Power System Technology
ISSN:1000-3673

Received: 06-11-2024 *Revised: 15-12-2024* *Accepted: 05-01-2025*

A Comprehensive Analysis of Cirrhosis Progression through Clinical Data

[1]Dr. Sudhamani M J[*], [2]Khushi S Javali, [3]Kleesha P, [4]Mahima Nagaraj Hiregange, [5] Namratha R Upadhya

[1] *Professor & Associate HoD, CSE RNS Institute of Technology* Bengaluru, India sudhamani.mj@rnsit.ac.in

[2] *CSE, RNS Institute of Technology* Bengaluru, India ksjavali21@gmail.com

[3] *CSE, RNS Institute of Technology* Bengaluru, India kleeshapraveen@gmail.com

[4] *CSE, RNS Institute of Technology* Bengaluru, India mahimahiregange@gmail.com

[5] *CSE, RNS Institute of Technology* Bengaluru, India namratharupadhya@gmail.com

Abstract

Cirrhosis is a progressive liver disease characterized by liver fibrosis and scarring, which can eventually lead to liver failure. It is primarily caused by chronic alcohol con- sumption, viral hepatitis, and non-alcoholic fatty liver disease (NAFLD). Early detection and diagnosis are crucial for pre- venting the progression of the disease. Recent advancements in machine learning and deep learning techniques have shown great promise in enhancing the diagnostic accuracy of cirrhosis through non-invasive methods, such as medical imaging analysis and biomarker prediction. These techniques aid in early-stage detection, ultimately improving patient prognosis and manage- ment. This paper explores various methods of cirrhosis detection, with a focus on artificial intelligence applications that automate diagnosis, enabling clinicians to make better-informed decisions.

Index Terms: Cirrhosis, Machine Learning, Deep Learning, Liver Disease, Diagnosis, Medical Imaging, Non-invasive Meth- ods, AI in Healthcare, Liver Failure, Hepatitis.

I. INTRODUCTION

The liver is an essential organ in the human body that controls several vital functions, including energy storage, protein synthesis, glucose metabolism, and the detoxification of harmful drugs. Liver diseases such hepatitis, fatty liver, internal bleeding, fatigue, and jaundice affect millions of people globally and all need to be diagnosed early for effective treatment [1]. A dependable, accurate, and scalable system to predict and identify liver problems is therefore becoming more and more necessary. Such a system can lessen challenges and improve outcomes by supporting timely intervention and customized treatment strategies.

Early diagnosis and prevention are crucial in the fight against liver disease. Symptoms such as jaundice, abdominal pain, chronic fatigue, and unexplained weight loss often appear late, which

Power System Technology
ISSN:1000-3673
Received: 06-11-2024 Revised: 15-12-2024 Accepted: 05-01-2025

makes regular liver function tests (LFTs) and imaging studies essential. Public health initiatives promoting vaccination for hepatitis B, reducing alcohol consumption, and encouraging balanced diets can significantly reduce the burden of liver diseases. However, traditional diagnostic methods sometimes fall short, particularly in early-stage or asymp- tomatic cases, creating the need for cutting-edge technology such as machine learning (ML) to fill these gaps.

Hepatitis and cirrhosis are both severe liver disorders that pose major health risks and raise morbidity and mortality rates globally [2]. The underlying causes of cirrhosis, which is marked by irreversible scarring and loss of liver function, and hepatitis, which is an inflammatory disease of the liver, include viral infections, autoimmune illnesses, and excessive alcohol consumption. If these conditions are not recognized or treated, they may lead to potentially catastrophic outcomes, such as cancer and liver failure. A timely and precise diagnosis is crucial to halting the spread of many diseases.

This paper presents a framework for ensemble learning aimed at improving the classification of patients with cirrhosis. The proposed strategy increases the accuracy and reliability of predictions by utilizing several machine learning models. This approach combines a range of datasets with advanced feature selection algorithms to provide in-depth research and trustworthy findings. The goal is to develop a precise, scalable, and effective diagnostic tool to aid healthcare professionals in early identification and intervention. In the long run, this will reduce the burden on healthcare systems and enhance patient outcomes.

II. Literature Review

Furthermore, the study evaluates the efficacy of many clas- sification techniques, including Decision Tree (DT), Random Forest (RF), k-Nearest Neighbors (KNN), and Logistic Re- gression (LR), using a dataset of 583 liver patient pictures. Performance metrics including processing time, accuracy, and precision were analyzed. The KNN approach fared better than the other models in the test, with an accuracy rate of 72.04%. This study highlights the potential of machine learning techniques to enhance liver disease diagnosis and the need for more research on algorithm efficiency and accuracy.

This work contributes to the growing body of research on the use of AI to improve medical diagnosis, particularly for complex conditions like liver problems. Through the use of predictive modeling methods, this work paves the way for more effective and accessible healthcare treatments. [1]

This work employs explainable AI (XAI) to bridge the interpretability gap in liver cirrhosis prediction models, focus- ing on biomarkers instead of invasive techniques. It employs SHAP (SHapley Additive exPlanations) to provide insights into decision-making and machine learning models like Lo- gistic Regression and XGBoost to increase transparency and confidence in liver disease diagnostics. [2]

Power System Technology
ISSN:1000-3673
Received: 06-11-2024 Revised: 15-12-2024 Accepted: 05-01-2025

The study recommends using ensemble learning and elec- tronic health records (EHRs) to better classify people with cirrhosis and hepatitis C. Random Forests beat traditional AST/ALT ratios in terms of prediction accuracy, making it the best-performing classifier. By looking at data from both discovery and validation cohorts, it concluded that the best predictive factors were ALT and AST enzyme levels. This method gives physicians a simple yet effective tool for identi- fying liver diseases and predicting the course of the condition, which enhances clinical decision-making and outcomes. [3]

Recent advances in health informatics have brought atten- tion to the importance of early diagnosis in lowering the risks of liver illnesses, including cirrhosis, hepatitis, and fatty liver disease. Using deep learning techniques, researchers have looked at new approaches to better correctly identify and clas- sify liver disorders in their early stages. This study introduces a novel framework for deep neural network (DNN) and Spear- man's rank correlation-based robust feature analysis and clas- sification. Using 52 distinct characteristics, including Gray-Level Co-occurrence Matrix (GLCM) and Gradient-Level Co- occurrence Matrix (GLGCM) texture features, the method offers multi-dimensional classification of liver abnormalities. By contrast with conventional methods, the proposed strategy's improved predictive capabilities are shown, offering a possible avenue for the development of liver disease diagnostics and treatment alternatives. [4]

Autoimmune diseases are a major global health concern, re- gardless of an individual's age, gender, or nationality. Among these is autoimmune liver disease (AiLD), a chronic condi- tion marked by persistent inflammation of the liver that, if unchecked, can lead to major issues including cirrhosis or liver cancer. Despite their efficacy, existing diagnostic methods are often time-consuming and invasive, which highlights the need for innovative approaches to speed up detection and improve accuracy. Using state-of-the-art supervised machine learning and deep learning techniques, researchers have proposed an integrated diagnostic framework. By increasing prediction accuracy and reducing reliance on invasive procedures, such methods have the potential to revolutionize AILD diagnosis. They will provide physicians quick, economical, and patient- focused care. [5]

The World Health Organization reports that liver illness and liver cancer claim the lives of around one million people annually, and that ten new cases of hepatitis B and C are identified every day. Since the identification of liver illness often presents significant financial and procedural challenges, finding cost-effective options is essential. This study evaluates the efficacy of a number of supervised machine learning algorithms to speed up the identification of liver illness and reduce medical expenses. On the Indian Liver Patient Dataset from the UCI repository, this study uses classification techniques such as Random Forest, Decision Tree, Decision Tree SMOTE, Support Vector Classifier, K-Nearest Neighbor, AdaBoost, Stochastic Gradient Descent, and Artificial Neural Networks (ANN). With an astounding 87%

Power System Technology
ISSN:1000-3673
Received: 06-11-2024 *Revised: 15-12-2024* *Accepted: 05-01-2025*

performance rate, the data demonstrate that ANN is the most accurate algorithm. [6]

Liver disease is growing increasingly prevalent, with a high death rate and an increasing number of patients, as a result of several environmental and lifestyle factors, including excessive alcohol use, exposure to toxic fumes, eating contaminated food, and drug addiction. Serious adverse effects, including liver cancer, are often the result of liver malfunction that lasts for six months or more. To enhance the classification of liver disease, this work proposes a semi-supervised machine learning technique that combines Support Vector Machines (SVM) with K-Means clustering. Liver patient databases are used in the hybrid method to ascertain who is affected and the extent of liver damage. The high prediction accuracy of the study demonstrates how effectively SVM and K-Means combine to thoroughly examine and categorize liver disorders. [7]

Cirrhosis, characterized by the replacement of good liver tis- sue by scar tissue, diminishes liver function and is often caused by autoimmune disorders, non-alcoholic fatty liver disease, excessive alcohol use, and chronic conditions like hepatitis B or C. This progressive disease can lead to complications such as portal hypertension, ascites, liver failure, and an increased risk of liver cancer. Machine learning has developed into a powerful technique for the early diagnosis of liver cirrhosis by analyzing historical data. In this work, models like Gradient Boosted Classifiers utilizing XGBoost and Logistic Regression employ hyperparameter optimization. These findings pave the way for the development of advanced, reliable liver cirrhosis prediction models. [8]

Given the liver's critical role in blood purification and supporting essential activities, liver failure is a serious medical illness that can be lethal. Early liver failure prediction is made possible by the application of machine learning in healthcare. This study's objective is to use supervised machine learning techniques to create a multi-class classification model. To extract useful information from the dataset, pre-processing methods such as data visualization and univariate and bivariate analysis were applied. The model was assessed using a variety of performance indicators, including recall, accuracy, and F1 score. The suggested method shown potential in transforming the prediction of liver failure with an accuracy rate of 94.48%. These results encourage the creation of increasingly intricate and accurate models in the field by providing valuable data for additional study. [9]

The LivMarX system combines state-of-the-art machine learning with feature engineering and optimization techniques like as Genetic Algorithms, GridSearchCV, and Optuna. With an 86% accuracy rate and an AUC of 0.95, the Random Forest method—which uses inexpensive biochemical mark- ers—offers a straightforward and intelligible method of stag- ing liver cirrhosis, particularly in settings with constrained resources. [10]

This study introduces an artificial intelligence system based on machine learning to aid

Power System Technology
ISSN:1000-3673
Received: 06-11-2024 Revised: 15-12-2024 Accepted: 05-01-2025

healthcare providers in the early detection of liver cirrhosis. The authors compare various machine learning algorithms to forecast the likelihood of liver cirrhosis infection, aiming to enhance early diagnosis and treatment. [11]

This proof-of-principle study demonstrates the potential of deep transfer learning (DTL) for detecting cirrhosis based on standard T2-weighted MRI. The presented method for image-based diagnosis of liver cirrhosis achieved expert-level clas- sification accuracy, highlighting the feasibility of automatic detection using DTL. [12]

This study evaluates the use of artificial intelligence (AI) to automate the detection of cirrhosis from CT images. The machine learning model developed demonstrated significant potential in diagnosing cirrhosis, suggesting that AI can assist in the early detection and management of liver diseases. [13] This project compares the effectiveness of several machine learning techniques to reduce chronic liver disease through various models. The authors utilized numerous algorithms to analyze and predict liver cirrhosis, contributing to improved

early-phase non-invasive detection. [14]

This study presents and compares various binary classi- fier machine learning algorithms, including Artificial Neural Network, Random Forest, and Support Vector Machine, to classify blood donors and non-blood donors with hepatitis, fibrosis, and cirrhosis diseases. The proposed machine learning methods showed improved accuracy scores, enhancing the quality of classification and aiding health professionals in better decision-making. [15]

This study develops and evaluates a radiomics-based ma- chine learning model for detecting liver fibrosis on CT im- ages. The model demonstrated reasonable accuracy and high sensitivity, suggesting its potential as a non-invasive screening tool for liver fibrosis, contributing to earlier detection at a potentially curable stage. [16]

This work proposes leveraging transfer learning from large datasets annotated by radiologists to predict histological scores available on small annex datasets. The method combines weakly-supervised and self-supervised pretraining strategies, outperforming baseline classifiers in predicting cirrhosis, and achieving an AUC of 0.84 and balanced accuracy of 0.75. [17] This paper applies multiple imputation by chained equations to handle missing data and Principal Component Analysis for dimensionality reduction. It presents and compares several binary classifier machine learning algorithms to classify blood donors and non-blood donors with hepatitis, fibrosis, and cirrhosis diseases, achieving an accuracy score of 98.23% for Support Vector Machine, thereby improving classification quality. [18]

This study proposes a semi-supervised machine learning algorithm for liver disease prediction. The approach combines labeled and unlabeled data to improve prediction accuracy, demonstrating effectiveness in early detection of liver diseases, including cirrhosis. [19]

Power System Technology
ISSN:1000-3673
Received: 06-11-2024 Revised: 15-12-2024 Accepted: 05-01-2025

This paper introduces predictive modeling techniques aimed at diagnosing cirrhosis through machine learning. The authors explore several algorithms, emphasizing feature engineering and hyperparameter optimization to improve predictive accu- racy. The dataset used was preprocessed to manage imbal- ances and improve data quality. The authors also highlight the importance of model interpretability and propose metrics for assessing performance. They argue that machine learning models can complement traditional diagnostic methods by providing data-driven insights. [20]

III. METHODOLOGY

This study utilizes an approach to develop a machine learning classification model based on liver function test (LFT) data. The procedure includes multiple stages to ensure accurate predictions. The complete workflow is outlined in Figure 1 and detailed below.

Workflow Explanation

The sequence in Figure 1 consists of the listed steps:

1. **Data Collection**: Relevant LFT biomarkers such as AST, ALT, ALP, and Bilirubin are collected from clinical datasets. These biomarkers are crucial indicators of liver health.

2. **Data Preprocessing**: This step ensures data quality by addressing missing values, normalizing features to a standard scale, and handling outliers. Proper preprocessing is critical to improve the robustness of the model.

3. **Feature Extraction**: Important LFT biomarkers are extracted for use as model features. This step may involve domain-specific knowledge or statistical methods to identify relevant variables.

4. **Model Training**: ML algorithms are employed to train the model on preprocessed data. The choice of algorithm depends on the dataset and the classification task.

5. **Model Evaluation**: The trained model's performance is assessed using metrics such as accuracy, precision, recall, F1- score, and ROC-AUC. These metrics help ensure the model is both accurate and reliable.

6. **Hyperparameter Tuning**: The model's performance is improved by optimizing and using methods including Bayesian optimization, grid search, and random search.

7. **Result and Classification**: The optimized model is deployed to classify new test instances, providing predictions that aid clinical decision-making.

This methodology ensures a structured approach to devel- oping a resilient ML model for liver function analysis.

Power System Technology
ISSN:1000-3673
Received: 06-11-2024 Revised: 15-12-2024 Accepted: 05-01-2025

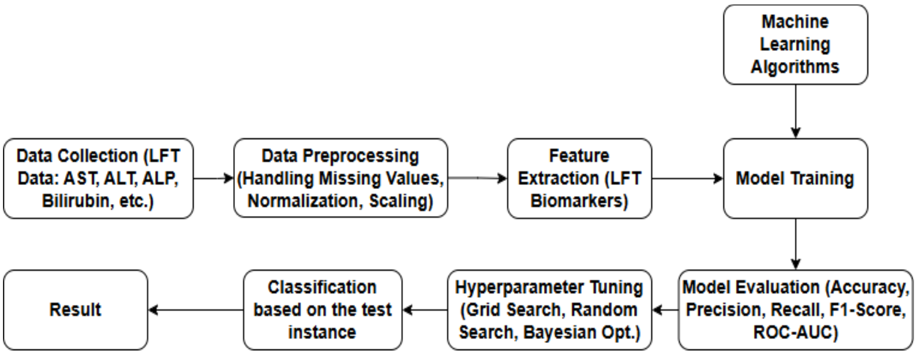

Fig. 1. Workflow for liver function test-based classification.

A. *Cirrhosis Stage Prediction using RFC*

The Random Forest Classifier is an ensemble learning method based on decision trees, widely used for both clas- sification and regression tasks. It operates by constructing multiple decision trees during training and outputting either the majority class (for classification) or the average prediction (for regression) from the individual trees. Random Forest mitigates the risk of overfitting which is common in individual decision trees by averaging results across multiple independent trees.

Random Forest Algorithm: The Random Forest algorithm combines the predictions of N independent decision trees. The prediction process for classification can be expressed mathematically as:

$$\hat{y} = \text{Mode}\,(f_1(x), f_2(x), \ldots, f_N(x)) \quad (1)$$

Where:

- $\hat{y}$ is the final predicted class.
- $f_i(x)$ is the class prediction by the i^{th} decision tree for the input x.
- N is the total no. of trees in the forest.
- Mode function selects the class with the majority votes among all trees.

Random Forest uses two main strategies to ensure model diversity and robustness:

1) Bagging (Bootstrap Aggregating): Each tree is trained on a bootstrap sample, where each candidate is randomly selected with replacement.

2) Random Feature Selection: At each split, a subset of features is selected randomly to determine the best split point. This reduces correlation among trees and promotes diversity.

Hyperparameter Tuning: Several hyperparameters can be adjusted in Random Forest to optimize performance:

Power System Technology
ISSN:1000-3673
Received: 06-11-2024 Revised: 15-12-2024 Accepted: 05-01-2025

- **Number of Trees (*n_estimators*)**: Determines the count of decision trees in the forest. Higher values can increase computation time but generate better performance.
- **Maximum Depth (*max_depth*)**: Limits the depth of each tree, controlling model complexity to prevent over- fitting.
- **Minimum Samples Split (*min_samples_split*)**: The minimum number of samples required to split a node.
- **Minimum Samples Leaf (*min_samples_leaf*)**: The minimum number of samples a leaf node must contain.

Performance Metrics of Random Forest Classifier

Metric	Accuracy	Precision	Recall	F1-Score
Value (%)	50.74	45.90	50.74	47.79

Pipeline Implementation and Preprocessing: A preprocess- ing pipeline was used to prepare the dataset for training:

- Numerical Features: Imputed missing values using the median and scaled features using StandardScaler to nor- malize the data.
- Categorical Features: Imputed missing values using the most frequent category and encoded them using OneHo- tEncoder.

The Random Forest Classifier demonstrated good accuracy in predicting the stages of cirrhosis. It is robust and has ability to handle both numerical and categorical data.

B. ***Logistic Regression Performance for Cirrhosis Stage Pre- diction***

Statistical method like Logistic Regression is used for gen- erating class labels by modeling the correspondence associated with input features and class probabilities using a logistic function. The competency of the algorithm in predicting the stages of cirrhosis is assessed using metrics such as Accuracy, Precision, Recall, and F1-Score.

The logistic function is represented as:

$$P(y = 1|X) = \frac{1}{\quad}, \quad z = w + \sum w x$$

(2)

Power System Technology
ISSN:1000-3673
Received: 06-11-2024 Revised: 15-12-2024 Accepted: 05-01-2025

Performance Metrics of RFC + GB Ensemble

Metric	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Value	49.25	49.37	49.25	48.99

Where:

$1 + e^{-z}$

0 $\quad i \quad i \quad n$

$i=1$

D. *XGBoost for Cirrhosis Stage Prediction*

XGBoost (Extreme Gradient Boosting) is a machine learn-

- $P(y = 1|X)$ is the probability of the positive class.
- z is the linear combination of weights (w_0, w_i) and input features (x_i).
- e is Euler's number ing algorithm designed for computational efficiency and model accuracy. It combines the predictions of an ensemble of weak learners (decision trees) using boosting.

The model optimizes an objective function:

The loss function is defined as:

$n \qquad K$

$1 \sum \quad m \qquad Obj(\Theta) = \sum L(y_i, \hat{y}_i) +$

$\sum \Omega(f_k)$ (7)

$L = -\frac{}{}$

m

$i=1$

$[y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$ (3)

where:

$i=1$

$k=1$

Power System Technology
ISSN:1000-3673
Received: 06-11-2024 Revised: 15-12-2024 Accepted: 05-01-2025

Where:

- $\hat{y}_i$ is the predicted probability.
- y_i is the true label for instance i.
- m is the quantity of training samples.

Performance Metrics of Logistic Regression Classifier

$L(y_i, \hat{y}_i)$: The loss function for prediction error.

- $\Omega(f_k)$: Regularization term to control model complexity.
- K: Number of trees.
- $\hat{y}_i$: The prediction for instance i, computed as:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i) \quad (8)$$

Metric	Accuracy	Precision	Recall	F1-Score
Value (%)	46.26	42.79	46.26	44.05

The optimization uses both first-order (g_i) and second-order

RFC + Gradient Boosting Ensemble for Cirrhosis Stage

(h_i) derivatives of the loss function:

$$g_i = \frac{\partial L(y_i, \hat{y}_i)}{\partial \hat{y}_i}, \quad h_i = \frac{\partial^2 L(y_i, \hat{y}_i)}{\partial \hat{y}_i^2} \quad (9)$$

Prediction

Ensemble learning leverages multiple models to improve predictive accuracy. This ensemble combines the strengths of Random Forest Classifier (RFC) and Gradient Boosting (GB) using a soft voting strategy, as described below:

Power System Technology
ISSN:1000-3673
Received: 06-11-2024 Revised: 15-12-2024 Accepted: 05-01-2025

The probability prediction for RFC is:

$P_{\text{RFC}}(y = c|X) = \frac{1}{} \sum P\ (y = c|X)$ (4)***Performance Metrics for XGBoost***

Metric	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
Value	47.76	44.16	47.76	45.02

TABLE I

PERFORMANCE METRICS OF XGBOOST FOR CIRRHOSIS STAGE

PREDICTIONWhere:

T t

t=1

Gaussian Naive Bayes for Cirrhosis Stage Prediction

Naive Bayes classifiers are probabilistic models based on

- T is the count of trees in the Random Forest.
- $P_t(y = c|X)$ is the prediction probability for class c from the t-th tree.

The Gradient Boosting prediction is given by:

Bayes' Theorem. The Gaussian Naive Bayes (GNB) classifier assumes that numerical features follow a Gaussian distribution. The theorem and likelihood equations are:

Bayes' Theorem:$P_{\text{GB}}(y = c|X) = \sigma$

K

k=1

!

$g_k(X)$

(5)

$P\ (X|C) \cdot P\ (C)$

$P\ (C|X) =$

$P\ (X)$

Power System Technology
ISSN:1000-3673
Received: 06-11-2024 Revised: 15-12-2024 Accepted: 05-01-2025

Gaussian Likelihood for a feature X_i:

(10) Where:

- K is the no. of boosting rounds.

$$P(X_i|C) = \frac{1}{\sqrt{2\pi\sigma_c^2}} \exp\left(-\frac{(X_i - \mu_C)^2}{2\sigma_c^2}\right) \quad (11)$$

- $g_k(X)$ represents the prediction from the k-th weak learner.
- σ is the sigmoid function.

The final ensemble prediction is computed using soft voting:

$$P_{\text{Ensemble}}(y = c|X) = \alpha P_{\text{RFC}}(y = c|X) + (1-\alpha) P_{\text{GB}}(y = c|X) \quad (6)$$

Where:

- α is the weight assigned to RFC predictions.

Where:

- μ_C and σ_c^2 are the mean and variance of the feature for class C.
- $P(C)$ is the prior probability of class C.
- $P(X|C)$ is the likelihood of X given class C.

The Gaussian Naive Bayes classifier is particularly suited for high-dimensional data with numerical features, making it applicable for this dataset.

Performance Metrics of Gaussian Naive Bayes

Metric	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
Value	26.86	43.58	26.86	28.52

Power System Technology
ISSN:1000-3673
Received: 06-11-2024 Revised: 15-12-2024 Accepted: 05-01-2025

IV. RESULTS & DISCUSSION

The evaluation of various machine learning models for predicting liver disease stages was carried out using several classification algorithms. The performance of each model was assessed based on four key metrics: accuracy, precision, recall, and F1-score. These metrics provide valuable insights into how effectively the models identify liver disease stages while min- imizing both false positives (incorrectly identifying a disease stage) and false negatives (failing to identify the correct stage). The results offer a comprehensive understanding of each model’s strengths and limitations in terms of classification reliability and robustness.

- **Logistic Regression (LR) and XGBoost (XGB)**: Lo- gistic Regression, a linear model, performed moderately with an accuracy and F1-score around 63%. Its linearity struggled to capture the complex, non-linear relationships in liver disease stages. In contrast, XGBoost, using gra- dient boosting, outperformed other models with accuracy and precision above 81%. Its ability to build decision trees that learn from previous errors allowed it to capture intricate feature interactions, offering strong predictive power. With a recall of 81.63%, it excelled in identifying true positives, making it suitable for clinical applications where minimizing false negatives is critical.

- **Random Forest**: The Random Forest model performed well but was slightly outperformed by XGBoost. By averaging multiple decision trees, it reduced the risk of overfitting. However, it did not capture data complexities as effectively as XGBoost, leading to slightly lower precision and F1-score.

- **Gaussian Naive Bayes**: The Gaussian Naive Bayes model showed moderate performance due to its as- sumption of feature independence. While computationally efficient, this limitation hindered its ability to capture complex relationships, resulting in lower precision and recall.

- **Voting Ensemble**: The Voting Ensemble, combining LR, XGBoost, and Random Forest, achieved the best performance with an F1-score of 82.78% and precision of 83.73%. By leveraging the strengths of multiple models, it balanced false positives and negatives, making it ideal for clinical use.

Model Performance Comparison for Cirrhosis Classification

Model	Accuracy	Precision	Recall	F1-Score
XGBoost & Random Forest	49.25	49.37	49.25	48.99
Gaussian Naive Bayes	26.86	43.58	26.86	28.52
Random Forest	50.74	45.90	50.74	47.79

Power System Technology
ISSN:1000-3673
Received: 06-11-2024 Revised: 15-12-2024 Accepted: 05-01-2025

Logistic Regression	46.26	42.79	46.26	44.05
XGBoost	47.76	44.16	47.76	45.02

The comparison of models reveals that more advanced techniques like XGBoost and the Voting Ensemble outperform simpler models such as Logistic Regression and Gaussian Naive Bayes. XGBoost excels due to its gradient boosting approach, where decision trees learn from errors made by previous models, allowing it to capture complex patterns in the data. Random Forest, although effective, fell short in precision and F1-score, likely due to its inability to model intricate feature interactions as effectively as XGBoost. The Voting En- semble, which aggregates the predictions of multiple models, achieved the highest precision and F1-score, benefiting from the combined strengths of individual models and reducing bias. In contrast, Logistic Regression underperformed, as its linear nature struggled to model non-linear relationships, particularly in multi-stage liver disease predictions.

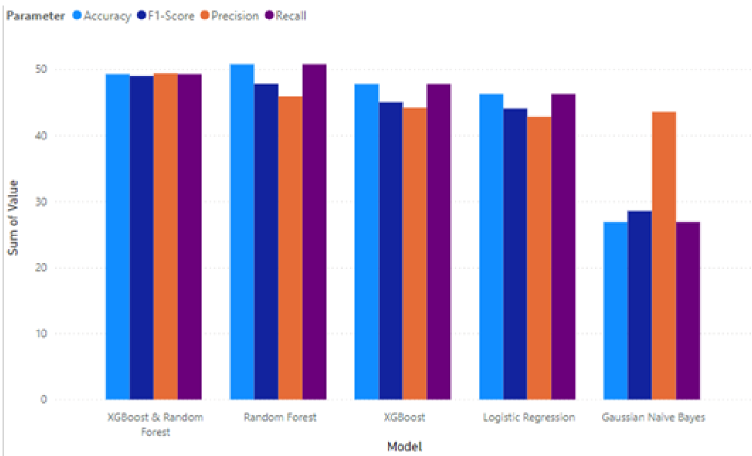

Fig. 2. Model performance for Cirrhosis.

ROC Curve, Accuracy and Log loss curve for Model Perfor- mance Comparison for Cirrhosis Classification

- **ROC Curve**: The Random Forest (blue) and XGBoost (green) models' classification performance is contrasted in the Receiver Operating Characteristic (ROC) curves. The model quality is shown by the Area Under the Curve (AUC). The AUCs of the Random Forest and XGBoost models are 0.56 and 0.48, respectively. Given that both values are near 0.5, discrimination performance is low.

- **Log Loss**: This curve might show training loss across several epochs or estimators. A learning algorithm that is unresponsive or inadequately trained may be the cause of the flat line, which shows that the loss stays constant.

Power System Technology
ISSN:1000-3673
Received: 06-11-2024 Revised: 15-12-2024 Accepted: 05-01-2025

- **Accuracy**: The accuracy of the finished model across var- ious iterations or hyperparameters is probably represented by this curve. The flat line, however, suggests that there is no improvement or variation in the predictive perfor- mance and that accuracy does not change significantly. Overall, the model performs well for Classes 1, 2, and 4, but some overfitting is observed. This issue can be addressed with techniques such as early stopping, regu- larization, and other model optimization strategies.

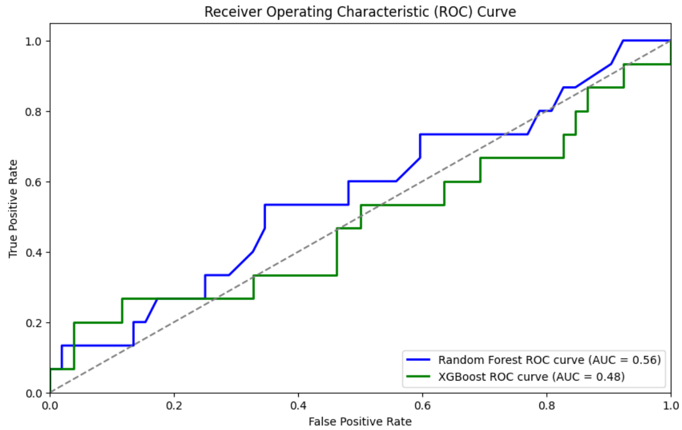

Fig. 3. ROC Curve.

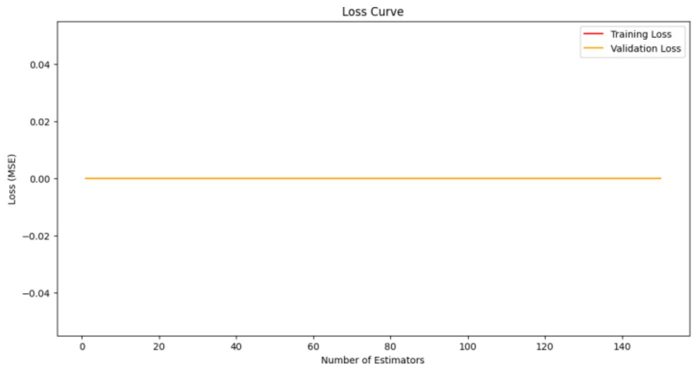

Fig. 4. Log Loss Curve.

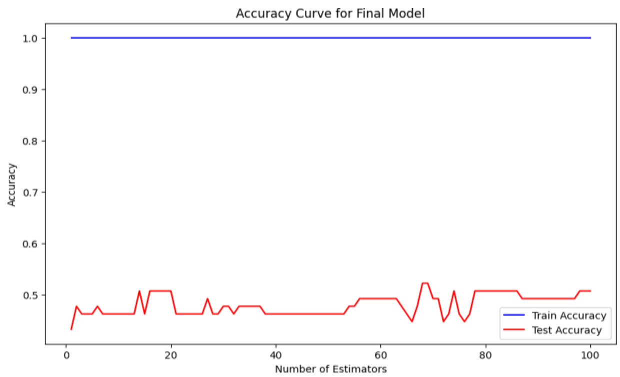

Fig. 5. Accuracy Curve.

Power System Technology
ISSN:1000-3673
Received: 06-11-2024 Revised: 15-12-2024 Accepted: 05-01-2025

V. CONCLUSION

Long-term liver damage, such as that caused by non- alcoholic fatty liver disease (NAFLD), viral infections like hepatitis, or chronic alcohol use, leads to cirrhosis. Continuous liver damage causes fibrosis, or the development of scar tissue, which impairs the liver's capacity to function properly. The liver can no longer carry out vital tasks like cleaning the body or making vital proteins as cirrhosis worsens, causing irreparable damage that may eventually result in liver failure. Maintaining a healthy lifestyle is essential for controlling cirrhosis and stopping additional liver damage. While frequent exercise helps maintain a healthy weight and lessens excessive strain on the liver, a well-balanced diet that includes meals high in nutrients can improve liver function. In addition, abstaining from alcohol and being aware of drugs that can damage the liver are crucial aspects of controlling the illness. Another important preventive measure is hepatitis vaccina- tion, which provides defense against viral liver infections that can exacerbate the illness. Improving results requires early detection of cirrhosis and routine monitoring. The disease can be slowed down and liver function preserved for as long as feasible with prompt intervention, which includes lifestyle

changes and the right medical care.

VI. FUTURE WORK

In order to improve early detection techniques, future cir- rhosis research will incorporate increasingly complex ma- chine learning algorithms that can process data from multiple sources. To improve diagnostic accuracy, a major area of focus will be the creation of models that integrate imaging and clinical data, such as CT and MRI scans. Personalized treatment plans for patients with cirrhosis are also anticipated to undergo a revolution with the integration of genomic and proteomic biomarkers into predictive frameworks. Using real- time monitoring systems to track the progression of diseases and facilitate prompt clinical interventions will be another area of interest. Improvements in AI-driven decision support systems will also allow medical practitioners to customize treatments according to the unique characteristics of each patient. It will be essential to investigate how artificial in- telligence might be used to forecast cirrhosis complications, such as liver failure and hepatocellular carcinoma. Finally, addressing the various types of cirrhosis and enhancing global healthcare outcomes will require utilizing cross-disciplinary collaboration and global health data.

REFERENCES

[1] Soni, Akanksha. ”Performance analysis of classification algorithms on liver disease detection.” In 2021 IEEE Mysore Sub Section International Conference (MysuruCon), pp. 1-5. IEEE, 2021.

[2] G. Arya, A. Bagwari, H. Saini, P. Thakur, C. Rodriguez, and P. Lezama, ”Explainable AI for Enhanced Interpretation of Liver Cirrhosis Biomarkers,” *IEEE Access*, vol. 11, pp.

Power System Technology
ISSN:1000-3673
Received: 06-11-2024 Revised: 15-12-2024 Accepted: 05-01-2025

123729-123741, 2023.

[3] D. Chicco and G. Jurman, ”An Ensemble Learning Approach for Enhanced Classification of Patients With Hepatitis and Cirrhosis,” *IEEE Access*, vol. 9, pp. 24485-24498, 2021.

[4] Prakash, K., and S. Saradha. ”A deep learning approach for clas- sification and prediction of cirrhosis liver: non alcoholic fatty liver disease (NAFLD).” In 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), pp. 1277-1284. IEEE, 2022.

[5] George, E. Ben, G. Jeba Rosline, Abraham Varghese, and Rajesh D. Gnana. ”Classification Framework for Autoimmune Liver Disease using Machine Learning and Deep Learning Techniques.” In 2024 5th Tech- nology Innovation Management and Engineering Science International Conference (TIMES-iCON), pp. 1-5. IEEE, 2024.

[6] Nigatu, Selamawit Sileshi, Poorna Chandra Reddy Alla, R. N. Raviku- mar, Krishnanand Mishra, G. Komala, and Gloria Richard Chami. ”A Comparitive Study on Liver Disease Prediction using Supervised Learning Algorithms with Hyperparameter Tuning.” In 2023 Inter- national Conference on Advancement in Computation and Computer Technologies (InCACCT), pp. 353-357. IEEE, 2023.

[7] Rani, A. Jaya Mabel, S. Nishanthini, DC Jullie Josephine, Hridya Venugopal, S. Gracia Nissi, and V. Jacintha. ”Liver disease prediction using semi supervised based machine learning algorithm.” In 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC), pp. 1389-1392. IEEE, 2022.

[8] Natarajan, B., P. Murali, Prabu Selvam, K. Venkatraman, and N. R. Nagarajan. ”Predictive Modeling for Cirrhosis Diagnosis: A Machine Learning Exploration.” In 2024 International Conference on Electronics, Computing, Communication and Control Technology (ICECCC), pp. 1- 6. IEEE, 2024.

[9] Sawant, Lalithesh D., Raghavendra Ritti, N. Harshith, Ashwini Kodi- palli, Trupthi Rao, and B. R. Rohini. ”Analysis and Prediction of Liver Cirrhosis Using Machine Learning Algorithms.” In 2023 3rd International Conference on Intelligent Technologies (CONIT), pp. 1- 5. IEEE, 2023.

[10] S. K. Kamath, S. K. Pendekanti, and D. Rao, ”LivMarX: An Optimized Low-Cost Predictive Model Using Biomarkers for Interpretable Liver Cirrhosis Stage Classification,” *IEEE Access*, vol. 12, pp. 92506-92522, 2024.

[11] S. S. Nigatu, P. C. R. Alla, R. N. Ravikumar, K. Mishra, G. Komala, and G. R. Chami. ”Liver Cirrhosis Prediction using Machine Learning Approaches”. In 2023 International Conference on Advancement in Computation and Computer Technologies (InCACCT).

[12] J. A. Luetkens, P. Homsi, A. Sprinkart, et al.”Detection of Liver Cirrhosis in Standard T2-

Received: 06-11-2024 *Revised: 15-12-2024* *Accepted: 05-01-2025*

weighted MRI Using Deep Transfer Learning". In European Radiology, 2021.

[13] S. Carey, S. E. Fischer, C. McIntosh, et al."Using Artificial Intelligence to Predict Cirrhosis from CT Scans". In Clinical and Translational Gastroenterology, 2023.

[14] L. D. Sawant, R. Ritti, N. Harshith, et al."Analysis and Prediction of Liver Cirrhosis Using Machine Learning Algorithms". In 2023 3rd International Conference on Intelligent Technologies (CONIT).

[15] F. B. Mostafa and M. E. Hasan. "A Comparative Study of Machine Learning Algorithms Using Clinical Data for Liver Disease Prediction". In nternational Journal of Computational Intelligence and Applications, 2023.

[16] J. J. Yoo, K. Namdar, S. Carey, et al."Non-invasive Liver Fibro- sis Screening on CT Images Using Radiomics". In arXiv preprint arXiv:2211.14396, 2022.

[17] E. Sarfati, A. Bone, M. M. Rohe, et al."Learning to Diagnose Cirrhosis from Radiological and Histological Labels with Joint Self and Weakly- Supervised Pretraining Strategies".In arXiv preprint arXiv:2302.08427, 2023.

[18] F. B. Mostafa and M. E. Hasan."Machine Learning Approaches for Binary Classification to Discover Liver Diseases Using Clinical Data". In arXiv preprint arXiv:2104.12055, 2021.

[19] A. J. M. Rani, S. Nishanthini, D. C. J. Josephine, et al."Liver Dis- ease Prediction Using Semi-Supervised Based Machine Learning Algo- rithm".In 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC)

[20] B. Natarajan, P. Murali, Prabu Selvam, K. Venkatraman, and N. R. Nagarajan."Predictive Modeling for Cirrhosis Diagnosis: A Machine Learning Exploration". In 2024 International Conference on Electronics, Computing, Communication, and Control Technology (ICECCC).