Application of Machine Learning Methods in Liver Cirrhosis Prediction

Rehana Parvin, Mst. Rashida Pervin, Mohammed Motaher Hossain* and Raffat Arman Islam

[1,2,3,4]International University of Business Agriculture and Technology
Uttara, Dhaka-1230, Bangladesh
motaher@iubat.edu

Abstract. Liver cirrhosis is typically described as the end stage of chronic hepatic disease, in which progressive and irreversible deposition of fibrotic tissue leads to the progressive hepatic dysfunction. Early diagnosis of cirrhosis at an early stage gives a significant benefit, as it allows timely therapeutic interventions and the introduction of a complex of health management measures. In order to assess the opportunities of machine learning methods in this scenario, we designed an automated predictive model and compared the performances of various algorithms, such as the Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), Extreme Gradient Boosting (XGBoost), Gradient Boosting Machine (GBM), and Light Gradient Boosting Machine (LGBM). The empirical study was based on a dataset of 584 patient records that were annotated with eleven salient clinical variables. The findings showed that XGBoost had the best performance measures with an accuracy of 82.32%, an F1-score of 83.78%, a sensitivity of 91.46% and a precision of 77.32%. Both LGBM and GBM also demonstrated good predictive abilities. These results highlight the potential usefulness of machine learning tools to complement the accuracy of early diagnosis and clinical management of patients with liver cirrhosis.

Keywords: Liver Cirrhosis, SMOTE, RF, XGBoost, Machine Learning.

1 Introduction

Liver cirrhosis is a systemic hepatic vasculature disease, which still remains a formidable public health issue, plaguing millions of people around the globe and causing a significant morbidity and mortality burden [1].One of the most annoying aspects of cirrhosis is its insidious nature in the prodromal stage, which often complicates the ability to detect it in time and prevents early treatment. According to the estimates provided by the World Health Organization, millions of people are getting heaptologic disorders every year, but a significant percentage of them do not get diagnosed until the disease progresses [2]. This kind of undercover development highlights the need to have powerful, timely diagnostic models. The inability to identify cirrhosis at an early stage may trigger irreversible liver damage, which will result in severe complications and high mortality rates [3]. In the absence of a vaccine that is universally effective, careful and timely evaluation of the state of hepatic functioning is crucial, not only to maximize the prognostics of patients, but also to prevent further transmission of the disease [4].The introduction of artificial intelligence (AI) into the

M. S. Arefin et al. (eds.), *Proceedings of the International Conference on Intelligent Data Analysis and Applications (IDAA 2025)*, Advances in Intelligent Systems Research 206,
https://doi.org/10.2991/978-94-6239-664-7_21

healthcare industry has significantly improved the work of diagnostic systems, especially in recent years. Machine learning (ML) and other artificial intelligence (AI) technologies offer powerful pattern recognition approaches and can be used to predict the early stages of prognostication in clinical settings. This ability is of utmost significance in life threatening cases of cirrhosis, hepatitis and hepatocellular carcinoma. Combined, these hepatic diseases cause over two million deaths each year, which is about 4 percent of the total death worldwide, which is a terrifying figure that can be translated to one out of twenty five people dying due to liver related diseases [5].The liver is considered to be a central organ in the context of human physiology, and it balances a range of vital functions, such as detoxification, protein synthesis, bile production, and an overall metabolism of nutrients. It is also crucial in blood clotting, immune response and storage of vitamins and minerals. The liver has many different responsibilities, so it is prone to many different disorders, including fascioliasis, alcoholic liver disease, primary sclerosing cholangitis (PSC), fatty liver disease, Gilbert syndrome, hepatitis, cirrhosis, and liver cancer. Chronic liver diseases such as cirrhosis are among them and are major causes of increased mortality rates around the world [6].Although there has been a significant surge in academic research on hepatic fibrosis and other forms of chronic hepatic diseases, most of the modern studies are still based on the traditional methods of statistical analysis, especially the use of logistic regression [7, 8].These traditional approaches are often not sufficient to produce strong predictive power despite their relatively high interpretability. On the other hand, modern machine learning algorithms have demonstrated better performance in a variety of clinical fields. However, the blackbox, or opaque nature of these models still remains a barrier to their general acceptance in clinical practice and broad implementation [9].

The paper is outlined into various parts as follows: Section 2 critically reviews the relevant literature and Section 3 provides a comprehensive description of the dataset used. Section 4 elaborates on the machine-learning approach to be applied in this investigation, followed by Section 5, which outlines the criteria to be employed in the evaluation of the model. Section 6 introduces the empirical results and makes a substantive discussion and finally Section 7 gives a conclusive synthesis of the main findings, and also gives the future research directions.

2 Related Work

Liver cirrhosis is a progressive and chronic hepatic disease, which develops due to the continued hepatic damage. Among etiological factors, there are often prolonged drinking, viral hepatitis, and non-alcoholic fatty liver disease, and each of them causes a sustained pathogenic load on the liver parenchyma. During the disease progression, the patients usually develop serious complications like hepatic failure, portal hypertension, and hepatic cellular carcinoma [10]. These complications are multifaceted and increasing in severity, which highlights the need to diagnose them early to reduce morbidity and mortality. Machine learning (ML) has become a useful resource in predicting the development of cirrhosis in recent years. Machine learning algorithms explain complex patterns that predetermine the development of hepatic degradation

through the combination of various clinical and laboratory attributes, thus offering a promising complement to the existing diagnostic procedures [11].Various types of data have been implemented in machine-learning applications of liver disease, including demographic, clinical, molecular, and radiologic. Machine learning technologies promise to revolutionize hepatology and liver transplantation as emphasized by modern researchers [12], meaning that prediction algorithms will become infiltrated in clinical practice. Some studies have used ML models to detect liver disease at a generalized level and avoid differentiating among specific hepatic pathology [13]. Some of the used algorithms include Decision Trees (DT), Linear Discriminant Analysis (LD), Fine-Gaussian Support Vector Machines (SVM), and Logistic Regression (LR) which have proven to be significant in terms of prognostics in hepatological settings.

A number of methodological frameworks have been developed to enhance predictive accuracy especially in high-dimensional data as typified situations. As an example, dimensionality-reduction algorithms, including Principal Component Analysis (PCA), Factor Analysis (FA), and Linear Discriminant Analysis (LDA), have been used to improve the accuracy of the model by isolating significant features [14]. Moreover, feature-selection pipelines have been optimized by incorporation of optimization methods especially Particle Swarm Optimization (PSO) which is followed by classification using various machine-learning algorithms- such as Bayesian networks and J-48 decision trees and have delivered promising results in liver disease detection [15].Several academic studies have highlighted the usefulness of Logistic Regression (LR) in the prognostication of chronic hepatic diseases, especially in the backdrop of resource constraints where simple and understandable models are of direct clinical importance [16].More so, specific research on the Taiwanese cohort utilized a set of machine learning paradigms to predict Fatty Liver Disease (FLD). In the said study, the XGBoost algorithm significantly outperformed other traditional ones such as neural networks, random forests, and support vector machines, thus, testifying to its strength in the hepatic disease predictive analytics [17, 18].

A study by Victor Anthonysamy and others (2023) compared several supervised learning algorithms, which included: support vector machines, k -nearest neighbours, a hybrid voting classifier and a multi-layer perceptron neural network as applied to a dataset of liver disease cases. The experimental results indicated that the hybrid voting classifier achieved the highest classification rate of 78.62, and thus, the classifier had superior prediction compared to those of the rest of the methods considered [19].Likewise, Anil Utku (2023) assessed a set of machine learning architectures, such as multilayer perceptron, decision trees, k-nearest neighbours, logistic regression, naive Bayes, random forests, and support vector machines, to identify cirrhosis using the deep learning paradigm. The multilayer perceptron proved to be the most capable model with an accuracy of 0.8048 and, as such, highlights its effectiveness in identifying cirrhotic pathology [20].

3 Dataset Description

3.1 Clinical Feature

The data used in this study was obtained with the Kaggle open-source system (https://www.kaggle.com/datasets/uciml/indian-liver-patient-records) and includes clinical data of 583 individual patients.

The dataset has also been created in a manner that it is easy to carry out the rigorous examination of hepatic pathology through the combination of the demographic and biochemical variables that are relevant to liver health examination. All of the observations are associated with a specific patient record that includes a collection of laboratory indices that cannot be omitted in clinical diagnostics. Attributes of the repository consist of eleven items, among which the response variable, as presented in Table 1. Some of these predictors include the age of the patient, their sex, and a list of biochemical indexes such as serum bilirubin, hepatic enzymes, alanine transaminase (ALT), aspartate transaminase (AST), and alkaline phosphatase (ALP), total protein, and albumin-globulin ratio. The outcome variable, which is denoted as Liver Disease, describes the existence or possibly lack of hepatic disease, whereby the value of 1 denotes a positive diagnosis whereas 2 indicates the negative diagnosis.

Table 1. Overview of clinical features in the liver disease dataset

SI. No	Attribute Name	Type	Description
1	Age	Numerical	Age of the patient in years
2	Gender	Categorical	Biological sex of the patient (0 = Male, 1 = Female)
3	Total_Bilirubin	Numerical	Total bilirubin concentration in the blood
4	Direct_Bilirubin	Numerical	Direct bilirubin concentration in the blood
5	Alkaline_Phosphotase	Numerical	Level of alkaline phosphatase enzyme
6	Almine_Aminotransferase	Numerical	Level of ALT (Alanine Aminotransferase) enzyme
7	Aspartate_Aminotransferase	Numerical	Level of AST (Aspartate Aminotransferase) enzyme
8	Total_Protiens	Numerical	Total protein concentration in the blood
9	Albumin	Numerical	Albumin concentration in the blood
10	Albumin_and_Globulin_Ratio	Numerical	Ratio of albumin to globulin in the blood
11	Liver_Disease	Categorical	Class label (1 = Liver Disease Present, 2 = Liver Disease Absent)

The data includes four clinically recognised phases of hepatic health hence giving a delicate and extensive view of the process of cirrhosis. The first, which can be called Healthy, is people who demonstrate normal hepatic functioning and do not have any identifiable tissue damage. The second phase, which is termed Fatty Liver, indicates early hepatic impairment, which is often explained by the retention of lipids in hepatocytes. Move a step higher, the Fibrosis stage becomes more characterized by intense scarring and structural destruction therefore beckoning further hepatic destruction. Finally, the Cirrhosis stage also represents the most critical and often irreversible form of hepatic injury, thus making it significantly challenging to provide therapeutic treatment and progress to recovery. Figure 1 below is a graphic representation of the stages mentioned above that shows a percentage representation of each condition within the dataset.

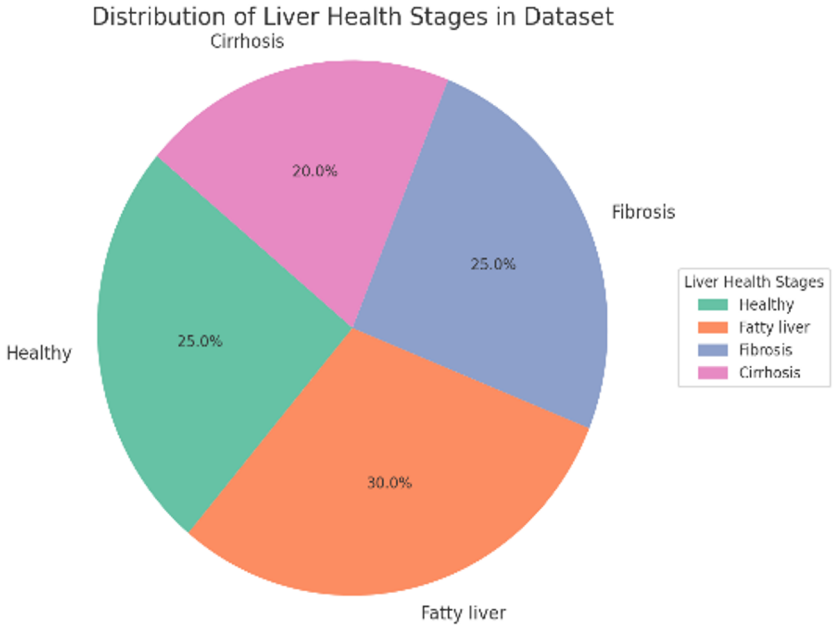

Fig. 1. Percentage Distribution of Liver Health Stages Representing Cirrhosis Progression

3.2 Feature Analysis

Table 2 provides a statistical summary of the five significant features of liver disease dataset, i.e., Age, Gender, Total Bilirubin, Direct Bilirubin, and Alkaline Phosphotase. The patients aged range between 4 and 90 years and the age mean 44.75 with a standard deviation of 16.19 that shows a wide age range among the patients. The gender is coded as a binary variable with a mean of 0.76 that indicates that one gender category is greatly represented in the dataset. The values of both Total Bilirubin and

Direct Bilirubin exhibit a significant degree of variability and skewness in their distributions with maximum values of 75.0 and 19.7 accordingly, thus emphasizing the presence of the potential outliers or severe hepatic issues. Alkaline Phosphotase also exhibits a high dispersion with the highest values of 2110 and the standard deviation of 242.94. More importantly, there is no observed value missing under these variables hence further supporting the validity of the dataset to be used in the analysis.

Table 2. Summary statistics of key features

Features	Mean	Std Dev	Min	Max
Age	44.75	16.19	4.0	90.0
Gender	0.76	0.43	0.0	1.0
Total Bilirubin	3.30	6.21	0.4	75.0
Direct Bilirubin	1.49	2.81	0.1	19.7
Alkaline Phosphotase	290.58	242.94	63.0	2110.0

An initial statistical analysis was performed to determine the distribution, variability and completeness of the data. The age of patients was between 4 and 90 years with an average age of about 44.7 years. The gender variable was coded as the number, the distribution is strongly skewed, as approximately 75.6% of the records belong to the category of class 1, which we assume is the female patients, and the rest 24.4% to the category of 0, which is probably the male patients. This imbalance between sexes is graphically summed up in Figure 2 that represents a bar chart of the proportional representation of each gender category. Total bilirubin and direct bilirubin are some of the biochemical markers that have high variability and skewed distributions. Their highest values, 75.0 and 19.7 respectively, indicate the occurrence of outliers or cases of hepatic dysfunction of severe nature. In the same way, liverenzyme markers, including alkaline phosphatase, alanine aminotransferase, and aspartate aminotransferase, have wide ranges, which is evidence of heterogeneity in hepatic function within the group of patients. Another salient strength of this dataset is data completeness. All the attributes are filled to the brim, except the Albumin and Globulin Ratio that has a low percentage of missing values, approximately 0.69. This level of completeness attained herein can be used to increase the accuracy of the data repository making it especially susceptible to classification activities as well as predictive modelling activities.

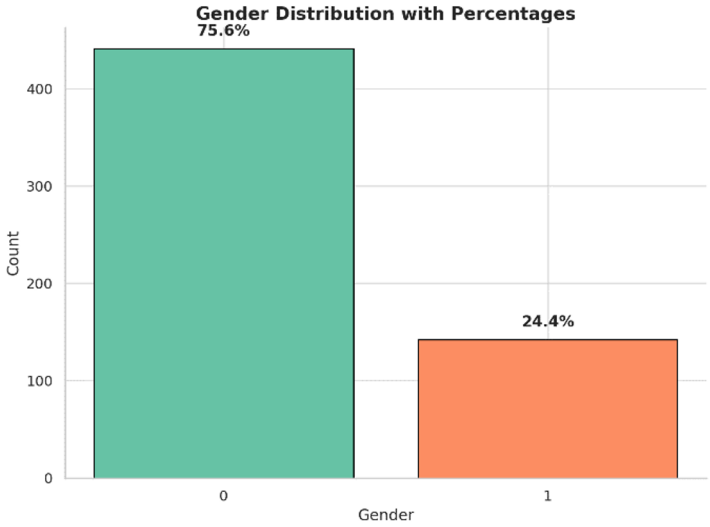

Fig. 2. Gender Distribution with Percentages

It performed an extensive evaluation of data completeness to identify any possible gaps that can pose a threat to further analyses or modelling activities. As shown in Table 3, only one feature, i.e., Albumin and Globulin_ Ratio, has missing values, which is only four records, or 0.69% of the whole dataset. There are no missing data in all the other attributes, such as Age, Gender, Total_Bilirubin, Albumin and Liver_ Disease _Status

Table 3. Missing values summary

Feature	Missing Values	Missing Percentage
Albumin_and_Globulin_Ratio	4	0.69%
Age	0	0.00%
Gender	0	0.00%
Total_Bilirubin	0	0.00%
Direct_Bilirubin	0	0.00%
Alkaline_Phosphotase	0	0.00%
Alamine_Aminotransferase	0	0.00%
Aspartate_Aminotransferase	0	0.00%
Total_Protiens	0	0.00%
Albumin	0	0.00%
Dataset	0	0.00%

This finding is graphically summarized in Figure 3 that provides a bar chart that outlines the percentage of missing observations of each feature.The high percentage of zero values in most bars testifies to an extremely high completeness of data sets. The fact that there is only one variable with a negligible level of missingness allows declaring the dataset incredibly trustworthy, which requires only a minimum of preprocessing, including simple imputation or row deletion.

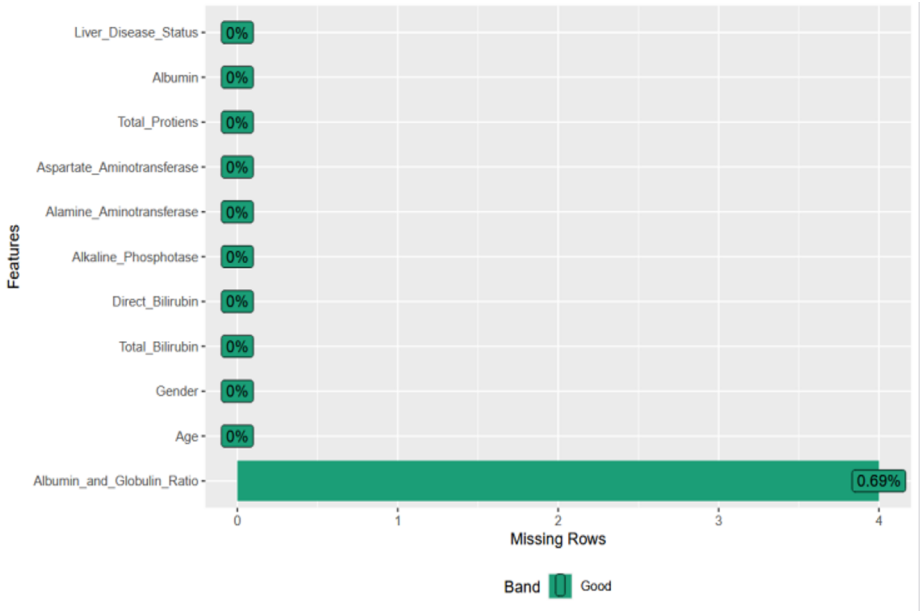

Fig. 3. Missing Data Overview in Medical Dataset

Figure 4 shows the age-based distribution of the stratified people based on liver disease status. The histogram has two different cohorts; the clinically diagnosed liver disease (Status 1, represented in red) and those who are not (Status 2, represented in cyan). Most of the subjects fall between the age groups of 30 and 60 with a steep peak around age 50. Patients with liver disease are overrepresented in most age groups indicating a higher prevalence in the dataset. The overlaid bars highlight the age groups in which both groups intersect each other, hence, providing a clear comparative age distribution picture among the two health conditions.

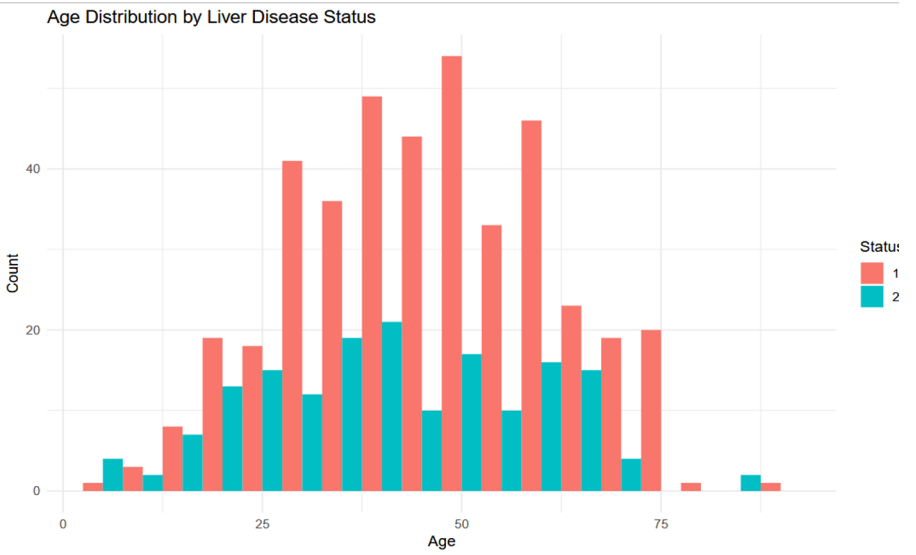

Fig. 4. Overview of Age Distribution by Liver Disease Status

Figure 5 shows a sequence of box plots, which compare the distribution of major liver functioning biomarkers in two groups distinguished by the presence of liver disease. The subplots represent each of the biomarkers individually, including Total Bilirubin, Alkaline Phosphatase, Albumin, and so on, and visualise the results of individuals with liver disease (Status 1) and without (Status 2). The line in the middle of each box represents the median and the boundary of the box represents the interquartile range (IQR). Whiskers are drawn to the data spread and the outliers are represented as single points outside the whiskers. Different colours are used to increase visual differences between the two groups. Such plots indicate that there is a great disparity in various biomarkers between the two statuses, which highlights their possible use as discriminative variables of liver health conditions.

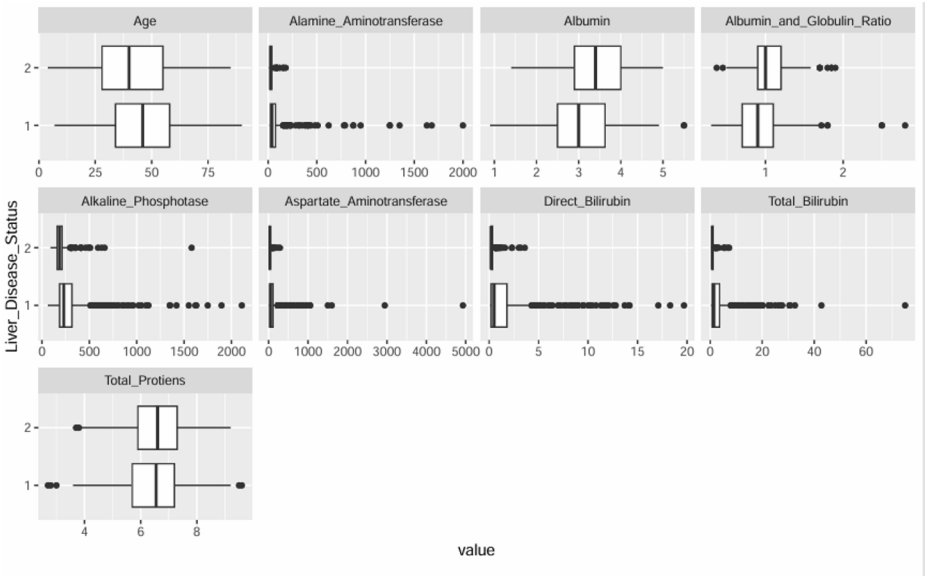

Fig. 5. Box Plot Comparison of Liver Biomarkers by Liver Disease Status.

Figure 6 shows a heatmap of correlation, which identifies the relationships between a range of clinical variables, such as Age, Total Bilirubin, Albumin, and so on. In the correlation matrix, the pairwise correlation coefficients are represented visually in each cell by a chromatic scale: deep blue colors indicate strong positive relationships, deep red colors indicate strong negative relationships, and white color indicates insignificant or null relationships. Most notably, the Total Bilirubin, Direct Bilirubin, Albumin, the Albumin_and_Globulin_Ratio, have a strong positive correlation; however, the Albumin_and_Globulin_Ratio and Total_Proteins have a strong negative correlation. In turn, this visualization provides the salient insight into the co-movement of variables, thus informing the selection of features and allowing further research.

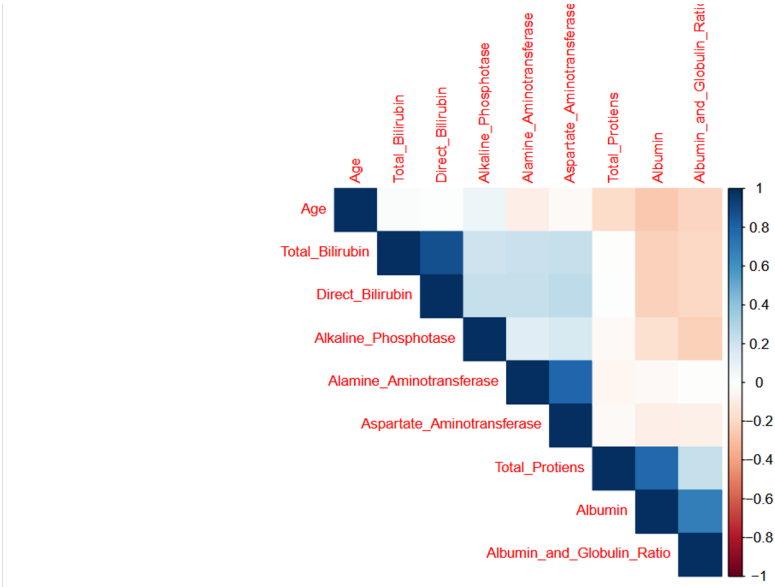

Fig 6. Heatmap of Liver Disease.

Figure 7 shows a set of histograms, which depict the frequency distributions of some of the most important laboratory measurements, including Age, Total_Bilirubin, Direct_Bilirubin, Alkaline_Phosphatase, Albumin, and other variables. The subplots show the frequency of specific numerical values, and the horizontal axis represents the range of the variable and the vertical axis represents the number of observations. Based on these representations it is clear that most of the variables have a right-skewness: they have a larger sample at the lower values and a smaller tail at the higher values. This graphical analysis helps to have a clear picture of the variability and central tendencies of the data set, helps identify the possible outliers, and helps to determine whether the distributions follow a normal distribution.

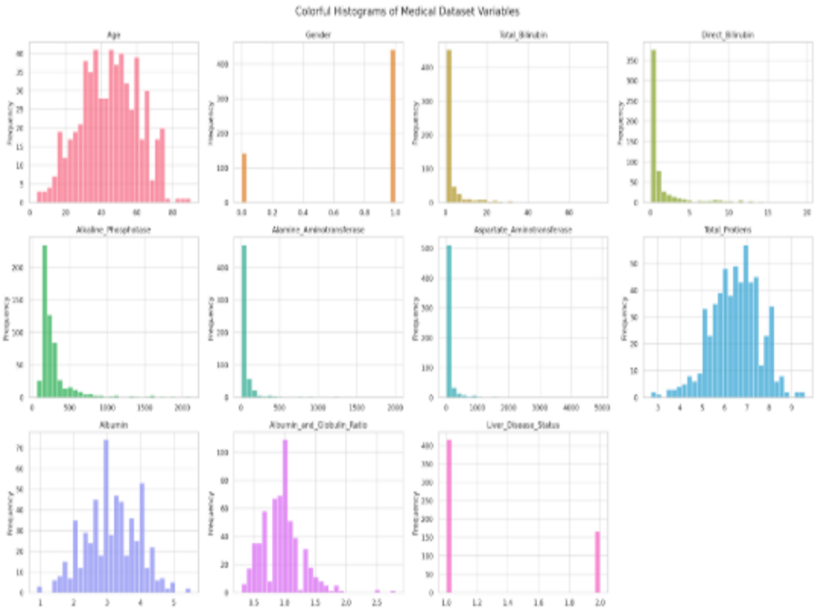

Fig 7. Distribution Analysis of Medical Variables through Histograms.

Figure 8 shows a pair plot which illustrates scatter plots, histograms, and correlation coefficients between six medical variables, Age, Total_Bilirubin, Direct_Bilirubin, Alkaline_Phosphatase, Alamine_Aminotransferase and Aspartate_Aminotransferase. The panels on the diagonal show the distribution of each variable, and the lower triangle is made of scatterplots that visually show the relationship between two variables. The correlation coefficients are reported in the upper triangle, and the stronger relationships are highlighted with the help of asterisks. Specifically, Total_Bilirubin and Direct_Bilirubin have a very strong positive correlation (0.875) and Alamine_Aminotransferase and Aspartate_Aminotransferase are also correlated (0.792), which means that they may have similar clinical implications. The visualization will help to detect strong linear relationships and multicollinearity in the data.

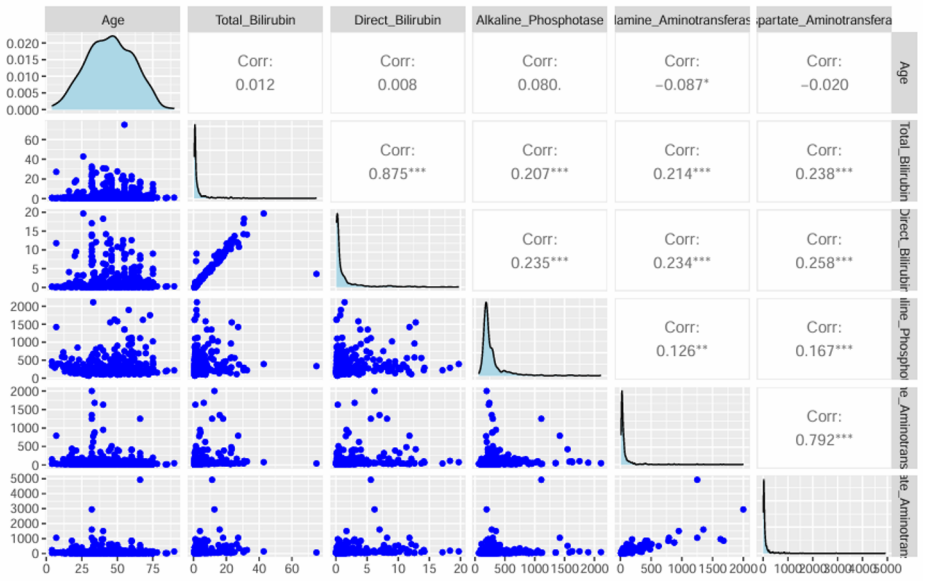

Fig 8. Pairwise Relationship Analysis of Key Medical Variables.

Figure 9 below shows violin plots of the distribution of several medical features stratified by the two dataset classes denoted by 1 and 2. The density profile and the degree of a specific variable, e.g., Age, Albumin, Alkaline_Phosphatase, etc., divided by class, is depicted in each subplot. The width of the individual violin indicates the concentration of the observations in the various ranges of values. As an example, such variables as Albumin, Total_Proteins, and the Albumin_and_Globulin Ratio have significant differences in dispersion and central tendency between the two groups, suggesting that these measures could have class-specific behaviour. Such visualisations are therefore indispensable in identifying the distributional differences and in identifying potential biomarkers that are relevant in classification undertakings.

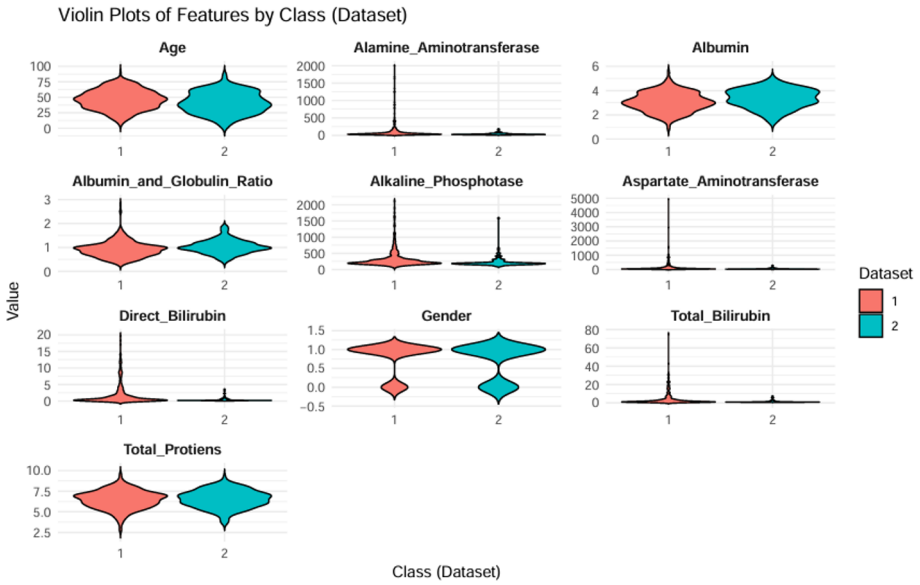

Fig 9. Violin Plot Comparison of Medical Features by Dataset Class

3.3 Data Preprocessing

Regarding our study, the data preprocessing pipeline, as shown in Figure 10, was carefully designed to ensure the reliability, consistency, and analytical appropriateness of the clinical dataset before the predictive model was developed.The raw dataset contained 584 patient records in the first place, and each record contained 11 features that were clinically salient. Systematic operations were applied to cleanse the dataset thoroughly and make it fit the further machine-learning processes.In the first preprocessing stage, the label encoding method was used to encode the categorical variables into numerical values that could be used in the later learning algorithms. This process maintains the inherent semantic connotations of categorical attributes and thus allows them to be dealt with computationally with relative ease.In order to deal with missing data, a strong approach was employed i.e., median imputation. The method reduces the effect of outliers and maintains the central tendency of each feature, which increases the accuracy of the data. Median imputation is especially beneficial to work with clinical data, in which gaps in data are often the result of differences in diagnostic practices or laboratory reporting gaps. In these settings, the approach provides the strength against occasional data lapses without affecting the validity of the analytical results.After the first preprocessing, all the numerical variables were scaled to standardize the ranges. This was necessary to avoid the influence of features with higher numerical values on the learning process, which is especially relevant to distance-based and gradient-based algorithms. Standardization also promotes better convergence and stability during the training process. This was followed by feature selection which was done in order to isolate the most informative predictors and hence mini-

mize model complexity. The judiciousness of the selection of features is used to improve the interpretability and computational efficiency of predictive models, mainly by the systematic removal of extraneous or redundant variables. The polished feature set was then split into training and testing subsets with an 80:20 split, so that the evaluation of the models would be performed on previously unknown data. In the training sub-component, the issue of class imbalance was resolved through SMOTE, and different machine-learning algorithms were implemented. Thereafter, the models were tested on an untouched test set to determine that the performance metrics were obtained in an objective and unbiased way. This hierarchical preprocessing process forms a strong base of correct prediction and generalizable behaviour of the model.

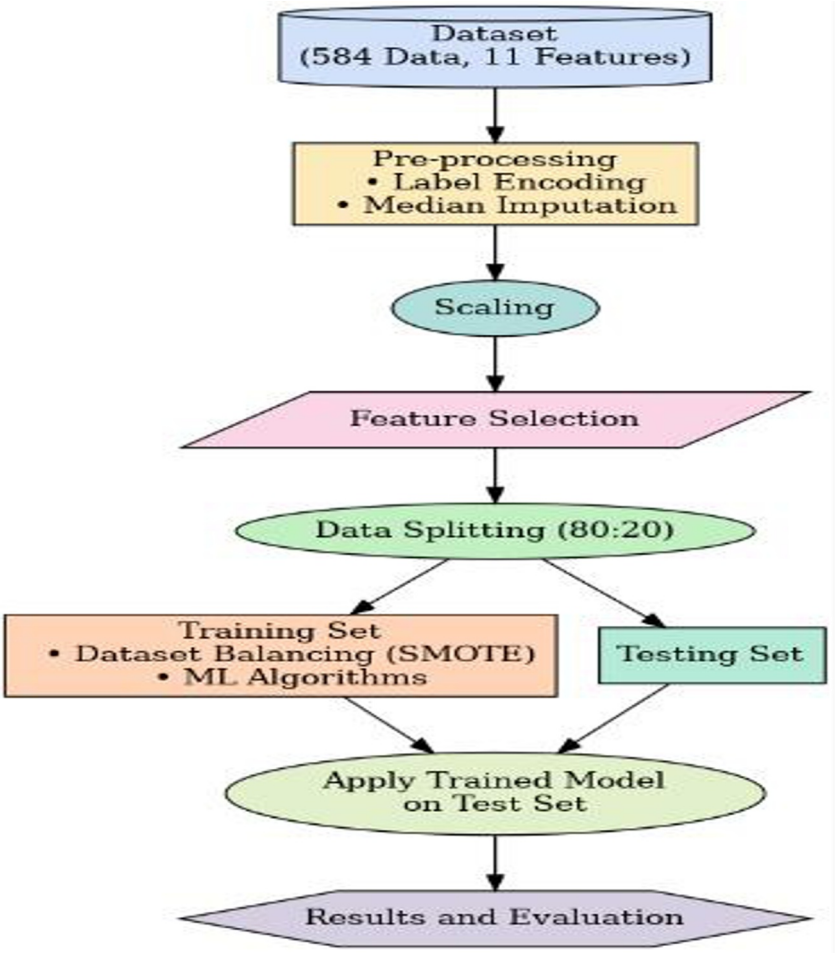

Fig 10. Flowchart of Proposed Algorithm

4 Machine Learning Technique

Logistic Regression (LR)

One of the simplest binary classification algorithms, which is famous due to its interpretability. The logistic regression makes approximations of the probability of classes by estimates to a logistic regression. It works very well when the underlying data is linearly separable, but is still vulnerable to multicollinearity amongst predictors.

Support Vector Machine (SVM)

An adaptive control learning model which finds the best hyperplane to divide classes. It is applicable in linear and non-linear classification, using radial basis function (RBF) as its kernel functions. The SVM is especially appropriate when dealing with complex data that are small to medium.

Random Forest (RF)

Random Forest is a collection of decision trees which are collectively built by bootstrapping aggregation (bagging).It averages predictions to enhance accuracy and reduce over-fitting besides giving information on the importance of features.

Extreme Gradiant Boosting (XGBoost)

A gradient-boosting model which builds trees one by one to rectify the errors of the previous models. It allows the use of regularisation words, missing values are introduced uniformly, and the algorithm is acclaimed as being fast and highly accurate.

Gradient Boosting Machine (GBM)

Gradient-Boosted Machines (GBM) are an effective ensemble algorithm, which builds one tree at a time with the purpose of reducing the residual errors of the previous tree. It is very useful in regression and classification by using gradient descent to minimise a loss function. However, GBM is computationally expensive and, unless carefully tuned, it is highly likely to over-fit.

Light GBM (LGBM)

An efficient, fast gradient-boosting algorithm based on histogram-based learning and leaf-wise tree-growth. It can be scaled to large datasets and it can be accelerated with the use of a GPU.

Feature Importance Identified by LASSO

The result of the feature selection using the LASSO regression model is a brief but informative depiction of the predictors that have the most significant contribution to the classification of liver cirrhosis. LASSO can be used to effectively remove redundant variables and only keep variables that have significant discriminatory power, thus improving the interpretability and generalization. Direct bilirubin as shown in Fig. 11 has the highest negative coefficient meaning that it has a strong negative relationship with the predicted class. This observation can be clinically attributed, as an increased bilirubin level is a proven indicator of defective hepatic clearance and progressive liver failure.

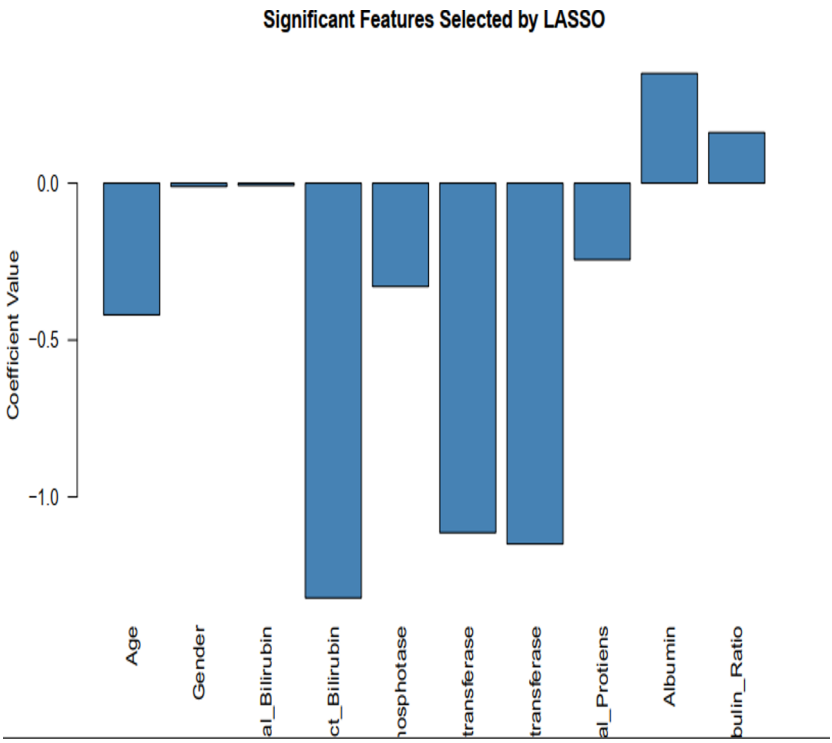

Fig.11 Predictive Features Identified via LASSO Regression

Similarly, alkaline phosphatase and the transaminase enzymes, which are also characterized by large negative coefficients, also become significant biochemical indicators. Their high levels are usually indicative of hepatocellular damage or cholestatic blockage, which further supports their usefulness in determining people at

increased risk of cirrhosis. Conversely, the coefficients of albumin and total protein are positive implying the protective effect. Higher levels of albumin normally indicate intact hepatic synthetic function and good nutritional condition- both of which decline significantly as the disease progresses to cirrhosis. These variables have positive contributions hence emphasizing their relevance in the differentiation of compensated and decompensated liver states. On the other hand, other demographic factors like age and gender are assigned relatively smaller magnitudes of coefficients. Their low contribution implies that, in this data, biochemical indicators are significantly better predictors than demographic characteristics. This result demonstrates the ability of LASSO to focus on physiologically meaningful predictors and inhibit predictors with low predictive value. In general, the feature set derived by LASSO highlights the superiority of the core hepatic biochemical parameters in the initial diagnosis of liver cirrhosis. The model increases the level of interpretability, and it is also consistent with the available clinical evidence, which further builds up the level of confidence in the applicability of the model to the real-life diagnostic setting.

Class Imbalance Handling

Unbalanced classes are a major problem in clinical prediction applications because learning algorithms are biased towards the majority class leading to poor performance in the detection of minority-class examples that are frequently clinically important conditions. The analysis of the original dataset in the current study showed that Class 0 had a significant imbalance, as it had about 71.6 percent of samples to Class 1 which had only 28.4 percent (Fig. 12). This type of disproportionate representation may negatively impact the model training by biasing the decision boundary in favor of the majority group, thereby decreasing sensitivity and increasing the chances of misclassification to patients in the minority group. In order to reduce this problem, the Synthetic Minority Oversampling Technique (SMOTE) was used. Synthetic Minority Over-Sampling Technique (SMOTE) gives the minority class artificial examples, created by interpolating between real minority examples, which is a richer sampling technique that avoids repetitive sampling of features. After SMOTE was applied, the number of observations in both classes was made equal (414 observations each), and a 50:50 distribution of the classes was obtained. This modification will ensure that the learning algorithms are equally represented by both outcome classes, and they will be able to better describe the underlying patterns that are related to the minority-class cases. It is especially important to balance the dataset in the clinical setting, where a false classification of the minority group, which can be interpreted as disease presence in most cases, can have serious diagnostic consequences. SMOTE does not only increase the fairness of model training, but also increases key performance measures, including sensitivity, specificity, and F1-score. The study will be able to correct the class imbalance, making the predictive models work more effectively in both classes, which will eventually make the predictive models more robust in terms of their clinical applicability and reliability.

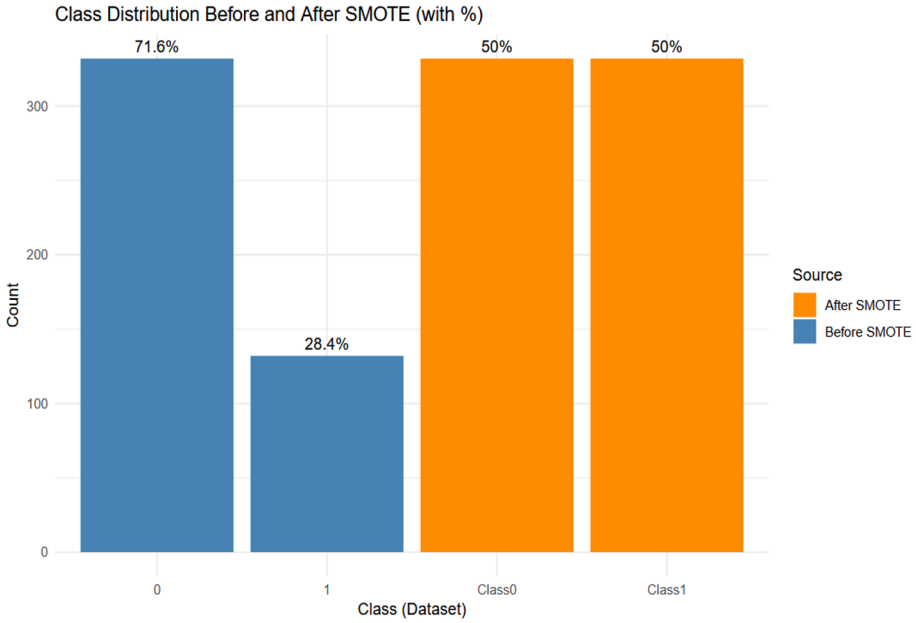

Fig.12. Class Distribution Comparison Before and After SMOTE for Liver Disease Classification

5 Model evaluation criteria

5.1 Confusion matrix

In order to measure the performance of the classification model based on liver disease prediction, we computed the canonical measures of precision, recall, accuracy, and specificity, which are based on the confusion matrix including true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN).

The confusion matrix can be employed to assess the accuracy of the proposed model. It will utilize just 1 and 0 values to characterize the real and anticipated values in matrix form.

1. If the model predicted a value of 1 and its actual value was 1, the end result would be a true positive.

2. If the model predicted a value of 1 and its actual value was 0, the end result would be a false positive.

3. If the model predicted a value of 0 and its actual value was 1, the end result would be a false negative.

4. If the model predicted a value of 1 and its actual value was also 1, the end result would be a true positive.

Precision measures the proportion of correctly predicted positive cases out of all cases predicted as positive. It is especially useful when the cost of false positives is high. Precision is defined as:

$$\mathrm{P}_{\omega} = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \tag{1}$$

Also known as the True Positive Rate, recall calculates the proportion of actual positive cases that were correctly identified by the model. It evaluates how well the model captures the true positive instances. The formula for recall is:

$$\mathrm{R}_{\omega} = \frac{\mathrm{TP}}{\mathrm{TP + FN}} \tag{2}$$

Accuracy represents the overall effectiveness of the model by measuring the ratio of correctly predicted observations (both positive and negative) to the total number of observations. It is calculated as:

$$A_{\omega} = \frac{TP + TN}{TP + FP + TN + FN} \tag{3}$$

Specificity evaluates the model's ability to correctly identify negative cases, also known as the True Negative Rate. It is useful when the cost of false positives is high and is given by:

$$S_\omega = \frac{TN}{TN+FP} \tag{4}$$

To compare the performance of different classification models, the F1-score is utilized as a balanced evaluation metric. The F1-score combines both Precision and Recall into a single value by calculating their harmonic mean. This approach ensures that both false positives and false negatives are taken into account, making the F-score especially useful in scenarios involving imbalanced datasets.

The F1-score is defined as:

$$F1 - score = \frac{2*P_\omega*R_\omega}{P_\omega+R_\omega} \tag{5}$$

Where P_ω is precision and R_ω is recall.

The F1-score ranges between 0 and 1, with values closer to 1 indicating a more accurate and well performing model. This metric is particularly beneficial when a trade-off between precision and recall must be considered, such as in medical diagnosis tasks like liver disease prediction.

6 Result and discussion

The models analyzed in the framework of liver cirrhosis prediction are Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), XGBoost, Gradient Boosting Machine (GBM), and Light GBM (LGBM). They were evaluated based on standard metrics: accuracy, sensitivity (recall), specificity, precision, and the F1 -score, which are presented in Table 4.Two of the traditional algorithms, SVM and LR, obtained moderate sensitivity levels (82.93% and 80.49%, respectively), which means that they are capable of detecting the instances of cirrhosis positive cases with a reasonable level of efficiency. However, their specificity values were relatively low (52.44% in the case of SVM and 56.10% in the case of LR), which means that the number of false positives was relatively high. Furthermore, the F1-scores of 71.75% (SVM) and 71.66% (LR) indicate a limited trade-off between precision and recall, which is especially problematic in clinical settings where both values are of utmost importance.

The traditional models were far inferior to the ensemble models. The overall performance of XGBoost was the best with the accuracy of 82.32%, the sensitivity of 91.46%, and the specificity of 73.17%, and the highest F1-score was 83.78%. It means that XGBoost is especially suitable to handle the liver cirrhosis classification

tasks, and it is highly reliable in not only identifying diseased patients but also preventing misclassification of healthy people. Light GBM and GBM also reached impressive and stable results with F1 -scores of 82.05% and 80.51%, respectively. Their accuracies of 80.49% (LGBM) and 78.66% (GBM) and sensitivities of 89.02% and 87.80% indicate their effectiveness in identifying cases of cirrhosis as well as maintaining reasonable specificity. Random Forest, though slightly worse than the boosting methods, still outperformed the baseline models with an F1 -score of 78.59%.

In general, the experimental data show that ensemble-based algorithms, and especially XGBoost, exhibit significantly high ability to cope with the challenges that liver cirrhosis classification presents. They are more precise, have better recall and higher F1-scores, which makes them highly applicable in clinical decision support, where the reduction of false positives and false negatives is vital to reliable diagnosis and prompt therapeutic intervention. The confusion matrix of the XGBoost model used to predict the stages of liver cirrhosis is shown in Figure 13. The model made the right classification of 60 class 0 (true negatives) and 75 class 1 (true positives). It falsely classified 7 cases of class 0 as class 1 (false positives) and 22 cases of class 1 as class 0 (false negatives).The confusion matrix highlights the excellent performance of XGBoost, and in particular, with regard to the rightful identification of positive cases, a facet of critical clinical significance.

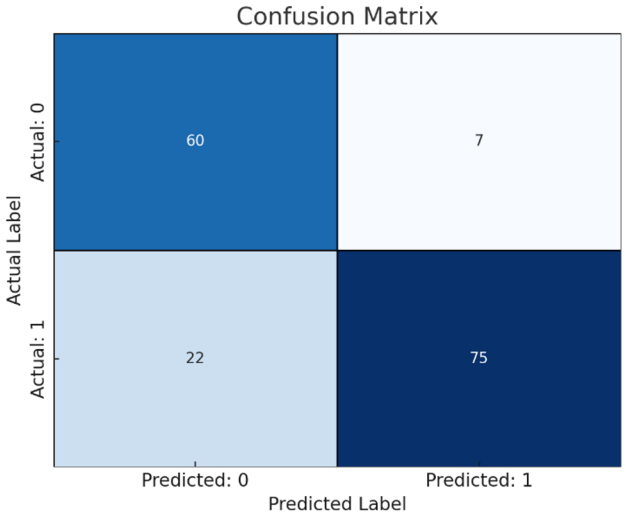

Fig.13. Classification performance illustrated by confusion matrix (XGBoost)

Figure 14 shows a simulated comparison of Receiver Operating Characteristic (ROC) curves of a few machine-learning models, namely, Logistic Regression, Support Vector Machine, Random Forest, Gradient Boosting Machine, Light GBM, and XGBoost. The plot shows the true positive rate and false positive rate to evaluate the performance of the classifier by using the accuracy and the estimat-

ed AUC values. This diagram shows the discriminative power of each model in a binary classification system.

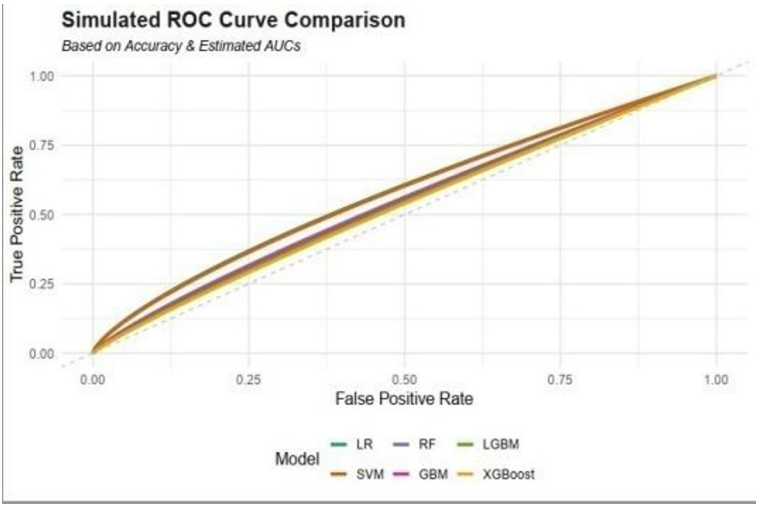

Fig.14. Performance comparison of machine learning models for liver cirrhosis stage prediction

When comparing the findings of Tables 4 and 5, which are dedicated to using machine-learning classifiers to predict liver diseases, they are based on different methodological and performance indicators. Table 4 provides a list of different studies and their highest classifier performance and the percentage of accuracy. An example is a decision tree with an accuracy of 81% was selected as the best classifier in the study by Allenki and Soni (2024) and 88.10% was the accuracy of a random forest model in Amin et al. (2023).

In the more detailed analysis of the work by Anil Utku (2023), the table V offers a variety of classifiers, including Logistic Regression, Support Vector Machine, Random Forest, and others, with the explicit measures of the accuracy, sensitivity, specificity, precision, and F1-score. The random forest classifier in this dataset had an accuracy of 77.44% with a sensitivity of 82.93% and specificity of 71.95%, and XGBoost outperformed it with an accuracy of 82.32% and maximum F1 -score of 83.78%. Overall, both tables demonstrate that different machine learning algorithms are effective in predicting liver diseases and random forest and XGBoost can be considered the best performing ones in the research. Nevertheless, Table 5 provides a

more detailed assessment, and it includes several measures other than accuracy, thus, allowing a more subtle comparison of model performance.

Table 4. Performance Comparison of Machine Learning Classifiers recent related work

Author & Reference	Title	Year	Classifiers	Best result by	Accuracy
J. Allenki, H.K. Soni [1]	Analysis of chronic liver disease detection using machine learning techniques	2024	DT, SVM, RF, AI, NB	DT	81.00 %
R.Amin, R. Yasmin, S. Ruhi, M. H. Rahman, M. S. Reza [11]	Prediction of chronic liver disease patients using integrated projection-based statistical feature extraction with machine learning algorithms	2023	LR,RF,KNN,SVM, MLP, EVC	RF	88.10 %
JulianA. Luetkens, et.al., [18]	Deep Learning Supports the Differentiation of Alcoholic and Other-than-Alcoholic Cirrhosis Based on MRI	2022	ResNet50, DenseNet121	ResNet50	75.00 %
VICTOR ANTHONYSAMY et. al. [19]	Multi Perceptron Neural Network and Voting Classifier for Liver Disease Dataset	2023	SVM,KNN, HVC, Deep NN	HVC	78.62 %
Anil Utku [20]	Deep Learning Based Cirrhosis Detection	2023	MLP, DT, KNN, LR, NB, RF, SVM	MLP	80.48 %
Proposed work			LR, SVM, RF, XGBoost, GBM, LGBM	XGBoost	82.32 %

Table 5. Performance comparison of machine learning classifiers for liver cirrhosis prediction

Classifier	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1-Score (%)
LR	68.29	80.49	56.10	64.71	71.66
SVM	67.68	82.93	52.44	63.55	71.75
RF	77.44	82.93	71.95	74.73	78.59
XGBoost	82.32	91.46	73.17	77.32	83.78
GBM	78.66	87.80	69.51	74.23	80.51
LGBM	80.49	89.02	71.95	76.04	82.05

A review of previous studies on the chronic liver disease prognostication shows that there is a significant heterogeneity in the methodological framework, feature extraction paradigm, and subsequent predictive performance. In earlier supervised learning projects, e.g., Allenki and Soni used canonical classifiers like decision trees, support vectors and random forests, achieving a peak accuracy of eighty one percent. On the other hand, Amin and others gave a larger number (88.10) by combining random forests with statistical feature extraction, highlighting the benefit of enriched feature representations. Later research that followed the approach of deep learning, such as that of Luetkens et al. and Utku, used architectures such as ResNet -50 and multilayer perceptrons to obtain receiveroperator characteristics and accuracies of 75% and 80.48%, respectively, but their performance depended on large imaging datasets and complex model design. The hybrid or ensemble strategies, such as the voting classifier suggested by Anthonysamy et al., achieved relatively small gains, up to 78.62% accuracy. Conversely, the current study uses a broader range of modern classifiers, and gradient boosting algorithms like XGBoost, which scored the highest of 82.32% herein, were used. These results indicate that, without the use of complex deep learning models or complex hybrid models, carefully optimized boosting models can provide strong and competitive results on tabular clinical data, thus making the proposed approach an effective and interpretable alternative to the existing ones.

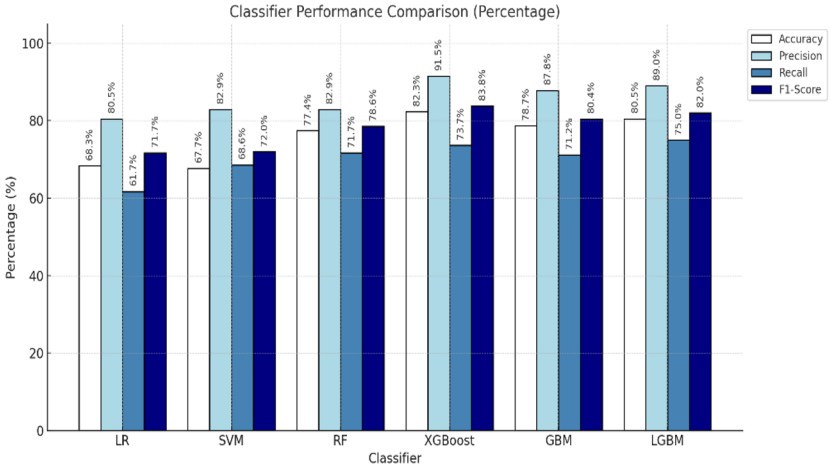

Fig. 15. Performance comparison of ML classifiers based on evaluation metrics

Figure 15 provides a comparative study of six machinelearning classifiers, including Logistic Regression, Support Vector Machine, Random Forest, XGBoost, Gradient Boosting Machine, and LightGBM using four major evaluation metrics: accuracy, precision, recall (sensitivity), and F1-score, all in percentages. XGBoost has the best performance in most of the metrics with an accuracy of 82.3% and a recall of 91.5% which means that it can accurately detect positive cases. LightGBM and GBM are the

next in line, which also show balanced and strong performance. On the contrary, LR and SVM have lower total scores, especially precision and accuracy. This value demonstrates the better predictive power of ensemble-based models, particularly, XGBoost, in the prediction of liver cirrhosis stage.

Clinical Interpretability

The present manuscript highlights the need to integrate clinically relevant variables into machinelearning systems into the healthcare system, specifically in liver disease management. The contributions of featureimportance studies through LASSO and tree-based ensembles were considered as the key to making the interpretability more powerful.The most notable predictors of the severity of cirrhosis were the level of bilirubin, transaminase enzymes, alkaline phosphatase, albumin, and total proteins, which are in line with the previous clinical understanding of hepatic function. The agreement between the model results and the established biomarkers provides clinicians with clear and logical descriptions of predictions, which enhances confidence in the diagnostic accuracy. Moreover, the quantification of the individual feature contributions facilitates clinical decision making by allowing practitioners to focus on important biomarkers when tracking diseases. Finally, the researchers highlight the future potential of machine learning models as predictive models, as well as decision support systems that can close the gap between data driven knowledge and clinical care, contributing to the development of liver disease management.

7 Conclusion

This study provides a comprehensive evaluation of a range of machine learning classifiers that are supposed to forecast the degree of liver cirrhosis, based on such classic measures of performance as accuracy, sensitivity, specificity, precision, and the F1-score. The results prove the effectiveness of XGBoost with the best accuracy (82.32%) and F1 -score (83.78%), then Random Forest and Gradient Boosting Machines. The originality of the work lies in the comparative analysis of conventional and ensemble-based algorithms using clinically relevant measures, thus not only measuring prognostic fidelity but also providing information about the dynamics of the classifier, such as sensitivity and specificity, which are the most important in clinical decision support.

The significance of this study is based on the ability to optimize the early diagnosis and continuous monitoring of liver cirrhosis, which is a condition of paramount social health interest. The study establishes a concrete connection between machine learning procedures and clinical hepatology by defining the most significant biochemical indicators, i.e., transaminases, bilirubin levels, and albumin levels, as the priority biomarker model of patient evaluation. Accordingly, the work can contribute to the expanding body of literature on applied machine learning in healthcare, specifically in the area of managing liver diseases as it provides a clear and repeatable assessment system of clinical applicability.

Despite the promising outcomes, there are a number of methodological limitations that should be mentioned. First of all, the research failed to apply hyperparameter tuning, which might have a positive impact on the performance of the classifier. In addition, the training data used in the training of the models was based on a particular cohort, which can limit the extrapolation of the models to wider populations. Lastly, the analysis did not consider potentially informative modalities like imaging and genomic data, which might provide a deeper understanding of the progression of cirrhosis, but only considered biochemical markers.

These shortcomings should be tackled in future studies through the combination of hyperparameter optimization methods like grid search or Bayesian optimisation to further improve the predictive accuracy. It would be reasonable to increase the dataset to cover a variety of populations and use multimodal data, such as biochemical, imaging, and genetic data, to increase the model robustness and generalizability. In its turn, such improvements would increase the clinical utility of machine learning models in the real-world liver disease diagnostics and treatment planning, which would eventually lead to the improved patient outcomes.

Overall, the current research illustrates that it is possible to implement machinelearning classifiers, particularly, XGBoost and RF, to identify the severity of liver cirrhosis. This study provides the foundation of future studies that will enhance the accuracy and usefulness of predictive models in the treatment of hepatic diseases by providing clinically interpretable results.

Author contribution statement

The contributions of the authors are as follows: Mohammed Motaher Hossain conducted formal analysis, conceptualization, methodology, writing of the original manuscript and editing of the manuscript. Rehana Parvin was involved in data visualization, drafting of the initial manuscript, and methodology. Mst. Rashida Pervin was instrumental in reviewing, editing, and developing concepts of manuscripts. Raffat Arman Islam provided support in reviewing the manuscript and editing it.

Acknowledgement

We want to thank our great mentor Mohammed Motaher Hossain of the Department of Quantitative Sciences (Statistics) whose help an**d guidance were of invaluable assistance during the course of this research. His support** and sharp criticism played a key role in the success of this research.

References

1. J. Allenki and H. K. Soni, "Analysis of chronic liver disease detection by using machine learning techniques," in 2024 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS). IEEE, 2024, pp. 1–8.
2. I. Hanif and M. M. Khan, "Liver cirrhosis prediction using machine learning approaches," in 2022 IEEE 13th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON). IEEE, 2022, pp. 0028–0034.
3. R. Manjunath, A. Ghanshala, and K. Kwadiki, "Deep learning algorithm performance evaluation in detection and classification of liver disease using ct images," Multimedia Tools and Applications, vol. 83, no. 1, pp. 2773–2790, 2024.
4. K. Pal, S. Panwar, and D. Choudhury, "A pragmatic approach of heart and liver disease prediction using machine learning classifiers," in 2024 International Conference on Emerging Systems and Intelligent Computing (ESIC). IEEE, 2024, pp. 728–734.
5. S. K. Asrani, H. Devarbhavi, J. Eaton, and P. S. Kamath, "Burden of liver diseases in the world," Journal of hepatology, vol. 70, no. 1, pp. 151–171, 2019.
6. J. M. Pezzuto, M. E. Johnson, and H. R. Manasse, Biotechnology and pharmacy. Springer Science & Business Media, 2013.
7. R. Wei, J. Wang, X. Wang, G. Xie, Y. Wang, H. Zhang, C.-Y. Peng, C. Rajani, S. Kwee, P. Liu et al., "Clinical prediction of hbv and hcv related hepatic fibrosis using machine learning," EBioMedicine, vol. 35, pp. 124–132, 2018.
8. K. Saboo, N. V. Petrakov, A. Shamsaddini, A. Fagan, E. A. Gavis,M.Sikaroodi, S. McGeorge, P. M. Gillevet, R. K. Iyer, and J. S. Bajaj, "Stool microbiota are superior to saliva in distinguishing cirrhosis and hepatic encephalopathy using machine learning," Journal of hepatology,vol. 76, no. 3, pp. 600–607, 2022.
9. A. M. Musolf, E. R. Holzinger, J. D. Malley, and J. E. Bailey-Wilson, "What makes a good prediction? feature importance and beginning to open the black box of machine learning in genetics," Human Genetics, vol. 141, no. 9, pp. 1515–1528, 2022.
10. M. Elsheikh, A. El Sabagh, I. B. Mohamed, M. Bhongade, M. M. Hassan, and P. K. Jalal, "Frailty in end-stage liver disease: Understanding pathophysiology, tools for assessment, and strategies for management," World journal of gastroenterology, vol. 29, no. 46, p. 6028, 2023.
11. R. Amin, R. Yasmin, S. Ruhi, M. H. Rahman, and M. S. Reza, "Prediction of chronic liver disease patients using integrated projection based statistical feature extraction with machine learning algorithms," Informatics in Medicine Unlocked, vol. 36, p. 101155, 2023.
12. A. Spann, A. Yasodhara, J. Kang, K. Watt, B. Wang, A. Goldenberg, and M. Bhat, "Applying machine learning in liver disease and trans plantation: a comprehensive review," Hepatology, vol. 71, no. 3, pp. 1093–1105, 2020.
13. V. J. Gogi and M. Vijayalakshmi, "Prognosis of liver disease: Using ma chine learning algorithms," in 2018 International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE). IEEE, 2018, pp. 875–879
14. M. B. Priya, P. L. Juliet, and P. Tamilselvi, "Performance analysis of liver disease prediction using machine learning algorithms," Int. Res. J. Eng. Technol, vol. 5, no. 1, pp. 206–211, 2018.
15. A. S. Rahman, F. Shamrat, Z. Tasnim, J. Roy, and S. A. Hossain, "A comparative study on liver disease prediction using supervised machine learning algorithms," International Journal of Scientific & Technology Research, vol. 8, no. 11, pp. 419–422, 2019.
16. [16] Y.-Y. Chen, C.-Y. Lin, H.-H. Yen, P.-Y. Su, Y.-H. Zeng, S.-P. Huang, and I.-L. Liu, "Machine-learning algorithm for predicting fatty liver disease in a taiwanese population," Journal of Personalized Medicine, vol. 12, no. 7, p. 1026, 2022.

17. K. Gupta, N. Jiwani, N. Afreen, and D. Divyarani, “Liver disease prediction using machine learning classification techniques,” in 2022 IEEE 11th International conference on communication systems and network technologies (CSNT). IEEE, 2022, pp. 221–226.
18. JulianA. Luetkens, et.al.,“Deep Learning Supports the Differentiation of Alcoholic and Other-than-Alcoholic Cirrhosis Based on MRI”,https://doi.org/10.1038/s41598-022-12410-2.
19. Victor Anthony Samy et. al., “Multi Perceptron Neural Network and Voting Classifier for Liver Disease Dataset
https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10254220.
20. Anil Utku, “Deep Learning Basesd Cirrhosis Detection”, https://oresta.org/menuscript/index.php/oresta/article/view/347/216.