Literature Review for XAI [Abdulla]

Explainable AI for Enhanced Interpretation of Liver Cirrhosis Biomarkers

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)

This work is funded by Universidad Nacional Mayor de San Marcos (UNMSM) according to R.R.No 00898-R-17.

ABSTRACT Liver cirrhosis is a terminal pathological result of chronic liver damage, illicit drugs, hepatotoxicity, 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.

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.

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.

Full Paper 👇

https://prod-files-secure.s3.us-west-2.amazonaws.com/62d68f87-b746-4b9f-9970-9e86c58cbcb9/b899fe60-11de-429d-b1a8-e6bba13a9d36/LR_DONE.pdf

1.0 Dataset Configuration

1.1 Dataset information

The dataset used in this study was sourced from Kaggle and is based on the **Mayo Clinic's primary biliary cirrhosis (PBC) study**. It contains data from 424 patients, with a gender distribution of 89% female and 11% male. The study focused on 14 clinical parameters and one target class, comprising numerical variables like albumin, bilirubin, copper, and platelets, and categorical variables such as sex, ascites, and edema.

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 represents four stages of liver health progression:

- **Healthy**: No signs of damage or scarring.
- **Fatty Liver**: Early-stage liver damage with fat accumulation.
- **Fibrosis**: Significant scarring and progressing damage.
- **Cirrhosis**: Advanced, irreversible liver damage.

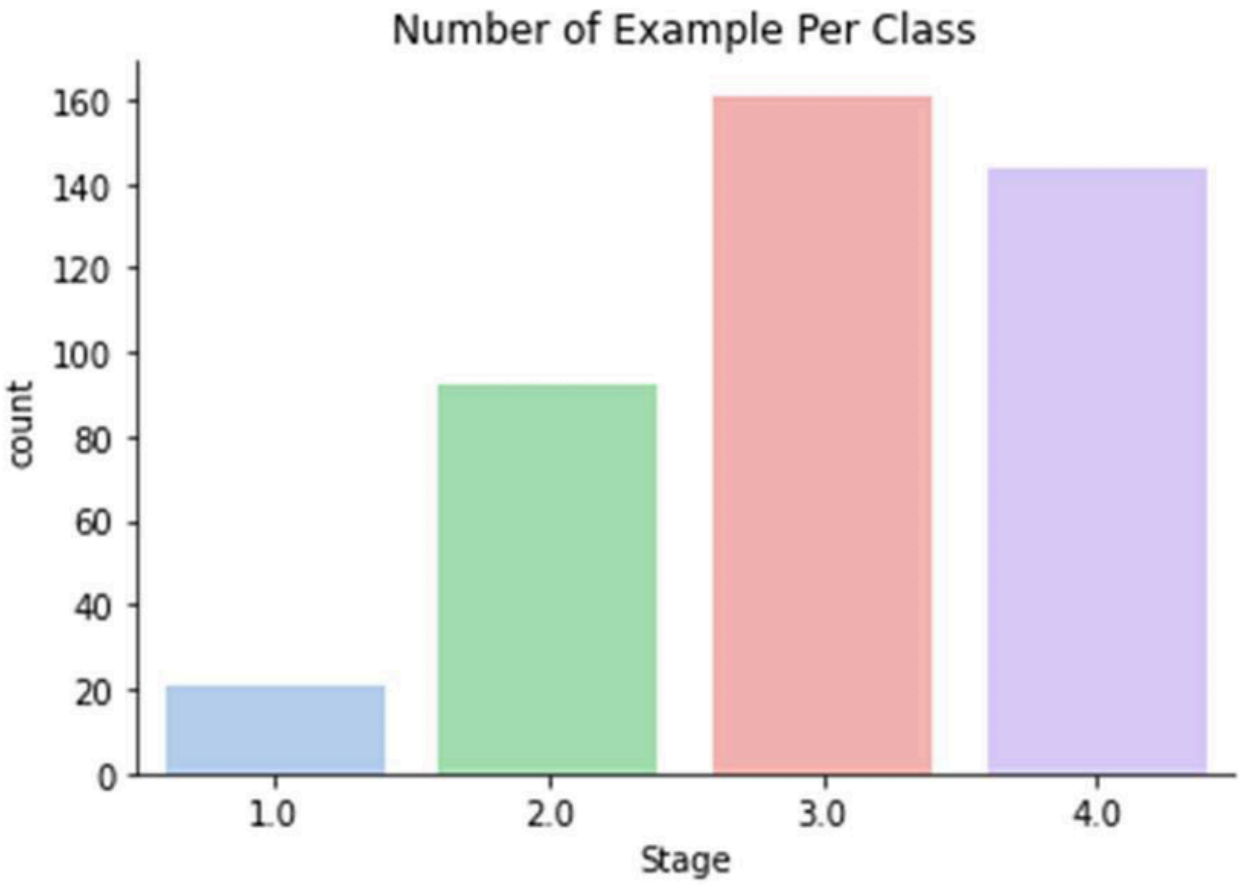

FIGURE 1. **Different stages of liver cirrhosis.**

2.0 Feature Analysis

2.1 Key Biomarkers in Liver Cirrhosis

Biomarkers are critical in assessing liver function and diagnosing cirrhosis, offering insights into disease progression and severity. This section provides an overview of the most significant features analyzed in this study, highlighting their clinical importance and role in understanding liver health.

2.1.1 Ascites

Fluid accumulation in the abdomen, linked to a poor prognosis and complications like bacterial peritonitis and kidney dysfunction. Its prevalence increases with disease severity.

Ascites prevalence increases with cirrhosis stages, indicating disease severity.

2.1.2 Hepatomegaly

Liver enlargement often signals progressive damage and fibrosis, commonly observed in intermediate and advanced cirrhosis stages.

Frequently observed in intermediate and advanced stages of cirrhosis.

2.1.3 Bilirubin

A marker of bile duct dysfunction, elevated bilirubin is associated with jaundice and oxidative stress. Levels rise consistently with disease progression.

Bilirubin levels progressively rise across cirrhosis stages, making it a reliable marker of liver function decline.

2.1.4 Spider Angiomas

Visible blood vessels linked to liver dysfunction, more common in advanced stages due to hormonal changes and vascular growth factors.

More prominent in advanced cirrhosis stages, highlighting its diagnostic relevance.

2.1.5 Edema

Fluid retention in the legs caused by portal hypertension, ranging from mild to severe. Its presence signals advanced liver damage.

Severe edema becomes increasingly common in later cirrhosis stages.

2.1.6 Prothrombin

Measured through prothrombin time, it indicates reduced clotting factor production and correlates with severe liver damage and advanced fibrosis.

Higher prothrombin times were observed in cirrhotic patients, reinforcing its value as a non-invasive diagnostic marker.

2.1.7 Albumin

Low levels of this vital protein reflect disease progression and are associated with poor outcomes in cirrhotic patients.

Albumin deficiency strongly correlates with advanced cirrhosis stages and poor patient outcomes.

2.1.8 Triglycerides

Linked to fatty liver disease, elevated levels are more evident in early liver damage stages, signaling impaired fat metabolism.

High triglyceride levels were prominent in earlier liver damage stages, especially fatty liver.

2.1.9 Platelets

A reduced platelet count (thrombocytopenia) is a sensitive marker for chronic liver disease and complications like portal hypertension and bleeding risks.

Platelet counts decline progressively with disease severity, making it a sensitive marker for cirrhosis diagnosis.

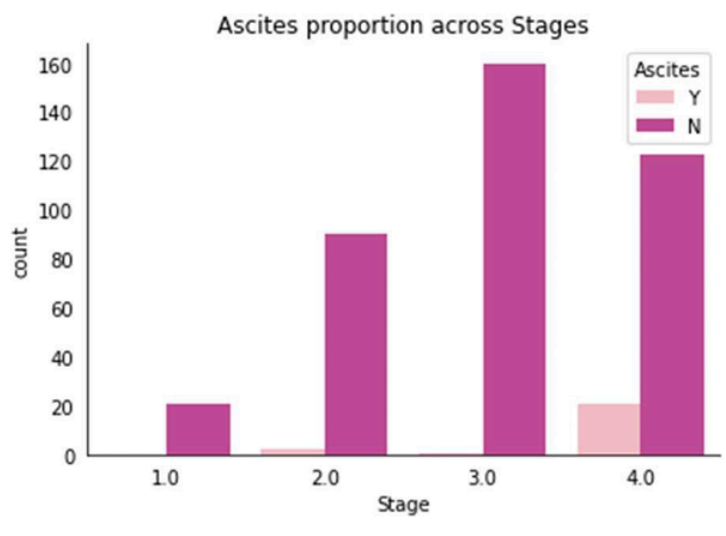

FIGURE 2. Ascites proportion across different stages.

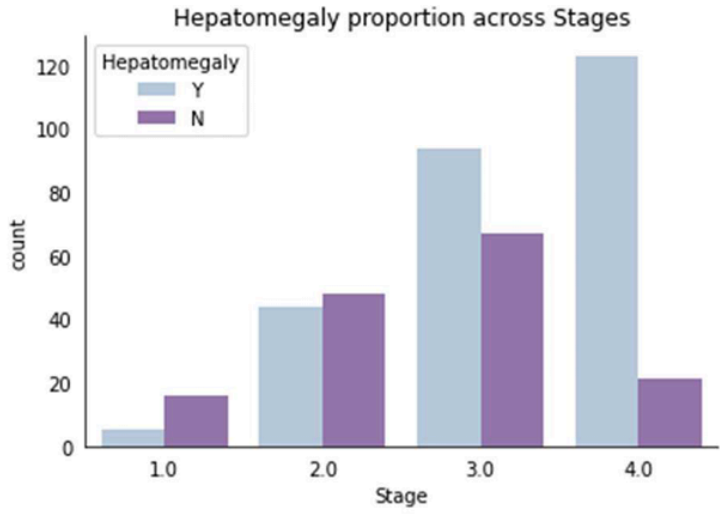

FIGURE 3. Hepatomegaly proportion across stages.

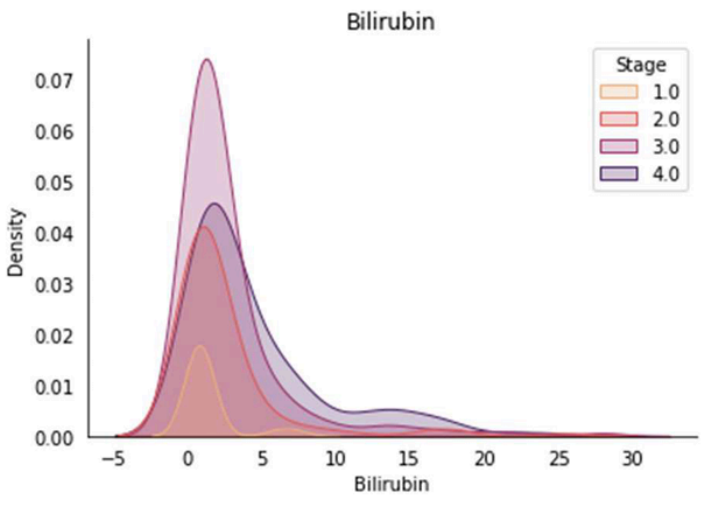

FIGURE 4. Bilirubin [mg/dl] across different stages.

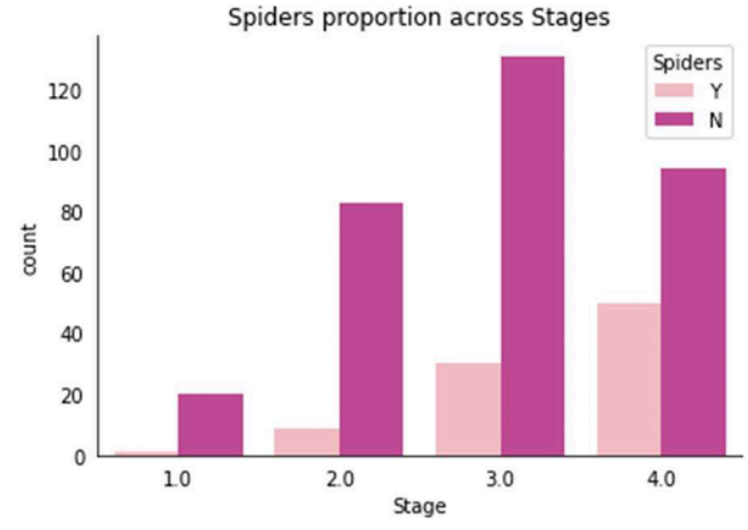

FIGURE 5. Patients with Spider angiomas across different stages.

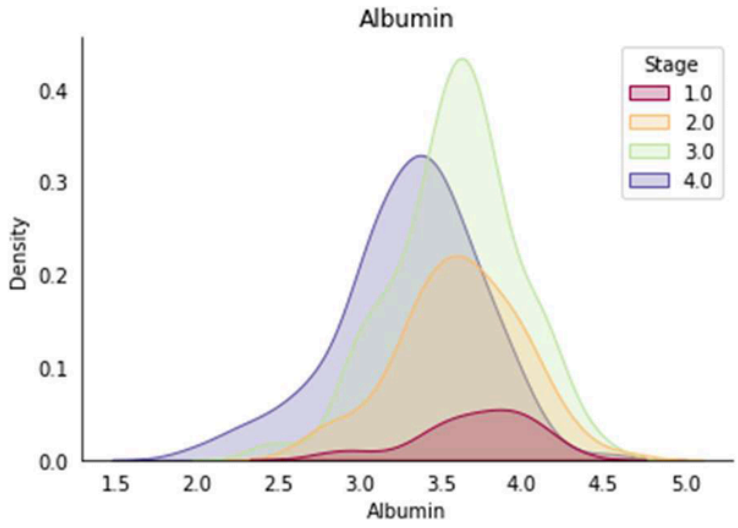

FIGURE 7. Albumin [g/dl] across different stages of liver damage.

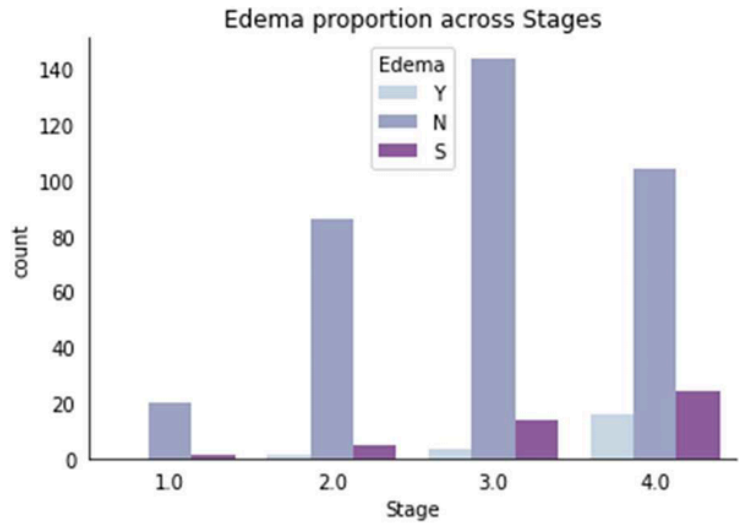

FIGURE 6. Patients with edema across different stages.

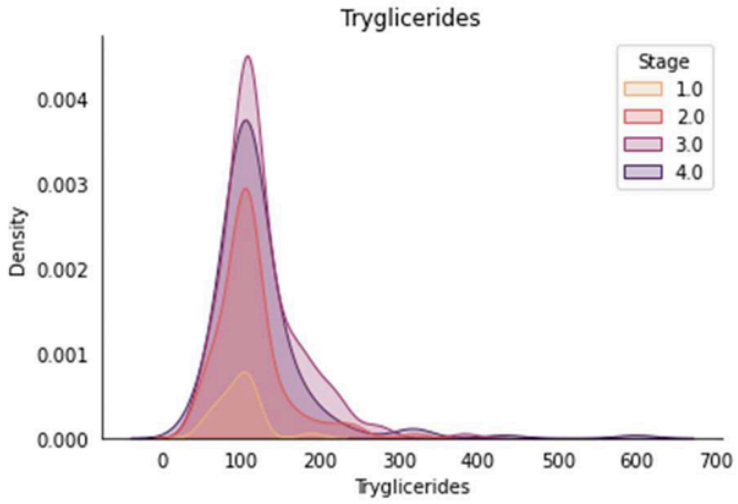

FIGURE 8. Triglyceride [mg/dl] across different stages.

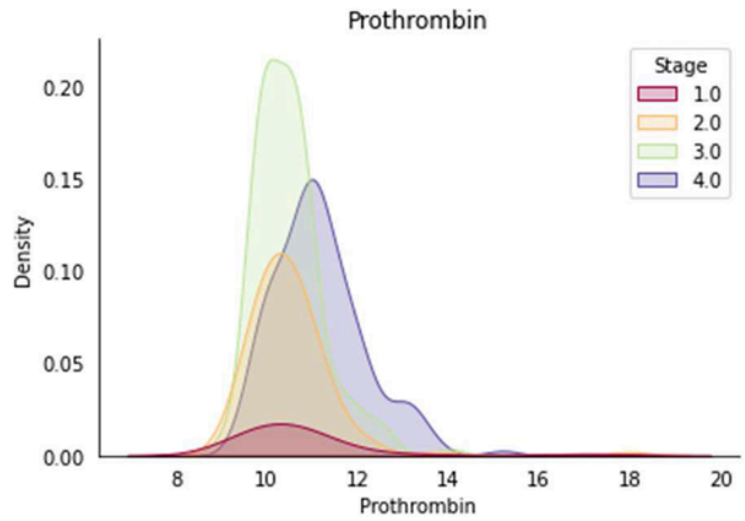

FIGURE 10. Prothrombin time in seconds [s] across different stages.

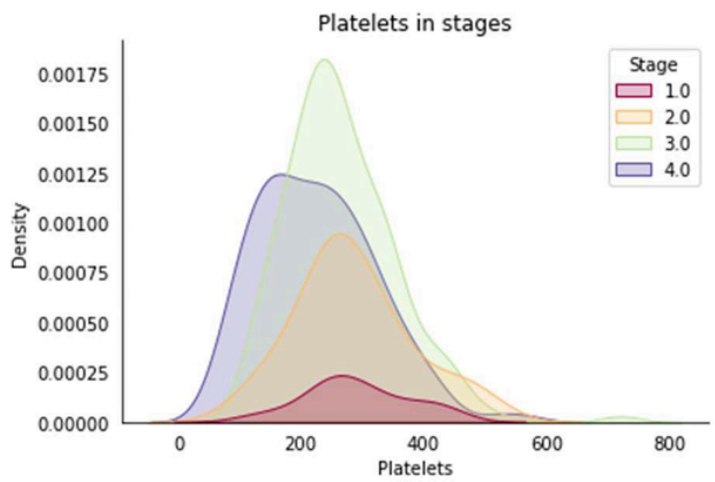

FIGURE 12. Platelets per cubic [ml/1000] across different stages.

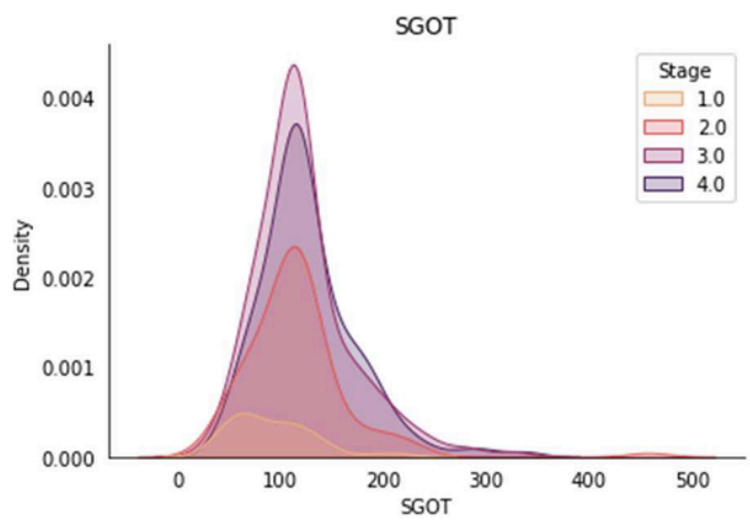

FIGURE 13. SGOT levels [U/ml] across different stages.

2.2 Additional Features

Several additional features also play a critical role in assessing liver function and cirrhosis progression:

- **Alkaline Phosphatase (ALP)**: This enzyme is primarily found in the bile ducts, and elevated levels indicate damage to biliary cells or obstruction in the bile ducts. High ALP is often observed in cholestatic liver diseases and serves as a marker for biliary inflammation or fibrosis.
- **Cholesterol**: In cases of non-alcoholic cirrhosis, serum cholesterol levels often reflect liver damage severity. A decline in cholesterol synthesis can indicate compromised liver function, while abnormal levels in fatty liver disease stages highlight metabolic imbalances.
- **Copper**: The liver regulates copper metabolism, and excessive accumulation of copper can cause toxicity, leading to acute liver damage. Elevated urinary copper levels are particularly significant in diseases like Wilson's disease but also serve as an indirect marker of cirrhosis.

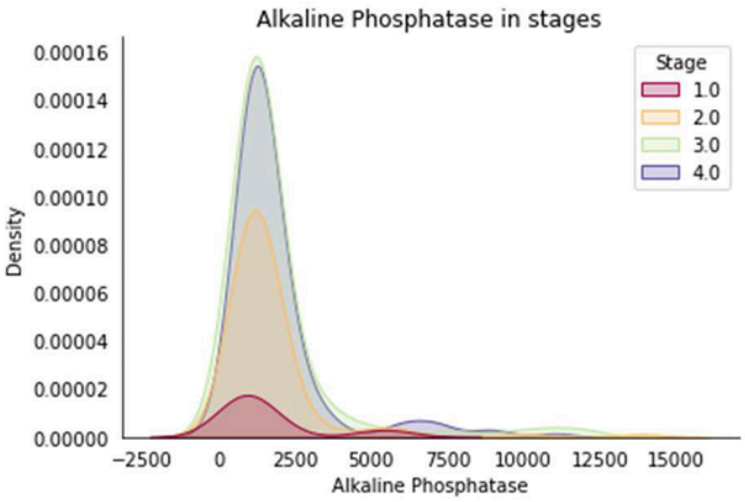

FIGURE 14. Alkaline phosphatase level [U/liter] across different stages.

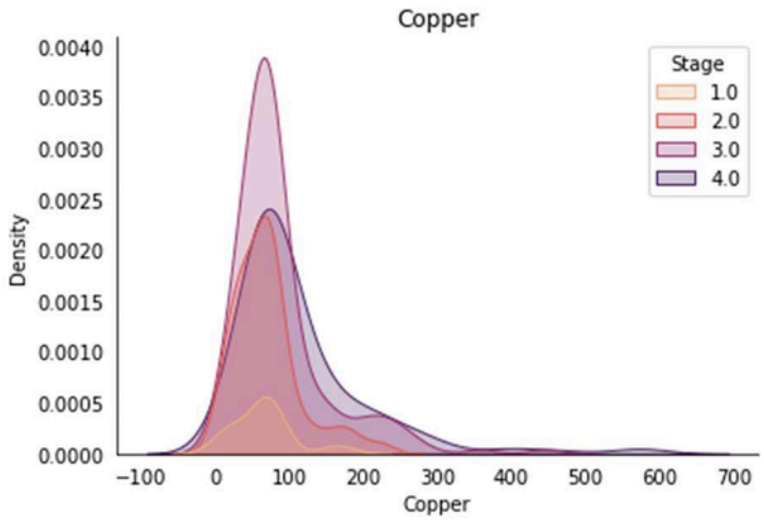

FIGURE 11. Urine copper level [ug/day] across different stages.

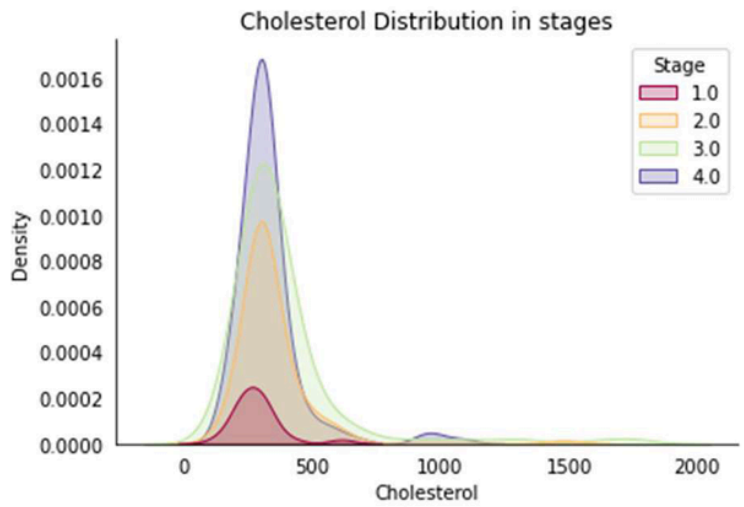

FIGURE 9. Cholesterol [mg/dl] across different stages.

3.0 Machine Learning Algorithms

This study employs Logistic Regression, XGBoost, and Explainable AI (SHAP) to analyze and predict liver cirrhosis progression.

3.1 Logistic Regression

Logistic regression is a statistical method used to estimate the probability of an event, such as distinguishing between cirrhotic and non-cirrhotic states. It applies a logit transformation to the odds ratio of the dependent variable.

$$\text{Logit}(p_i) = \ln\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_m X_m$$

Where:

- p_i: Probability of the event occurring for instance i
- $\beta 0, \beta 1, \ldots, \beta m$: Model coefficients
- $X1, \ldots, Xm$: Feature values

3.2 XGBoost

XGBoost is a scalable tree-boosting system that optimizes Gradient Boosting techniques. It builds regression trees to minimize an objective function.

Objective Function:

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

Where:

- $l(y_i, \hat{y}_i)$: Loss function measuring the difference between true and predicted values
- $\omega(fm)$: Regularization term to prevent overfitting

Prediction Equation:

$$\hat{y}_i = \sum_{m=1}^{M} g_k(x_i), \quad g_k \in E$$

Where:

- $\hat{y}_i$: Predicted value for the i^{th} data point
- $g_k(x_i)$: Prediction made by the k^{th} tree for the i^{th} data point
- E: Function space containing all possible regression trees

Additive Strategy (To minimize the loss function iteratively):

$$\hat{y}_j^{(s)} = \sum_{m=1}^{s} g_m(x_j) = \hat{y}_i^{(s-1)} + g_s(x_j)$$

Where:

- $\hat{y}_j^{(s)}$: Prediction at step s
- $g_m(x_j)$: Output of the newly added tree at step s
- $\hat{y}_j^{(s-1)}$: Prediction from the previous step

As you can see, this uses an additive approach.

> 💡 It’s not really important to understand the math behind these algorithms, what we care about is the Explainable AI (SHAP) which is the next.

3.3 Explainable AI, SHAP (SHapley Additive exPlanations)

SHAP is a game-theoretic approach to explain the output of machine learning models by attributing the contribution of each feature to the prediction. It computes Shapley values to explain the difference between the model’s prediction for a given instance and the baseline.

Shapley Value Equation:

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

Thus:

- T: Subset of features excluding the i^{th} feature
- Q: Total number of features
- $|T|$: Size of the subset T
- $F_X(T)$: Prediction function based on the subset T
- $F_X(T \cup \{i\})$: Prediction function for the subset T combined with the i^{th} feature
- $|\ T\ |!$: Accounts for all possible permutations of features in subset T before the i^{th} feature
- $(Q - |T| - 1)!$: Represents the permutations of the features after the i^{th} feature
- $Q!$: Total permutations of all features
- **Marginal Contribution** $F_X(T \cup \{i\}) - F_X(T)$: The difference in prediction when the i^{th} feature is added to the subset T, quantifying its contribution

Summary:

This equation calculates the fair contribution of each feature to a model's prediction by considering all possible combinations of features and how the inclusion of a specific feature impacts the output.

It's important to note that primary effect influencing the prediction outcome is the disparity between the Shapely values and the sum of SHAP interaction values for a given feature

$$\Phi_{i,j} = \Phi_i - \sum_{j \neq i} \Phi_{i,j}$$

4.0 Methodology

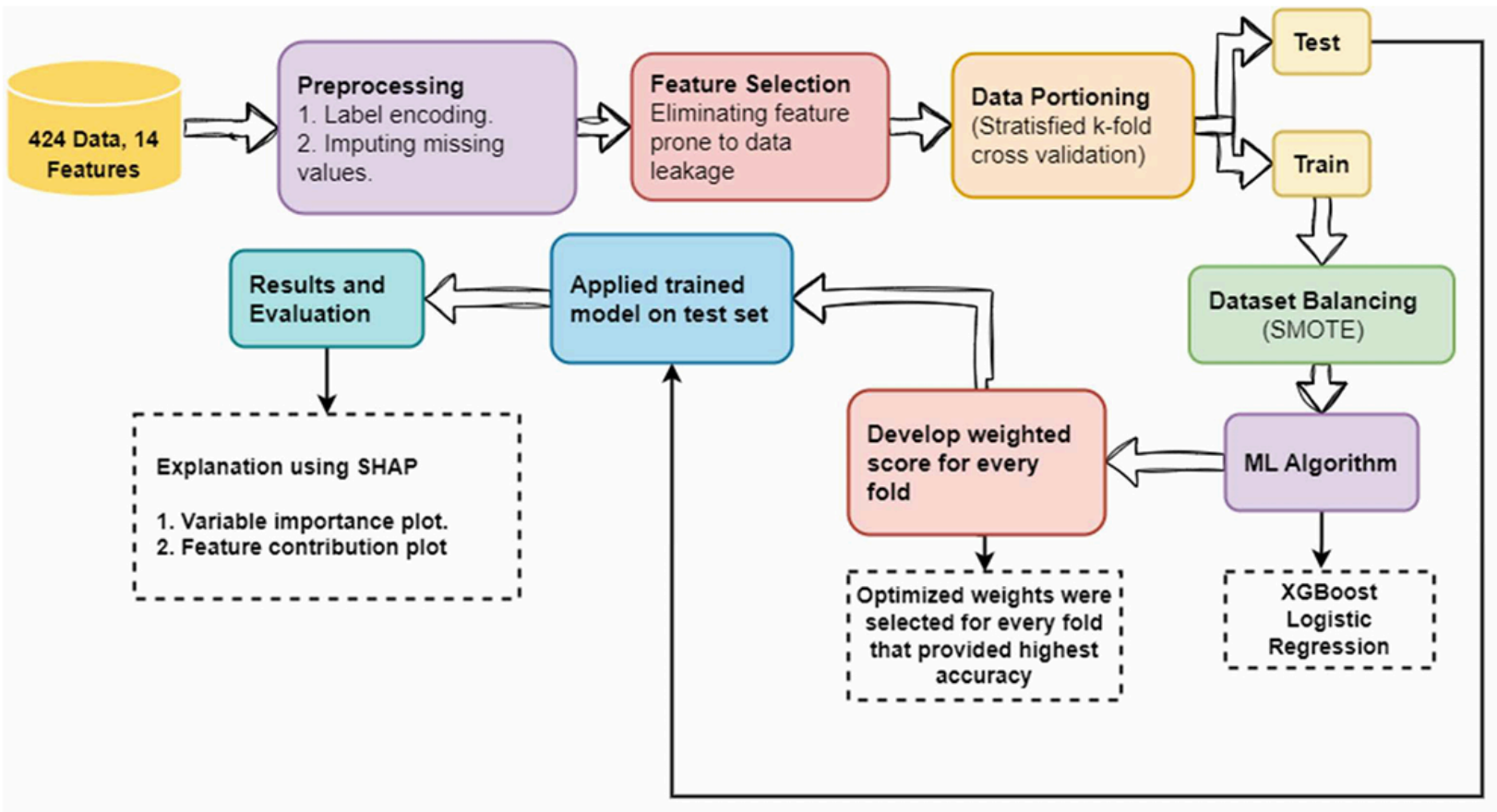

FIGURE 15. Workflow of the proposed algorithm.

The proposed methodology consists of five key steps.

4.1 Data Pre-processing

Missing values were replaced with the median of their respective columns for integer data, and missing values were imputed with the mode of the respective column for the categorical data, the dataset comprised three distinct datatypes: `float`, `int`, `object` in which each datatype has required unique pre-processing.

4.2 Categorical Data Transformation

Categorical data were converted to numerical format for analysis:

- **Mapping of categories**:
  - Sex:
    - Male (`M`) → `0`
    - Female (`F`) → `1`
  - **Ascites**:
    - Present → `1`
    - Absent → `0`
  - **Hepatomegaly**:
    - Present → `1`
    - Absent → `0`
  - **Spiders**:
    - Present → `1`
    - Absent → `0`
  - **Edema**:
    - No edema (`N`) → `0`
    - Edema with diuretics (`Y`) → `1`
    - Edema without diuretics (`S`) → `1`

As this needs to be done in every experiment to ensure that the model correctly understands and comprehends the labels and classes provided. This transformation ensured compatibility with machine learning algorithms while preserving the dataset's structure.

4.3 Feature Selection

To prevent data leakage, the following features were removed:

- `Status`
- `N_days`

4.4 Class Balancing

Imbalanced data distribution was addressed using oversampling techniques (e.g. SMOTE) to ensure that all classes were equally represented. This step was necessary to prevent the model from favoring majority classes over minority ones.

4.5 Cross-Validation

By maintaining class proportions in training and testing splits, stratified 20-fold cross-validation is used to address class imbalance and guarantee robust evaluation. To allow for thorough analysis of the entire dataset, it is split into 20 folds, of which one is used as the test set and the other 19 are used for training. The model is guaranteed to generalize well to new data thanks to this iterative process, which also lessens overfitting. Reliable performance estimates are obtained by averaging performance metrics such as:

- **Accuracy**

- **Precision**
- **Recall**
- **F1-score**

over a total of 20 iterations. To improve model performance, hyperparameter optimization within each fold adjusts parameters like

- `learning_rate`
- `max_depth`
- `random_state`
- `gamma`

In order to improve interpretability and match predictions with clinical relevance, SHAP values are calculated after training to quantify feature importance. The results usability is improved by this explain ability integration.

Algorithmic Pseudocode 👇

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. $M(\Theta) \leftarrow \arg\min_{\Theta} \left(\frac{1}{N_{train}} \sum_{i=1}^{N_{train}} L_M(\Theta),\ X^i_{\text{train}},\ Y^i_{\text{train}} \right)$
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 representations.**
19. **Exit**

💡 **Θ could represent parameters like `max_depth`, `n_estimators`, `min_child_weight` and others for the XGBoost model's GridSearch**

5.0 Analysis & Results

5.1 Metrics Results

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

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

→ The proposed XGBoost model outperforms Logistic Regression in all metrics, achieving 90.5% accuracy, 87% recall, 85% precision, and 89.9% F1-score.

5.2 Clinical Feature Analysis [Important]

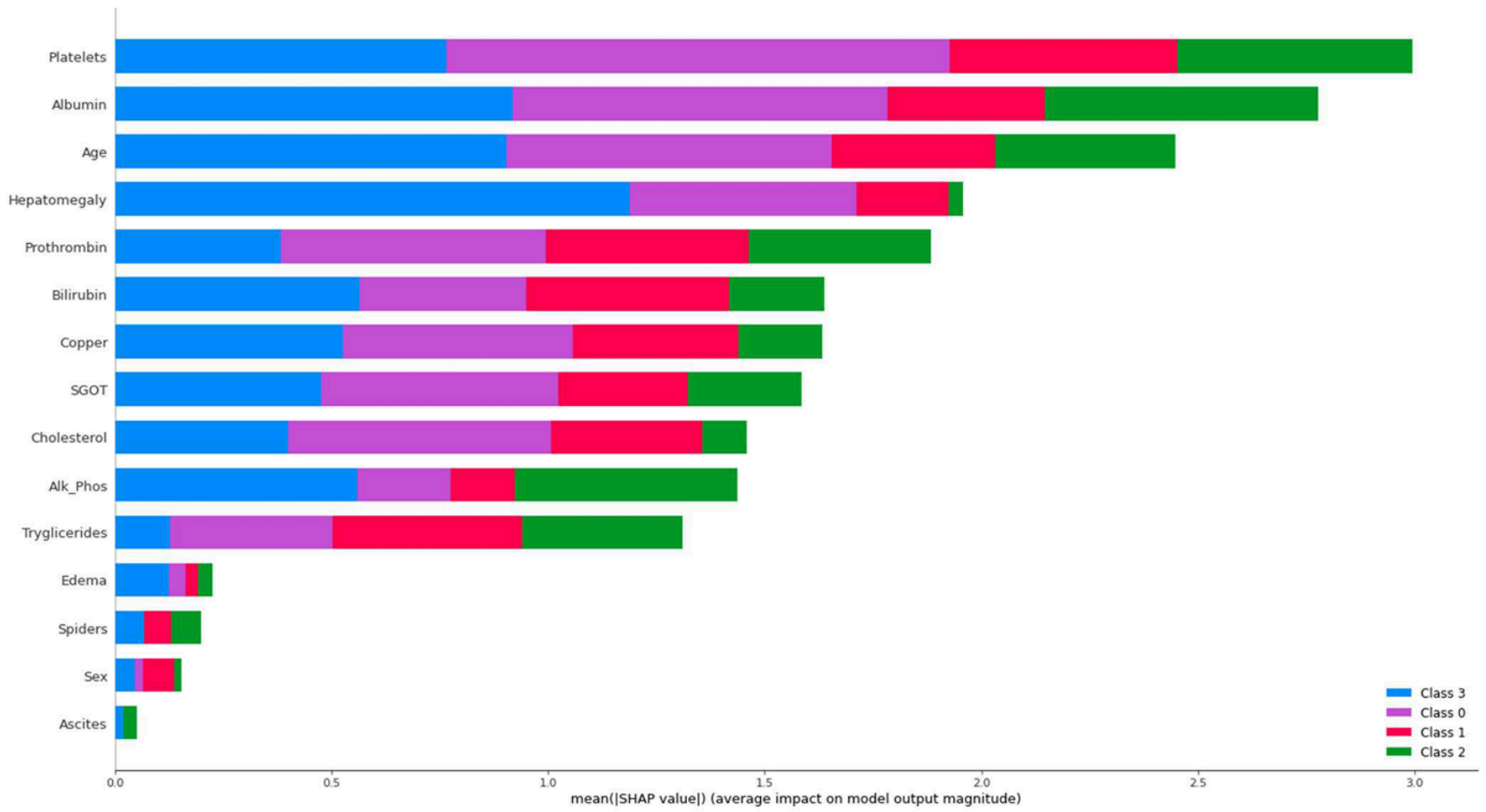

FIGURE 16. Variable importance plot. Features listed at the top (Platelets) contribute most to the model and have high predictive power. However, features listed at the bottom have less predictive power (contribute less to the model).

FIGURE 17. Feature contribution graph.

> 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.

> 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.
>
> 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.

To summarize those graphs:

Figure 16 provides a global view of feature importance across the entire dataset, emphasizing which features have the highest predictive power in the model. In contrast, Figure 17 offers a localized explanation of how individual features influence a single prediction. Together, these visualizations enhance the interpretability of the model by breaking down predictions into meaningful components, making it easier to understand the model’s decision making process and its clinical relevance.

Key Points (what these graphs are, what they demonstrate, how to read them and whats their use):

- **Figure 16: Variable Importance Plot**:
  - Highlights the most and least influential features in the model.
  - `Platelets` is identified as the most predictive feature, contributing significantly to the model’s performance.
  - Features at the bottom (e.g., `Sex`, `Ascites`) have lower predictive power and contribute less to the model.
  - SHAP values quantify the impact of each feature by showing how the prediction changes when a feature assumes specific values.
- **Figure 17: Feature Contribution Graph**:
  - Demonstrates the contribution of each feature to a single, randomly selected prediction.
  - Features in **pink** (e.g., `Cholesterol`) increase the prediction, while features in **blue** (e.g., `Age`, `Platelets`) reduce it.
  - The magnitude of the SHAP values visually indicates the strength of each feature's influence on the prediction.
  - `Cholesterol` is the most influential feature in this specific case, while other features like `Age` and `Platelets` reduce the prediction.

> 💡 We can use these graphs after implementing SHAP to provide transparency and explainability for the model by identifying key features driving predictions and understanding how individual features influence specific predictions making the results interpretable and clinically actionable.

6.0 Extra Findings [Not Important]

💡 This section is not very important as it speaks about liver cirrhosis, not related to HCC. However some of these facts about the feature analysis could be interesting.

> 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.

> 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.

> 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.

Engineer Ahmed's Literature Review 👇

https://prod-files-secure.s3.us-west-2.amazonaws.com/62d68f87-b746-4b9f-9970-9e86c58cbcb9/719bf642-d3b1-4775-8bd0-8dc5f35c78f3/LR.pdf

Senior Group Progress Report 👇

https://prod-files-secure.s3.us-west-2.amazonaws.com/62d68f87-b746-4b9f-9970-9e86c58cbcb9/f267d680-e8f9-4053-8f90-512544068628/Progress_Report_senior_.pdf