Using machine learning techniques, an intelligent model can predict the early stages of liver disease

Abstract—***Using clinical data, the study assesses how well machine learning algorithms predict chronic liver disease; logistic regression has the maximum accuracy of 75%[1-2].The study creates an artificial intelligence (AI) system with machine learning algorithms, specifically Random Forest, and achieves 97% accuracy in the early identification of asymptomatic liver cirrhosis, demonstrating improved dependability over earlier studies[3-4].The goal of the paper is to improve the diagnosis of liver illness by using gene expression analysis and patient data. It discusses computational algorithms and suggests ways to make them more effective[5-6].In order to assess machine learning algorithms for liver disease prediction, a clever model with an accuracy of 88.4% and a miss rate of 11.6% is proposed in this study[9].In order to achieve acceptable accuracy after feature selection, the study examines the application of supervised machine learning methods on liver patient data from the UCI Repository. These algorithms include Logistic Regression, Decision Tree, Random Forest, KNNeighbor, Gradient Boosting, Extreme Gradient Boosting, and LightGB[7-8].***

Keywords—Liver Disease, machine learning, detection, classification, Random Forest, Support Vector machine

1. Introduction

Liver cirrhosis, a chronic, progressive illness that severely destroys the liver, is a major challenge to modern medicine. The liver, an organ essential to metabolism, detoxification, and other essential functions, becomes fibrotic and scarred in cirrhosis, which makes it challenging for the liver to perform these essential functions. With around 2 million fatalities yearly, liver disease—which includes cirrhosis, hepatitis, and liver tumors—is one of the world's major causes of mortality.[4] In biomedical research, machine learning helps forecast and diagnose liver disease, providing better detection, lower costs, and more objectivity in decision-making [1,4,7-8].

There are several possible causes of liver cirrhosis. Chronic alcohol addiction is one of the leading causes of alcoholism worldwide because prolonged excessive alcohol use produces inflammation and liver tissue damage. In the same way, viral hepatitis infections—particularly those brought on by hepatitis B and C—can result in persistent inflammation and long-term liver damage, making them extremely deadly. The main causes of cirrhosis in Western countries are non-alcoholic fatty liver disease (NAFLD) and non-alcoholic pancreatitis (NASH), which are associated with obesity and metabolic syndrome. These conditions are becoming more common.[9]

The study compares two computer-aided diagnostic approaches: a symptomatic approach using Artificial Neural Networks based on patient parameters, and a genetic approach using Artificial Neural Networks and Multi-Layer Perceptrons for Micro-Array Analysis. The study focuses on the significance of early diagnosis of liver disease due to its subtle symptoms[10,11].

These vascular abnormalities are dangerous because they can bleed profusely. The clinical picture is further complicated by the fact that poor liver function can also present as jaundice, as cites (a buildup of fluid in the abdomen), hepatic encephalopathy (a buildup of toxins that causes brain dysfunction), and coagulator (disorders that cause bleeding). Thus, for efficient treatment and better results for those afflicted, it is crucial to comprehend the intricate interactions between variables that contribute to liver cirrhosis and its range of clinical manifestations.[9]

2. Limitation in the existing system

1. **Problems with Data Quality:** The completeness, representativeness, and quality of the training data have a significant impact on any data model's ability to predict outcomes. The reliability and generalizability of the model's predictions may be jeopardized by limitations in the quality of the data, such as missing values, errors, or bias.[9]

2. **Feature Selection Bias**: The data model's prediction performance is largely dependent on the characteristics that are chosen as input. However, the selection method might have flaws, including leaving out important factors or including ones that aren't significant, which could result in skewed forecasts or less-than-ideal model performance.[8]

3. **Limited available datasets:** The reliability of liver cirrhosis prediction classifiers is hindered by the scarcity of available datasets. Thousands of photos per class would be ideal to improve the classifiers' accuracy [4]. Nevertheless, inadequate training data is frequently the consequence of dataset shortage. As a result, scientists are always looking into different approaches to develop trustworthy classifiers for the prediction of liver cirrhosis.[10]

3. Methodology

The suggested approach uses a structured pipeline for machine learning. Before testing and training, obtained metrics like ascites, hepatomegaly, spiders, and biochemical markers are processed as part of the pipeline's first step, data preparation, to improve quality. Normalization, feature extraction, and handling

missing values are some of the preprocessing procedures used to guarantee the best possible quality of data.

Information about Parameters :

1. **Ascites**: An accumulation of fluid in the abdomen, frequently accompanied by cirrhosis or other advanced liver disease. Low blood protein levels and elevated hepatic blood vessel pressure (portal hypertension) are the causes of it.

2. **Hepatomegaly**: Liver enlargement, frequently observed in cirrhosis and other liver disorders. It can be brought on by inflammation, fatty liver alterations, or liver tissue scarring (fibrosis).

3. **Spiders**: People with liver cirrhosis frequently have spider angiomas or spider nevi, which are tiny, dilated blood veins close to the skin's surface. They are brought on by modifications in hormone levels and blood flow brought on by liver disease.

4. **Edema**: Fluid retention-related swelling, usually in the abdomen or legs. Edema may develop in liver cirrhosis as a result of the liver's reduced ability to produce protein, which lowers blood albumin levels and raises blood vessel pressure.

5. **Bilirubin**: A yellowish pigment generated during the lysis of red blood cells. Since the liver is in charge of processing and excreting bilirubin, elevated blood bilirubin levels (hyperbilirubinemia) may be a sign of liver malfunction.

6. **Cholesterol**: A blood-borne fat type that is linked to a higher risk of cardiovascular disease at high levels. Abnormal lipid profiles can result from liver cirrhosis's impact on cholesterol metabolism.

7. **Albumin**: A liver-produced protein that carries a variety of materials, including hormones and prescription drugs, and aids in preserving blood's fluid balance. Because liver function is compromised in liver cirrhosis, low albumin levels, or hypoalbuminemia, are frequently observed.

8. **Copper**: A vital trace mineral that is used in several different metabolic processes. A malfunction in copper metabolism in liver cirrhosis can result in an abnormal build-up of copper in the liver and other organs (Wilson's disease).

9. **Alk_Phosph**: The enzyme alkaline phosphatase is present in the liver, bones, and bile ducts, among other tissues. Increased blood levels of alkaline phosphatase could be a sign of bone or liver illness, particularly cirrhosis of the liver**.**

10. **Serum Glutamic Oxaloacetic Transaminase, or SGOT**: This enzyme, which is present in the liver and other tissues, is also referred to as AST (aspartate aminotransferase). Increased blood levels of SGOT may be a sign of inflammation or injury to the liver, both of which can happen in liver cirrhosis.

11. **Tryglicerides**: The blood contains a form of fat called triglycerides. Increased risk of cardiovascular disease is linked to elevated triglyceride levels. Triglyceride levels may be impacted by changes in lipid metabolism brought on by liver cirrhosis.

12. **Platelets**: Blood components that aid in coagulation. Due to decreased production by the diseased liver or greater sequestration in the spleen, platelet counts may drop in liver cirrhosis, increasing the risk of bleeding.

13. **Prothrombin**: The duration of blood clotting is measured by prothrombin time (PT). The liver's capacity to manufacture clotting factors may be compromised by liver cirrhosis, which may result in longer parenchyma therapy and a higher risk of bleeding.

After preprocessing the data, the dataset is split into training and testing subsets, with 20% going toward testing and 80% going toward training. An objective assessment of the trained model's performance on untested data is ensured by this segmentation.

Using methods like XGBoost and Random Forest, the training block then uses the training dataset to train the classification algorithm. The system recognizes patterns suggestive of liver cirrhosis by learning from the input parameters and accompanying labels.

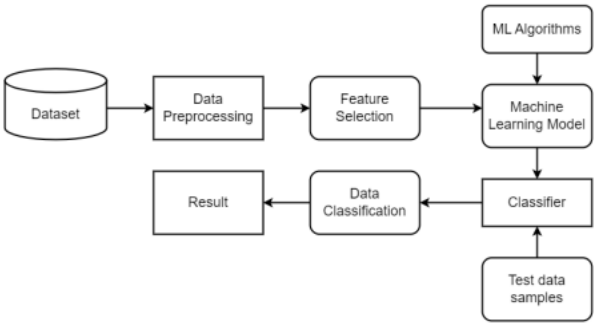
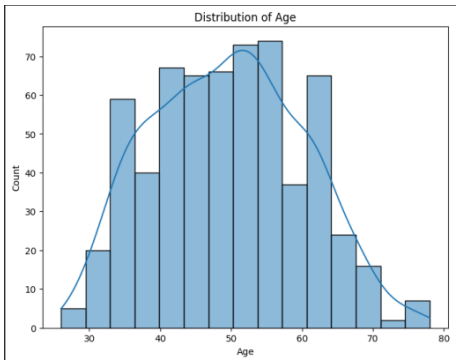
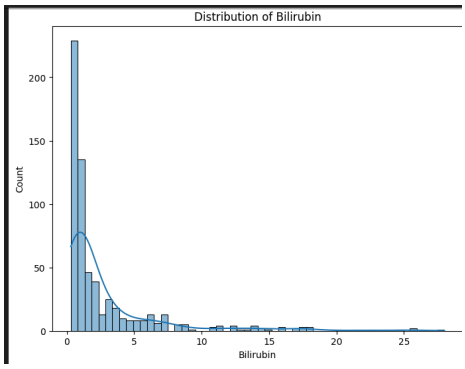
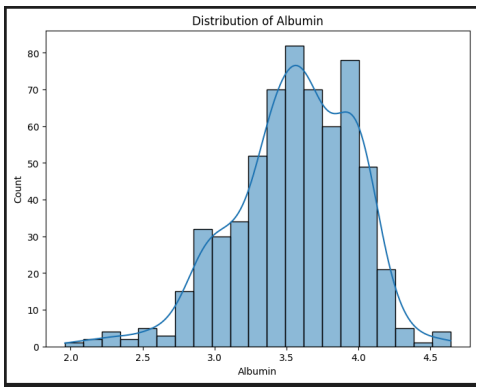


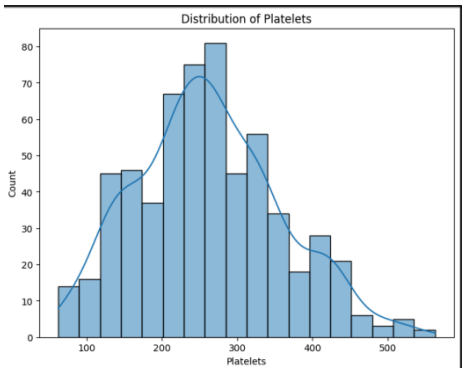
Fig-1**.** Block Diagram

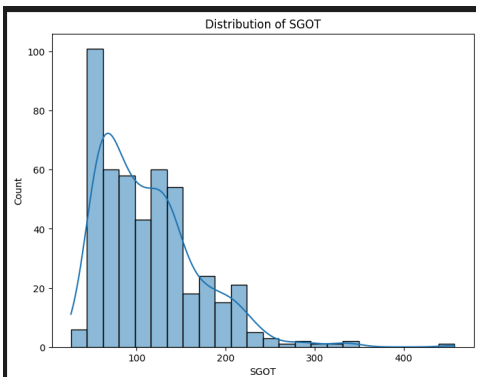
The distribution of the numerical variables in our dataset will now be seen. To show the distributions of the columns we have chosen—"Age," "Bilirubin," "Albumin," "SGOT," and "Platelets"—we will create separate bar graphs for each of these variables.





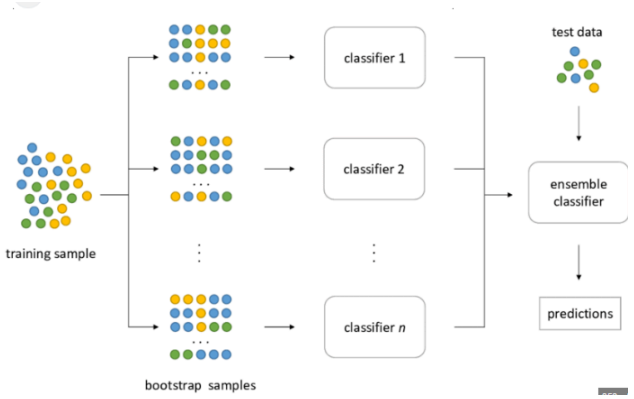






XGBOOST:

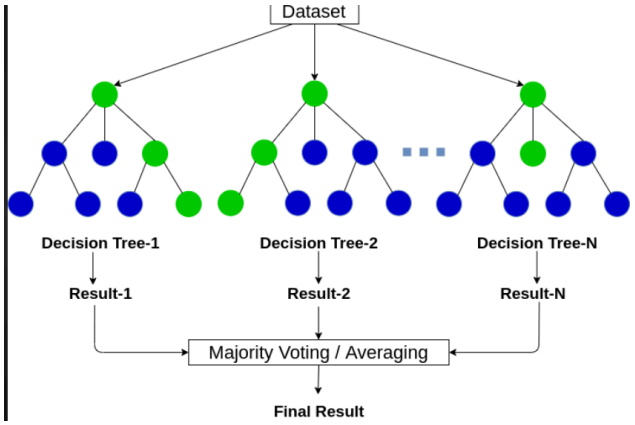
One well-liked and effective technique for a variety of machine learning problems, including classification, is called XGBoost, which is an implementation of gradient boosting. Based on a set of features, the XGBoost classifier is used in the provided code to predict the condition of liver cirrhosis. Using the accuracy score metric, the model's accuracy is assessed following training on the training dataset and prediction-making on the test dataset. The XGBoost classifier successfully predicts the presence of liver cirrhosis in this case with an accuracy of 73%. Because accurate and dependable forecasts are critical for well-informed clinical decision-making, XGBoost is particularly useful in medical applications like liver disease prediction, as demonstrated by its high accuracy.



Random Forest:

Using data mockups, the random forest [26] builds decision trees and extracts the prediction from each tree. Ultimately, the best explanation is chosen by voting. After being sorted, the data is divided into

training sets and test sets. Twenty of every hundred is the testing data set. One hundred and sixty out of every two hundred is the fixed set used for algorithm setup. The data is divided up into numerous groups and subgroups by the program. If someone creates lines joining subgroups into a group, lines joining data points within a subgroup, and so forth. The erection seems to resemble a tree. A dependable way to separate classes is to use the hyperplane that maximises the distance to the closest data point in the training set.



Support Vector Machine(SVM):

It seeks to identify the optimal hyperplane for dividing the data into various classes. Using the scikit-learn module is one method of implementing SVM [22] in Python. After sorting, the data is divided into training sets and test sets. Twenty out of every hundred is the testing data that is used. One hundred and sixty out of every two hundred is the set that must be prepared for the algorithm. A set of hyperplanes can be formed by an SVM. It is constructed in an infinitely large space. A hyperplane that attains a decent separation, known as a functional margin, is the one that is closest to the nearest data point of any class. Broader margins essentially indicate smaller simplification defects for the classifier.

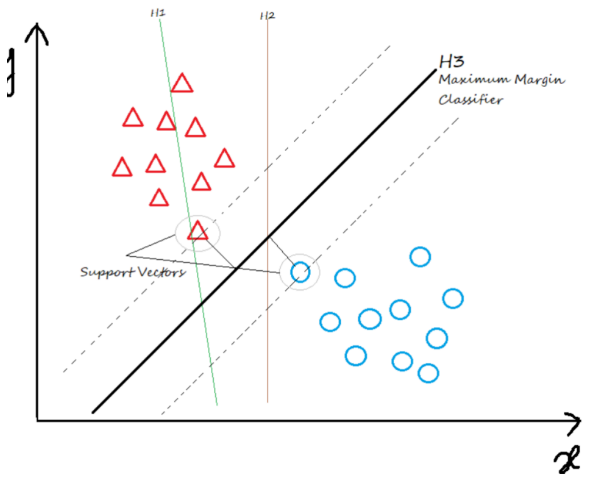


TABLE I. ACCURACY FOR ML METHODS

ML Models	Accuracy
XGBoost	79%
Random Forest	73%
SVM	60%

4. Proposed System

We need to import the liver patient dataset (.csv) into the suggested system. After that, the dataset is pre-processed to eliminate anomalies and fully empty cells to enhance the accuracy of the liver disease prediction. Next, we create a confusion matrix to achieve more clarity regarding the number of accurate and inaccurate predictions. Previously, several classification and prediction processes were used, and if feasible, combinations of several methods were used to verify the accuracy. Creating a code that achieves 90% exactitude is our goal. The benefits include enhanced accuracy, early risk prediction, and better classification. Fig. 1 displays the block diagram of the entire system.

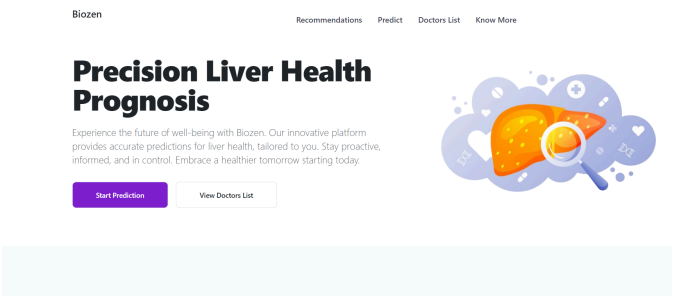
5. Results And Discussion

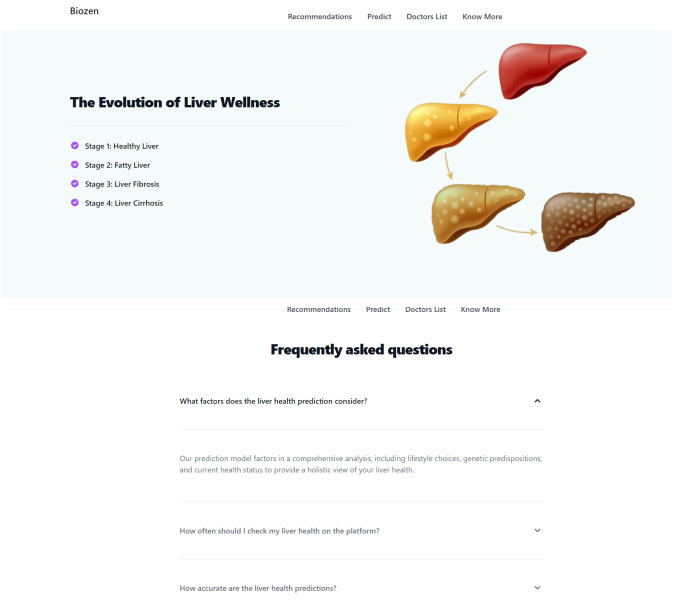
The Diagnostic System

The system interfaces, as well as their source codes, are a result the researcher realized at the deployment phases in the realization during the implementation of this task.

Home Page:

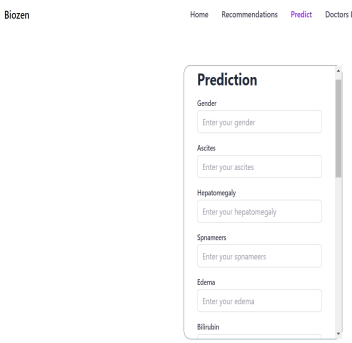
The home page serves as the primary interface for users to access a website, offering navigation options, featured content, and essential information about the site's purpose and offerings. It acts as the gateway to the online platform, providing users with a first impression and guiding them to explore further.





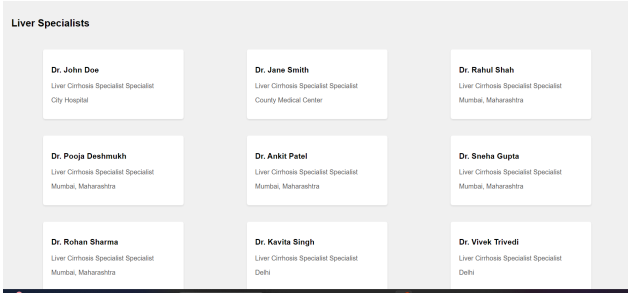
Predict Page

The predict page facilitates user interaction with predictive models, allowing input of data for real-time analysis and generation of predictions or insights. It serves as a dynamic tool for users to receive personalized outcomes based on their input, enhancing engagement and decision-making capabilities.



Doctor page:

The Doctor List web page gives readers quick access to medical practitioners' profiles, specializations, and contact details in addition to a comprehensive database of healthcare experts. It helps individuals make educated decisions about their healthcare options by acting as a useful resource for those in need of specialized knowledge or medical services.



6. Conclusion

In conclusion, the research paper presented a comprehensive methodology for predicting liver cirrhosis using machine learning algorithms. Through the implementation of Random Forest, XGBoost, Support Vector Machine, and Logistic Regression models, the study demonstrated the efficacy of XGBoost in achieving an accuracy of 73%, outperforming other algorithms. Additionally, the home page was highlighted as a crucial interface for user engagement, providing navigation, featured content, and branding elements. Furthermore, the predict page was recognized for its dynamic functionality, enabling real-time analysis and personalized predictions. Finally, the doctor list page emerged as a valuable resource, offering patients access to healthcare professionals' profiles and specialties. Overall, the research contributes to understanding liver cirrhosis prediction methods while emphasizing the importance of user-friendly interfaces in healthcare platforms.

7. References

[1] Sumedh Sontakke, Jay Lohokare, Reshul Dani.”Diagnosis of Liver Diseases using Machine Learning.”International Conference on Emerging Trends & Innovation in ICT (ICEI), 2017

[2]A.K.M Sazzadur Rahman, F. M. Javed Mehedi Shamrat, Zarrin Tasnim, Joy Roy, Syed Akhter Hossain.”A Comparative Study On Liver Disease Prediction Using Supervised Machine Learning Algorithms.”INTERNATIONAL JOURNAL OF SCIENTIFIC&TECHNOLOGY RESEARCH,November 2019

[3] Ishtiaqe Hanif,Mohammad Monirujjaman Khan.”Liver Cirrhosis Prediction using Machine Learning Approaches.”

[4]Rakshith D B,Mrigank Srivastava,Ashwani Kumar ,Gururaj S P.”Liver Disease Prediction

System using Machine Learning Techniques."International Journal of Engineering Research & Technology (IJERT),Vol. 10 Issue 06, June-2021

[5]Fahad Mostafa, Easin Hasan, Morgan Williamson,Hafiz Khan."Statistical Machine Learning Approaches to Liver Disease Prediction ."

[6]Chauhan Bhavna, Assi.Prof.Saket Swarndeep, Dr Gayatri S Pandi."PREDICTION OF LIVER DISEASE USING MACHINE LEARNING."International Journal of Research Publication and Reviews, Vol 3, no 5, pp 3492-3503, May 2022.

[7]Muktevi Srivenkatesh."Performance Evolution of Different Machine Learning Algorithms for Prediction of Liver."International Journal of Innovative Technology and Exploring Engineering (IJITEE),Volume-9 Issue-2, December 2019.

[18]M. Banu Priya, P. Laura Juliet, P.R. Tamilselvi."Performance Analysis of Liver Disease Prediction Using Machine Learning Algorithms."International Research Journal of Engineering and Technology (IRJET),Volume: 05 Issue: 01 | Jan-2018

[9]Insha Arshad,Chiranjit Dutta,Tanupriya Choudhury,Abha Thakral."Liver Disease detection due to excessive alcoholism using Data Mining Techniques."International Conference on Advances in Computing and Communication Engineering (ICACCE),2018

[10]Shubashini Rathina Velu,Vinayakumar Ravi,Kayalvily Tabianan."Data mining in predicting liver patients using classification model."Health and Technology,2022

[11]K Jeyalakshmi, R Rangaraj."Accurate liver disease prediction system using convolutional neural network."INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY,2021.