2022 International Conference on Emerging Smart Computing and Informatics (ESCI)
AISSMS Institute of Information Technology, Pune, India. Mar 9-11, 2022

Comparative Analysis of Liver diseases by using Machine Learning Techniques

D. Palanivel Rajan,
Department of CSE,
CMR Engineering College,
Hyderabad, Telangana, India,
*palanivelrajan.d@gmail.com.

K S Kannan,
Department of CSE,
CMR Engineering College,
Hyderabad, Telangana, India,
saikannan2012@gmail.com.

P. Divya
Department of CSE
Bannari Amman Institute of Technology
Sathyamangalam, Erode, TamilNadu
divisrecme@gmail.com

Velliangiri. S
Department of CSE
BV Raju Institute of Technology
Narasapur, Telangana, India
velliangiris@gmail.com

***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 whether the patient has been affected by liver disease or not, based on 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. The liver dataset is retrieved from the Kaggle dataset.**

Keywords—Liver Diseases, Machine Learning algorithms, Physicians, Patients

I. 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.[1] 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 [2]

A. Functions of Liver:

These are some of the functions of the Liver:

- It produces a component in the immune system that can combat illness.
- Producing the proteins that aid in blood coagulation.
- Red blood cells that are old or damaged are broken down.
- Excess blood sugar is stored as glycogen

The liver and its activities can be harmed by a variety of disorders. Some people respond well to treatment, while others do not. Fig. 1 shows the condition of Normal and affected liver with diseases. Some of the common conditions that affect the liver are discussed below:

1) Autoimmune hepatitis

The immune system of the body attacks itself and destroys healthy liver tissue in this disease. Cirrhosis and other liver damage can result from autoimmune hepatitis.

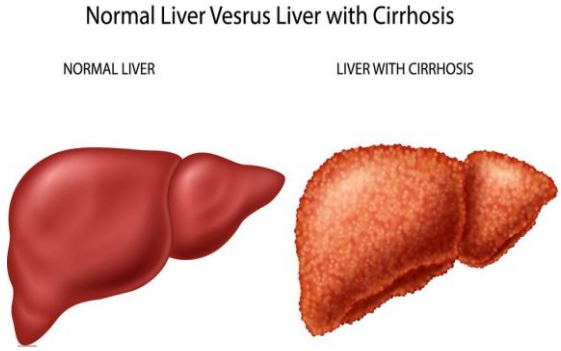

Fig.1. Difference between normal and affected liver

2) Cirrhosis

In this healthy liver, tissues are affected and changed as scar tissue due to chronic hepatitis, Long-term excessive alcohol consumption, and rare hereditary disorders such as Wilson's disease are all examples that might cause this problem.

3) Hemochromatosis

An overabundance of iron builds up in the body as a result of this disorder. The liver might be harmed by too much iron.

4) Hepatitis A

viral infection which causes swelling in the liver is known as Viral hepatitis. There are some types of hepatitis, like A, B, C, D, and E. Each has its own set of causes and consequences.

Hepatitis A is most common in underdeveloped nations with poor sanitation and access to clean drinking water. Hepatitis A is usually treatable without causing liver damage or long-term consequences

5) Hepatitis B

It can be an infection either short-term or long-term. It can also be contracted by sharing the needles with others or inadvertently injecting oneself with a contaminated needle. These serious complications, which include a cause of liver

failure and cancer, can occur as a result of illness. There is a vaccine available to prevent the sickness.

6) Hepatitis C

Hepatitis C is a viral infection that can be either acute or persistent. It's disseminated most usually through coming into touch with hepatitis C virus-infected blood, such as by using dirty needles to inject drugs or apply tattoos. Due to this liver failure, and liver cancer are all possible.

7) Non-alcoholic fatty liver disease and NASH

The excess fat that builds up in the liver will damage the liver, which causes swelling. Fatty liver disease may cause scarring or fibrosis due to non-alcoholic steatohepatitis. Type 2 diabetes-related diseases may cause due to this problem.

B. Symptoms of liver conditions

There are so many types of liver disorders, which show symptoms like flu- and cause more serious damage in the liver which includes jaundice and dark-colored urine.

The following are some of the signs and symptoms of liver disease:

- fatigue
- a decrease in appetite
- vomiting
- Pain in joint
- stomach ache or discomfort
- Bleeds in the nose
- aberrant blood vessels on the surface of the skin (spider angiomas)

Symptoms that are more severe include:

- The skin and eyes turn a yellowish color (jaundice)
- bloating in the stomach (ascites)
- Leg swollenness (edema)
- Gynecomastia is a term used to describe a condition in which a man develops (when males start to develop breast tissue)
- Enlarged liver (hepatomegaly)

II. LITERATURE SURVEY

Nazmun Nahar, et al. [3] 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.

A Saranya, et al. [4] 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.

S. Dhamodharan [5] 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.

Kemal Akyol, Yasemin Gultepe [6] 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.

Shambel Kefelegn, Pooja Kamat [7] 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

Fadl Mutaher Ba-Alwi, et al. [8] 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.

K. Thirunavukkarasu, et al. [9] 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.

Bendi Venkata, et al. [10] used different classifications algorithms, they checked the accuracy, precision, sensitivity, and specificity on liver datasets.

Tapas Ranjan Baitharua et al. [11] 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

A diagnostic support system [12] was developed with the support of a number of models with help of neural, which is helped to the physicians for diagnosis on the liver in the medical field.

M. Banu Priya P, et al. [13] Using a root mean error value, root mean square value, the accuracy is calculated, and better accuracy is produced with the support of the PSO features selection technique.

Dietterich, Thomas G [14] states that the ensemble learning technique produces a better performance than the other single classifier techniques with the Bayesian averaging, error-correcting output coding, boosting, and bagging. In this paper, the author analyses existing ensemble approaches with some novel experiments to determine why Adaboost does not overfit quickly.

III. CLASSIFICATION TECHNIQUES AND METHODOLOGY USED FOR IMPLEMENTATION

A. System Architecture:

The flow chart (Fig. 2) of the system shows the working methodology for finding better results with different techniques [15].

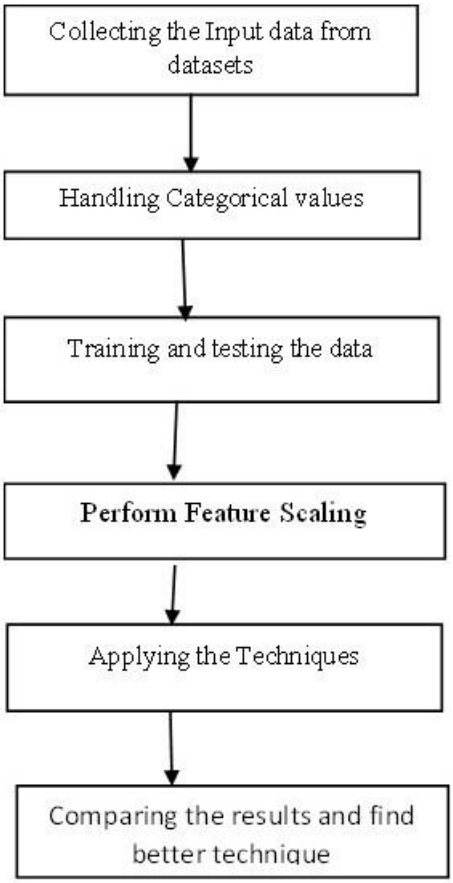

Fig 2 Flow chart for system architecture

To analysis about liver disease, the data is retrieved from the Indian Liver Patient from the Kaggle. Here the patient has been characterized based on diseases as either 1 or 2. The values used in the dataset are given in Table I. The gender attribute is transformed to an integer value during the data pre-processing stage (0 and 1). The overall procedure of the proposed system is depicted in Fig.2. The workflow of the proposed system is as follows a collection of data sets, Handling Categorical values, Splitting the data for Training, and Testing. Perform feature selection and apply the machine learning techniques and compare the predicted result and find better accuracy [16].

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)

B. *Logistic Regression*

It is a supervised Machine learning technique applied for both classification and regression kinds of problems, but it is used for classification types of problems. this model is applied to predict the categorical dependent variable with support of independent variables the output should be 0 or 1. It can be used in situations where the probability of two classes is needed. For example, if it will rain today or not, 0 or 1, true or false, and so on. The concept of Maximum Likelihood estimation supports logistic regression. The observed data should, in this estimation, be the most expected one.

C. *Naive Bayes:*

This Classifier technique is effective when only a small amount of training data is required to derive approximation parameters. With highly scalable model creation, it can tackle a wide range of challenging real-world problems. Because it can consider the complete property in each class, this classifier can solve zero conditional probability issues. The formula for the Naïve Bayes theorem is as follows:

$$P(A|B) = P(B|A)P(A)/P(B)$$

D. *KNN:*

This is a pattern recognition system that involves the training datasets for finding the k closest relatives in new conditions. When using k-NN for classification, we must calculate the location data within the nearest neighbor's category. If k = 1, then it will be allocated to the class closest to the value 1. A plurality vote of its neighbors classifies the K.

E. *Decision Tree:*

This is a supervised learning approach for classification tasks that can structure the classes accurately. The data will be sorted in two related categories at a time, beginning with the "tree stem" and progressing to "branches" and "leaves," where the categories become increasingly finite in a similar manner.

F. *Random Forest:*

In this method first, we need to build a multitude of decision trees with the training data, then we need to fit our new data within one of the trees as a "random forest". By averaging the value, it connects data to the nearest tree. This strategy is advantageous for resolving the issue of decision trees excessively "pushing" data points into a category.

G. *Kernel SVM:*

In this technique, a two-dimensional plane is placed in a higher-dimensional space. This technique is difficult to classify non-linear data. To overcome this issue kernel project the Non-linear data into a high dimensional space and make it is easy to classify the data where it needs to be linearly divided by a plane. This is accomplished mathematically through the use of Lagrangian multipliers and the Lagrangian formula.

IV. 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:

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

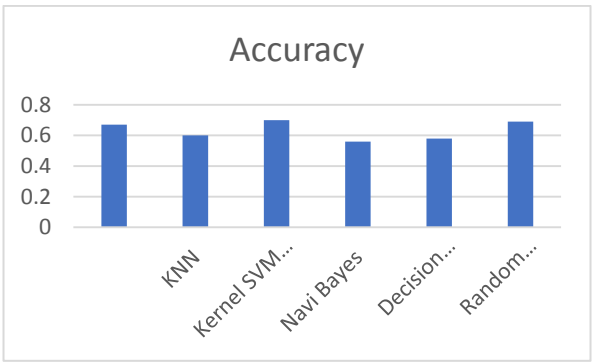

Fig.3. Analysis of Accuracy values

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. It's the weighted average of the absolute discrepancies between predicted and actual observation over the test sample. In Table III shows the MSE value of all techniques and Fig. 4 shows the analysis of the MSE values of various Machine Learning Techniques.

$$MAE = 1/n \sum |yj - y\hat{}j|a$$

TABLE III. MAE VALUES OF DIFFERENT ALGORITHMS

Techniques	MAE
Logistic Regression	.33
KNN	.40
Kernel SVM Approach	.30
Navi Bayes	.44
Decision Tree	.42
Random forest	.31

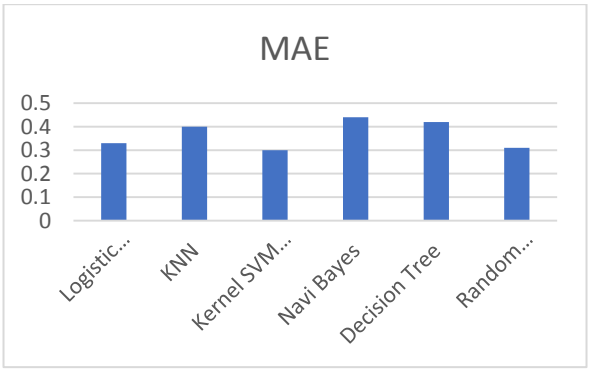

Fig.4. Analysis of MAE Accuracy values

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

TABLE IV. RMSE VALUES OF DIFFERENT ALGORITHMS

Techniques	RMSE
Logistic Regression	.27
KNN	.63
Kernel SVM Approach	.55
Navi Bayes	.66
Decision Tree	.65
Random forest	.55

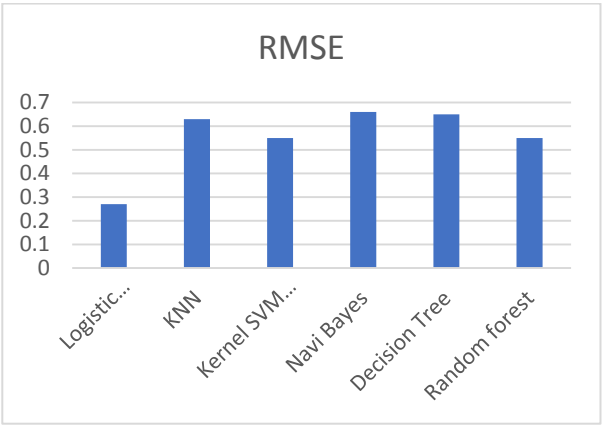

Fig.5. Analysis of RMSE Accuracy values

Relative Squared Error (RSE): The relative squared error (RSE) in this simple predictor is used. It takes the sum of the 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. It compares the mistakes between the models whose mistakes are measured in the different units. Table V shows the RSE value of all techniques and Fig. 6 show the analysis of the RSE values of various Machine Learning Techniques.

TABLE V. RSE VALUES OF DIFFERENT ALGORITHMS

Techniques	RSE
Linear Regression	.026
KNN	.028
Kernel SVM Approach	.024
Navi Bayes	.030
Decision Tree	.029
Random forest	.025

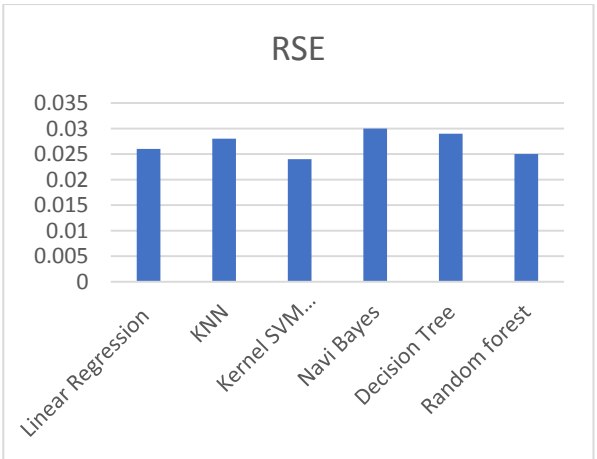

Fig.6. Analysis of RSE Accuracy values

V. CONCLUSION

In the above system, we have studied some classification algorithms like SVM, Random Forest, KNN, LR, Naive Bayes, Kernel SVM, Decision tree to predict the patent has a 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 to predict more accurate results. To get better results we need to use some advanced techniques for predicting the diseases

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 Decision Support System Using Liver Disorder Dataset", International Conference on Computational Modeling and Security, 2016.

[12] BUPA Liver Disorder Dataset. UCI repository machine learning databases

[13] M. Banu Priya P, et al.," Performance Analysis of Liver Disease Prediction using Machine learning Algorithms", International Research Journal of Engineering and Technology, Vol. 05 Issue:01(Jan. 2018).

[14] Dietterich, Thomas G. "Ensemble methods in machine learning." International workshop on multiple classifier systems. Springer, Berlin, Heidelberg, 2000

[15] Velliangiri, S., Karthikeyan, P., Joseph, I. T., & Kumar, S. A. (2019, December). Investigation of Deep Learning Schemes in Medical Application. In 2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE) (pp. 87-92). IEEE.

[16] Joseph, S. I. T., Sasikala, J., Juliet, D. S., & Velliangiri, S. (2021). Hybrid spatio-frequency domain global thresholding filter (HSFGTF) model for SAR image enhancement. Pattern Recognition Letters, 146, 8-14.