RESEARCH ARTICLE

Explainable AI for Enhanced Interpretation of Liver Cirrhosis Biomarkers

GREESHMA ARYA [1], **ASHISH BAGWARI** [2], **(Senior Member, IEEE), HITESHI SAINI** [1],
PRACHI THAKUR [1], **CIRO RODRIGUEZ** [3], **(Senior Member, IEEE), AND PEDRO LEZAMA** [4]

[1] Department of Electronics and Communication Engineering, Indira Gandhi Delhi Technical University for Women, New Delhi 110006, India
[2] Department of Electronics and Communication Engineering, Uttarakhand Technical University, Dehradun 248007, India
[3] Department of Software Engineering, Universidad Nacional Mayor de San Marcos UNMSM, Lima 15081, Peru
[4] Department of Computer Science, Universidad Nacional Mayor de San Marcos UNMSM, Lima 15081, Peru

Corresponding author: Ciro Rodriguez (crodriguezro@unmsm.edu.pe)

ABSTRACT Liver cirrhosis is a terminal pathological result of chronic liver damage, illicit drugs, hepato-toxicity, and non-alcoholic steatohepatitis. Assessment of liver cirrhosis via non-invasive methods in order to circumvent the limitations of liver biopsy. This research builds upon the growing body of knowledge in liver cirrhosis assessment by examining a dataset comprising cases of primary biliary cirrhosis. Prior research has primarily concentrated on comparative analyses and development of machine learning models integrating imaging modalities such as ultrasound, magnetic resonance imaging (MRI), and elastography, with limited focus on serum biomarkers. This research endeavors to address the neglected aspects of liver cirrhosis assessment by leveraging Explainable AI algorithm to bridge the gap between AI models and human comprehension via providing insights into the intricate decision-making process of the proposed machine learning model, thus enhancing transparency and trustworthiness. This novel approach aims to overcome the limitations of previous works and contribute to improved liver cirrhosis diagnosis.

INDEX TERMS Explainable AI, extreme gradient boosting, feature visualization, invasive technique, liver cirrhosis, shapley additive explanations, shapley feature importance.

I. INTRODUCTION

The histological development of regenerative nodules surrounded by fibrous bands defines cirrhosis, a reaction to persistent liver injury [1]. Viral hepatitis, continued alcohol misuse, and non-alcoholic fatty liver disease (NAFLD) are the major causes of this disorder, a prominent cause of mortality and morbidity globally.

The onset of cirrhosis may be silent and asymptomatic for considerable years. However, symptoms like jaundice, edema, spider navy, Hepatomegaly, Hematemesis, and further malignant concerns include liver failure, portal hypertension, and hepatocellular carcinoma. Portal hypertension is a pathological state characterized by the anomalous elevation of blood pressure in the portal venous system, which elicits the development of varices, namely ectatic and aberrant blood vessels that may give rise to lethal haemorrhage,

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

notably in the esophageal and gastric domains. Hepatocellular carcinoma, a type of malignant neoplasm affecting liver parenchyma, could be a result of underlying cirrhotic condition. It is of utmost significance to underscore that cirrhosis constitutes a substantial risk factor for the emergence of hepatocellular carcinoma. Therefore, effective management of portal hypertension and cirrhosis is crucial to avert the progression to hepatocellular carcinoma.

Regional differences exist in the etiology of liver cirrhosis. However, chronic hepatitis B is the primary culprit in the Asia Pacific region [2]; virus infection, chronic hepatitis C, alcohol use, and NAFLD are the most prevalent causes in Western countries.

The diagnosis of cirrhosis demonstrates a higher risk of mortality and morbidity. The standard for cirrhosis diagnosis and fibrosis stage is liver biopsy. However, there are disadvantages to this invasive technique. It involves extracting a core of liver tissue under local anesthetic for pathologic examination. Several grading methods are available to categorize

the level of fibrosis in biopsy specimens. The METAVIR, Scheuer, Batts-Ludwig, and Ishak scores are a few of them. The most popular scores are METAVIR and Ishak [3].

Nevertheless, various variables affect how accurately a liver biopsy can stage fibrosis. For instance, the length of the biopsy specimen is related to diagnostic accuracy [4]. The research found that the specimen sample impacted the accuracy rate; an increase of 10 mm in the liver specimen might result in a 10% variation in the outcome [5].

To circumvent the constraints of liver biopsy, numerous non-invasive methods have been developed to assess liver fibrosis, including radiological procedures based on ultrasound, magnetic resonance imaging (MRI), and elastography Non-invasive methods have a lower risk of complications, and can be repeated frequently, which is useful for monitoring patients with liver disease over time.

Accessing the severity of cirrhosis utilizing cirrhosis-related serum-based indicators plays a pivotal role. These indicators are instrumental in assessing the extent of the condition. Additionally, these serum-based indicators find application in the training of AI models by implementing machine learning algorithms. While these machine learning models have demonstrated commendable predictive accuracy, their inherent inability to elucidate the rationale behind their decision-making procedures characterizes them as black box models.

In healthcare, gaining insights into AI models' decision-making processes is paramount. This insight facilitates a deeper understanding of how these models arrive at their conclusions, which is particularly critical for decision-making in clinical settings. To bridge the gap between AI models and human comprehension, the development of Explainable AI (XAI) emerges as a valuable tool. XAI serves as a mechanism to unravel and comprehend the intricate decision-making processes of these models, thereby enhancing their interpretability and fostering trust in their applications within the healthcare domain. This insight facilitates a deeper understanding of how these models arrive at their conclusions, which is particularly critical for decision-making in clinical settings.

The key focal points of this study encompass:
- Facilitating a connection between AI models and human comprehension within the healthcare domain instils trust and transparency in model outputs.
- Pioneering the adoption of XAI to augment the interpretability of biomarkers associated with liver cirrhosis.
- Harnessing the potential of biomarkers for the precise identification of early-stage liver cirrhosis and aiding clinicians in pinpointing the root cause.
- Conducting a thorough examination of biomarkers and employing various machine-learning methodologies for comprehensive analysis.

The subsequent sections are organized as follows: Section II explores existing models and related works. Section III represents an exposition of feature analysis and machine learning algorithms. Section IV introduces the proposed model and elucidates the workflow, offering insights into the conceptualization and operationalization of the research framework. Sections V delve into the model's confidence and provide a comparative analysis. Finally, Section V-D serves as the conclusion, summarizing key findings and outlining directions for future research.

II. RELATED WORKS

Conventional methods for detecting liver cirrhosis, such as liver biopsy and blood tests like liver function tests, play a crucial role in diagnosing this condition. Liver biopsy involves the microscopic examination of liver tissue to assess fibrosis and cirrhosis levels, and its success relies on the expertise of surgeons and the sample of tissue taken from the liver. [6] Concurrently, blood tests like liver function tests quantify enzyme and protein levels in the blood to indicate liver damage or dysfunction.

Additionally, imaging modalities such as ultrasound, computed tomography (CT) scans, and magnetic resonance imaging (MRI) offer visual insights into the liver's condition, facilitating the identification of cirrhosis-related indicators, including liver nodules, scarring, and alterations in size and shape [7], [8], [9].

On the other hand, machine learning approaches present a compelling alternative for early liver cirrhosis detection. These algorithms possess the capacity to analyze extensive datasets and discern intricate patterns that may elude conventional methodologies, thereby enabling the timely identification of liver cirrhosis [10], [11]. Through training on diverse datasets, machine learning models acquire the ability to recognize subtle features and variations associated with early-stage cirrhosis, resulting in heightened accuracy and sensitivity. Furthermore, integrating machine learning with non-invasive imaging techniques, such as ultrasound shear wave elastography (SWE), synergistically enhances the precision and efficiency of liver fibrosis diagnosis. Thus, adopting machine learning in this context offers a promising avenue for improved early detection and diagnosis of liver cirrhosis, addressing critical healthcare needs.

Hyun et al. [12] have introduced a sophisticated machine-learning approach to accurately identify and classify histological subtypes of lung cancer. However, the retrospective nature of the work and the lack of external validation may affect the findings' generalizability. Despite internal validation, it is imperative to conduct external validation using a larger sample size to confirm the robustness and reliability of the results.

In a complementary effort, Sato et al. [13] have proposed a predictive machine learning model for diagnosing hepatocellular carcinoma, leveraging state-of-the-art techniques and achieving an accuracy of 87.34%. The study also examined the importance of the variables in the optimal predictive model, employing gradient boosting as the methodology for analysis.

Several studies have explored the potential of Electronic Medical Record (EMR) based algorithms for diagnosing

NAFLD and have reported promising results. For instance, Corey et al. [14] proposed an EMR-based algorithm for identifying NAFLD patients. They demonstrated that unsupervised clustering can be utilized to identify clinically relevant disease subtypes with distinct patterns of adverse outcomes. This finding suggests that EMR-based algorithms are promising to improve the diagnosis and management of NAFLD, which is becoming increasingly prevalent worldwide.

Kuzhippallil et al. [15] analyzed several machine-learning models using various metrics. Prior to feature selection and outlier elimination, the performance metrics of several classification algorithms were collected. The feature selection process was carried out through a combination of Genetic Algorithm and XGBoost classifier, while outlier elimination was accomplished using isolation forest. This approach allowed for identifying the most relevant features and removing data points that could potentially skew the results, thus improving the accuracy and reliability of the classification models.

The study conducted by [16] Wilson disease (WD), a rare autosomal recessive disorder arising from ATP7B gene mutations. WD has substantial implications for the clinical management and prognosis of affected individuals, primarily due to the looming risk of liver cirrhosis. In the quest to identify biomarkers for predicting cirrhosis in WD patients, particular attention has been given to the potential of routine blood examinations. Utilizing a dataset comprising comprehensive clinical information, including blood routine examination, urine copper levels, and serum ceruloplasmin measurements, a meticulous machine learning model was constructed to prognosticate the likelihood of cirrhosis occurrence among 346 WD patients, of whom 246 remained cirrhosis-free. Key predictors for cirrhosis were elucidated, including Platelet Large Cell Count (P-LCC), Red Cell Distribution Width - Coefficient of Variation (RDW-CV), serum ceruloplasmin levels, age at diagnosis, and mean corpuscular volume (MCV).

Nevertheless, it is essential to acknowledge that the study's reliance on a case-control design may restrict the broader applicability of the findings. Additionally, while the predictive model adeptly incorporates crucial general clinical variables, it may inadvertently overlook other viable predictors. Despite its commendable performance during the training phase, the model exhibited relatively diminished robustness in the testing phase (accuracy = 0.76), emphasizing the necessity for subsequent validation and refinement efforts to bolster its reliability and applicability.

In consonance with previous investigations, the study by [17] primarily focused on predicting fatty liver disease (FLD) within the Taiwanese populace, employing a range of machine-learning algorithms. Notably, the XGBoost model emerged as the frontrunner, demonstrating the highest performance compared to alternative algorithms such as neural networks, logistic regression, random forest, and support vector machines. [18].

Building on previous approaches, our approach emphasizes the incorporating an extended set of clinical features in liver cirrhosis detection. Leveraging a comprehensive array of clinical attributes, we aim to enhance the accuracy and depth of our model's learning. In this context, the selection of the XGBoost model, recognized for its vital performance metrics, aligns seamlessly with our objectives. However, our study takes a step further by recognizing the importance of Explainable Artificial Intelligence (XAI). By integrating XAI techniques, we aim to establish trust and foster transparency in bridging the interpretability gap between model's outputs and human understanding, thereby advancing liver cirrhosis detection for improved patient outcomes.

III. MATERIALS AND METHODS
A. CLINICAL FEATURES
The datasets utilized for this experiment were sourced from Kaggle. The initial dataset for the primary biliary cirrhosis (PBC) study conducted at the Mayo Clinic served as the foundation [19]. The dataset comprises 424 PBC patients, with a gender distribution of 89% female and 11% male. Initially, there were 18 clinical variables in this dataset; however, this study focused on 14 specific parameters and 1 parameter representing the target class.

The dataset encompasses two distinct categories of parameters: numerical values like albumin, bilirubin, copper, Cholesterol, aspartate aminotransferase (SGOT), alkaline phosphatase, platelets, triglycerides, and prothrombin, categorical parameters like sex, ascites, hepatomegaly, spiders, and edema (see Table 1). These data have been subjected to comprehensive analysis, with further elaboration provided in subsequent sections, considering the nature of each parameter.

TABLE 1. Numerical features of the dataset.

Features	$\bar{m} \pm \sigma^a$
Bilirubin	3.22 ± 4.42
Cholesterol	369.51 ± 194.46
Albumin	3.49 ± 0.49
Copper	97.64 ± 74.99
Alkaline Phosphatase	1982.65 ± 18887.61
Aspartate Aminotransferase (SGOT)	122.55 ± 49.43
Triglycerides	124.70 ± 54.42
Platelets	257.02 ± 94.46
Prothrombin	10.73 ± 1.02

The dataset comprises four distinct phases corresponding to various levels of cirrhosis progression, providing a comprehensive representation of liver health states. These stages are classified as follows:

1. 'Healthy': This stage indicates individuals with normal liver health, showing no signs of damage or scarring.

2. 'Fatty liver': In this stage, individuals are in the early phases of liver damage, often associated with the accumulation of fat in the liver.
3. 'Fibrosis': At this intermediate stage, individuals exhibit significant liver scarring, indicating progressing liver damage.
4. 'Cirrhosis': The fourth stage represents advanced and often irreversible liver damage, posing significant challenges for effective treatment and recovery.

It is important to emphasize that 'Cirrhosis', the fourth stage, represents a condition that is challenging to treat effectively. Detailed statistics for each of these classes are illustrated in Fig. 1, providing valuable insights into the dataset's composition.
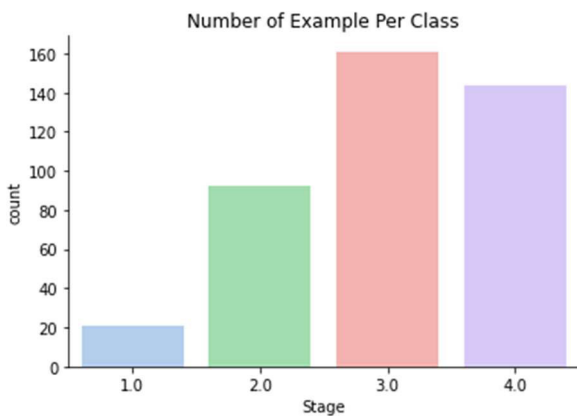
FIGURE 1. Different stages of liver cirrhosis.

B. FEATURE ANALYSIS

Ascites is a medical term that describes the accumulation of fluid in the abdominal cavity. The presence of ascites is a significant indicator of chronic liver failure, and it is associated with a grim prognosis, carrying a 3-year mortality rate of 50%. Patients with ascites are prone to complications such as spontaneous bacterial peritonitis, hyponatremia, and worsening renal function, which are frequently observed in individuals with ascites. Fig. 2 represents the number of cases of ascites across different stages.

Hepatomegaly, also known for an enlarged liver. The underlying symptoms of hepatomegaly are feeling heavy on the right side of the body or discomfort, fatigue, change in bowel habits, and unexplained weight loss. Hepatomegaly is one of the primary markers for diagnosing the liver since it is more susceptible to developing as the liver deteriorates (see Fig. 3).

Bilirubin is produced during the breakdown of hemoglobin when old or damaged red blood cells are recycled. Elevated levels of bilirubin can serve as an indicator of various liver or bile duct issues (as depicted in Fig. 4). On occasion, increased bilirubin levels may be associated with heightened oxidation of red blood cells. Symptoms that may accompany elevated bilirubin levels include stomach discomfort, nausea,
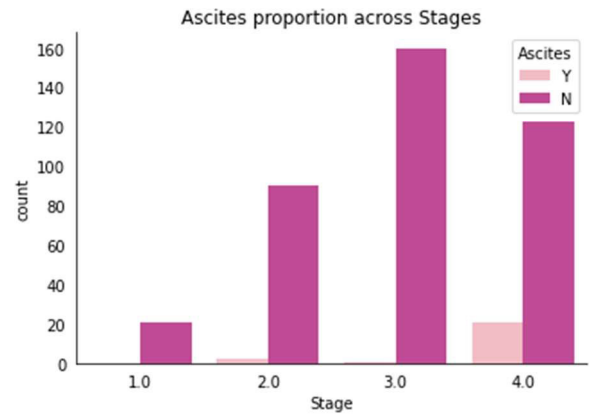

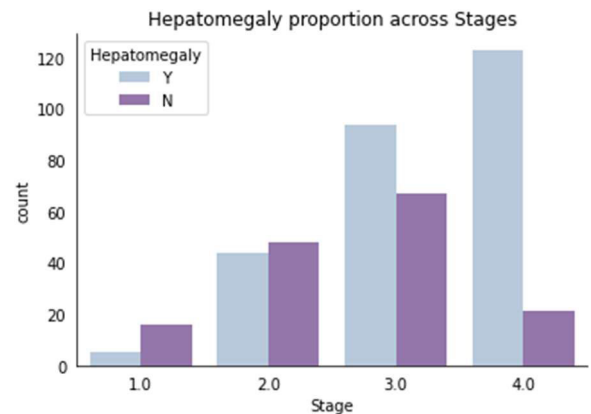**FIGURE 2.** Ascites proportion across different stages.

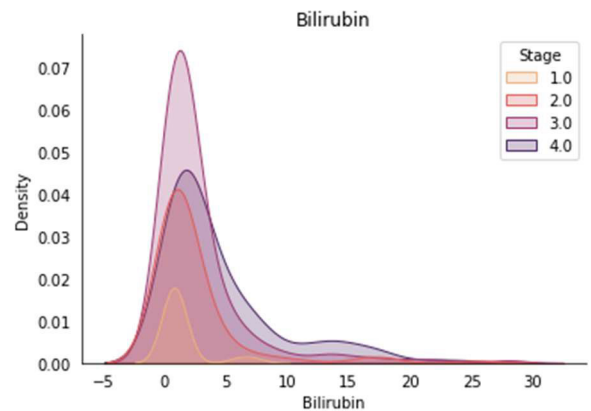**FIGURE 3.** Hepatomegaly proportion across stages.

FIGURE 4. Bilirubin [mg/dl] across different stages.

vomiting, and the presence of small visible blood vessels known as spider angiomas.

Patients diagnosed with liver cirrhosis can have visible spider angiomas. However, it is not just associated with cirrhosis; people taking oral estrogen or expecting mothers also develop visible spiders. One research indicated that 27% of non-cirrhotic individuals and 50% of cirrhotic patients had spider naevi [20]. Only 20% of the individuals in the dataset

shown in Fig. 5 developed liver scarring and had visible spider angiomas.
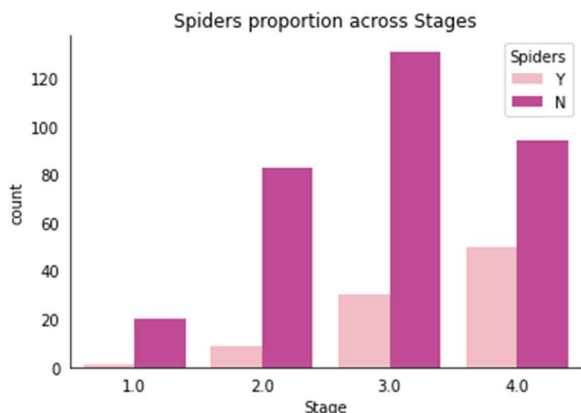

FIGURE 5. Patients with Spider angiomas across different stages.

Fibrosis slows down the blood flow through the liver, resulting in increased pressure in the vein, which is responsible for returning blood to the liver from the intestines and spleen. This elevated pressure in portal veins causes the fluid to accumulate in the legs (edema), causing the legs to swell. In Fig. 6, N represents the presence of no edema, and S represents edema without diuretic or edema resolved by diuretic, which can be observed increasing throughout the different level stages. In contrast, Y represents edema despite diuretic therapy's significant increase in this category, which can be seen throughout the stages.
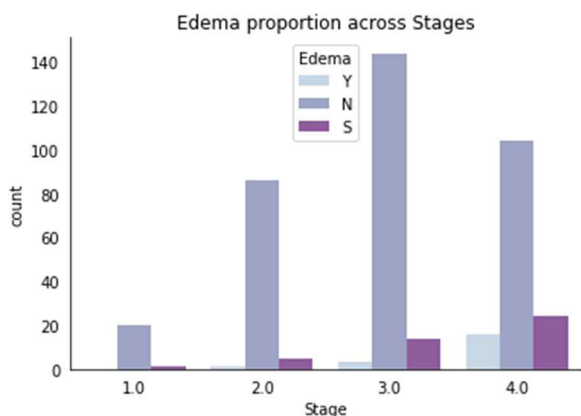

FIGURE 6. Patients with edema across different stages.

Patients diagnosed with liver cirrhosis experience a significantly reduced 5-year survival rate when their blood albumin level is low (less than 3.5g/dL). In cases of liver cirrhosis, the average annual decline in blood albumin levels is approximately 0.15g/dL [21]. Conventionally, patients were prescribed a high-protein diet to elevate their blood albumin levels to 3.5g/dL or higher. However, when the condition has progressed to the level of decompensated cirrhosis, a high-protein diet and concomitant rise in blood ammonia value may raise the chance of developing hepatic encephalopathy

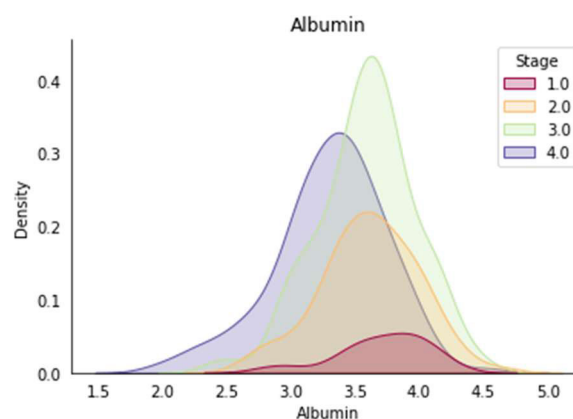(see Fig. 7). In such situations, a low-protein diet is recommended as the more optimal dietary approach.

FIGURE 7. Albumin [g/dl] across different stages of liver damage.

Adipose tissue and blood glucose can be exchanged between the liver and the bloodstream in both directions when triglycerides are present in the blood. Frequently, fatty liver does not exhibit symptoms, but when left untreated, it can lead to cirrhosis and chronic liver damage. However, the serum triglyceride level (as depicted in Fig. 8) appears to exhibit a distinct correlation with alcoholic liver cirrhosis and serves as an indicator of the severity of liver damage. In cases of non-alcoholic cirrhosis, serum cholesterol levels (as seen in Fig. 9) may function as markers for assessing the extent of liver damage.
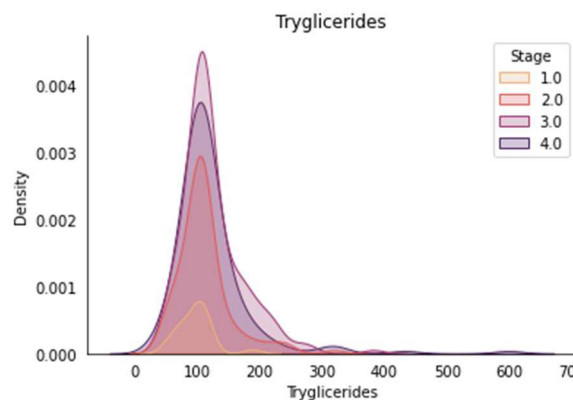

FIGURE 8. Triglyceride [mg/dl] across different stages.

Prothrombin plays a crucial role in facilitating blood clotting. One method for assessing blood clotting time is by measuring 'prothrombin time' (PT), which is expressed in seconds. When PT is prolonged (as illustrated in Fig. 10), it indicates slower blood clotting. This is often attributed to the liver producing fewer blood clotting proteins, leading to an extended clotting process. Elevated PT levels are typically indicative of cirrhosis or severe liver disease. Prothrombin time is closely associated with the pathological liver fibrosis score, making it a direct and non-invasive marker for detecting cirrhosis.
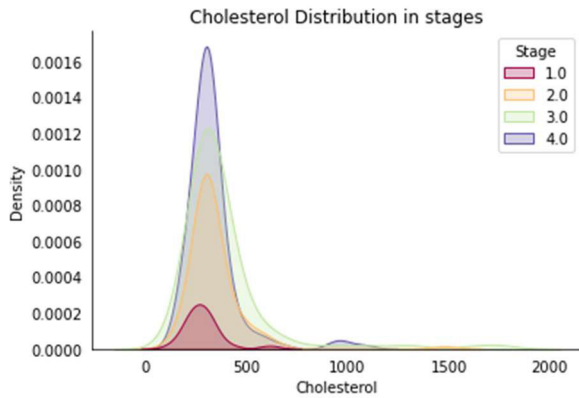
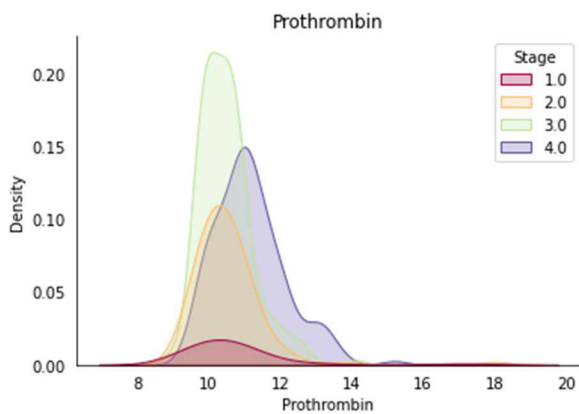**FIGURE 9.** Cholesterol [mg/dl] across different stages.

FIGURE 10. Prothrombin time in seconds [s] across different stages.

Due to their involvement in numerous metabolic pathways, essential trace elements play critical roles in maintaining health. Since the liver controls a significant proportion of the metabolism of trace elements, liver function problems can contribute to a wide range of metabolic disorders. Chronic ingestion of copper in excessive amounts can cause copper overdose, resulting in acute liver damage (see Fig. 11).
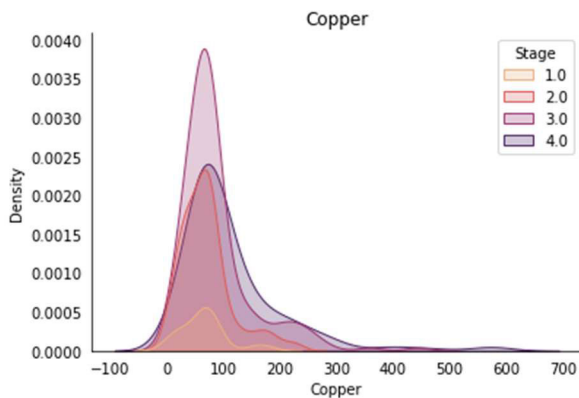

FIGURE 11. Urine copper level [ug/day] across different stages.

In as many as 76% of cirrhotic patients, Chronic liver disease (CLD) frequently leads to the complication of thrombocytopenia (platelet counts of 150,000/L) [22] (see Fig. 12).
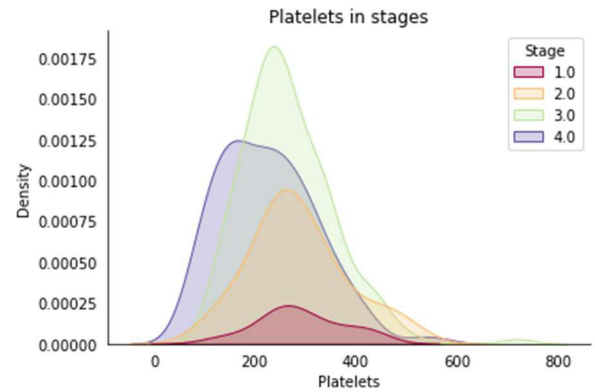

FIGURE 12. Platelets per cubic [ml/1000] across different stages.

Mild thrombocytopenia (>75,000/L to 150,000/L) has little clinical significance and typically has no impact on management or treatment choices. About 13% of patients with cirrhosis have moderate thrombocytopenia (50,000/L-75,000/L). Patients with advanced liver cancer, disease, immune thrombocytopenic purpura (ITP) [23], chronic hepatitis C virus (HCV) infection, and other diseases are frequently complicated by severe thrombocytopenia (50,000/L), which is associated with significant morbidity.

Aminotransferases are among the liver enzymes known for their high sensitivity. Aspartate aminotransferase (AST or SGOT) and alanine aminotransferase (ALT or SGPT) are primarily present in liver cells and, to a lesser extent, in muscle cells. Another crucial liver enzyme often examined is alkaline phosphatase, which is located in the walls of the bile duct. Elevated levels of alkaline phosphatase may indicate damage to the biliary cells. In cases of liver damage, these enzymes are released from liver cells into the bloodstream, resulting in increased serum levels of AST and ALT enzymes, which are indicative of liver disease. It is important to note that elevated AST levels can also be caused by damage to other tissues, making it a less specific marker for liver injury (e.g., in the case of a heart attack).

Increased levels of AST (as seen in Fig. 13), ALT, or ALP (as depicted in Fig. 14) activity are indicative of liver and kidney damage, often associated with lipid peroxidation following the generation of free radicals.

C. MACHINE LEARNING TECHNIQUES
Machine learning algorithms, specifically logistic regression and XGBoost, played a pivotal role in the experiment. An Explainable AI algorithm known as SHapley Additive exPlanations (SHAP) is for in-depth analysis to enhance the understanding of predictive class outcomes.

1) LOGISTIC REGRESSION
Logistic regression estimates the probability of an event occurring, such as voting in favor of cirrhotic and non-cirrhotic liver [24]. Since the outcome is a probability, the dependent variable is bounded between '0' and '1'. In logistic regression, a logit transformation is applied to the
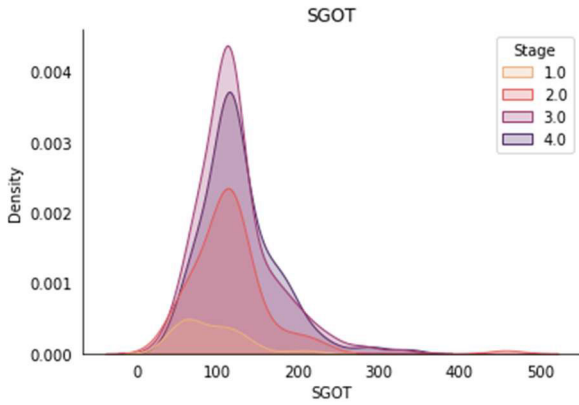
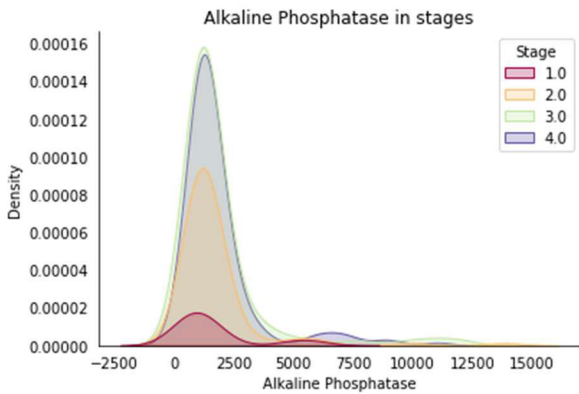**FIGURE 13.** SGOT levels [U/ml] across different stages.

FIGURE 14. Alkaline phosphatase level [U/liter] across different stages.

odds, refer (1).

$$\text{Logit}(pi) = \frac{1}{1 + \exp(-pi)} \tag{1}$$

$$\ln\left(\frac{pi}{1 - pi}\right) = B_0 + B_1 X_1 + \ldots + B_m X_m \tag{2}$$

Equations (1) and (2), *logit(pi)* is the dependent variable, and X is the independent variable. The B parameter is estimated via maximum likelihood estimation (MLE) by testing different values of B through multiple iterations to optimize for the best fit of log odds [25].

2) XGBOOST

XGBoost is a scalable machine learning system for tree boosting, written in C++, which optimizes the training of Gradient Boosting. The tree-boosting model is developed for a sample space S with b samples and d features by adding the scores from each decision tree:

$$\hat{y}_i = \sum_{m=1}^{M} g_k(x_i), g_k \epsilon E \tag{3}$$

where, E is the space of regression trees, g_k corresponds to a tree structure 't' leaf weight 'w'.

The objective function for the above model is given by:

$$obj(\emptyset) = \sum_{j=1}^{b} l(y_i, \hat{y}_i) + \sum_{m=1}^{M} \omega(f_m) \tag{4}$$

where, $S = \{(x_j, y_j)\}$, M is the number of regression trees.

E is a function map from the sample point to the fraction. $l(y_j, \hat{y}_i)$ represents a convex and differentiable loss function utilized for gauging the dissimilarity between the target, y_j, $\hat{y}_i$. Additionally, $\omega(f_m)$ serves as a regularization term employed to prevent over-fitting [26].

Applying additive strategy, minimizing the loss. The model is summarized in (5), (6), (7), and (8):

$$\widehat{y_j}^{(0)} = 0, \tag{5}$$

$$\widehat{y_j}^{(1)} = g_1(x_j) = \widehat{y_j}^{(0)} + g_1(x_j), \tag{6}$$

$$\widehat{y_j}^{(2)} = g_1(x_j) + g(x_j) = \widehat{y_j}^{(1)} + g_2(x_j), \tag{7}$$

$$\widehat{y_j}^{(s)} = \sum_{m=1}^{s} g_m(x_j) = \widehat{y_i}^{(s-1)} + g_s(x_j) \tag{8}$$

3) SHAP

SHAP is a game theoretic approach to explain the output of any machine learning model by calculating shapely values [27]. Shapely values explain prediction for every instance as the sum of contributions from its feature values, refer (9).

$$\Phi_i = \sum_{T \subseteq K \setminus \{i\}} \frac{|T|!(Q - |T| - 1)!}{Q!} [F_X(T \cup \{i\}) - F_X(T)] \tag{9}$$

where, T = subset of features excluding i^{th} features,
$|Q|$ = total features,
$|T|$ = subset size,
$F_x()$ = prediction function of subset T.

Equation (8), $|T|!$ represents the permutations of the feature values before the ith feature. Similarly, $(Q -|T|-1)!$ represents the feature value after the ith feature.

The computational complexity of the algorithm is $O(LS2^M)$.

Where, S is the number of trees,
L is the maximum number of leaves among S trees,
M is a number of features.

In [28], authors propose a modified version of this algorithm that keeps track of the number of subsets 'T' that flow into each tree node. A slight modification in the algorithm has the computational complexity as $O(LSd^2)$.

Where, d is the maximum depth of the tree.

Considering pairwise feature attributions and computing interaction between every combination of pairs of features will lead to a matrix of attribution values representing their impact on the model. This interaction effect is computed from the Shapley interaction index from game theory and is given in (9),

$$\emptyset_{i,j} = \sum_{T \subseteq K\{i\}} \frac{|T|!(Q - |T| - 1)!}{2(Q - 1)!} \nabla_{ij}(T), \tag{10}$$

where,

$$\nabla_{ij}(T) = f_x(T \cup (i, j) - f_x(T \cup \{i\}) - f_x(T \cup \{j\}) + f_x(T), \tag{11}$$

$$\nabla_{ij}(T) = f_x(T \cup (i, j) - f_x(T \cup \{j\}) - f_x(T \cup \{i\}) - f_x(T) \tag{12}$$

Equations (10), (11), and (12) illustrate that the SHAP interaction value pertaining to the ith feature in relation to the jth feature can be interpreted as the difference between the Shapely values of the ith feature considering and not considering the presence of the jth feature. This enables the use of Shapely values for computing SHAP interaction values.

The SHAP interaction effect between the ith and j^{th} features is evenly divided ($\emptyset_{i,j} = \emptyset_{j,i}$), resulting in a total interaction effect. Consequently, the primary effect influencing the prediction outcome is the disparity between the Shapely values and the sum of SHAP interaction values for a given feature (13).

$$\emptyset_{i,j} = \Phi_i - \sum_{j \neq i} \emptyset_{i,j} \qquad (13)$$

IV. PROPOSED METHODOLOGY

This study proposes a machine learning method for the early detection of liver cirrhosis, conducting a comprehensive and empirical analysis. The experiment consists of five steps, as illustrated in Fig. 15. After acquiring the data, data processing procedures are executed as follows:

1) DATA PRE-PROCESSING

The dataset initially contained missing values, required preprocessing. The data comprised three distinct datatypes: 'float', 'int', and 'object'. Therefore, each data type required unique pre-processing. For 'object' datatypes, missing values were imputed with the mode of the respective column. 'int' and 'float' data types had missing values replaced with the median of their respective columns.

2) CATEGORICAL DATA TRANSFORMATION

To facilitate a thorough analysis, categorical data in the dataset were transformed into integer form while preserving normalization. For instance, the 'Sex' column was encoded as '0' for 'M' (Male) and '1' for 'F' (Female). Similarly, the presence of 'Ascites' was encoded as '1', while its absence was encoded as '0'. The presence of 'Hepatomegaly' and 'Spiders' were also encoded as '1', while their absence was encoded as '0'. In the case of 'Edema', since there were three different values, 'N' (no edema), 'Y' (edema with diuretic), and 'S' (edema with no diuretic), they were replaced with '0', '1', and '-1', respectively.

3) FEATURE SELECTION

The 'Status' and 'N_days' features were removed to prevent data leakage during model training.

4) CLASS BALANCING

To ensure the balance of all classes, oversampling was employed on the minority classes. The presence of imbalanced classes in the dataset prompted the use of oversampling technique to balance all the classes.

5) CROSS-VALIDATION

Stratified 20-Fold cross validation was employed to mitigate the impact of uneven class distribution. This method provides train/test indices and ensures that the percentage of each class in the dataset is preserved in each fold.

The utilization of 20-fold cross-validation can lead to improved model performance by reducing overfitting.

1) Minimized the overfitting impact.
2) Improved the reliability of the findings.

Algorithm 1 outlines the proposed approach to enhance biomarkers analysis, with primary focus on delivering predictions and providing graphical insights into feature contributions. During training, hyperparameters ($Learning_{rate}$, Max_{depth}, $Random_{state}$, and $Gamma$) facilitate iterative optimization of the XGBoost model. This optimization minimizes the loss function, resulting in superior model accuracy. Subsequently, the best predictions are obtained using the trained model. Computing Shapely values, refer (9), quantifies the impact of each feature in the test data, offering valuable insights into feature importance. Additionally, graphical representations are generated to enhance dataset comprehension.

Algorithm 1 SHAP – Enhanced Cross-Validated XGBoost Algorithm

---

INPUT: Dataset containing serum information
OUTPUT: Dataset containing *prediction* and graphical representation of feature's contribution.

1. **for** *fold_no=1* to *fold_no <=20* **do**
2.   Generate cross-validation splits *train_index* and *test_index* randomly and store labels in *X_train* and *X_test*.
3.   For *fold_no* **train**
4.   **Define** hyperparameters :
5. $Learning_{rate} = 0.3$
6. $Max_{depth} = 2$
7. $Random_{state} = 1$
8. $Gamma = 0$
9. $M(\Theta) \leftarrow argmin_{\Theta}\left(\frac{1}{N_{train}}\sum_{i=1}^{N_{train}} LM(\Theta), X_{train}^i, Y_{train}^i\right)$
10. $M\rightarrow$ represents XGBoost model,
11. $\Theta \rightarrow$ represents model's parameters (including weights and biases),
12. $L\rightarrow$ loss function.
13. Display *accuracy*.
14. **End for**
15. Generate the best *prediction* using model M.
16. **Compute** Φ_i for each case in *X_test*:
17. $\Phi_i = \sum_{T \subseteq K \setminus \{i\}} \frac{|T|!(Q-|T|-1)!}{Q!} [F_X(T \cup \{i\}) - F_X(T)]$
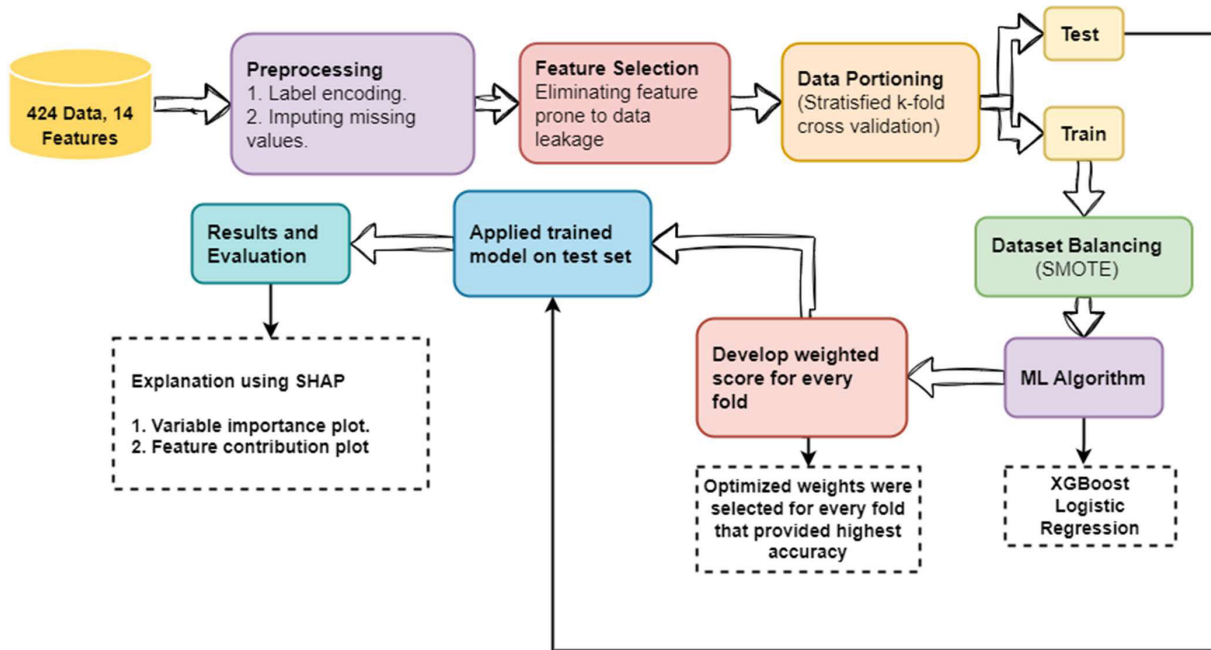18.   Generate graphical representation.
19. **Exit**

FIGURE 15. Workflow of the proposed algorithm.

V. RESULTS AND ANALYSIS

A. HARDWARE SPECIFICATIONS

For research and stimulation process, Python programming language and essential libraries such as Numpy, Pandas, and Scikit-learn. Computational resources including an Intel(R) Core (TM) i5-7200U CPU @ 2.50GHz and 16 GB of RAM, complemented by Kaggle's GPU for model training. The initial phase involved constructing a foundational model using a logistic regression classifier. Subsequently, in the second experiment employed an XGBoost classifier, incorporating Explainable Artificial Intelligence (XAI) techniques to enhance model transparency and interpretability.

B. COMPLEXITY ANALYSIS

The line-by-line analysis was conducted to examine the computational complexities inherent in the proposed algorithm. Stratified k-Fold cross-validation with the *n_splits* parameter set to 20 yields a linear time complexity of *O(n_splits)* (refer Algorithm 1). Within each cross-validation fold, the primary operation entailed training an XGBoost model, with its complexity primarily hinging on the number of training samples (*N_train*), the count of features (14 features), and the number of estimators in the XGBoost model (*num_estimators*). The complexity is approximately *O(N_train. × 2. × 14. × num_estimators)*. Accuracy scores were recorded and tabulated in Table 2 at specific intervals for both the proposed and base models. Other operations, such as hyperparameter configuration, accuracy display, best prediction generation, and graphical representation, exhibit constant time complexities *(O(1))*. The computation of Shapely values, reflecting feature impact, is contingent on the size of the test dataset (*N_test*) and the specific employed method, varying from

**TABLE 2.** Comparison of accuracy between base and proposed model.

	Logistic Regression (Base model)	XGBoost (Proposed Model)
5th fold	51	74
10th fold	58	78
15th fold	59	77.3
20th fold	61	78

(O(1)) to O($N\_test. \times Q$), where Q represents the complexity of *TreeExplainer()* (refer Eq. 10) (discussed in Section III-C).

The evaluation metrics, including accuracy, recall, precision, and F1-score were computed for both models, and the results were compared and presented in Table 3. The precision, recall, and F1-score were determined individually for each of the four classes, and the average value was calculated.

TABLE 3. Evaluation metrics.

Evaluation Parameter	Logistic Regression (Base Model) (In percentage)	XGBoost (Proposed Model) (In percentage)
Accuracy	68	90.5
Recall	51	87
Precision	59	85
F1-score	67	89.9
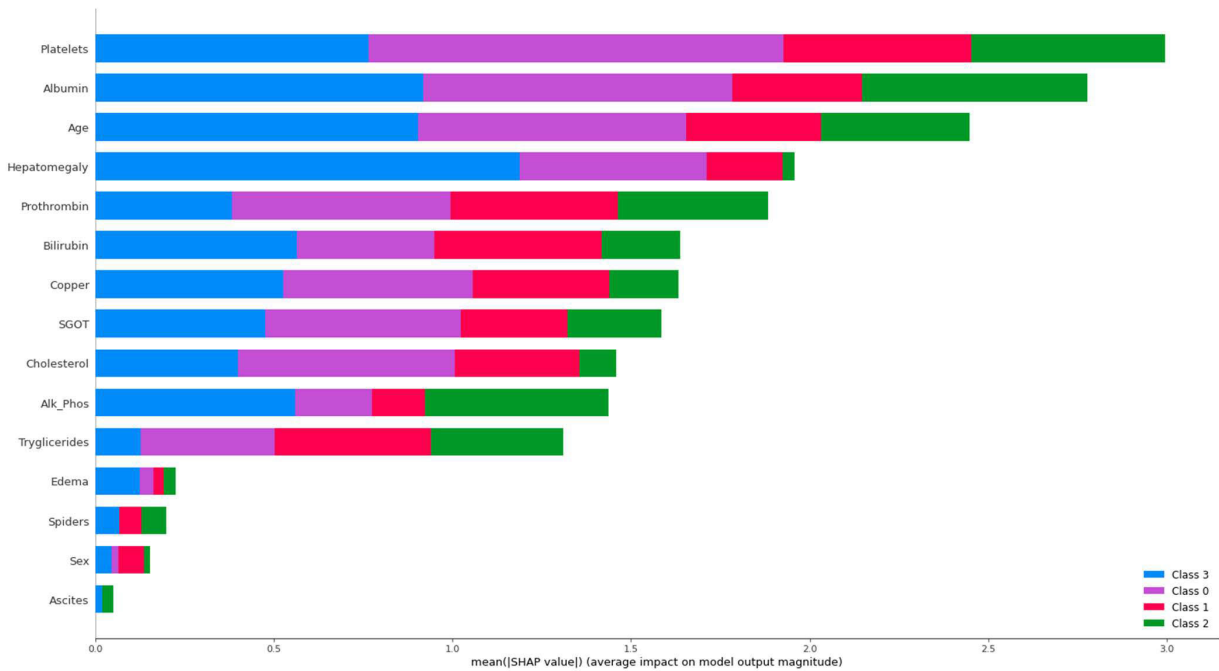
FIGURE 16. Variable importance plot. Features listed at the top (Platelets) contribute most to the model and have high predictive power. However, features listed at the bottom have less predictive power (contribute less to the model).
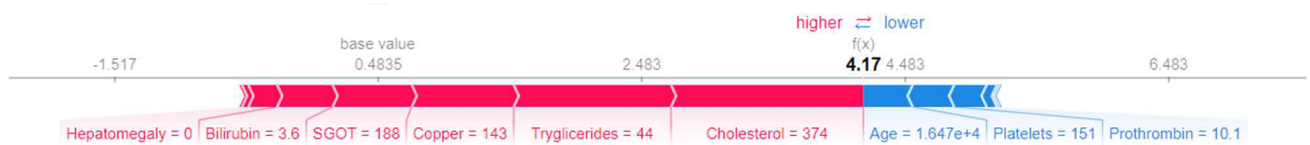
FIGURE 17. Feature contribution graph.

C. CLINICAL FEATURE ANALYSIS

With the rapid advancements in computing power and learning algorithms, artificial intelligence has gained significant prominence in the field of healthcare and medicine [23]. Researchers can gain valuable insights by repeatedly training learning algorithms on subsets of the training data. However, this approach often lacks the depth required to fully comprehend how the model reaches its conclusions.

The introduction of Explainable-AI has facilitated the ability to understand the model and the impact of each feature on outcome predictions. Consequently, healthcare providers can more accurately identify individuals in need of special care and provide tailored recommendations for each person [20].

Fig. 16 represents the variable importance plot, with the feature listed at the top ('*Platelets*') contributing most to the model and demonstrating high predictive power. While features at the bottom exhibit lower predictive power, (contribute less to the model). The interpretation of a feature's impact on predictions is facilitated by Shapley values, which consider how the model's prediction would change when a feature assumes a specific baseline value. The cumulative sum of Shapley values for each feature elucidates the deviation between the prediction and the baseline, allowing for a breakdown of predictions, as demonstrated in Fig. 17. Feature contribution graph of a prediction taken randomly from the predictions made by the model by Shapley values, which consider how the model's prediction would change when a feature assumes a specific baseline value. The cumulative sum of Shapley values for each feature elucidates the deviation between the prediction and the baseline, allowing for a breakdown of predictions, as demonstrated in Fig. 17, which is a feature contribution graph of a prediction taken randomly from the predictions made by the model.

In contrast to the XGBoost model's prediction of 4.17, the base value is 0.4835. Fig. 17 visually represents the magnitude of feature values that increase predictions in pink, indicating the extent of their impact, while blue signifies feature values that diminish the prediction. In this specific case, '*Cholesterol*' exerts the most significant influence on the prediction, whereas '*Age*' and '*Platelets*' notably diminish the impact. It is essential to acknowledge that '*Age*', as a variable, is typically beyond control in real world scenarios. Interestingly, while Fig. 16 highlights '*Platelets*' as the most influential feature in the model, this specific case underscores the significance of '*Cholesterol*' as a contributing factor to the prediction.

In contemporary times, the burgeoning influence of artificial intelligence and its subfield, machine learning, is becom-

TABLE 4. Comparison of the proposed method with recent publication.

S No.	Author	Year	Disease	Conclusion
1.	Ke Chen[16]	2022	Liver Cirrhosis	The study aims to construct a predictive model for the occurrence of liver cirrhosis in Wilson disease patients by analyzing general clinical information. The proposed model showed high accuracy in the training set, while the performance deviated in the testing set, resulting in an accuracy of 78.7%
2.	Ketan Gupta [18]	2022	Liver Disease	The study focuses on machine learning classification techniques to predict liver disease, achieving better results after performing feature selection.
3.	Yang-Yuan Chen[17]	2022	Fatty Liver Disease	It aims to predict FLD in a Taiwanese population using XGBoost, demonstrating the highest performance of 88.2%
4.	Maria Alex Kuzhippallil [15]	2020	Liver Disease	The study highlights the importance of feature selection while comparing different machine learning algorithms and proposes a method using a stacking estimator and genetic algorithm, producing an accuracy of 85%.
5.	Chieh-Chen Wu [30]	2018	Fatty Liver Disease	It proposed a method using Random Forest classifier, achieving an accuracy of 86.4%.
6.	Vasan Durai [31]	2019	Liver Disease	This study compared KNN and AVM for improved prediction and performance accuracy.
7.	**Proposed Work**		**Liver Cirrhosis**	**SHAP – Enhanced Cross-Validated XGBoost model, achieving an accuracy of 90.5%.**

ing increasingly noteworthy within healthcare and medicine. The significant advancements in the field of healthcare can be primarily attributed to several factors, including the exponential growth of computing power, the development of more advanced learning algorithms, and the establishment of extensive databases sourced from a diverse range of electronic health records, pharmaceutical research, health monitoring systems, genomic sequencing, and medical imaging. These factors collectively enabled researchers and practitioners to analyze and interpret complex medical data more effectively, leading to a better understanding of various diseases and more precise treatment options [29].

D. COMPARATIVE ANALYSIS

Conventional methods for detecting liver cirrhosis rely on clinical evaluation, liver function tests, imaging studies (such as MRI, CT scans, and Fibro Scan), and liver biopsy. The effectiveness of assessing these physiological parameters, imaging techniques, and biopsies depends on the expertise of clinicians and the quality of the liver tissue sample obtained. The choice of diagnostic method varies depending on the patient's specific clinical presentation and risk factors.

Machine learning methodologies possess the capability to analyze extensive datasets and identify patterns that might elude conventional approaches. The repeated training of these

algorithms on subsets of the training data has provided researchers unparalleled insights into therapeutic options, diagnostics, and patient outcomes. Incorporating big data and artificial intelligence applications has enabled clinicians to gain and achieve a comprehensive understanding of etiology, epidemiology, and treatment options for various diseases. This, in turn, has facilitated the identification of patients who require additional care, thereby allowing healthcare professionals to provide personalized recommendations for each individual.

Recent advancements in machine learning have significantly influenced biomarker discovery and disease diagnosis within the medical field. Various machine learning algorithms have been established to predict thyroid malignancy, yielding performance comparable to that of an expert cytopathologist [32].

The results of this study demonstrate that the combination of the XGBoost model with SHAP analysis can achieve an accuracy rate of 90.5% in diagnosing liver cirrhosis. This represents a substantial improvement compared to previous studies that employed machine learning methods for liver cirrhosis diagnosis.

As shown in the Table 4, the proposed XGBoost model with SHAP analysis outperformed all other models in terms of accuracy. The Random Forest classifier achieved the

second-highest accuracy of 86.4%, followed by the stacking estimator with genetic algorithm at 85%. The KNN and AVM models had lower accuracies of 83% and 81%, respectively.

In contrast with previous works, implementation of SHAP analysis, allowed us to identify the essential features in the XGBoost model and gain insights into the underlying decision-making mechanisms. This is consistent with previous research highlighting feature selection's importance in machine learning-based liver disease diagnosis [14]. By removing irrelevant features and focusing on the most important ones, we were able to improve the accuracy and robustness of our model.

The advancements highlighted in this research underscore the increasing utility and potential of machine learning in the medical field, offering valuable insights into the practical application of these techniques for enhancing disease diagnosis and treatment. A comparison with recent studies is presented in Table 4.

The proposed method boasts several advantages, including:

1) Achieving the highest level of performance in terms of accuracy and F1-Score.
2) Demonstrating robustness and accuracy, as evidenced by the implementation of k-fold cross-validation.
3) Effectively reducing the occurrence of overfitting.

VI. CONCLUSION
This research aims to develop a novel architecture for detecting liver cirrhosis at an early phase noninvasively. The XGBoost classifier is utilized and achieves an accuracy of 90.5%. This study also proposed a method to obtain insights and detailed explanations of features contributing to prediction, which can help healthcare professionals diagnose patients using the Explainable-AI algorithm.

Future research endeavors should prioritize building a multimodel framework incorporating a broader spectrum of clinical data and biomarkers, such as radiomics data from MRI and CT scans combined with Electronic Health Records (EHR). This holistic approach has the potential to lead to even more effective and precise detection of liver cirrhosis. Expanding the range of clinical data sources and integrating radiomics will advance our understanding and capabilities in this critical area of healthcare. Additionally, exploring the potential of advanced machine learning techniques in handling multimodal data could further improve the accuracy and reliability of cirrhosis detection.

CONFLICT OF INTERESTS
The authors declare that there is no conflict of interests regarding the publication of this paper

REFERENCES
[1] D. Schuppan and N. H. Afdhal, "Liver cirrhosis," *Lancet*, vol. 371, no. 9615, pp. 838–851, 2008.

[2] M. A. Odenwald and S. Paul, "Viral hepatitis: Past, present, and future," *World J. Gastroenterol.*, vol. 28, no. 14, pp. 1405–1429, Apr. 2022.

[3] A. B. Chowdhury and K. J. Mehta, "Liver biopsy for assessment of chronic liver diseases: A synopsis," *Clin. Experim. Med.*, vol. 23, pp. 1–13, Feb. 2022.

[4] P. Bedossa, "Sampling variability of liver fibrosis in chronic hepatitis C," *Hepatology*, vol. 38, no. 6, pp. 1449–1457, Dec. 2003.

[5] S. Sharma, K. Khalili, and G. C. Nguyen, "Non-invasive diagnosis of advanced fibrosis and cirrhosis," *World J. Gastroenterol.*, vol. 20, no. 45, p. 16820, 2014.

[6] N. Bao, Y. Chen, Y. Liu, and C. Chakraborty, "Multi-objective path planning for lung biopsy surgery," *Multimedia Tools Appl.*, vol. 81, no. 25, pp. 36153–36170, Oct. 2022, doi: 10.1007/s11042-021-11476-w.

[7] European Association for Study of Liver, "EASL-ALEH clinical practice guidelines: Non-invasive tests for evaluation of liver disease severity and prognosis," *J. Hepatol.*, vol. 63, no. 1, pp. 237–264, 2015, doi: 10.1016/j.jhep.2015.04.006.

[8] M. Giuffrè, S. Fouraki, M. Campigotto, A. Colombo, A. Visintin, M. R. Buonocore, A. Aversano, M. Budel, F. Tinè, C. Abazia, F. Masutti, and L. S. Crocè, "Alanine aminotransferase and spleno-portal dynamics affect spleen stiffness measured by point shear-wave elastography in patients with chronic hepatitis C in the absence of significant liver fibrosis," *J. Ultrasound*, vol. 24, no. 1, pp. 67–73, Mar. 2021, doi: 10.1007/s40477-020-00456-9.

[9] F. Ravaioli, M. Montagnani, A. Lisotti, D. Festi, G. Mazzella, and F. Azzaroli, "Noninvasive assessment of portal hypertension in advanced chronic liver disease: An update," *Gastroenterology Res. Pract.*, vol. 2018, pp. 1–11, Jun. 2018, doi: 10.1155/2018/4202091.

[10] W. Jeberson, A. Kishor, and C. Chakraborty, "Intelligent healthcare data segregation using fog computing with Internet of Things and machine learning," *Int. J. Eng. Syst. Model. Simul.*, vol. 1, no. 1, p. 188, 2021, doi: 10.1504/IJESMS.2021.115533.

[11] A. Sarkar, M. Z. Khan, M. M. Singh, A. Noorwali, C. Chakraborty, and S. K. Pani, "Artificial neural synchronization using nature inspired whale optimization," *IEEE Access*, vol. 9, pp. 16435–16447, 2021, doi: 10.1109/ACCESS.2021.3052884.

[12] S. H. Hyun, M. S. Ahn, Y. W. Koh, and S. J. Lee, "A machine-learning approach using PET-based radiomics to predict the histological subtypes of lung cancer," *Clin. Nucl. Med.*, vol. 44, no. 12, pp. 956–960, 2019.

[13] M. Sato, K. Morimoto, S. Kajihara, R. Tateishi, S. Shiina, K. Koike, and Y. Yatomi, "Machine-learning approach for the development of a novel predictive model for the diagnosis of hepatocellular carcinoma," *Sci. Rep.*, vol. 9, no. 1, p. 7704, May 2019.

[14] K. E. Corey, U. Kartoun, H. Zheng, and S. Y. Shaw, "Development and validation of an algorithm to identify nonalcoholic fatty liver disease in the electronic medical record," *Digestive Diseases Sci.*, vol. 61, no. 3, pp. 913–919, Mar. 2016.

[15] M. A. Kuzhippallil, C. Joseph, and A. Kannan, "Comparative analysis of machine learning techniques for Indian liver disease patients," in *Proc. 6th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS)*, Mar. 2020, pp. 778–782.

[16] K. Chen, Y. Wan, J. Mao, Y. Lai, G. Zhuo-Ma, and P. Hong, "Liver cirrhosis prediction for patients with Wilson disease based on machine learning: A case–control study from Southwest China," *Eur. J. Gastroenterology Hepatology*, vol. 34, no. 10, pp. 1067–1073, Oct. 2022, doi: 10.1097/MEG.0000000000002424.

[17] 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," *J. Personalized Med.*, vol. 12, no. 7, p. 1026, Jun. 2022, doi: 10.3390/jpm12071026.

[18] K. Gupta, N. Jiwani, and N. Afreen, "Liver disease prediction using machine learning classification techniques," in *Proc. IEEE 11th Int. Conf. Commun. Syst. Netw. Technol. (CSNT)*, Apr. 2022, pp. 221–226.

[19] Fedesoriano. (Aug. 2021). *Cirrhosis Prediction Dataset*. [Online]. Available: https://www.kaggle.com/fedesoriano/cirrhosis-prediction-dataset

[20] C.-P. Li, F.-Y. Lee, S.-J. Hwang, R.-H. Lu, W.-P. Lee, Y. Chao, S.-S. Wang, F.-Y. Chang, J. Whang-Peng, and S.-D. Lee, "Spider angiomas in patients with liver cirrhosis: Role of vascular endothelial growth factor and basic fibroblast growth factor," *World J. Gastroenterology*, vol. 9, no. 12, p. 2832, 2003.

[21] J. R. Carvalho and M. V. Machado, "New insights about albumin and liver disease," *Ann. Hepatology*, vol. 17, no. 4, pp. 547–560, Jul. 2018.

[22] S. Rawi and G. Y. Wu, "Pathogenesis of thrombocytopenia in chronic HCV infection: A review," *J. Clin. Transl. Hepatol.*, vol. 8, no. 2, pp. 184–191, Apr. 2020.

[23] J. N. George, "Idiopathic thrombocytopenic purpura in adults: Current issues for pathogenesis, diagnosis and management," *Hematology J.*, vol. 5, pp. S12–S14, Jul. 2004.

[24] V. Bewick, L. Cheek, and J. Ball, "Statistics review 14: Logistic regression," *Crit. Care*, vol. 9, no. 1, p. 112, 2005, doi: 10.1186/cc3045.

[25] H. Yang, X. Li, H. Cao, Y. Cui, Y. Luo, J. Liu, and Y. Zhang, "Using machine learning methods to predict hepatic encephalopathy in cirrhotic patients with unbalanced data," *Comput. Methods Programs Biomed.*, vol. 211, Nov. 2021, Art. no. 106420.

[26] X. Zhang and W. Chen, "Stock selection based on extreme gradient boosting," in *Proc. Chin. Control Conf. (CCC)*, Jul. 2019, pp. 8926–8931.

[27] S. Z. Alparslan-Gök, S. Miquel, and S. H. Tijs, "Cooperation under interval uncertainty," *Math. Methods Operations Res.*, vol. 69, no. 1, pp. 99–109, Mar. 2009, doi: 10.1007/s00186-008-0211-3.

[28] S. M. Lundberg, G. G. Erion, and S.-I. Lee, "Consistent individualized feature attribution for tree ensembles," 2018, *arXiv:1802.03888*.

[29] A. S. Ahuja, "The impact of artificial intelligence in medicine on the future role of the physician," *PeerJ*, vol. 7, p. e7702, Oct. 2019.

[30] C.-C. Wu, W.-C. Yeh, W.-D. Hsu, M. M. Islam, P. A. A. Nguyen, T. N. Poly, Y.-C. Wang, H.-C. Yang, and Y.-C. J. Li, "Prediction of fatty liver disease using machine learning algorithms," *Comput. Methods Programs Biomed.*, vol. 170, pp. 23–29, Mar. 2019.

[31] V. Durai, S. Ramesh, and D. Kalthireddy, "Liver disease prediction using machine learning," *Int. J. Adv. Res. Ideas Innov. Technol.*, vol. 5, no. 2, pp. 1584–1588, 2019.

[32] D. D. E. Range, D. Dov, S. Z. Kovalsky, R. Henao, L. Carin, and J. Cohen, "Application of a machine learning algorithm to predict malignancy in thyroid cytopathology," *Cancer Cytopathology*, vol. 128, no. 4, pp. 287–295, 2020.
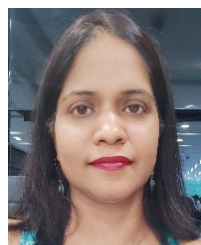
GREESHMA ARYA received the B.Tech. and M.Tech. degrees (Hons.) from Dr. A. P. J. Abdul Kalam University, Lucknow, India, and the Ph.D. degree in electronics and communication engineering from Uttarakhand Technical University, Dehradun, India. She is currently an Associate Professor with the Department of Electronics and Communication Engineering, Indira Gandhi Delhi Technical University for Women (IGDTUW) (State Government University), Delhi, India. She has more than 17.5 years of experience in academics and research. She has published more than 35 research articles in various international journals (including SCI, ESCI, Scopus, and ISI indexed) that also include IEEE international conferences. Her research interests include wireless sensor networks, wireless communication, network security, the Internet of Things, 5G network technology, artificial intelligence, deep learning, cloud computing and distributed networks, data communication and networks, and big data computing.

ASHISH BAGWARI (Senior Member, IEEE) received the B.Tech. (Hons.), M.Tech. (Hons.), and Ph.D. degrees in electronics and communication engineering. He is currently the Head of the Department of Electronics and Communication Engineering, Women Institute of Technology (WIT) (Institute of State Government), affiliating institution of Uttarakhand Technical University, Dehradun, India. He has more than 11.5 years of experience in industry, academics, and research. He has published more than 110 research articles in various international journals that also include IEEE international conferences. His research interests include cognitive radio networks, mobile communication, sensor networks, wireless, 5G communication, and digital communication.

He is an active member of various professional societies like IEEE, USA, a Senior Member of the Institute of Electronics and Telecommunication Engineers (IETE), India, a lifetime member, a professional member of ACM, and a member of the Machine Intelligence Research Laboratory Society. He received the Best WIT Faculty Award, in 2013 and 2015, and Best Project Guide Award, in 2015. He received the Gold Medal for the M.Tech. degree.
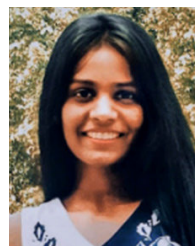
HITESHI SAINI received the B.Tech. degree in electronics and communication engineering from Indira Gandhi Delhi Technical University, Delhi, India, where she is currently pursuing the Ph.D. degree. She is currently a Researcher. Her current research interests include computer vision, artificial intelligence, and neural networks.

PRACHI THAKUR received the B.Tech. degree in electronics and communication engineering from Indira Gandhi Delhi Technical University for Women, Delhi, India. She is currently a Researcher. Her research interests include machine learning, computer vision, cybersecurity, and blockchain.

CIRO RODRIGUEZ (Senior Member, IEEE) received the Ph.D. degree in system engineering and the dual degree from the Institute of Theoretical Physics (ICTP), Italy, United States Particle Accelerator School (USPAS), and Information Technology Development Policy Studies Korea Telecom (KT), South Korea. He is currently a Professor with National University Mayor de San Marcos, Lima, Peru, where he is also associated with the Department of Software Engineering and also associated with the Department of Informatics Engineering and Electronics, National University Federico Villarreal. He has published over 100 research articles in reputed journals indexed in Scopus, WoS, and IEEE, and filed two patents in engineering fields. Recently, he has published the book "Variables in the research methodology." His research interests include artificial intelligence, health-social welfare, environment, cybersecurity, and photonics.

PEDRO LEZAMA received the Ph.D. degree in systems engineering and the master's degree in systems engineering and in information technology management. He was a Systems Engineer and Certified PMP, PMI-ACP, SAMC, SPOC, SMC, SSGB, and RENACYT (P0109942). He is currently a Researcher and an University Professor with National University Mayor de San Marcos having expertise in project management under traditional and agile approaches, with more than 15 years of general experience in the implementation and development of large-scale information software and the Project Manager of Business Intelligence and Data Analytics Projects in the implementation and development of large-scale information software and the Project Manager of Business Intelligence and Data Analytics Projects.

• • •