IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 28 May 2024, accepted 29 June 2024, date of publication 3 July 2024, date of current version 12 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3422451

RESEARCH ARTICLE

LivMarX: An Optimized Low-Cost Predictive Model Using Biomarkers for Interpretable Liver Cirrhosis Stage Classification

SUDIKSHA KOTTACHERY KAMATH, SANJEEV KUSHAL PENDEKANTI, AND DIVYA RAO
Department of Information and Communication Technology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, India
Corresponding author: Divya Rao (divya.r@manipal.edu)

ABSTRACT Liver cirrhosis, a progressive and irreversible condition characterized by the replacement of healthy liver tissue with scar tissue, presents a persistent challenge in healthcare. Leveraging a comprehensive dataset of 424 patients, including controlled trial data from 312 patients and real-world follow-up information from an additional 112 patients sourced from the Mayo Clinic trial on primary biliary cirrhosis, this study refines patient demographics and integrates biomarkers for a detailed analysis. Our paper introduces LivMarX, a novel model that utilises advanced machine learning algorithms within an interdisciplinary framework and strives to stage Liver Cirrhosis using biomarkers instead of images. Our study employs meticulous feature engineering techniques, the creation of synthetic variables and categorizations to unveil critical relationships between patient characteristics and disease stages. To further refine performance, hyperparameter optimization is implemented, combining Genetic Algorithm, Optuna, and GridSearchCV. The Random Forest Classifier in LivMarX outperformed other models with an accuracy of 84.33%, and post-optimization, an accuracy improvement of 86%. LivMarX demonstrates an AUC of 0.95, showing that it provides reliable staging of liver cirrhosis. By relying on common blood tests instead of expensive imaging, this offers a cost-effective and comfortable approach to Liver Cirrhosis diagnosis. This study positions LivMarX as a potential model for accurately classifying Liver Cirrhosis stages, particularly in areas with limited access to complex imaging equipment.

INDEX TERMS Liver cirrhosis, machine learning, health determinants, healthcare, disease prediction.

I. INTRODUCTION

Liver cirrhosis represents a progressive and chronic pathology characterized by fibrosis and the formation of scar tissue, leading to the deterioration of liver function [1]. Various factors, including excessive alcohol consumption, chronic viral infections, and fatty liver disease, result in scarring of the liver. This leads to the liver struggling to perform its vital tasks, such as processing nutrients and filtering toxins from the blood. Consequently, symptoms like fatigue, jaundice, and complications like liver failure may arise [2]. It is influenced by various factors, including age, gender, and clinical manifestations such as ascites, hepatomegaly, spider veins, and edema. The risk of cirrhosis increases with age, with older individuals more susceptible to liver damage. Men are generally at a higher risk than women, potentially due to hormonal differences. Observable signs like ascites and hepatomegaly are indicative of advanced liver disease [3]. According to findings from the Global Burden of Disease Study in 2017, the worldwide prevalence of compensated cirrhosis, a stage of liver disease where the liver is damaged but still partially functional, was estimated at 112 million individuals [4]. This corresponds to an age-standardized frequency rate of 1,395 cases per 100,000 people globally, highlighting the significant global impact of cirrhosis and chronic liver diseases [4]. Biochemical markers play a crucial role in assessing liver function. Elevated serum bilirubin levels indicate impaired bile metabolism, while reduced

The associate editor coordinating the review of this manuscript and approving it for publication was Wojciech Sałabun.

albumin levels reflect compromised synthetic liver function [5]. Cholesterol abnormalities, copper accumulation, prolonged prothrombin time, and increased alkaline phosphatase levels are also associated with liver dysfunction.

One of the central aspects of the research is the histologic stage of Liver Cirrhosis, which categorizes the disease into four distinct stages based on tissue examination [6]. Liver histologies provide a wealth of data that includes extensive histologic staging information. This abundance of data facilitates the development of predictive models that can significantly enhance the recognition and management of Liver Cirrhosis. Imaging methods are now widely utilized in the diagnosis and treatment of Liver Cirrhosis. These include ultrasound, elastography computed tomography (CT) and magnetic resonance imaging (MRI), which are pivotal in assessing liver damage and guiding therapeutic decisions. They allow visualization of structural changes in the liver, detection of nodules, and evaluation of fibrosis severity. These methods are also essential in determining the disease's histologic stage, aiding in the characterization of liver tissue and identification of specific features indicative of cirrhosis [7]. The diagnostic process typically begins with non-invasive imaging, such as ultrasound, to identify liver abnormalities. More advanced techniques like CT or MRI may be employed for detailed insights into the liver's architecture and cirrhotic changes. Elastography, measuring liver stiffness, provides a quantitative assessment of fibrosis severity [8].

The current limitations in the diagnosis and treatment of Liver Cirrhosis have prompted the exploration of innovative solutions. Systems incorporating data mining and machine learning (ML) algorithms aim to address these limitations, providing precise and timely diagnoses. These advancements empower both doctors and patients, enabling informed decisions about treatment options [9]. The rapid advancement of artificial intelligence (AI) and big data analytics has enabled the harnessing of ML potential to extract valuable insights from extensive medical datasets. This integration of technology facilitates early disease diagnosis, enhancing the prognosis for patients with liver diseases [10]. This integration of technology promises to enhance the overall quality of care in the field of Liver Cirrhosis diagnosis and treatment. Timely diagnosis, made possible by AI and ML, allows for the implementation of lifestyle modifications and medical interventions. This, in turn, reduces the risk of complications such as variceal bleeding and hepatic encephalopathy [11]. It also facilitates the selection of appropriate treatment modalities which mitigates the need for advanced and costly interventions like liver transplantation [12]. Early detection equips doctors with the tools to intervene quickly, improving patient outcomes and reducing the overall burden of liver-related illness and mortality.

The aim of this research work is to improve patient outcomes and healthcare decision-making in the context of liver disease. The ultimate goal is to contribute to early detection and accurate staging of Liver Cirrhosis, a development that has the potential to significantly benefit a large number of individuals.

The following are the contributions made by this paper:

1) Obtain and refine a suitable dataset for the analysis.
2) Introduction of LivMarX, an interpretable and accurate predictive model for Liver Cirrhosis progression prediction.
3) Enhance performance using a combination of machine learning techniques to improve healthcare decision-making in limited resource settings.

This paper begins with a review of the literature that provides a comprehensive overview of related existing studies, followed by the methodology includes the dataset description, pre-processing, and feature extraction steps. The methodology outlines the application of various machine learning algorithms for predicting Liver Cirrhosis stages. The paper then presents results, with visualizations of the features and outputs. It concludes by summarising the key findings of the work and suggesting avenues for future research.

II. RELATED WORKS

This section provides a review of recent studies that have tackled Liver Cirrhosis classification and prediction (refer Table 1). In the table, "Acc" refers to "Accuracy", "Sen" refers to "Sensitivity" and "Spec" refers to "Specificity". In the last study, "Val" and "Data Test" refer to the AUC values obtained. Studies 1-5 in Table 1 presents studies leveraging imaging modalities such as MRI and ultrasound scans for predicting the stage of Liver Cirrhosis disease. A dominant trend is the use of deep learning algorithms CNNs appearing in 3 out of 5 studies. While the accuracy achieved is promising, ranging from 94% to 99%, most recent studies have not made their data available publicly, hindering reproducibility and generalizability. Only one study reports metrics beyond accuracy [13], making it difficult to fully compare the performance of different methods. Studies using publicly available data achieved higher accuracy [14], [15]. Studies 6-8 in Table 1 focus on diagnosing Liver Cirrhosis using non-imaging approaches, specifically through biochemical markers and body symptoms. Compared to imaging-based methods, these studies generally achieve lower accuracy. However, they offer a more cost-effective and accessible alternative, especially in resource-limited settings. One study in this table utilizes a publicly available dataset from the Mayo Clinic, potentially enhancing generalizability.

The review of literature reveals the ongoing research efforts in Liver Cirrhosis diagnosis using diverse approaches. While imaging modalities achieve higher accuracy, their reliance on specialized equipment and non-public datasets limits their accessibility. Biochemical markers and body symptoms offer a promising alternative but require further research to improve their accuracy and identify optimal combinations of markers for reliable diagnosis. This observation highlights the importance of data size and accessibility for generalizable models. A key gap in the review of literature is achieving high accuracy in Liver Cirrhosis diagnosis using non-imaging

TABLE 1. Related literature on liver Cirrhosis using a variety of datasets.

Sl. No.	Author	Year	Aim	Methodology	Input Data	Dataset Size	Availability	Metrics
1	Xu et al. [14]	2017	Fully automatic diagnosis method for Liver Cirrhosis	Deep Neural Network, Partition Clustering	Ultrasound Images	122	On Request	Acc: 96.06%
2	Ribeiro et al. [16]	2013	Utilizing Multimodal Data for the Categorization and Staging of Chronic Liver Disease	Clinical Based Classifier, KNN, Bayes, Parzen, SVM	Liver Images	100	On Request	Acc: 98.67%
3	Prakash et al. [13]	2022	Novel Deep Learning Method for Categorization and Forecasting Cirrhosis Liver	Deep Neural Network, Spearman's Rank Correlation	MRI Images	418	On Request	Acc: 99%, Sen: 94%, Spec: 94.2%
4	Zhang et al. [17]	2021	Classify liver fibrosis stage of chronic liver disease using ultrasound pictures utilizing machine learning-based approaches	Convolutional Neural Network, Radiomics model	Ultrasound Images	394	On Request	Acc: 94%
5	Randhawa et al. [15]	2018	Utilizing the Weighted Fisher Discriminant Ratio Algorithm to Predict Liver Cirrhosis	Weighted Fischer Discriminant Ratio Algorithm, FDR Classifier	Ultrasound Images	45	Public	Acc: 96.28%, Sen: 98.33%
6	Wang et al. [18]	2008	Correlation between Traditional Chinese Medicine (TCM) symptoms and the stages, classification, and Child-Pugh grade of Liver Cirrhosis	Information Gain, Logistic, BayesNet, NaiveBayes, RBF, C4.5	Symptoms	294	On Request	Acc: 81%
7	Goyal et al. [19]	2022	Primary biliary cirrhosis of the liver experiment by Mayo Clinic	XGBoost, Tree Structure, ML Algorithms	Biochemical Markers	424	Public	Acc: 78%
8	Bostan et al. [20]	2015	Non-invasive technique for Liver Cirrhosis diagnosis	Artificial Neural Network (ANN) using MATLAB	Biochemical Markers	390	On Request	Val: 0.89, Data Test: 0.73

datasets and publicly available data. Our research builds upon the gap in the literature by utilizing a large, publicly available dataset and exploring alternative methods to address potential limitations of purely deep learning approaches.

III. METHODOLOGY

The methodology section of this paper provides a thorough explanation of the steps taken in the study to address the staging of Liver Cirrhosis as illustrated in Figure 1.

This section specifies the crucial stages involved in the model development and evaluation of LivMarX, the proposed method for Liver Cirrhosis stage prediction. It commences with an explanation of data acquisition procedures, followed by a detailed discussion of the preprocessing methods employed to prepare the data for further analysis. Subsequently, the architectural design and operational flow of the LivMarX model are meticulously described, highlighting both its machine learning and optimization submodules. Lastly, a thorough analysis with a range of performance metrics is carried out to determine the effectiveness of the suggested model. Each stage is segmented into distinct subsections to ensure a comprehensive understanding of the image classification process and its intricacies. Every individual phase is further elaborated upon in the subsequent subsections.

A. DATASET DESCRIPTION

This research makes use of information gathered from the 1974–1984 Mayo Clinic trial on Primary Biliary Cirrhosis (PBC) of the liver [21]. This dataset incorporates details from 424 PBC patients, all satisfying eligibility criteria for a randomized placebo-controlled trial evaluating the effectiveness of the drug D-penicillamine. The inclusion criteria for participants involved being diagnosed with PBC, having a certain level of liver function, and not having any other major liver diseases. Patients were excluded if they had advanced liver disease beyond a specific threshold, other concurrent major illnesses, or were unable to provide informed consent. The inclusion criteria required patients to be diagnosed with PBC and referred to the Mayo Clinic during the specified period, fit for the clinical trial, and consent to participation or follow-up. Exclusion criteria likely included patients with other liver conditions or those unwilling or unable to provide informed consent. Among these patients, the initial 312 actively engaged in the clinical trial, contributing extensive data. Utilizing a public dataset offers several advantages for research and analysis. The results have more validity because of the large number of people who fit the requirements for a randomized placebo-controlled study. In addition to the initial 312 trial participants, the dataset includes data from 112 patients who did not participate in the clinical trial. These patients represent a broader spectrum of PBC. Their inclusion is particularly valuable as it brings diverse stages of the disease, treatment histories, and demographic profiles that are essential for understanding the full scope of PBC. These additional data points are crucial for enhancing the robustness and applicability of the findings which offer insights into a wider array of patient experiences and outcomes and for understanding how the drug may impact individuals who did not actively engage in the clinical trial.

The following details the distribution of stages and the demographic composition of the patient population:

- **Stage Distribution and Age:**
  - Stage 1: 21 patients, Average Age: 46.8 ± 9.5 years
  - Stage 2: 92 patients, Average Age: 49.5 ± 9.6 years
  - Stage 3: 155 patients, Average Age: 49.0 ± 10.1 years
  - Stage 4: 144 patients, Average Age: 53.8 ± 10.8 years
- **Gender Distribution:**
  - Female: 374 patients
  - Male: 44 patients

B. DATA PREPROCESSING AND FEATURE ENGINEERING

The dataset, its features, visualization of spread of the important variables that could be potential contributors to prediction of stage of Liver Cirrhosis were explored in this step.

The dataset includes information on patient ID, the number of days spent in the hospital (N_Days), patient status, prescribed drugs, age, gender, ascites presence, hepatomegaly, spider angiomas, edema, bilirubin levels, cholesterol levels, albumin levels, copper levels, serum glutamic oxaloacetic transaminase (SGOT) levels, triglycerides levels, platelet counts, alkaline phosphatase levels, prothrombin time, and the disease stage. Before data preprocessing, the dataset contained several attributes that required transformation and cleansing. Initially, the patient ID attribute was removed, as it provided no valuable insight for the needed analysis. The 'Age' attribute was also transformed from days to a more appropriate format. Then, the numerical variable 'Stage' which represents the disease stage of patients was converted into a categorical variable and resolved any type of errors. Several data preprocessing steps such as outlier detection and handling were also performed. Notably, the 'Cholesterol' and 'Prothrombin' attributes were identified to have outliers and were addressed by capping or truncating extreme values to ensure the data remained within a specified range. Before applying any preprocessing steps, the presence of missing values in some variables was observed using a heat map (Fig. 2), with 'Stage' having six missing values. The median of the attributes were used to fill in the missing values and subsequently the six rows with missing values in 'Stage' were removed. The choice of using median values for imputation is supported by its robustness to outliers and skewed distributions, making it a reliable method for handling missing data in medical datasets. By using the median, it is ensured that the central tendency of the data is not altered which minimizes the bias introduced by missing values and enhancing the reliability of the model.

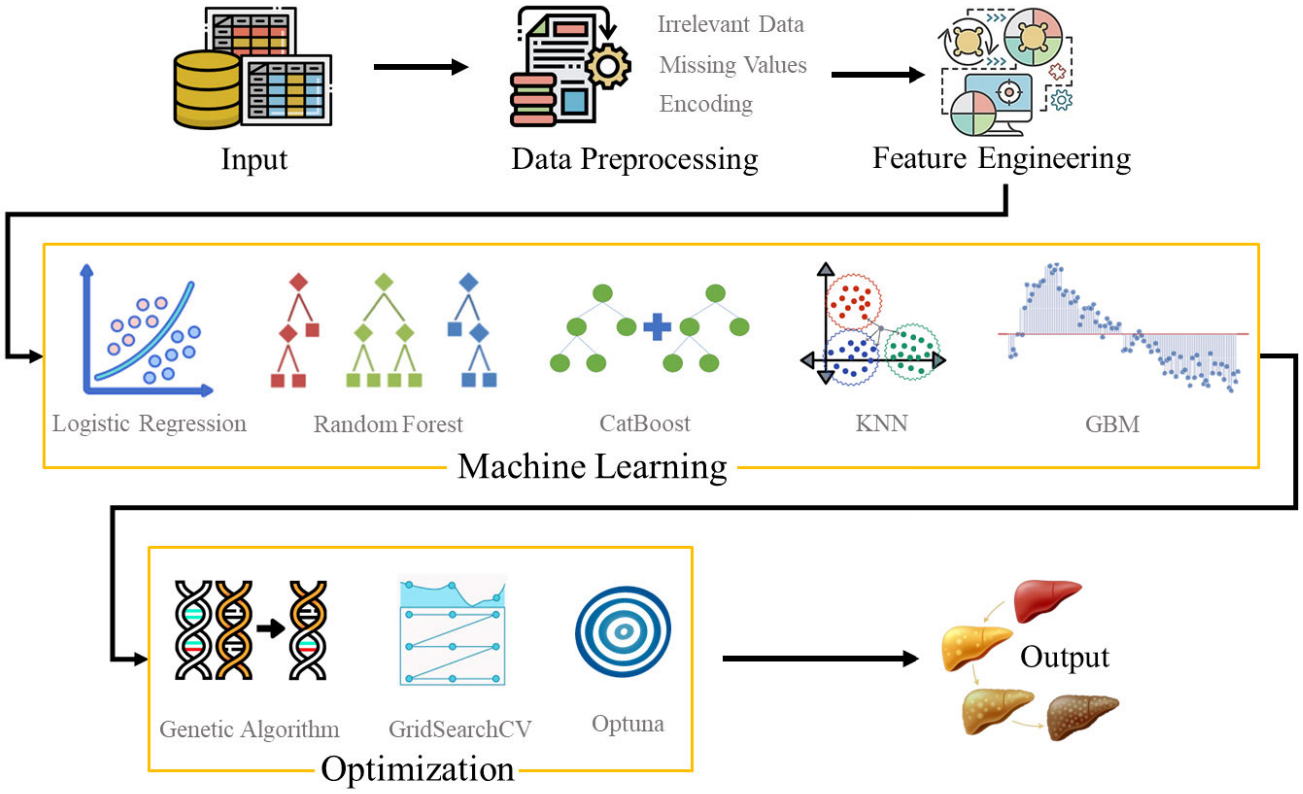


FIGURE 1. Workflow of the proposed LivMarX model.

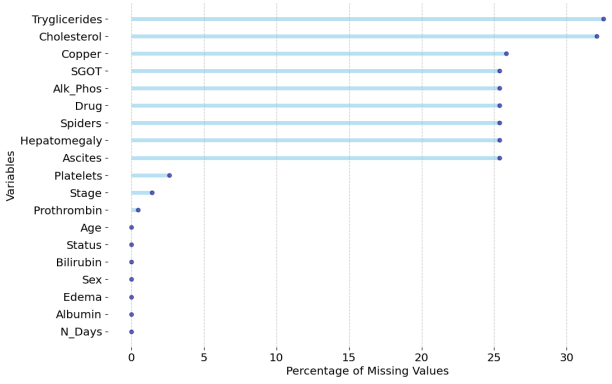


FIGURE 2. Graphical representation of the percentage of missing values across variables.

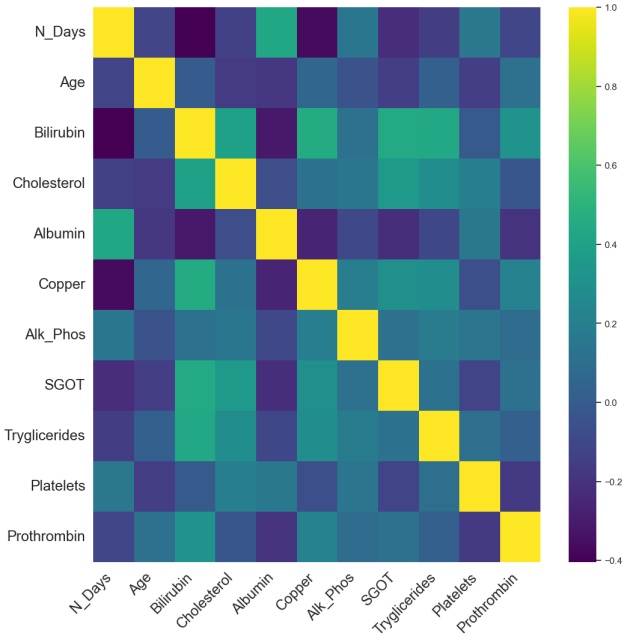


FIGURE 3. Visualization of the correlation between various biochemical markers.

After identifying missing values and outliers, utilizing the correlation matrix in Figure 3 the pairwise correlations between various variables associated with predicting liver cirrhosis were examined. Each cell in the matrix displays a correlation coefficient, ranging from -1.0 to 1.0 where 1.0 signifies a perfect positive linear relationship, -1.0 represents a perfect negative linear relationship, and 0 indicates no linear correlation. The matrix uses a color scale to visually distinguish these relationships: yellow highlights strong positive correlations, dark purple indicates strong negative correlations, and lighter shades such as blue or light purple denote weaker or non-existent correlations. The scale on the right-hand side of the matrix that ranges from -0.4 to

1.0, acts as a legend. It helps in the quick identification of the strength and direction of the correlations, which is crucial for recognizing significant predictors of liver cirrhosis and understanding the data patterns for further analysis. This analysis highlighted significant correlations, notably between bilirubin and copper levels, as well as with other blood components such as triglycerides and SGOT. This finding emphasized the importance of bilirubin in determining Liver Cirrhosis stages. This understanding guided subsequent preprocessing steps, ensuring that the approach considered not only individual variables but also their interconnections. Then, feature engineering was done to enhance the predictive accuracy of the analysis. For instance, a new variable was created based on 'N_Days' to group patients into different categories, allowing the exploration of the relationship between the length of hospital stays and other variables. Categorical variables related to albumin levels, copper levels, alkaline phosphatase levels, SGOT levels, triglycerides levels, platelet counts, prothrombin time, age and sex combinations, drug and disease stage combinations, edema and disease stage combinations, and ascites and disease stage combinations were also generated. After completing these preprocessing steps, the dataset was transformed, resulting in new attributes that reflected the cleaned data. These attributes included 'N_Days_New,' 'Bilirubin_High,' 'Cholesterol_High,' 'Albumin_High,' 'Albumin_Low,' 'Copper_Low,' 'Copper_High,' 'Alk_Phos_Low,' 'Alk_Phos_High,' 'SGOT_Normal,' 'SGOT_High,' 'Triglycerides_Normal,' 'Triglycerides_High,' 'Platelets_Abnormality,' 'Platelets_Normality,' 'Prothrombin_Low,' 'Prothrombin_High,' 'New_Sex_Age_Cat,' 'drugstage,' 'edemastage,' and 'ascitestage.' These preprocessing steps were crucial to ensure the dataset's quality and prepare it for comprehensive analysis, including investigative data analysis, modeling, and interpretation of results. The transformed dataset enabled the investigation of the relationships between various patient attributes and the disease stage more effectively, shedding light on potential factors contributing to liver disease progression.

In attribute value classification for liver health assessment, feature engineering was employed to create meaningful categories based on key biochemical markers. In Table 2 this is illustrated. Bilirubin levels were categorized as ''Bilirubin High'' if greater than or equal to 1.2 mg/dL [22], indicating potential liver dysfunction. Cholesterol levels were dichotomized into ''Cholesterol High'' for values equal to or exceeding 200 mg/dL [23], pointing towards elevated cholesterol levels associated with liver complications. Albumin values were divided into ''Albumin High'' for readings equal to or exceeding 5.4 g/dl, while values below 3.4 g/dl were classified as ''Albumin Low,'' [24] indicating potential liver synthetic dysfunction. Copper levels were categorized as ''Copper High'' for values equal to or exceeding 50 ug/day, and ''Copper Low'' for values below 20 ug/day [25], suggesting potential abnormalities in copper metabolism. Alkaline Phosphate levels were stratified into ''Alk Phos High'' for values greater than 147 U/liter and ''Alk Phos Low'' for values below 44 U/liter [26], indicative of potential liver or bone-related issues. SGOT levels were classified as ''SGOT Normal'' if less than or equal to 36 U/ml, otherwise as ''SGOT High,'' suggesting potential liver damage. Triglyceride levels were categorized as ''Triglycerides High'' for values exceeding 199 mg/dl [27], reflecting elevated triglyceride levels associated with liver dysfunction. Platelet counts below 180 ml/1000 were labeled as ''Platelets Abnormal,'' [28] potentially signaling liver-related abnormalities. This approach simplifies complex biochemical data, aiding in the identification of potential liver health issues through interpretable and clinically relevant categories.

TABLE 2. Biomarker thresholds and attribute value classifications for dimensionality reduction in Liver Cirrhosis risk assessment.

Attribute	Value	Classification
Bilirubin	≥ 1.2	Bilirubin_High
Cholesterol	≥ 200	Cholesterol_High
Albumin	≥ 5.4	Albumin_High
	< 3.4	Albumin_Low
Copper	≥ 50	Copper_High
	< 20	Copper_Low
Alkaline Phosphate	> 147	Alk_Phos_High
	< 44	Alk_Phos_Low
SGOT	> 36	SGOT_Normal
	≤ 36	SGOT_High
Triglycerides	> 199	Triglycerides_High
	≤ 199	Triglycerides_Normal
Platelets	< 150	Platelets_Abnormality
	≥ 150	Platelets_Normality

Feature engineering was applied to create meaningful classifications based on the given tables. In Table 3, three categories were derived. If the entry was 'N,' it was classified as ''No Edema.'' For 'Y,' it was categorized as ''Edema no diuretic,'' and for 'S,' it was labeled as ''Edema diuretic'' based on the stage. This classification provides insights into the presence of edema and its potential management with diuretic treatment.

In Table 3, four categories were established. 'N' entries for all four instances were classified as ''No illness but ascites,'' indicating the absence of a specific illness but the presence of ascites. 'Y' entries were categorized as ''Illness ascites'' or ''Illness no ascites,'' reflecting the presence or absence of ascites in the context of an illness. This classification helps distinguish between different health conditions based on the presence or absence of ascites.

Table 4 was processed to create age and sex-related categories. For males, individuals aged 21 or below were classified as ''youngmale,'' those aged between 21 and

TABLE 3. Liver Cirrhosis staging and medical actions for dimensionality reduction based on clinical conditions.

Condition	Value	Stage	Action
Edema	N	1	No Edema
	Y	1	Edema no diuretic
	S	1	Edema diuretic
	N	2	Early no edema
	Y	2	Early edema no diuretic
	S	2	Early edema diuretic
	N	3	Late no edema
	Y	3	Late edema no diuretic
	S	3	Late edema diuretic
	N	4	Late no edema
	Y	4	Late edema no diuretic
	S	4	Late edema diuretic
Ascites	Y	1	No illness but ascites
	N	1	Normal
	Y	2	Illness ascites
	N	2	Illness no ascites
	Y	3	Illness ascites
	N	3	Illness no ascites
	Y	4	Illness ascites
	N	4	Illness no ascites

TABLE 4. Dimensionality reduction through classification labels assigned based on sex and age features.

Sex	Age	Classification
M	≤ 21	youngmale
	> 21 & < 40	maturemale
	≥ 40	seniormale
F	≤ 21	youngfemale
	> 21 & < 40	maturefemale
	≥ 40	seniorfemale

40 as "maturemale," and those aged 40 and above as "seniormale." Similarly, for females, individuals aged 21 or below were categorized as "youngfemale," those aged between 21 and 40 as "maturefemale," and those aged 40 and above as "seniorfemale." This classification enables the analysis of demographic patterns based on age and gender.

C. MACHINE LEARNING METHODS

After completing data preprocessing and feature engineering, the dataset was partitioned into 70% training, 20% test, and 10% validation sets. The validation set gives access to hyperparameter tuning and model selection ensuring that the models are not overfitted to the training data and generalize well to unseen data. Subsequently, several machine learning algorithms were applied to predict the Liver Cirrhosis stage.

1) LOGISTIC REGRESSION

This algorithm is a linear classification model that is well-suited for binary classification tasks. Logistic Regression was chosen because of its robustness in handling binary outcomes, important for the objective of categorizing patients into the different cirrhosis stages. This model is an excellent baseline model in the field of machine learning. It has a straightforward methodology and is easy to understand making it a standard choice for comparison to other, more complex models. Logistic regression models the probability of a binary outcome (1 or 0) using the logistic function [29]. The formula for the sigmoid function is given in equation 1:

$$P(Y=1) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_2+\beta_2 X_2+\ldots+\beta_n X_n)}} \tag{1}$$

2) DECISION TREE

The Decision Tree Classifier is a non-linear model that divides the data into subgroups according to the values of the attributes in order to provide predictions [30]. Decision Trees were utilized for their transparency and ease of interpretation, which is essential in medical applications for validating model logic by clinical experts. They are especially useful for handling the non-linear relationships and interactions between the various clinical parameters in our dataset, such as enzyme levels and bilirubin counts, which are critical for staging liver disease.

$$\hat{y}_i = \text{CART}(X_i) = \sum_{j=1}^{J} c_j \cdot \mathbf{1}(X_i \in R_j) \tag{2}$$

- $\hat{y}_i$ represents the predicted output for the i-th sample.
- $\text{CART}(X_i)$ represents the CART (Classification and Regression Trees) model applied to the input features X_i.
- J is the number of terminal nodes (leaves) in the tree.
- c_j is the predicted value associated with the j-th terminal node.
- R_j represents the j-th region (subset of the feature space) defined by the decision tree.
- $\mathbf{1}(X_i \in R_j)$ is an indicator function that equals 1 if X_i falls into region R_j, and 0 otherwise.

3) RANDOM FOREST

Random Forest Classifier where several decision trees are built using an ensemble learning technique, which are then aggregated to predict. [31]. This model was chosen as it helps to overcome overfitting, a common challenge in medical datasets with many parameters but relatively fewer instances. This model also helped get a better accuracy when predicting using the varied biochemical markers in our dataset. Random Forest is particularly effective in handling high-dimensional datasets and capturing the importance of various features.

$$\hat{y}_i = \text{RF}(X_i) = \frac{1}{B}\sum_{b=1}^{B} f(x_i;\theta_b) \tag{3}$$

- $\hat{y}_i$ represents the predicted output for the i-th sample.
- $\text{RF}(X_i)$ represents the Random Forest model applied to the input features X_i.
- B is the number of trees in the forest.

- $f(x_i; \theta_b)$ represents the b-th decision tree in the forest, with parameters θ_b.

4) GRADIENT BOOSTING

A boosting approach called the Gradient Boosting Classifier which combines the predictions of several weak learners to produce a strong ensemble model was applied [32]. This model was chosen for its ability to fine-tune the model by focusing on mistakes made in previous rounds of training. Gradient Boosting is known for its high predictive accuracy and ability to handle complex relationships in the data. This feature is particularly beneficial for this dataset, which includes subtle variations in cirrhosis indicators across different patient groups.

$$\hat{y}_i = \text{GBM}(X_i) = \sum_{k=1}^{K} f_k(X_i) \tag{4}$$

where $\hat{y}_i$ represents the predicted output for the i-th sample, and $\text{GBM}(X_i)$ represents the Gradient Boosting model applied to the input features X_i. K is the number of weak learners (usually decision trees) in the ensemble, and $f_k(X_i)$ represents the prediction of the k-th weak learner.

5) CATBOOST

The CatBoost Classifier, a gradient-boosting variant specifically designed for categorical data was applied [33]. This model was chosen for its superior ability to handle categorical variables efficiently without needing extensive pre-processing. This was applied since the dataset includes critical categorical variables such as drug type, gender, and the presence or absence of clinical features like ascites, hepatomegaly, and edema. These factors are vital for modeling liver cirrhosis progression and treatment outcomes. CatBoost's approach minimizes prediction errors compared to traditional encoding methods, making it especially suitable since a combination of clinical and demographic data is used to predict disease progression.

Each algorithm was strategically picked out to take advantage of their unique strengths, matching them to the specific needs and challenges of the liver cirrhosis dataset. This careful selection ensured that the approach is was methodologically sound making sure the predictions were accurate and reliable.

D. OPTIMIZATION

A thorough analysis was carried out to determine the best machine learning method for predicting the stages of Liver Cirrhosis, taking into account the advantages and disadvantages of different classifiers. Each ML method was selected based on its unique qualities and suitability for the properties of the dataset. The Random Forest Classifier performed better than expected, demonstrating its capacity to manage high-dimensional datasets and weigh the significance of different variables. By building several decision trees and combining their predictions, this ensemble learning method has shown significant improvements in accuracy and resistance to overfitting. Decision trees are highly interpretable, and Random.

1) GENETIC ALGORITHM

Algorithm 1 Genetic Algorithm Optimization for Random-Forest Hyperparameters

Define evaluation function *evaluate*
(θ) where $\theta = \{n\_estimators, max\_depth, min\_samples\_split, min\_samples\_leaf\}$.
Ensure parameter validity: $\theta_i = \max(\theta_i^{min}, \min(\theta_i^{max}, \text{int}(\theta_i)))$.
Configure RandomForestClassifier: $RF(\theta)$.
Define Fitness as negative mean of cross-validation scores:
$Fitness \leftarrow -1 \times \text{mean}(\text{CrossValScore}($
$RF(\theta), X_{train}, y_{train},$
$cv = 5, scoring =' accuracy'))$.
Problem $\leftarrow$ Maximize(fitness(θ)).

Define genetic space:
$\Theta_{n\_estimators} \sim \mathcal{U}(10, 200)$.
$\Theta_{max\_depth} \sim \mathcal{U}(1, 100)$.
$\Theta_{min\_samples\_split} \sim \mathcal{U}(2, 20)$.
$\Theta_{min\_samples\_leaf} \sim \mathcal{U}(1, 10)$.
Create GA individual:
$I = (\Theta_{n\_estimators}, \Theta_{max\_depth}, \Theta_{min\_samples\_split}, \Theta_{min\_samples\_leaf})$.

Register genetic operations in toolbox:
Mate $\leftarrow$ Blend Crossover with $\alpha = 0.5$.
Mutate $\leftarrow$ Gaussian Mutation with $\mu = 0, \sigma = 10, indpb = 0.2$.
Select $\leftarrow$ Tournament Selection with $size = 3$.
Evaluate $\leftarrow$ $evaluate(\theta)$.

Initialize population P with $N = 10$ individuals.
Run GA: $P_{next} \leftarrow \text{EA}(P, \mu = 10, \lambda = 50, cxpb = 0.7, mutpb = 0.2, ngen = 10)$.
Select best individual I_{best} based on fitness.
Output I_{best}: Best Hyperparameters.

Instantiate $RF(I_{best})$.
Fit $RF(I_{best})$ on X_{train}.
Predict y_{pred} on X_{test} using $RF(I_{best})$.
Evaluate accuracy $\leftarrow$ Accuracy(y_{test}, y_{pred}).
Return accuracy: Test Set Performance.

Forest's accuracy improvements make them an excellent choice for more optimization [34].

To comprehensively explore the hyperparameter space and ensure the identification of an optimal configuration for the Random Forest Classifier, a hybrid approach combining three distinct optimization techniques—Genetic Algorithm (GA),

Optuna, and GridSearchCV—was implemented. This synergistic strategy aimed to capitalize on the unique strengths of each optimization method, leveraging their complementary features. The Genetic Algorithm introduced diversity and adaptability through its evolutionary process, allowing for a broad exploration of the hyperparameter landscape. The GA effectively identified advantageous areas within the search space by developing a population of potential solutions over several generations. The hyperparameter optimization of the Random Forest Classifier was performed using a custom genetic algorithm implemented with the DEAP (Distributed Evolutionary Algorithms in Python) library [35]. The algorithm employs a '(μ, λ)-evolution strategy' (mu comma lambda), where a population of μ individuals is evolved over generations to produce λ offspring. The genetic operators applied include blend crossover and Gaussian mutation, with tournament selection for reproductive success. This strategy iteratively refines the hyperparameters based on cross-validated model accuracy, aiming to discover an optimal set that maximizes predictive performance. This approach allowed for a thorough investigation of hyperparameters, uncovering configurations and paving the way for subsequent different kinds of optimization.

2) OPTUNA

Following this, Optuna, a Bayesian optimization library, was used [36]. Employing a more targeted exploration, Optuna adapted its search strategy based on the outcome of previous trials. This Bayesian optimization, steered by probabilistic models, demonstrated efficiency in narrowing down the search space. Optuna played a key role in refining the model's hyperparameters and enhancing its overall performance.

3) GRIDSEARCHCV

To further refine the model and validate the findings from the evolutionary and Bayesian approaches, GridSearchCV [37] was incorporated. This exhaustive search method systematically evaluated hyperparameter combinations in a predefined grid, ensuring no stone was left unturned in the pursuit of optimal configurations. While computationally more demanding, GridSearchCV provided a comprehensive and exhaustive exploration, serving as a final validation step. GridSearchCV provided a comprehensive and exhaustive exploration, serving as a final validation step, ultimately delivering the best results in terms of model performance and hyperparameter optimization.

E. EVALUATION METRICS

The performance of these algorithms was evaluated using key evaluation metrics, including F1-score, recall, precision, and accuracy. The F1 score balances precision and recall and is particularly useful when dealing with imbalanced datasets. Recall calculates the ratio of true positives to all real positives, whereas precision calculates the ratio of true positive predictions to all positive predictions. Accuracy provides an overall measure of the model's correctness in its predictions. The expression for the different metrics for performance are given in equations 5 - 8:

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{6}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{7}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

where TP represents True Positives, TN represents True Negatives, FP represents False Positives, and FN represents False Negatives.

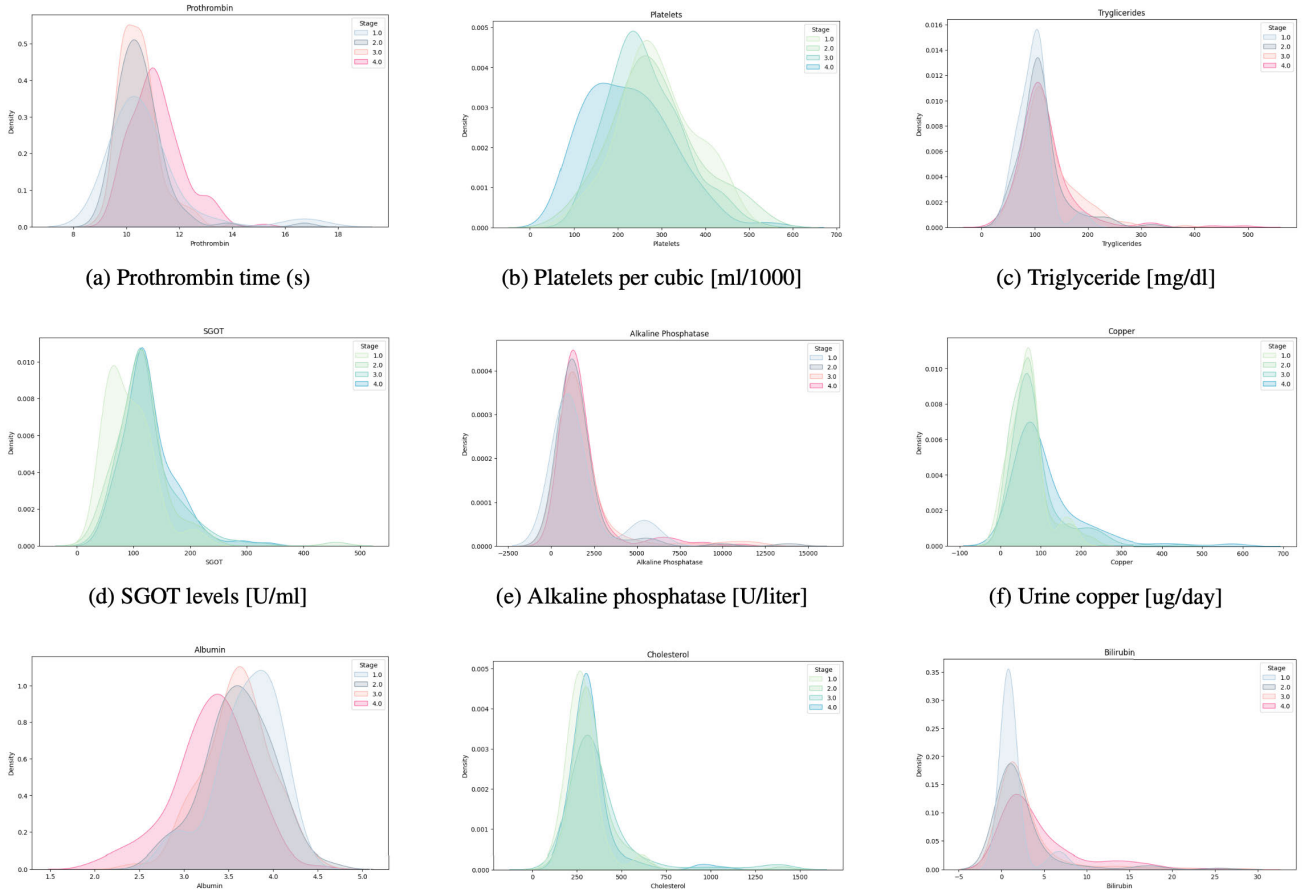
Algorithm 2 Optimization and Evaluation of RandomForest for Liver Cirrhosis Prediction Using GridSearchCV

Initialize RandomForestClassifier with a fixed random state of 17 for reproducibility.
Define hyperparameter grid:
  Number of estimators, $N_e \in \{50, 100, 200\}$
  Maximum depth, $D_{max} \in \{\text{None}, 10, 20\}$
  Minimum samples split, $S_{min} \in \{2, 5, 10\}$
  Minimum samples leaf, $L_{min} \in \{1, 2, 4\}$
Configure GridSearchCV with RandomForestClassifier, parameter grid, accuracy scoring, and 5-fold cross-validation.
Execute hyperparameter tuning via GridSearchCV to find the best combination of parameters.
Retrieve best hyperparameters from GridSearchCV and **print** the optimal parameter set.
Instantiate and **train** RandomForestClassifier with optimal parameters obtained.
Evaluate the optimized model on the test set: compute accuracy and generate a classification report.
Calculate feature importance for model interpretability:
  $I(f_i) = \frac{1}{N_e} \sum_{j=1}^{N_e} \Delta accuracy(T_j, f_i)$
  Where $I(f_i)$ is the importance of feature f_i, N_e is the number of estimators, and $\Delta accuracy(T_j, f_i)$ is the change in model accuracy when feature f_i is shuffled in tree T_j.
Present results including best parameters, accuracy on the test set, detailed performance metrics from the classification report, and feature importances.

IV. RESULTS

This section describes in detail the results obtained for this study. Commencing with a detailed exploration of predictive analysis on Liver Cirrhosis stages, a thorough analysis of the cleaned data was conducted. This was followed by an extensive preprocessing phase that enhanced data quality and relevance.

The dataset contained attributes that included ID, N_Days, Status, Drug, Age, Sex, Ascites, Hepatomegaly, Spiders, Edema, Bilirubin, Cholesterol, Albumin, Triglycerides,



(a) Prothrombin time (s) (b) Platelets per cubic [ml/1000] (c) Triglyceride [mg/dl]

(d) SGOT levels [U/ml] (e) Alkaline phosphatase [U/liter] (f) Urine copper [ug/day]

(g) Albumin [g/dl] (h) Cholesterol [mg/dl] (i) Bilirubin [mg/dl]

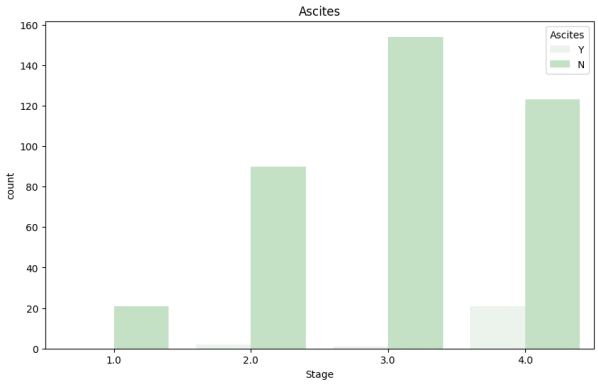
FIGURE 4. **Blood and urine markers across different stages of liver disease.**

Copper, Alk_Phos, SGOT, Prothrombin, Platelets, and Stage. Preprocessing steps applied were: dropping the ID attribute, transforming Age into a more appropriate by converting it from days to years. The numerical variable Stage was converted into a categorical one, addressing any type errors encountered in the process. Outliers in Cholesterol and Prothrombin were handled appropriately by calculating lower and upper outlier thresholds. Missing values were managed using the median. Categorical variables were encoded, and new features were introduced to enhance the accuracy of the model. Figures 4 and 5 offer detailed charts that clearly show the relationship between different types of variables and the progression of Liver Cirrhosis through its four stages. In Figure 4, the charts show kernel density estimates (KDE) for various biomarkers used in predicting the stages of Liver Cirrhosis. Density on these graphs represents the smoothed probability density function of the measured values across different disease stages. Each curve's area sums to one.

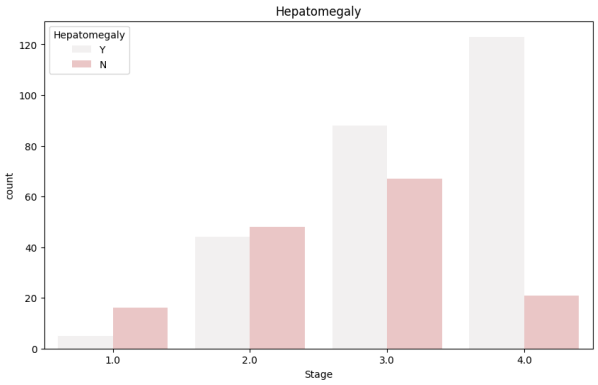
This method offers a clearer view of the distribution shapes which are essential for identifying distinct modes and the overall spread of data. The y-axis scales vary by biomarker due to the differences in their value ranges, distribution shapes, and their data spread. For example, biomarkers like platelets have a different numerical range and variance compared to bilirubin which influences the width and height of their respective density curves. These are insightful illustrations that accurately show the probability distribution of the biomarker values at various stages of Liver Cirrhosis. They help us see the increase or decrease in these markers, emphasizing their importance in understanding the disease and their usefulness in predicting how severe it is.

Figure 5, showcases discrete variables like Ascites, Hepatomegaly, Spider angiomas, and Edema. The figure assesses the frequency of each clinical sign in each stages of Liver Cirrhosis. These visuals provide a clear picture of how these biomarkers correlate with the disease's progression, highlighting patterns that become apparent as the condition worsens. By grouping patients according to whether they exhibit these signs, these figures help deepen our understanding of how Liver Cirrhosis affects the body, aiding in a more thorough approach to diagnosis.

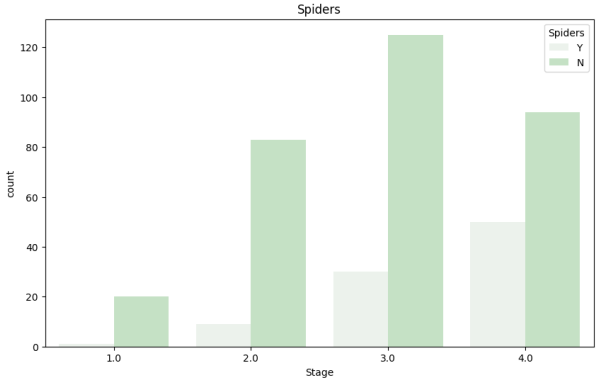
Novel variables were also established: Bilirubin_High, Cholesterol_High, Albumin_High, Albumin_Low, Copper_Low, Copper_High, Alk_Phos_Low, Alk_Phos_High,



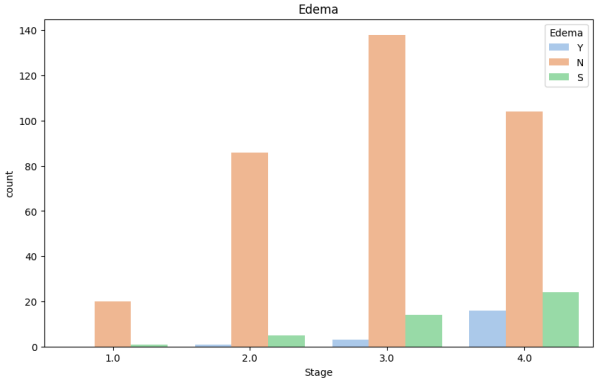
(a) Ascites proportion across different stages.



(b) Hepatomegaly proportion across stages.



(c) Patients with Spider angiomas across different stages.



(d) Patients with edema across different stages.

FIGURE 5. **Clinical signs across different stages of liver disease.**

SGOT_Normal, SGOT_High, Triglycerides_Normal, Triglycerides_High, Platelets_Abnormality, Platelets_Normality, Prothrombin_Low, Prothrombin_High, New_Sex_Age_Cat, drugstage, edemastage, and ascitestage. Each of these steps was crucial in preparing and refining the data for subsequent analyses.

The resulting dataset, post-preprocessing, encompasses all these attributes, providing a refined foundation for subsequent analysis. The preprocessed dataset was split into training and test sets for the analysis to evaluate machine learning algorithm performance. In order to provide a fair assessment of the models employed, 70% of the data was designated as the training set, 20% as the test set, and 10% as the validation set. This ensured that the test set was not visible during the training process and allowed for an additional validation phase to monitor the model's performance and prevent overfitting during training.

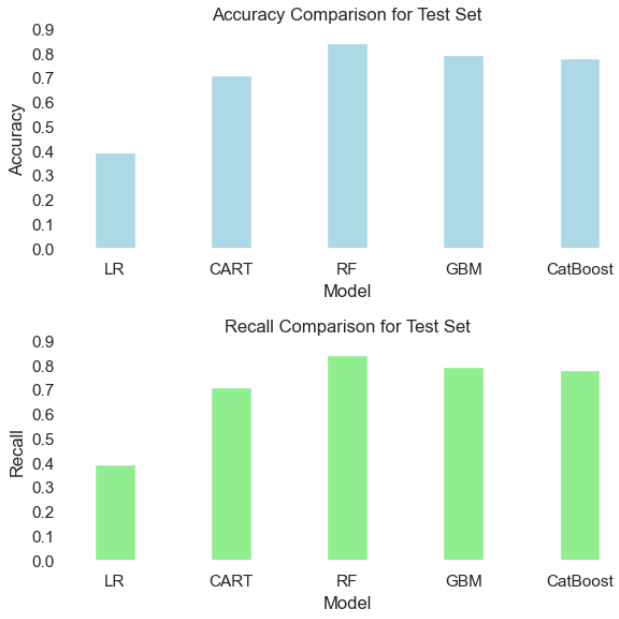
Performance evaluation of various machine learning models was done to understand their capability in accurately predicting the stages of Liver Cirrhosis. This section carefully outlines and describes the outcomes obtained from each model, highlighting their predictive accuracies and offering a comparative insight into their effectiveness. The Random Forest Classifier surpassed the performance of

TABLE 5. **Comparison of accuracy for the base models.**

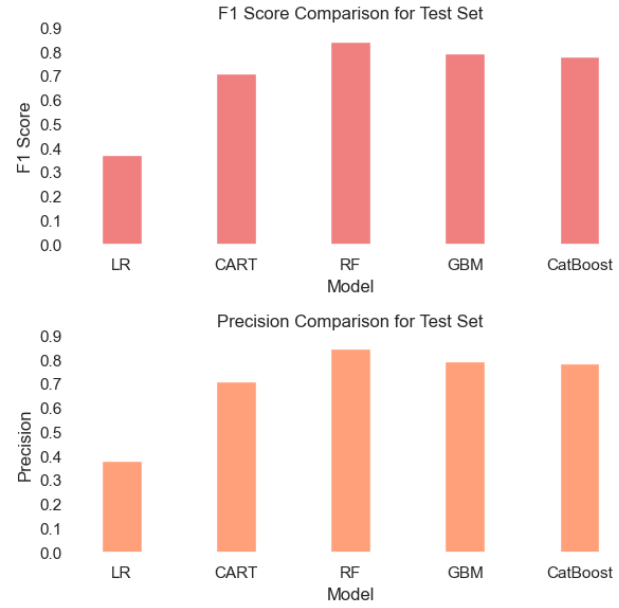
Algorithm	Accuracy (%)
Logistic Regression	39.75
Decision Tree Classifier	73.49
Random Forest Classifier	84.33
Gradient Boosting Classifier	79.51
CatBoost Classifier	78.31

Gradient Boosting Classifier, CatBoost Classifier, as well as Decision Tree Classifier, achieving an accuracy of 84.33%. LR exhibited the lowest accuracy at 39.75%. Random Forest successfully navigated the data relationships and patterns linked to liver disease progression. The effectiveness of Random Forest in providing accurate and reliable predictions is highlighted, particularly within the scope of the proposed study. These findings are detailed in Table 5.

The visualization of the evaluation metrics used in F1 score, recall, precision and accuracy is shown in Figure 6. Random Forest Classifier outperformed Gradient Boosting Classifier, CatBoost Classifier as well as Decision Tree Classifier by having an accuracy of 84.33%. LR performed the poorest with an accuracy of 39.75%. The accuracy and



(a) Comparison of Accuracy and Recall



(b) Comparison of F1 Score and Precision

FIGURE 6. Comparison of evaluation metrics for initial base models.

recall metrics obtained from Random Forest were 84.33% and 84.0%, respectively. GBM yielded an accuracy of 79.51% with a recall of 80%. LR exhibited the lowest performance, achieving an accuracy of 39.75% and a recall of 40%. The precision and F1 score of Random Forest were also the best with a precision of 85% and an F1 score of 84%. GBM gave a precision of 80% and F1 score of 80%. LR performed the poorest with a precision and recall score of 38%.

The performance differences among these models can be attributed to their inherent characteristics and suitability for the complexities of the dataset. Logistic Regression

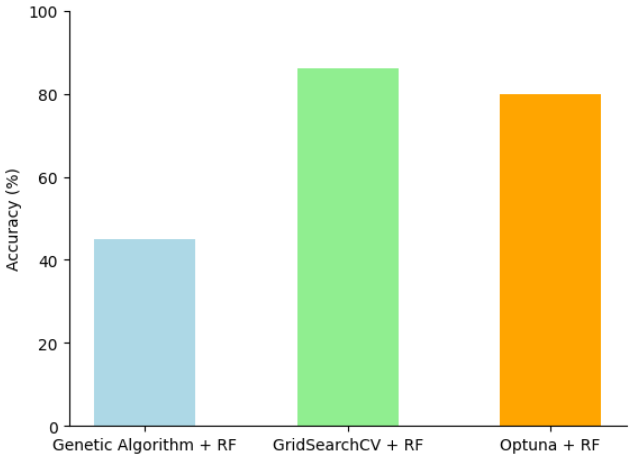


FIGURE 7. Comparison of model accuracies post optimization.

exhibited the lowest accuracy because its linear nature limits its ability to capture the complex, non-linear relationships in the liver cirrhosis dataset. The Decision Tree Classifier with an accuracy of 73.49%, performed better than LR but still lower than ensemble methods due to its tendency to overfit the training data. CatBoost, designed to handle categorical data efficiently, showed good performance with an accuracy of 78.31%, but it did not surpass Random Forest, possibly due to the specific nature of our dataset. The Gradient Boosting Classifier also performed well, achieving 79.51% accuracy, but its sensitivity to hyperparameters and extensive tuning requirements may have limited its performance compared to Random Forest. The Random Forest Classifier had the best performance because it combined multiple decision trees which reduced overfitting and effectively handled the variability and complexity in the data. This ensemble method allowed it to capture complex patterns leading to the highest accuracy.

In summary, the Random Forest Classifier's superior performance highlights its ability to generalize well across complex datasets, while the lower performance of Logistic Regression highlights the need for more complex models to capture the patterns in medical data.

Figure 7 showcases the effectiveness of various optimization strategies applied to enhance the Random Forest model's performance. By integrating the Random Forest algorithm with sophisticated optimization methods such as Genetic Algorithm, Optuna, and GridSearchCV, we observed diverse outcomes in terms of predictive accuracy. Optimising Random Forest with Genetic Algorithm performed poorly decreasing accuracy from 84% to 45%. Parameters and Hyperparameters for this method included a continuous range from 1 to 100 for n_estimators and max_depth, and from 2 to 100 for min_samples_split and min_samples_leaf. The genetic algorithm used evolutionary strategies involving a population of 10 individuals over 10 generations with mutation and crossover operations. The best hyperparameters identified were n_estimators=49, max_depth=83,

min_samples_split=19, and min_samples_leaf=4. When optimized with Optuna, the model outperformed existing models from other studies with an accuracy of 81%. Optuna's hyperparameter tuning involved parameters like n_estimators ranging from 50 to 300, max_depth options including None, 10, 20, 30, and a range from 2 to 20 for min_samples_split, 1 to 10 for min_samples_leaf, with max_features options of ['auto', 'sqrt', 'log2', None]. Bayesian optimization was utilized over 100 trials to find the optimal parameters.

However, our original LivMarX model with just Random Forest had a better accuracy. Random Forest optimized with GridSearchCV reached an accuracy of 86% which enhanced the original model by 2%. The optimization with GridSearchCV involved a 5-fold cross-validation to systematically search through the parameter grid of n_estimators: [50, 100, 200], max_depth: [None, 10, 20], min_samples_split: [2, 5, 10], and min_samples_leaf: [1, 2, 4]. The best parameters found were n_estimators=50, max_depth=None, min_samples_split=10, and min_samples_leaf=1. This optimized model was tested with a validation set of 10% and achieved an accuracy of 84%, which is only 2% less than the test set accuracy. This minor difference indicates that the model generalizes well to unseen data, showing its reliability. The close performance between the validation and test sets also suggest that the model is not overfitting. Original model optimized with GridSearchCV is the proposed LivMarX model in this paper.

Receiver Operating Characteristic curves are a plot of the true positive rate vs the false positive rate at various threshold settings. Figure 8 depicts the micro-averaged ROC curves for each predictive model we assessed. The Area Under the Curve (AUC) is a measure of the performance of a classification model at various threshold settings. Its value ranges from 0 to 1, where an AUC of 0.5 suggests a model with no discriminative ability and an AUC of 1 indicates a perfect model that correctly classifies all positives and negatives. It provides a more nuanced view of model performance, considering both the sensitivity and specificity of each classifier, which are crucial for the reliable diagnosis of Liver Cirrhosis stages. This ROC analysis complements the accuracy metrics presented in Table 5 and the evaluation metrics visualized in Figure 6. The proposed model LivMarX, outperformed the seven other models compared in this study with an AUC of 0.95. LivMarX demonstrates superior discrimination ability and provides a reliable diagnosis of Liver Cirrhosis stages. This analysis further reinforces the superiority of LivMarX over other models, each reflected by their respective AUC scores.

To enhance the interpretability of LivMarX, SHAP (SHapley Additive exPlanations) summary plots were computed for each feature across four distinct stages of Liver Cirrhosis, revealing the average impact of each feature on LivMarX's output magnitude. Figure 9a shows that in Stage 1, ascitestage_normal and edemastage_no edema exhibited the highest mean SHAP values. This indicates a strong association with the absence of ascites and edema,

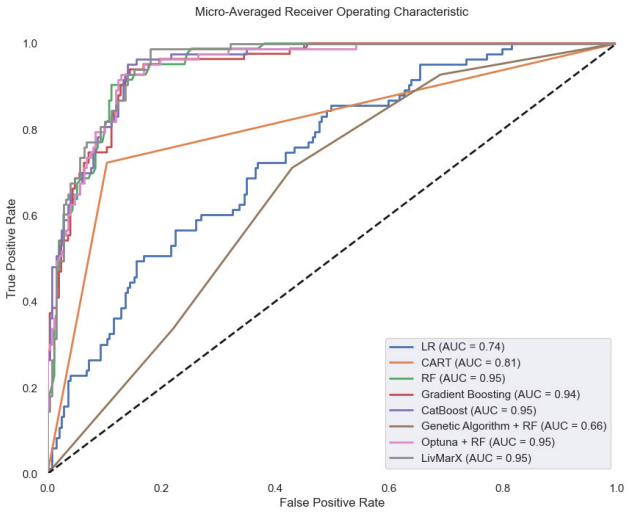


FIGURE 8. Micro-averaged ROC curves illustrating the performance of various models.

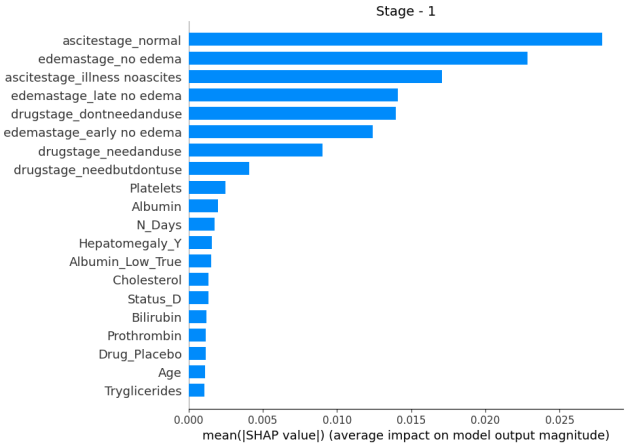
which suggests a milder progression of the disease. However, in Stage 2, edemastage_early no edema and edemastage_late no edema were among the most impactful features, reflecting a nuanced change in the disease's manifestation.This is reflected in Figure 9b along with other features that contributed to prediction.

Figure 9c indicates that Stage 3's SHAP values demonstrated a shift in feature importance where edemastage_late no edema and edemastage_early no edema continued to be significant but were accompanied by Prothrombin and N_Days, which could be indicative of more advanced liver dysfunction and its clinical timeline. The progression to Stage 4 showed in Figure 10 marked an increase in the importance of Prothrombin, N_Days, and edemastage_early no edema, alongside Albumin and Hepatomegaly_Y, pointing to the clinical relevance of liver enlargement and protein synthesis capacity in advanced cirrhosis.

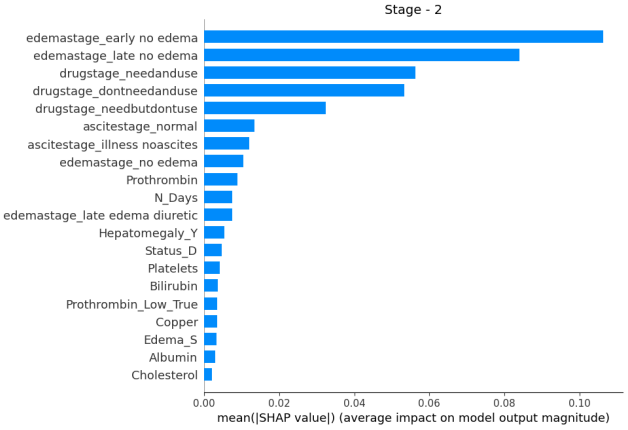
The SHAP summary plots also displayed a trend where certain features such as Platelets and Bilirubin maintained consistent importance across stages, underscoring their critical role in the pathology of Liver Cirrhosis. The features related to drug stages, namely drugstage_dontneedanduse and drugstage_needanduse, were less influential in the later stages, suggesting a potential shift in treatment approach as the disease progresses. The SHAP analysis indicate the features that drive the prediction of Liver Cirrhosis stages in LivMarX. This interpretability is crucial for clinicians to trust and effectively use predictive models in a medical setting.

V. DISCUSSIONS

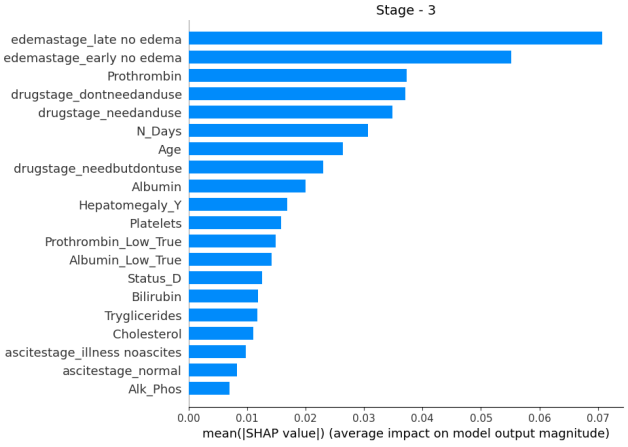
LivMarX as a predictive model for liver cirrhosis in clinical settings shows great potential because of its non-invasive nature, high accuracy, and efficiency. By using readily available clinical and laboratory data, LivMarX can provide valuable insights into the stages of liver cirrhosis, making sure healthcare professionals to make informed decisions



(a) SHAP summary plot for Stage 1 of Liver Cirrhosis.



(b) SHAP summary plot for Stage 2 of Liver Cirrhosis.



(c) SHAP summary plot for Stage 3 of Liver Cirrhosis.

FIGURE 9. **SHAP summary plots across stages 1 - 3.**

regarding patient management and treatment plans. The model's ability to accurately predict the stages of liver cirrhosis encourages early diagnosis and treatment, which can improve patient outcomes by preventing disease progression and reducing the risk of complications. LivMarX's reliance on purely blood samples reduces the need for expensive and complex imaging techniques, making it a cost-effective tool

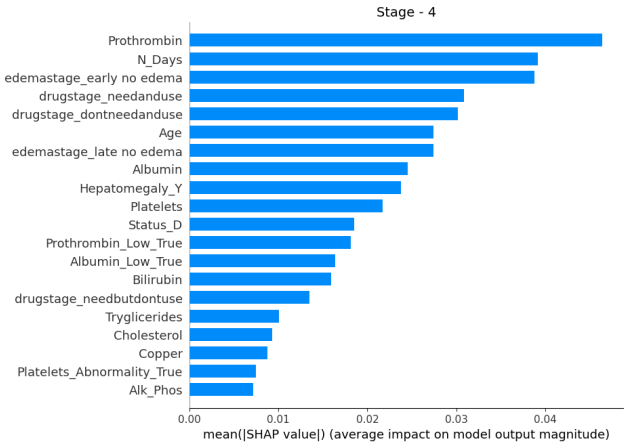


FIGURE 10. **Comparison of model accuracies post optimization.**

for healthcare settings with limited resources. This can also reduce the burden on healthcare facilities by optimizing the use of available resources.

The potential for integrating LivMarX with other diagnostic tools or healthcare technologies gives opportunities to enhance its use in clinical settings. Combining LivMarX with advanced imaging techniques, such as ultrasound elastography or MRI, can provide a better assessment of liver health. These imaging modes can offer more detailed information which when integrated with LivMarX's predictive capabilities, can lead to more accurate staging of liver cirrhosis. Incorporating data from wearable health devices that monitor physical activity and diet can also improve the dataset used by LivMarX, leading to more personalized and dynamic patient monitoring and management.

Integrating LivMarX with telemedicine platforms can help expand its reach and accessibility, particularly in remote or underserved areas. Telemedicine integration allows healthcare providers to remotely monitor patients' liver health using LivMarX's predictions, which will help with timely interventions and continuous care without the need for frequent in-person visits. This is particularly useful for patients with mobility issues or those living in regions with limited access to specialized healthcare services.

There are several challenges that LivMarkX pose that need to be addressed. Integrating the model with existing Electronic Health Records (EHR) systems is important for ensuring easy access to patient data and real-time analysis. This requires collaboration with EHR vendors to develop compatible interfaces and ensure data interoperability. Data privacy and security are also important, and robust encryption methods, secure data storage solutions, and strict access controls must be implemented to protect patient information. While LivMarX has demonstrated high accuracy in a controlled study environment, an extensive clinical validation is necessary to ensure its reliability and effectiveness in real-world settings. Conducting large-scale clinical trials and collaborating with healthcare institutions for pilot studies can provide the necessary validation. Training healthcare

professionals to effectively use LivMarX and interpret its predictions is another critical factor. Training programs and good User Interface can facilitate the adoption of the model in clinical practice. Involving clinicians in the development process can help make sure that the model meets their needs and expectations. The successful implementation of LivMarX in clinical settings can ease further advancements in predictive modelling for liver diseases. Future research can explore the integration of additional data sources, such as genetic information and lifestyle factors, to enhance the model's predictive power.

Comparing LivMarX to exisiting literature, we have outperformed the others with a notable accuracy of 86%. Wang et al. [18] explore the relationship between TCM symptoms and liver cirrhosis severity using data mining techniques and classifiers like logistic regression and Naive-Bayes. Their study emphasizes attribute selection to enhance classification accuracy based on subjective symptom assessments. In contrast, LivMarX uses standardized, non-invasive clinical and laboratory data, which reduces subjectivity and enhances reliability. This approach ensures that the data used is objective and consistent, making LivMarX more generalizable. LivMarX also leverages advanced machine learning algorithms such as Random Forest and Gradient Boosting, which are capable of capturing complex relationships within the data, offering significantly improved predictive accuracy. While Wang et al.'s model depends heavily on TCM expertise and the subjective interpretation of symptoms, LivMarX's methodology is more accessible and cost-effective for widespread clinical use.

Goyal et al. [19] classifies cirrhosis disease using machine learning techniques like polynomial features and XGBoosting, achieving a maximum accuracy of 78% with Random Forest. whereas LivMarX attains a superior accuracy of 86%. This is by utilizing advanced algorithms such as Random Forest and Gradient Boosting, coupled with extensive hyperparameter optimization using methods like Genetic Algorithms, Optuna, and GridSearchCV. LivMarX's reliance on standardized clinical and laboratory data enhances its practicality for widespread clinical use. While Goyal et al.'s model demonstrates robust methods, LivMarX's comprehensive approach offers higher predictive accuracy and reliability.

Bostan and Pantelimon [20] develops an ANN model to diagnose liver cirrhosis using non-invasive laboratory data, achieving a success rate of nearly 90%. The ANN model used inputs such as age, BMI, diabetes status, and various liver enzymes, demonstrating high specificity but moderate sensitivity. In contrast, LivMarX employs advanced algorithms on a dataset with attributes such as Bilirubin, Cholesterol and Copper. While their model focuses on neural networks and specific lab parameters, LivMarX utilizes a broader range of standardized clinical and laboratory data, enhancing its reliability. Additionally, LivMarX's integration with sophisticated optimization methods ensures superior performance.

Images integrated into the model can potentially contribute to more accurate stage predictions. Studies achieving higher accuracy in predicting cirrhosis stages have used image data. Further exploration of multimodal datasets could be a promising avenue for future research, potentially unlocking additional insights and refining predictive models for Liver Cirrhosis diagnosis.

VI. CONCLUSION

In conclusion, this study introduces LivMarX, a novel approach to Liver Cirrhosis prediction, leveraging a dataset with unique strengths. The combination of controlled trial data from 312 patients and real-world follow-up information from an additional 112 patients allowed for diverse insights into broad patient populations. The dataset is public accessible, promoting transparency and thereby enhancing the research's reliability and generalizability in Liver Cirrhosis prediction. Without the need for complex imaging equipment and storage facilities, LivMarX becomes more accessible to healthcare settings with limited resources and offers a cost-effective alternative to traditional imaging approaches. Using novel meticulous feature engineering steps proposed in this study, we uncovered critical connections between patient characteristics and disease stages. Engineered variables were created from the native dataset, categorization of features was performed such as albumin and cholesterol, and categorical combinations representing drug-stage and edema-stage relationships were generated. The study utilized a variety of machine learning algorithms to arrive at a stage prediction. Hyperparameter optimization was employed in refining LivMarX's performance. This optimization step further elevated the accuracy to 86%. LivMarX accuracy surpassed previous studies utilizing the same dataset by over 8%, underscoring the efficacy of the approach.

The findings of this study offer promise in the field of Liver Cirrhosis prediction and demonstrate the potential to significantly aid medical professionals in diagnosing and managing this complex disease. The performance of LivMarX model in accurately pinpointing disease stages, coupled with its efficiency and non-invasive nature, positions it as a valuable tool for improving patient care. This work paves the way for further enhancements, including integrating clinical data for even more robust predictions, and broader applications in clinical settings. This work holds the potential to significantly impact the well-being of patients suffering from Liver Cirrhosis and offer them personalized and timely intervention for better health outcomes.

AUTHOR CONTRIBUTIONS

The conceptualization, methodology, and investigation for this project were collectively carried out by Sudiksha Kottachery Kamath, Sanjeev Kushal Pendekanti, and Divya Rao. Sudiksha Kottachery Kamath, Sanjeev Kushal Pendekanti, and Divya Rao contributed to the formal analysis, resources, and data curation. Sudiksha Kottachery Kamath and Sanjeev Kushal Pendekanti played roles in

software development, while Divya Rao was involved in validation. Writing tasks, including the original draft, were undertaken by Sudiksha Kottachery Kamath and Sanjeev Kushal Pendekanti, while Divya Rao focused on writing, review, and editing. Visualization responsibilities were shared by Sudiksha Kottachery Kamath and Sanjeev Kushal Pendekanti, and Divya Rao Supervision and project administration was carried out by Divya Rao.

COMPETING INTERESTS

The authors confirm that there are no known conflicts of interest associated with this publication and here have been no financial gains for this work that could have influenced its outcome.

FUNDING INFORMATION

This research was conducted independently without external financial support. The authors acknowledge the kind assistance of Manipal Academy of Higher Education, Manipal, in facilitating the publication process.

DATA AVAILABILITY STATEMENT

This study prioritizes reproducibility by utilizing a publicly available dataset and modest computational resources. This study does not involve any human or animal subjects.

REFERENCES

[1] C. Huang and R. Ogawa, "The vascular involvement in soft tissue fibrosis—Lessons learned from pathological scarring," *Int. J. Mol. Sci.*, vol. 21, no. 7, p. 2542, Apr. 2020, doi: 10.3390/ijms21072542.

[2] J. Trebicka, P. Bork, A. Krag, and M. Arumugam, "Utilizing the gut microbiome in decompensated cirrhosis and acute-on-chronic liver failure," *Nature Rev. Gastroenterology Hepatology*, vol. 18, no. 3, pp. 167–180, Nov. 2020, doi: 10.1038/s41575-020-00376-3.

[3] N. V. Bergasa, *Approach to Patient With Liver Disease*. London, U.K.: Springer, 2022, pp. 5–26, doi: 10.1007/978-1-4471-4715-2_2.

[4] D. Q. Huang, N. A. Terrault, F. Tacke, L. L. Gluud, M. Arrese, E. Bugianesi, and R. Loomba, "Global epidemiology of cirrhosis—Aetiology, trends and predictions," *Nature Rev. Gastroenterology Hepatology*, vol. 20, no. 6, pp. 388–398, Jun. 2023.

[5] J. F. Creeden, D. M. Gordon, D. E. Stec, and T. D. Hinds, "Bilirubin as a metabolic hormone: The physiological relevance of low levels," *Amer. J. Physiol.-Endocrinol. Metabolism*, vol. 320, no. 2, pp. E191–E207, Feb. 2021, doi: 10.1152/ajpendo.00405.2020.

[6] P. Ginès, A. Krag, J. G. Abraldes, E. Solà, N. Fabrellas, and P. S. Kamath, "Liver cirrhosis," *Lancet*, vol. 398, no. 10308, pp. 1359–1376, 2021.

[7] L. Fabris and M. Strazzabosco, "Rare and undiagnosed liver diseases: Challenges and opportunities," *Transl. Gastroenterol. Hepatol.*, vol. 6, p. 18, Apr. 2021, doi: 10.21037/tgh-2020-05.

[8] M. Yin, J. Woollard, X. Wang, V. E. Torres, P. C. Harris, C. J. Ward, K. J. Glaser, A. Manduca, and R. L. Ehman, "Quantitative assessment of hepatic fibrosis in an animal model with magnetic resonance elastography," *Magn. Reson. Med.*, vol. 58, no. 2, pp. 346–353, Aug. 2007, doi: 10.1002/mrm.21286.

[9] N. Tanwar and K. F. Rahman, "Machine learning in liver disease diagnosis: Current progress and future opportunities," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 1022, no. 1, Jan. 2021, Art. no. 012029. [Online]. Available: https://iopscience.iop.org/article/10.1088/1757-899X/1022/1/012029

[10] D. W. Crabb, G. Y. Im, G. Szabo, J. L. Mellinger, and M. R. Lucey, "Diagnosis and treatment of alcohol-associated liver diseases: 2019 practice guidance from the American Association for the Study of Liver Diseases," *Hepatology*, vol. 71, no. 1, pp. 306–333, Jan. 2020, doi: 10.1002/hep.30866.

[11] M. P. Diaz-Soto and G. Garcia-Tsao, "Management of varices and variceal hemorrhage in liver cirrhosis: A recent update," *Therapeutic Adv. Gastroenterology*, vol. 15, Jan. 2022, Art. no. 175628482211017, doi: 10.1177/17562848221101712.

[12] M. Serper, T. Bittermann, M. Rossi, D. S. Goldberg, A. M. Thomasson, K. M. Olthoff, and A. Shaked, "Functional status, healthcare utilization, and the costs of liver transplantation," *Amer. J. Transplantation*, vol. 18, no. 5, pp. 1187–1196, May 2018, doi: 10.1111/ajt.14576.

[13] K. Prakash and S. Saradha, "A deep learning approach for classification and prediction of cirrhosis liver: Non alcoholic fatty liver disease (NAFLD)," in *Proc. 6th Int. Conf. Trends Electron. Informat. (ICOEI)*, Apr. 2022, pp. 1277–1284, doi: 10.1109/ICOEI53556.2022.9777239.

[14] Z. Xu, X. Liu, X. E. Cheng, J. L. Song, and J. Q. Zhang, "Diagnosis of cirrhosis stage via deep neural network," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Nov. 2017, pp. 745–749, doi: 10.1109/BIBM.2017.8217748.

[15] S. K. Randhawa, R. K. Sunkaria, and A. K. Bedi, "Prediction of liver cirrhosis using weighted Fisher discriminant ratio algorithm," in *Proc. 1st Int. Conf. Secure Cyber Comput. Commun. (ICSCCC)*, Dec. 2018, pp. 184–187, doi: 10.1109/ICSCCC.2018.8703204.

[16] R. T. Ribeiro, R. T. Marinho, and J. M. Sanches, "Classification and staging of chronic liver disease from multimodal data," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 5, pp. 1336–1344, May 2013, doi: 10.1109/TBME.2012.2235438.

[17] Y. Zhang, Y. Zhang, Y. Zhang, D. Wang, F. Peng, S. Cui, and Z. Yang, "Ultrasonic image fibrosis staging based on machine learning for chronic liver disease," in *Proc. IEEE Int. Conf. Med. Imag. Phys. Eng. (ICMIPE)*, Nov. 2021, pp. 1–5, doi: 10.1109/ICMIPE53131.2021.9698912.

[18] Y. Wang, L. Ma, P. Liu, and X. Liao, "Research on the relationship between symptoms of TCM and stage, class, child-pugh grade about liver cirrhosis," in *Proc. 2nd Int. Conf. Bioinf. Biomed. Eng.*, May 2008, pp. 1968–1970, doi: 10.1109/ICBBE.2008.821.

[19] A. Goyal, A. Zafar, M. Kumar, S. Bharadwaj, B. K. Tejas, and J. Malik, "Cirrhosis disease classification by using polynomial feature and XGBoosting," in *Proc. 13th Int. Conf. Comput. Commun. Netw. Technol. (ICCCNT)*, Oct. 2022, pp. 1–5, doi: 10.1109/ICCCNT54827.2022.9984293.

[20] V. M. Bostan and B. Pantelimon, "Creating a model based on artificial neural network for liver cirrhosis diagnose," in *Proc. 9th Int. Symp. Adv. Topics Electr. Eng. (ATEE)*, May 2015, pp. 295–298, doi: 10.1109/ATEE.2015.7133783.

[21] E. Dickson, P. Grambsch, T. Fleming, L. Fisher, and A. Langworthy, "Cirrhosis patient survival prediction," *UCI Mach. Learn. Repository*, 2023, doi: 10.24432/C5R02G.

[22] M.-S. Kwak, D. Kim, G. E. Chung, S. J. Kang, M. J. Park, Y. J. Kim, J.-H. Yoon, and H.-S. Lee, "Serum bilirubin levels are inversely associated with nonalcoholic fatty liver disease," *Clin. Mol. Hepatol.*, vol. 18, no. 4, p. 383, 2012, doi: 10.3350/cmh.2012.18.4.383.

[23] Cleveland Clinic. (2022). *Cholesterol: Understanding Levels and Numbers*. Accessed: Sep. 16, 2023. [Online]. Available: https://my.clevelandclinic.org/health/articles/11920-cholesterol-numbers-what-do-they-mean

[24] UR Medicine—University of Rochester Medical Center. *Albumin Blood Levels*. Accessed: Sep. 16, 2023. [Online]. Available: https://www.urmc.rochester.edu/encyclopedia/content.aspx?contenttypeid=167&contentid=albumin_blood

[25] M. Stanojevic. (2019). *Copper Deficiency: Causes, Symptoms, Health Effects & Tests—SelfDecode Labs*. Accessed: Sep. 16, 2023. [Online]. Available: https://labs.selfdecode.com/blog/copper-deficiency-blood-test-diseases/

[26] UR Medicine—University of Rochester Medical Center. *Alkaline Phosphatase*. Accessed: Sep. 16, 2023. [Online]. Available: https://www.urmc.rochester.edu/encyclopedia/content.aspx?contenttypeid=167&contentid=alkaline_phosphatase

[27] Mayo Clinic. (2022). *Triglycerides*. Accessed: Sep. 16, 2023. [Online]. Available: https://www.mayoclinic.org/diseases-conditions/high-blood-cholesterol/in-depth/triglycerides/art-20048186

[28] NHLBI, NIH. (Mar. 2022). *Thrombocytopenia*. Accessed: Sep. 16, 2023. [Online]. Available: https://www.nhlbi.nih.gov/health/thrombocytopenia

[29] D. R. Cox, "The regression analysis of binary sequences," *J. Roy. Stat. Soc. B, Stat. Methodol.*, vol. 20, no. 2, pp. 215–232, Jul. 1958.

[30] J. R. Quinlan, "Induction of decision trees," *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, Mar. 1986, doi: 10.1007/bf00116251.

[31] M. Pal, "Random forest classifier for remote sensing classification," *Int. J. Remote Sens.*, vol. 26, no. 1, pp. 217–222, Jan. 2005, doi: 10.1080/01431160412331269698.
[32] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," *Frontiers Neurorobot.*, vol. 7, 2013, doi: 10.3389/fnbot.2013.00021.
[33] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: Unbiased boosting with categorical features," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018.
[34] J. Ali, R. Khan, N. Ahmad, and I. Maqsood, "Random forests and decision trees," *Int. J. Comput. Sci. Issues*, vol. 9, no. 5, p. 272, 2012.
[35] F.-A. Fortin, F.-M. De Rainville, M.-A. G. Gardner, M. Parizeau, and C. Gagné, "DEAP: Evolutionary algorithms made easy," *J. Mach. Lang. Res.*, vol. 13, pp. 2171–2175, Jul. 2012.
[36] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proc. KDD*. New York, NY, USA: Association for Computing Machinery, Jul. 2019, pp. 2623–2631, doi: 10.1145/3292500.3330701.
[37] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Nov. 2011.

SUDIKSHA KOTTACHERY KAMATH is currently pursuing the bachelor's degree in information technology with Manipal Institute of Technology, Manipal, with a focus on a minor in business management. She has a deep interest in machine learning and its implementations in the healthcare industry and aspires to contribute to the integration of information technology and the healthcare sector.

SANJEEV KUSHAL PENDEKANTI is currently pursuing the bachelor's degree in information technology with the Information and Communication Technology Department, Manipal Institute of Technology. He is deeply interested in the field of data mining and machine learning and their applications in the healthcare domain. He aspires to contribute to impactful research in this domain and use technology to make significant improvements to the existing medical sector.

DIVYA RAO received the Ph.D. degree in artificial intelligence with a focus on laryngeal oncology. She is currently a Senior Assistant Professor with the Department of Information and Communication Technology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, with eight years of teaching experience. Her research interests include applied machine learning, natural language processing, healthcare informatics, and human–computer interaction, specifically UI-UX design.