In this,several classification techniques are used to compare the effectiveness, correction rate, and also accuracy for the data is analysed with different scenarios.

3.CLASSIFICATION ALGORITHMS

3.1. Logistics Regression (LR)

Calculated Regression was for the most part utilized in natural research and applications in the mid-20th century [8]. Logistic regression can deal with any number of numerical as well as absolute factors. In addition, it introduces a discrete parallel item somewhere in the range of 0 and 1.

3.2. Naive Bayes (NB)

Naive Bayes is one of the basic, best and ordinarily utilized, AI techniques. It is a probabilistic classifier that classifies utilizing the speculation of restrictive freedom with the pre-trained datasets. From this time forward, Naive Bayes classifiers are procedures for finding the conventional arrangement of grouping issues.

3.3. K-Nearest Neighbors (KNN)

KNN is one of the most fundamental occasion-based classification algorithms in Machine Learning. In any case, theKNN takes a shot at the idea that examples are near fit in similar examples class. A KNN sorts an example to the class that is most decided among K neighboring. K is a limitation for adjusting the classification algorithms

3.4. Decision Tree (DT)

Decision Tree calculation has a place with the supervised learning algorithms. In contrast to other supervised learningalgorithms, a decision tree algorithm can be utilized for taking care of regression and classification issues as well.

3.5. Random Forest (RF)

Random forests or random decision forests are an ensemble learning technique for classification, regression and different assignments that works by developing a huge number of decision trees at training time and yielding the class that is the method of the classes (classification) or mean forecast (regression) of the individual trees.

3.4. Kernel Support Vector Machine (SVM)

Kernel SVM is a supervised learning calculation. Kernel Function is a method used to take data as input and transform it into the required form of processing data. “Kernel” is used due to a set of mathematical functions used in Support Vector Machine providing the window to manipulate the data.

It can utilize for both grouping or relapse issues however generally it is utilized in characterization issues. SVM function admirably for some, human services issues and can comprehend both linear and non-linear issues. SVM grouping strategy which is an endeavor to pass a linearly separable hyperplane to order the dataset into two classes. At long last, the model can without a doubt gauge the objective groups (labels) for new cases.

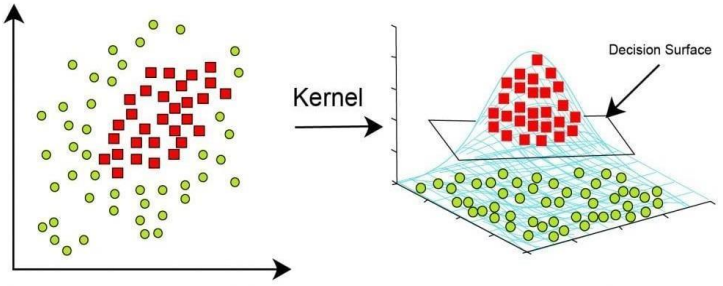

Fig -2: Kernel SVM

TABLE I. VALUES USED IN THE DATASET

No	Data	Data Type
1	Age	int64
2	Gender	Object
3	Total Bilirubin	float64
4	Direct Bilirubin	float64
5	Alkaline_Phosphotase	int64
6	Alamine_Aminotransferase	int64
7	Aspartate Aminotransferase	int64
8	Total_Protiens	float64
9	Albumin	float64
10	Albumin_and_Globulin_Ratio	float64
11	Result (Datasets)	Int64(1or2)

4.PERFORMANCE EVALUATION AND RESULTS

In this, we predict whether the patient has a liver disease or not using different machine learning techniques. By using Random Forest, KNN, LR, Naive Bayes, Kernel SVM, Decision tree in the above-said algorithms we evaluated the following parameters Accuracy, Root Mean Square Error, Mean Absolute Error, and Root Relative Squared Error, Table II shows the accuracy value of all techniques and Fig.3 shows the analysis of the accuracy values of various Machine Learning Techniques.

TABLE II. ACCURACY VALUES OF DIFFERENT ALGORITHMS

Techniques	Accuracy
Logistic Regression	.67
KNN	.60
Kernel SVM Approach	.70
Navi Bayes	.56
Decision Tree	.58
Random forest	.69

The Accuracy values of various algorithm is calculated as follows:

Accuracy = (TP + TN) / (TP + TN + FP + FN)

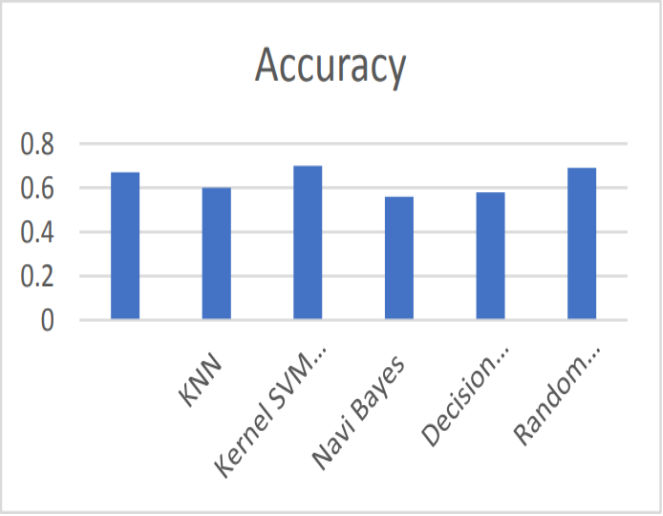

Fig -3: Analysis of Accuracy values

Mean Absolute Error MAE It calculates the average size of mistakes in a set of forecasts without taking the direction of the errors into account. Fig. 4 shows the analysis of the MSE values of various Machine Learning Techniques.

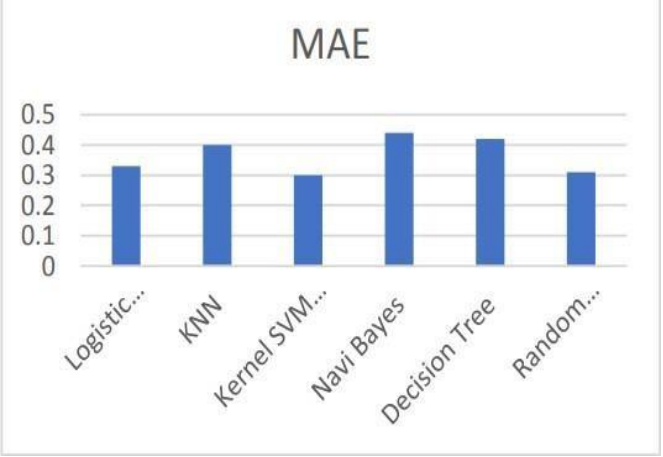

Fig -4: Analysis of MAE Accuracy values

Root Mean Squared Error (RMSE): The square root of the mean of the squared differences between actual and predicted outcomes is calculated. Fig. 5 shows the analysis of the RMSE values of various Machine Learning Techniques

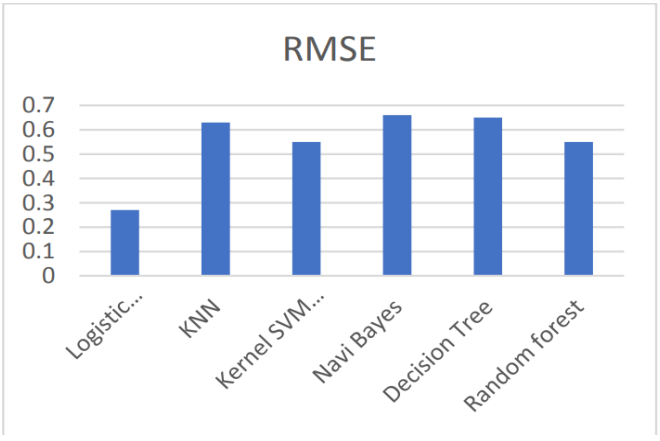

Fig -5: Analysis of RMSE Accuracy values

Relative Squared Error (RSE): The relative squared error (RSE) in this simple predictor is used. It takes the sum ofthe real values. Then the relative squared mistakes are taken as the total squared error and regularized by dividing by the total squared mistakes of the simple predictor.

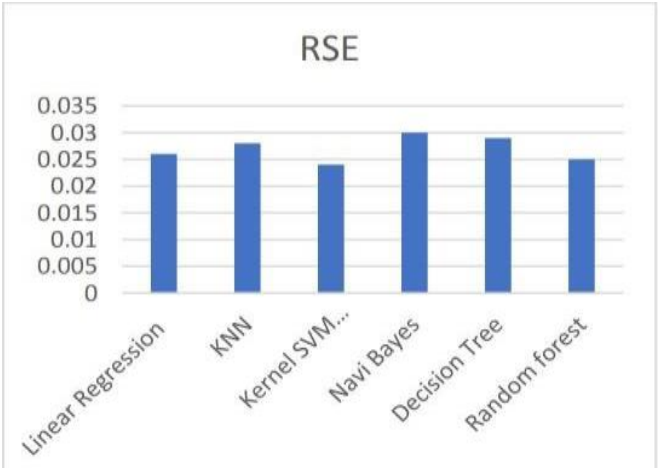

Fig -6: Analysis of RSE Accuracy values

5. CONCLUSION

In this paper we have studied some classification algorithms like SVM, Random Forest, KNN, LR, Naive Bayes, Kernel SVM, Decision tree to predict the patent hasa Live disease or not. The early prediction gives the physicians to take the necessary steps to save the life of the patient. In this, the Kernel SVM approach produces better accuracy results than the other techniques.But we need future enhsncement for real world applications.

ACKNOWLEDGEMENT

We are appreciative to our Division of Computer Science Engineering for their help and giving us a chance to make things simpler. While looking about this point we found out about different significant and fascinating realities. The tools and the web services provided are of much help.

REFERENCES

[1] Joel Jacob et al., "Diagnosis of Liver Disease Using Machine Learning Techniques", International Research Journal of Engineering and Technology (IRJET), Volume: 05, Issue: 04,2018, pp 4011-4014.

[2] Pragati Bhagat et al.," System for diagnosis of Liver Disease Using Machine Learning Technique", International Research Journal of Creative Research Thoughts", ISSN: 2320-2882, pp25-30.

[3] Nazmun Nahar and Ferdous Ara, "Liver Disease Prediction Using Different Decision Tree Techniques", International Journal of Data Mining & Knowledge Management Process Vol.8 No.2(March 2018).

[4] A Saranya, G.Seenuvasan, "A Comparative Study Of Diagnosing Liver Disorder Disease Using Classification Algorithm", International Journal Of Computer Science and Mobile Computing, Vol. 6 Issue 8mpage no 49-54(August 2017).

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

[6] Kemal Akyol, Yasemin Gultepe," A Study on Liver Disease Diagnosis based On Assessing the Importance Of Attributes", I.J. Intelligent systems, and Applications, Vol. 11, Page No 1- 9(2017).

[7] Shambel Kefelegn, Pooja Kamat, "Prediction and Analysis of Liver Disorder Diseases by using Data Mining Technique: a survey", International Journal of pure and applied mathematics, volume 118, No 9,765-770, 2018.

[8] Fadl Mutaher Ba-Alwi et al....," Comparative Study for Analysis the Prognostic in Hepatitis Data: Data Mining Approach", International Journal of Scientific & Engineering Research, Vol 4mIssue 8 ISSN 2229-5518(2013).

[9] Bendi Venkata Ramana, et al. ..., "A Critical Study of Selected Classification Algorithms for Liver Disease Diagnosis",International Journal of Database Management Systems, Vol 3,(May 2011).

[10] K. Thirunavukkarasu et al ," Prediction of Liver Disease using Classification Algorithms", International Conference on Computing Communication and Automation (ICCCA), https://doi.org/10.1109/CCAA.2018.87776552018.

[11] Tapas Ranjan Baitharua, Subhendu Kumar Panib, "Analysis of Data Mining Techniques For Healthcare DecisionSupport System Using Liver Disorder Dataset",

BIOGRAPHIES

Team Guide, B. Prasanna Rani is Assistant Professor – Computer Science Engineering in VSM College of Engineering, Ramachandrapuram.

Team Leader, K. Manasa is Computer Science Engineering student in VSM College of Engineering, Ramachandrapuram.

Team member, D.Bhargav Gowd is Computer Science Engineering student in VSM College of Engineering, Ramachandrapuram.

Team member, K.V.Padma Sai is Computer Science Engineering student in VSM College of Engineering, Ramachandrapuram.

Team member, V.Likitha Sri is Computer Science Engineering student in VSM College of Engineering, Ramachandrapuram.

Team member, N.Surya Sri Durgais Computer Science Engineering student in VSM College of Engineering, Ramachandrapuram.