



iJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** III **Month of publication:** March 2023

DOI: <https://doi.org/10.22214/ijraset.2023.49954>

www.ijraset.com

Call: ☎ 08813907089

E-mail ID: ijraset@gmail.com

A Comparison of Machine Learning Algorithms for Predicting Liver Disease

Shaik Anjum Niha¹, Kolla Jhansi Lakshmi², Pidathala Glory Bless³, Thamatham Sai Lakshmi⁴, Yarramasu Mounika Chowdary⁵, Mr. M. Purnachandra Rao⁶

^{1, 2, 3, 4, 5} Undergraduate Students, ⁶ Associate Professor Department of Information Technology, KKR & KSR Institute Of Technology And Sciences, Guntur-522017, India

Abstract: The liver is a vital internal organ of the human body that performs two primary functions: blood filtration and digestion. The inhalation of contaminated air, narcotics, excessive alcohol consumption, and obesity all contribute to liver damage. The most common cause of death in the world, chronic liver disease kills a lot of people. A variety of factors that affect the liver contribute to the development of this disease. It causes liver encephalopathy, jaundice, aberrant nerve function, bloody coughing or vomiting, renal failure, liver failure, and many other symptoms. The diagnosis of this illness is very expensive and difficult. Medical experts struggle to identify the side effects of liver disease because people experience pain even while suffering from the disease. Because of advancements in machine learning technologies, early detection of liver disease is now possible. so that the dangerous illness can be rapidly and easily diagnosed. In healthcare, this will be more helpful. A medical expert system and a department can both be employed in a remote location. Early-stage prediction will prevent it from being too late and help save lives. There are various types of machine learning for the diagnosis of liver disease, such as Logistic Regression, KNN, random forest, and they differ in accuracy, precision, and sensitivity. According to the analysis's findings, the Logistic regression had the most accuracy. Additionally, the primary goal of the current work is to use clinical data for the prediction of liver disease. Through our analysis, we explore several ways to describe such data.

Keywords: liver diagnosis, machine learning, logistic regression, KNN, random forest.

I. INTRODUCTION

The largest internal organ in a human is the liver. It has a football-like appearance on the right side of our body. It plays a crucial part in the elimination of dangerous compounds in addition to manufacturing a variety of chemicals to properly digest our meals. proteins are created, bile and albumin are made, bilirubin, carbs, and fat are metabolized. It possesses a special mechanism for self-regeneration, making it the only visceral organ in vertebrates. The rebuilding process normally takes 7 to 14 days to be completed without any functional loss if at least 35% of the tissue in the human body is still there. In terms of functionality, the liver is the most complex organ in the body. Since liver malfunction can cause a variety of problems, including fatty liver disease, which affects diabetics and the obese, maintaining liver health is essential. According to the most recent survey data from the World Health Organization, which was published in 2017, liver disease- related deaths represent 2.95% of all fatalities, ranking India as 63rd internationally. The liver, the largest organ in the body, is crucial for eliminating toxins from the body and for the digestion of meals. Alcohol consumption combined with virus use damages the liver and can be fatal. Hepatitis, cirrhosis, liver tumors, liver cancer, and many other conditions can affect the liver. Cirrhosis and liver disorders are the main killers. Consequently, one of the greatest health issues in the globe is liver disease. Scarring of the liver is referred to as cirrhosis. a soft, healthy tissue called connective tissue in its place. The liver will fail and be completely incapable of functioning if liver cirrhosis is not treated as it worsens. Cirrhosis can cause a variety of other issues in addition to liver damage. Cirrhosis of the liver symptoms in some individuals may be the primary cause of their condition.

Around 2 million people worldwide pass away from liver disease each year. According to the Global Burden of Disease (GBD) research, which was published in BMC Medicine, cirrhosis caused one million deaths in 2010 and liver cancer affects another million people. Deaths from chronic liver diseases are more common among minorities in the US. Early-stage liver disease can occasionally be difficult to diagnose. Studies on artificial immune systems and genetic algorithms have been described for the identification of liver disease. Diagnoses differ according on different classifiers and data sets. Liver disease is becoming a serious and potentially fatal problem on a global scale. Algorithms for machine learning can aid in early diagnosis to lower risk. Analysis of earlier studies showed poor performance. As a result, the goal of this research is to produce more pleasing performance.

The biomedical field's use of machine learning for the diagnosis and prognosis of liver disease has had a significant impact. Machine learning promotes the impartiality of the decision-making process and guarantees an improvement in disease detection and prediction, two areas of importance in the biomedical industry. Medical issues can be treated quickly and at a lower cost with the use of machine learning techniques. This study's main objectives are to improve prediction of outcomes and lower the price of diagnosis in the healthcare industry. As a result, we classified patients according to whether or not they had liver disease using several categorization systems. LR, KNN, and RF are three machine learning techniques that have been used, and the effectiveness of these techniques has been measured from a number of angles, including accuracy, precision, recall, and the f-1 score. Additionally, utilizing the receiver operating characteristic, the performance was compared (ROC).

II. LITERATURE REVIEW

Golmei Shaheamlung, Harsh Preet Kaur & Mandeep Kaur proposed a paper based on "Survey on machine learning techniques for the diagnosis of liver disease". In this, it checks the performance of various machine learning techniques based on their accuracy. In this Decision tree, J48 and ANN provide better accuracy in the detection and prediction of liver disease [1].

A manuscript titled "Evaluation-based Approaches for Liver Disease Prediction Using Machine Learning Algorithms" was proposed by C. Geetha and Dr. AR. Arunachalam. The accuracy of the Naive Bayes classifier was improved by the authors by 51.59 percent, that of the C4.5 algorithm by 55.94 percent, that of the BPNN method by 66.66 percent, and that of knowledge discovery in the data set by 62.6 percent. There are two primary machine learning algorithms used: SVM and Logistic Regression. All of the models have been used in the prediction study, and their effectiveness has been evaluated. The likelihood of liver disease was predicted with 96 percent accuracy [2].

In their proposed study, Varun Vats and Lining Zhang compare the prediction accuracy and computational complexity of three different machine learning algorithms, namely DBSCAN, K-Means, and affinity prediction. The ECG heartbeats in this were divided into 25 categories. When compared to traditional techniques utilized in the MIT-BIH database, the data set used for the analysis of arrhythmia and predicted high degree of precision was applied. Numerous variables, including the V-measure, completeness, homogeneity, adjusted rand index, adjusted mutual information, and silhouette coefficient, are considered when calculating performance [3].

A paper titled Diagnosis of Liver Disorder was proposed by Pushpendra Kumar and Ramjeevan Singh Thakur. They utilized two algorithms for this: Neighbor Weighted K-NN Method for LFT Imbalanced Data and Using Fuzzy Adaptive. Compared to CHAID, the boosted C5.0 algorithm produces accuracy levels of 93.75% versus 65.00%. In comparison to Fuzzy- ADP TKNN, Fuzzy- ANWKNN on the ILPD dataset has produced average outcomes of 84.29%, 65.93%, 88.81%, 91.36%, 90.07%, 34.07%, and 11.19% in terms of accuracy, specificity, sensitivity, and precision as well as F1-score, FPR, and FNR. According to the ILPD dataset, the average findings for accuracy, specificity, sensitivity, and precision and the F1-Score, FPR, and FNR were 76.34%, 42.86%, 81.50%, 90.24%, 85.65%, 57.14%, and 18.50%, respectively. The suggested model aims to improve the performance of a predictive model for liver disorders based on unbalanced LFT datasets [4].

Vijayalaxmi and Sateesh Ambesange The title of the proposed study is "Optimizing random forest prediction of liver disease by various data balancing strategies." Five models were utilized in this study, and each had a different accuracy that increased in ascending order from the others. The precision of the Model 5 is perfect. The Indian Liver Patient Dataset (ILPD), which is based on Indian patients, is used in this work to develop the machine-learning model, and the Random Forest (RF) algorithm is employed to forecast the disease with various preprocessing methods [5].

The paper "Prediction of Graft Dysfunction in Pediatric Liver Transplantation by Logistic Regression" was introduced by Krasimira Prodanova and Yordanka Uzunova. The models were built using data from patients at the University Hospital "Lozenets" in Sofia, Bulgaria. STATISTICA was the software package used for the statistical modeling of real-world data. For predictive purposes, a dichotomous regression model used. The model describes the phenomenon of the presence or absence of a specific complication (including disease) as a function of various factors accepted as independent continuous variables. Pediatric liver transplantation is one of the most recent and rapidly expanding areas of modern medicine. The goal of LT is not only to ensure the patient's survival but also to provide him or her with a state of health that allows for psychological and physical integrity [6].

Prognosis of Liver Disease: Using Machine Learning Algorithms is the title of the article that Vyshali J. Gogi and Dr. Vijayalakshmi M.N. proposed. Disease diagnosis is being studied using classification algorithms like decision trees, SVM, and logistic regression. The most efficient model is taken into consideration. Once the liver illness is suspected, the patient is then advised to undergo imaging tests to identify the presence of tumor lesions and stages. Support Vector Machine (SVM), Logistic Regression (LR), and Decision Tree are a few of the several learning strategies used by data mining techniques.

The accuracy of the SVM and decision tree were 82.7 and 94.9 respectively, while the accuracy of the logistic regression approach was 95.8% [7].

The authors of the study "Comparative Study of Artificial Neural Network based Classification for Liver Patient," Anil Kumar Tiwari, Lokesh Kumar Sharma, and G. Rama Krishna, proposed it. The classification of the patient data for the liver is done in this study using back propagations (BP) learning, radial basis function networks (RBF), self organising maps (SOM), and support vector machines (SVM). Medical researchers and practitioners are increasingly using predictive data mining as a key tool. In this work, feature selection strategies predicator qualities are established after univariate analysis of the liver data. Additionally, ANN- based classifiers were used on a few chosen attributes. It demonstrates how ANN classifiers can be used as a tool for patient prediction [8].

Pushpavanam Subramaniam and Babita K. Verma introduced a presentation titled "Characterizing different classes of patients based on their liver regeneration potential after hepatectomy and the prognosis of safe future liver volume for improved recovery." The Furchtgott model was developed into the Cook mathematical model, which is what we used. This idea explains how hepatocytes grow after a partial hepatectomy. Hepatocytes are assumed to enter the cell cycle and stay in one of the three phases of quiescence (Q), priming (P), or replication by a series of signals from cytokines and growth factors (R). We assessed the response distribution in a cohort of virtual patients in order to discriminate between the three patient classes of the normal recovery, hindered recovery, and liver failure. We found that overall cell death is higher in liver failure, despite the fact that proliferation and quiescence rates are the same in both classes, using a sample virtual patient from the recovery and failure classes [9].

The title "Liver Diseases Prediction Using KNN with Hyperparameter Tuning Techniques" was proposed by Ranjana Nadagoudar and Sushma Patil. In this, performance is evaluated using various metrics such as precision, recall, the f1-score, the precision-recall curve (PRC), and the receiver operating curve (ROC). The paper focuses on KNN algorithms, model optimization steps, and the development of several models step by step [10].

Prediction of Liver Disease using Classification Algorithms is a work that Thirunavukkarasu K. and Dr. Irfan suggested. The Indian Liver Patient Dataset serves as the source of the dataset (ILPD). The UCL Machine Learning Repository is where you may download this. 567 occurrences and 10 attributes make up this dataset. The following attributes are included: Age, Gender, DB, TB, ALB, SGOT, SGPT, TP, ALP, and A/G ratio. Three classification techniques—K-nearest Neighbors, Support Vector Machines, and Logistic Regression—have been utilized in this. Based on classification accuracy, which is discovered using a confusion matrix, all of these techniques have been compared. In the experiment, the two models with the highest accuracy and sensitivity were logistic regression and K-Nearest Neighbor. Thus, it can be said that Logistic Regression is suitable for forecasting liver illness [11].

In their study, "The Diagnosis of Chronic Liver Disease Employing Machine Learning Techniques," Golmei Shaheamlung and Harshpreet Kaur suggested using machine learning algorithms including SVM, K-mean clustering, KNN, Random Forest, and Logistic Regression. A patient with Chronic Liver Disease (CLD) in its early stages will live longer as a result of this. Results are assessed in terms of accuracy, precision, and recall once the proposed technique is put into practise in Python using the Spyder tool. By combining the three classifiers of logistic regression, random forest, and KNN method, the proposed model was enhanced. Python is used for the model's implementation, and the results show that it is accurate to within 77.58 percent of the target [12].

S. Gr. Mougiakakou, I. Valavanis, K. S. Nikita, A. Nikita, and D. Kelekis published a paper titled "Characterization of CT Liver Lesions Based on Texture Features and a Multiple Neural Network Classification Scheme." Data Acquisition, Feature Extraction, Feature Selection, Multiple Classifier System describes the proposed CAD system for the classification of focal liver lesions from CT images. The neural network has been classified into five types in this study. When compared to the other primary classifiers, "NN1" and "NN4", which used FOS and TEM feature vectors, respectively, produced the highest classification rates. The proposed system achieved an overall classification performance of 97% [13].

III. RELATED WORK

A. Data Collection

We get a dataset from the UCI Machine Learning for this experiment. Additionally, the initial dataset was gathered in Andhra Pradesh, India's northeast. Data from 583 liver patients, 75.64% of whom are men and 24.36% of whom are women, are included in this dataset. This dataset has 11 distinct parameters, however we only used 10 of them for further analysis and 1 of them as a target class. the like,

- 1) Age of the patient.
- 2) Gender of the Patients.
- 3) Total Bilirubin.
- 4) Direct Bilirubin.
- 3) Alkaline Phosphatase.

- 4) Alanine Aminotransferase.
- 5) Aspartate Aminotransferase.
- 6) Total Proteins.
- 7) Albumin.
- 8) Albumin-to-Globulin Ratio.
- 9) The data were divided into two sets by a selector field (labelled by the experts).

Table 1: Attribute description table

Data columns	Non-null count	Data type
Age	583 non-null	int64
Gender	583 non-null	object
Total bilirubin	583 non-null	float64
Direct bilirubin	583 non-null	float64
Alkaline phosphatase	583 non-null	int64
Alanine aminotransferase	583 non-null	int64
Aspartate aminotransferase	583 non-null	int64
Total proteins	583 non-null	float64
Albumin	583 non-null	float64
Albumin and globulin rate	579 non-null	float64
Dataset	583 non-null	int64

B. Data Preprocessing

In this study, we examined the data of 583 liver patients, 416 of whom were liver patients and 167 of whom were not. Figure 1 depicts the ratio of total liver patients. Furthermore, 441 male samples and 112 female samples were taken for analysis from the liver patient's dataset (Fig. 2)



Fig :1



Fig: 2

There are some associated characteristics in the heatmap in Fig. 3. These columns have a low correlation in some cases. In order to improve our ability to predict liver illness, we therefore removed some of the features.

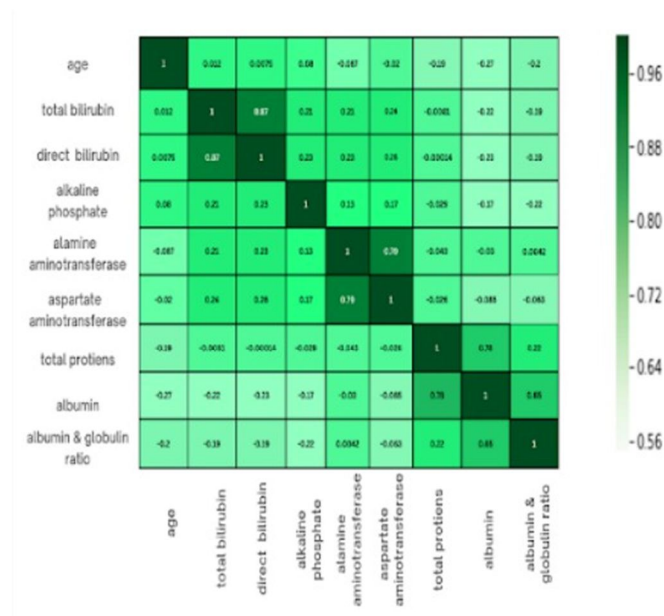


Fig 3: Heat map for the liver dataset's correlated columns.

C. Functional Requirements

The functional Requirements for the project are

- 1) Python 3.7 or later PIP for Windows.
- 2) Streamlit
- 3) Jupyter Notebook.

D. Representation of Classical Algorithms

- 1) **Logistic Regression:** By examining the correlation between one or more already present independent variables, a logistic regression model forecasts a dependant data variable. A logistic regression could be used, for instance, to forecast whether a candidate for office will win or lose, or if a high school student will be accepted into a particular institution or not. These simple choices between two options allow for binary outcomes. Many criteria for input can be taken into account using a logistic regression model. The logistic function can take into account the student's grade point average, SAT score, and number of extracurricular activities in the case of college acceptance. It then rates new cases according to their likelihood of falling into one of two outcome categories based on historical information about past outcomes involving the same input criteria. The technique of logistic regression has grown in significance in the field of machine learning. It enables machine learning algorithms to categorise incoming input based on previous data. The algorithms improve at predicting classes within data sets as more pertinent data is included. The technique of logistic regression has grown in significance in the field of machine learning. It enables machine learning algorithms to categorise incoming input based on previous data. The algorithms improve at predicting classes within data sets as more pertinent data is included. Logistic Regression model is implemented for the prediction of liver disease. The dependent variable 'is_patient' as '1' for yes and '0' for no respectively are divided based on gender and plotted against the count of the patients. Logistic Regression can be used to classify the observations using different types of data and can easily determine the most effective variables used for the classification. The below image is showing the logistic function. Because it can classify new data using both continuous and discrete datasets, logistic regression is a key machine learning approach.
- 2) **KNN:** One of the simplest machine learning algorithms based on the supervised learning method is K-Nearest Neighbour. The K- NN algorithm places the new instance in the category that is most similar to the available categories based on the assumption that the new case and the available cases are comparable.

When a new data point is added, the K-NN algorithm classifies it based on how similar the existing data is to it and stores it all. This means that utilising the K-NN method, fresh data can be quickly and accurately categorised into a suitable category. Although the K-NN technique can be applied to classification and regression problems, it is most frequently utilised for classification issues. Since K-NN is a non-parametric technique, it makes no assumptions about the underlying data. It is also known as a lazy learner algorithm since it saves the training dataset rather than learning from it immediately. Instead, it uses the dataset to perform an action when classifying data. The KNN method simply saves the information during the training phase, and when it receives new data, it categorises it into a category that is quite similar to the new data.

- 3) *Random forest*: Random forests, also known as random decision forests, are an ensemble learning technique for classification, regression, and other assignments that generates the class that is the mean forecast (regression) or method of the classes (classification) of the individual decision trees. When a new set of samples is input into a random forest, each decision tree in the forest makes a prediction on this set of samples separately and integrates the prediction results of each tree to get a final result. This is done after the random forest has established a large number of decision trees in accordance with a specific random rule. This technique works by developing a large number of decision trees during training time. For decision trees' tendency to overfit their training set, random decision forests are the best option. There is an immediate relationship between the united trees and the result that the forest of trees can produce. To obtain forecasts that are more accurate and effective, random forest adds an additional layer of irregularity to the stowing. Based on the distinctive qualities and classification outcomes of a given dataset, Random Forest is a great supervised learning method that can train a model to predict which classification results in a particular sample type belong to. The decision tree-based Random Forest uses the "bagging" (bootstrap aggregating method) to generate various training sample sets. The best attribute from a group of attributes chosen at random is chosen by the random subspace division approach to divide internal nodes. The voting method is utilised to categorise the input samples, and the many decision trees created are used as weak classifiers. Multiple weak classifiers combine to construct a robust classifier. Each decision tree in the forest makes a prediction on this fresh set of samples independently after a random forest has built up a sizable number of decision trees in accordance with a certain random rule. The final result is obtained by integrating the prediction results of all the trees.

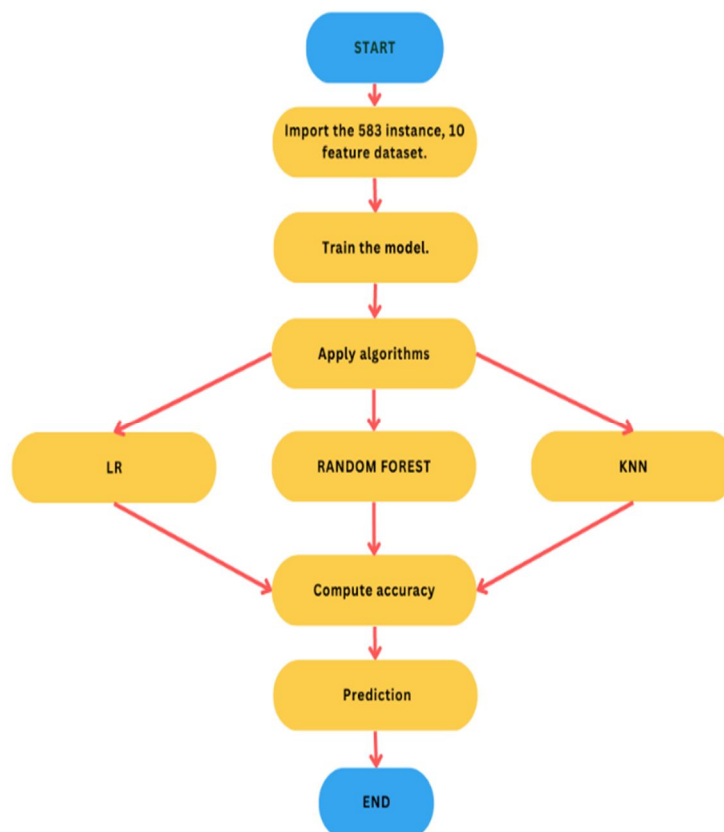


Fig 4: Flowchart of proposed model.

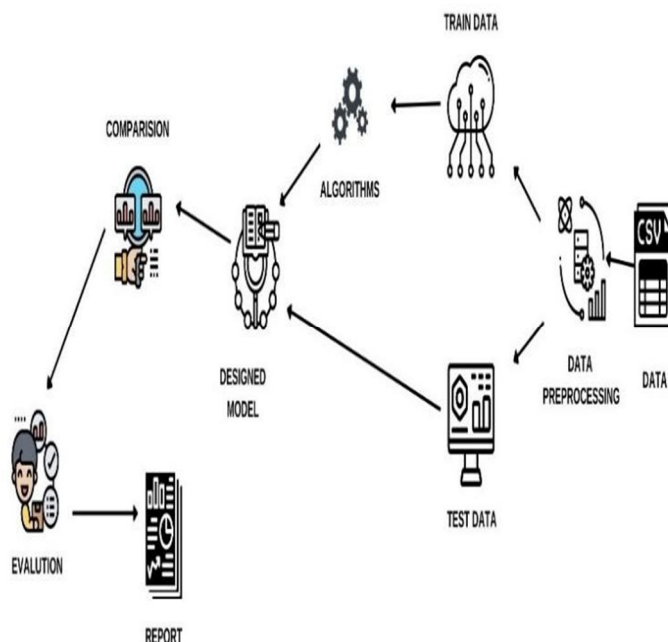


Fig 5: Machine learning model Architecture.

IV. RESULT

A. Classification of Evaluation Techniques

In our work, we employed certain genuine estimations to assess the test execution of various categorization algorithms. The various evaluation methodologies, such as accuracy, sensitivity, specificity, precision, and the f1 measure, were used to examine the performance of the classification systems. As a result, the confusion matrix determines the exhibition assessment variables. In this case, true positive (TP): The prediction properly identifies a patient as having liver illness. False Positive (FP): The forecast wrongly identifies a patient as having liver disease. True Negative (TN): The conclusion of the prediction correctly rejects the patient's allegation of liver illness. False Negative (FN): The prediction result mistakenly eliminates the possibility that a patient has liver disease. The precision of the prediction model offers the difference between sound and patient capacity ratio. To establish classification precision, one must first determine the true positive, true negative, false positive, and false negative.

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

This effectability test successfully distinguishes patients with liver illness. It mostly demonstrates the test's positive results. It is also known as recall and positive rate.

$$\text{Sensitivity} = (\text{TP}) / (\text{TP} + \text{FN})$$

The negative repercussions of the condition are underlined in detail. It indicates the amount of the patients' missing sickness. It is also known as the "real negative rate" (TNR).

$$\text{Sensitivity} = (\text{TN}) / (\text{FP} + \text{TN})$$

Positive predictive value is another name for precision. It displays the percentage of favourable outcomes that classifier algorithms accurately anticipated

$$\text{Precision} = (\text{TP}) / (\text{TP} + \text{FP})$$

F1 combines accuracy and recall to determine the model's precision. It provides the model's FP and FN proportions.

$$\text{F1} = (2 * (\text{R} * \text{P})) / (\text{R} + \text{P})$$

B. Analysis of the result

In this experiment, we examined three machine learning classifiers for the categorization of liver disease dataset using various techniques. After employing oversampling methods, KNN's accuracy value was 70.28%. RF received 74.28%. The highest result was 76.57% for LR. The LR classification algorithm outperforms the other classifiers for predicting liver disease based on these measurement criteria.

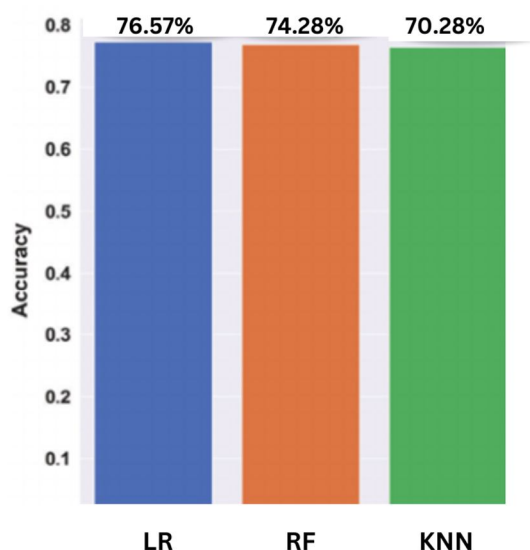
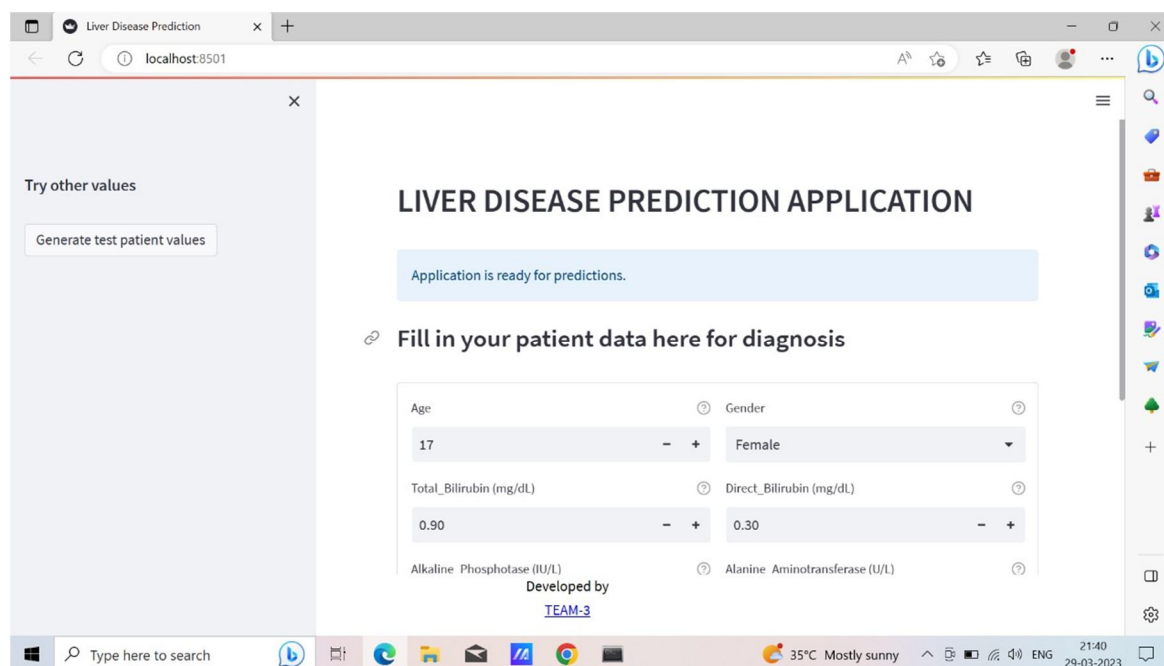


Table 2: Statistical results of this study

ALGORITHMS	ACCURACY	PRECISION	RECALL	F1_SCORE
LR	76.57%	75.00%	99.50%	87.16%
RF	74.28%	84.40%	77.90%	80.32%
KNN	70.20%	83.10%	84.90%	84.32%

A. Output Screens

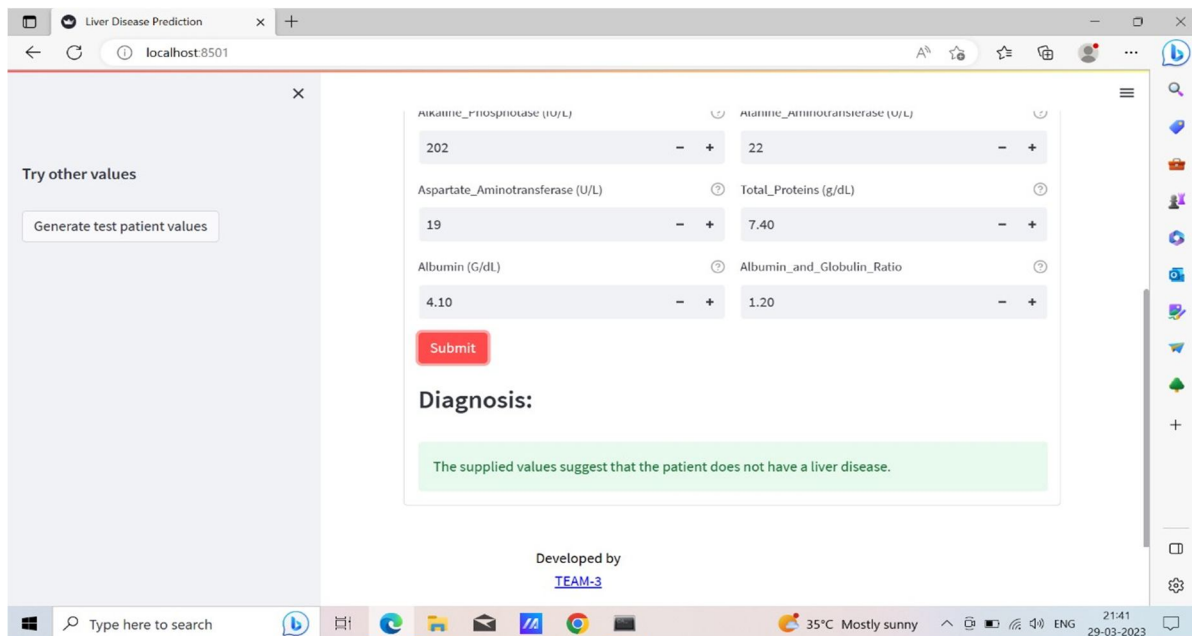


The screenshot shows a web application titled "LIVER DISEASE PREDICTION APPLICATION". It features a sidebar on the left with a "Try other values" section and a "Generate test patient values" button. The main content area has a status bar indicating "Application is ready for predictions." Below this, there is a section titled "Fill in your patient data here for diagnosis" with a form containing the following fields:

- Age: 17 (with minus and plus buttons)
- Gender: Female (dropdown menu)
- Total_Bilirubin (mg/dL): 0.90 (with minus and plus buttons)
- Direct_Bilirubin (mg/dL): 0.30 (with minus and plus buttons)
- Alkaline Phosphatase (IU/L): (empty field)
- Alanine Aminotransferase (IU/L): (empty field)

At the bottom of the form, it says "Developed by TEAM-3". The browser's address bar shows "localhost:8501". The Windows taskbar at the bottom shows the date as 29-03-2023 and the time as 21:40.

Fig 9: Liver Disease Prediction Application.

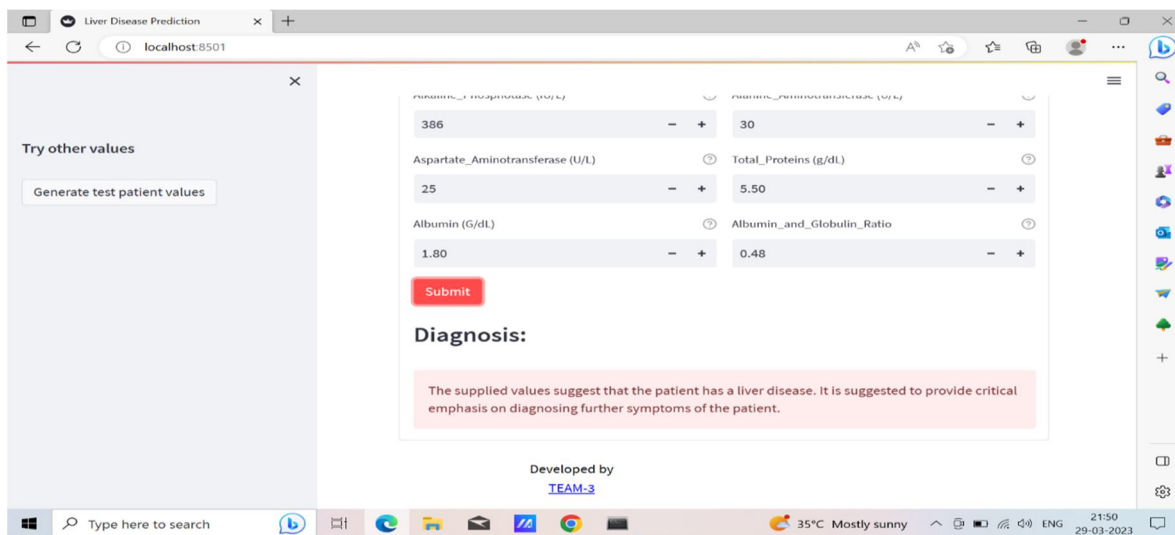


The screenshot shows a web browser window with the title "Liver Disease Prediction" and the URL "localhost:8501". On the left, there is a sidebar with a close button and a button labeled "Generate test patient values". The main area contains a form with six input fields, each with a minus and plus button for adjustment:

- Alkaline_Phenylalanase (U/L): 202
- Alkaline_Aminotransferase (U/L): 22
- Aspartate_Aminotransferase (U/L): 19
- Total_Proteins (g/dL): 7.40
- Albumin (G/dL): 4.10
- Albumin_and_Globulin_Ratio: 1.20

Below the form is a red "Submit" button. Underneath, the "Diagnosis:" section shows a green box with the text: "The supplied values suggest that the patient does not have a liver disease." At the bottom, it says "Developed by TEAM-3". The Windows taskbar at the bottom shows the time as 21:41 on 29-03-2023.

Fig 10: The fig shows patient does not have a liver disease.



This screenshot shows the same web application with different input values:

- Alkaline_Phenylalanase (U/L): 386
- Alkaline_Aminotransferase (U/L): 30
- Aspartate_Aminotransferase (U/L): 25
- Total_Proteins (g/dL): 5.50
- Albumin (G/dL): 1.80
- Albumin_and_Globulin_Ratio: 0.48

After clicking the "Submit" button, the "Diagnosis:" section now shows a pink box with the text: "The supplied values suggest that the patient has a liver disease. It is suggested to provide critical emphasis on diagnosing further symptoms of the patient." The rest of the interface, including the sidebar and taskbar, remains the same.

Fig 11: Fig shows patient has a liver Disease.

V. CONCLUSION

In order to anticipate and diagnose liver disease at an early stage, this research exhibited several prediction algorithms. The data set displayed various input parameters that were acquired, and we verified and trained the models using those input values. The main goal of this research is to develop a chorionic liver infection diagnosis system using six unique supervised machine learning classifiers. The LR classifier provided the highest order exactness, according to our analysis of how well each classifier performed using the patient's information parameters. depending on the F1 measure, which has the lowest accuracy , to predict liver illness The future outperforms A chronic disease diagnosis and decision-support system will be offered via the classification process. In low-wage nations with a dearth of medical institutions and experts, this application may be unexpectedly profitable. There are a few implications from our study for upcoming research in this area. We have only examined a few well-known supervised machine learning algorithms; other algorithms can be chosen to build an ever-more-accurate model of liver disease prediction, and performance can be steadily enhanced. Additionally, this work is positioned to have a significant impact on both therapeutic and medical research.

Be prepared for a liver infection. By assessing the algorithms using attribute collecting and data set training, the prediction of liver illness was tested more precisely. These findings suggest novel characteristics that classifiers can utilise specifically to diagnose liver illness at an early stage. To predict liver disease, LR, KNN, and RF are built. These results demonstrated that the 1 model correctly predicted patients with liver illness. While LR provided a good performance at every stage, some of the algorithms performed well at some particular parameters. LR is therefore regarded as the finest and most promising algorithm for predicting the course of liver disease.

REFERENCES

- [1] A Survey on Machine Learning Techniques for the Diagnosis of Liver Disease, Golmei Shaheamlung, Harshpreet Kaur, and Mandeep Kaur, International Conference on Intelligent Engineering and Management (ICIEM) (2020), pp. 337–341.
- [2] Geetha, Dr. AR.Arunachalam. "Evaluation-based Approaches for Liver Disease Prediction Using Machine Learning Algorithms," International Conference on Computer Communication and Informatics, January 27–29, 2021.
- [3] Varun Vats, Lining Zhang, Sreejit Chatterjee, Sabbir Ahmed, Elvin Enziama, and Kemal Tepe are among the participants. "A Comparative Analysis of Unsupervised Machine Techniques for Predicting Liver Disease." (2018).
- [4] Ramjeevan Singh Thakur and Pushpendra Kumar. "Liver Disorder Diagnosis Using Fuzzy Adaptive and Neighbor Weighted K-NN Method for Imbalanced LFT Data," ICSSS 2019.
- [5] Sateesh Ambesange, Vijayalaxmi A, Rashmi Uppin, Shruthi Patil, and Vilaskumar Patil "Optimizing liver disease prediction with random forest using various data balancing techniques," International Conference on Cloud Computing in Emerging Markets (2020), pp. 98-102.
- [6] K. Prodanova and Y. Uzunova, "Logistic Regression Prediction of Graft Dysfunction in Pediatric Liver Transplantation," 2020 International Conference on Mathematics and Computers in Science and Engineering (MACISE), pp. 260-263, doi:10.1109/MACISE49704.2020.00054.
- [7] V. J. Gogi and V. M.N., "Prognosis of Liver Disease: Using Machine Learning Algorithms," 2018 International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE), Bhubaneswar, India, 2018, pp.875-879, DOI: 10.1109/ICRIEECE44171.2018.9008482.
- [8] Comparative Study of Artificial Neural Network-based Classification of Liver Patient, Journal of Information Engineering and Application. B. K. Verma, P. Subramaniam, and R. Vadigepalli, "Characterizing different class of patients based on their liver regeneration capacity post hepatectomy and the prediction of safe future liver volume for improved recovery," 2018 International Conference on Bioinformatics and Systems Biology (BSB), 2018, pp. 152-156, DOI: 10.1109/BSB.2018.8770553.
- [9] S. Ambesange, R. Nadagoudar, R. Uppin, V. Patil, S. Patil and S. Patil, "Liver Diseases Prediction using KNN with Hyper Parameter Tuning Techniques," 2020 IEEE Bangalore Humanitarian Technology Conference (B-HTC), 2020, pp. 1-6, doi: 10.1109/B-HTC50970.2020.9297949.
- [10] k. Thirunavukkarasu, A. S. Singh, M. Irfan, and A. Chowdhury, "Prediction of Liver Disease using Classification Algorithms," 2018 4th International Conference on Computing Communication and Automation (ICCCA), Greater Noida, India, 2018, pp. 1-3, DOI: 10.1109/CCAA.2018.8777655.
- [11] Golmei Shaheamlung, Harshpreet Kaur, "Diagnosis of Liver Disease Using Machine Learning Techniques," International Research Journal of Engineering and Technology (IRJET), Volume: 05, Issue: 04, 2018, pp. 4011–4014.
- [12] S. G. Mougiakakou, I. Valavanis, K. S. Nikita, A. Nikita and D. Kelekis, "Characterization of CT liver lesions based on texture features and a multiple neural network classification scheme," Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No.03CH37439), Cancun, Mexico, 2003, pp.1287-1290 Vol.2, doi:10.1109/IEMBS.2003.1279.



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)