Early Prediction and Classification of Liver Cirrhosis Using Machine Learning and Clinical Biomarkers

Nguyen Thanh Quan*, Le Tu Nhan*, Tran Le Minh Thuy*, Nguyen Minh Dung*, Pham Khoi Nguyen*
Lecturer: PhD. Tran Van Hai Trieu
Teacher Assistant: Nguyen Minh Nhut
*University of Information Technology (UIT),
Vietnam National University Ho Chi Minh City (VNU-HCM)

***Abstract*—Liver cirrhosis is one of the most severe forms of chronic liver disease and remains a major contributor to global morbidity and mortality. Traditional diagnostic procedures, including ultrasound imaging and liver biopsy, are either expensive, invasive, or inefficiently suited for large-scale screening. Meanwhile, healthcare experts often experience diagnostic overload due to a ton of daily patients and the complexity of biochemical experiment testing. In this study, we systematically analyze clinical biomarkers combined with statistical evaluation and multiple machine learning models to classify cirrhosis using standard laboratory tests. Exploratory analysis incorporates Q–Q plots to assess distributional patterns, box and violin plots to characterize outliers, and statistical measures including the mean, variance, skewness, kurtosis, and quantile ranges to evaluate numerical properties of the data. For hypothesis testing, ANOVA is applied to continuous variables, while Chi-square tests are used for categorical features. We develop and compare several classifiers, including Softmax Regression, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Random Forest, and XGBoost. The experimental results show that the ensemble models, particularly Random Forest and XGBoost, achieve superior performance compared with linear and instance-based approaches and maintain consistently low error margins. Overall, the findings show that machine-learning methods can be helpful, affordable, and practical tools for the early detection of cirrhosis.**

***Index Terms*—Liver cirrhosis, machine learning, statistical analysis, clinical biomarkers, XGBoost, SVM, exploratory data analysis.**

I. Introduction

The liver is essential for metabolism, detoxification, and protein production. When chronic inflammation or ongoing liver damage happens, the organ slowly develops fibrotic scarring. This can eventually lead to liver cirrhosis. Common causes are alcoholic liver disease, chronic viral hepatitis (`HBV`, `HCV`), and non-alcoholic fatty liver disease (`NAFLD`). Cirrhosis significantly reduces quality of life, increases the risk of liver failure or `hepatocellular` carcinoma, and reason for millions of deaths worldwide [1], [6].

Despite its severity, early-stage cirrhosis are often unexpected productive detection, leading to delayed clinical detection. Traditionally scientific methods based on imaging techniques or liver biopsy methods are costly, invasive, or inaccessible in many regions [6]. Clinical blood indicators such as `ALT`, `AST`, `Bilirubin`, `Albumin`, and `ALP` can reflect liver dysfunction but are difficult to interpret individually due to non-normal distributions, outliers, and biological variability. As a result, clinicians face diagnostic overload, especially in high-volume healthcare settings [6].

To address these challenges, this study investigates a structured statistical and machine learning pipeline for cirrhosis classification. Exploratory tools such as Q-Q plots, histogram–violin plot comparisons, and box plots are used to characterize data distribution. Statistical tasks include the computation of mean, std, variance, quantile value, skewness, and kurtosis. Hypothesis testing uses ANOVA to examine differences between healthy and cirrhotic groups, and Chi-square tests for categorical variables.

We evaluate multiple supervised learning models including Softmax Regression, SVM, KNN, Random Forest, and XGBoost [3]. Our goal is to identify models that have high predictive and diagnostic accuracy. We also want to keep error levels consistent across classes. Key contributions include:

- Comprehensive visual and statistical analysis of liver cirrhosis biomarkers.
- Evaluation of distributional properties using Q-Q plots, boxplots, violin plots, and descriptive statistics.
- Hypothesis testing with ANOVA for numerical attributes and Chi-square for categorical ones.
- Comparative assessment of classical and ensemble machine learning models.
- Demonstration of XGBoost and Random Forest as stable, high-performing classifiers.

II. Related Work

A. Data Visualization in Biomedical Analysis

Visualization techniques play a fundamental role in biomedical data interpretation, especially when laboratory biomarkers exhibit heterogeneous scales, non-normal distributions, or extreme outliers. For example, Q-Q plots are extensively used in clinical analytics to assess the degree to which each biochemical variable deviates from Gaussian assumptions. This deviation is not only descriptive; it provides practical

information on the preprocessing strategy. When Q–Q plots show heavy tails, marked asymmetry, or clear deviations from linearity, they indicate that more robust scaling methods, such as the Robust Scaler, which relies on interquartile ranges as they are less influenced by extreme values. In contrast, variables with mild skewness or roughly symmetric distributions are generally well suited to Standard Scaler. MinMaxScaler is most appropriate when preserving relative distances within a fixed interval is important, as in distance-based models such as KNN or SVM with RBF kernels. In this way, Q–Q plots help inform the choice of normalization by revealing distributional features that directly affect model stability.

Box plots assist with this analysis by displaying the presence, severity, and clinical significance of outliers. These outliers often occur naturally in biomarkers such as ALT, AST, or Bilirubin. They can indicate actual pathological conditions rather than just measurement noise. Violin plots combine density estimation with summary statistics, allowing us to see multimodal patterns and changes in distribution between healthy and cirrhotic groups. Together, these visualization tools offer descriptive insights and lay the groundwork for informed preprocessing and improve model generalizability.

B. Statistical Analysis in Hepatology

Statistical analysis plays a central role in understanding biological variation between groups of patients. Descriptive statistics—mean, variance, quartiles, skewness, and kurtosis—offer essential quantitative summaries of liver biomarkers and often reveal asymmetries, long-tail effects, or compressed distributions that have direct implications for classification tasks. For example, high skewness in bilirubin values reflects clinically signigicant pathological escalation in cirrhotic patients, while kurtosis helps to identify whether extreme enzyme values are expected or anomalous.

More importantly, inferential statistical tests offer a method for finding out if the differences between healthy and cirrhotic groups are significant. One-way ANOVA is commonly used in hepatology studies as it can spot mean differences in biomarkers like `ALT`, `AST`, `ALP`, or `Albumin`. These results help identify which biomarkers possess strong discriminative potential for machine learning models. Likewise, Chi-square tests allow researchers to evaluate associations between categorical attributes (e.g., `Status`, `Drug`, `Sex`, `Ascites`, `Helaromegaly`) and cirrhosis status. Insights from these tests make sure that feature selection is based on biological evidence instead of just data-driven rules. Overall, statistical analysis supports machine learning by providing a clear and understandable way to identify which clinical variables have the highest predictive value.

C. Machine Learning for Liver Disease Prediction

A variety of machine learning methods have been studied for classifying liver disease. Early research used classic models like Logistic Regression, Naive Bayes, and Support Vector Machines. These models showed moderate success with straightforward decision boundaries. More recent studies highlight the strength of tree-based ensemble methods, including Random Forest and Gradient Boosting. These methods are very reliable against noise, non-linearity, and complex interactions among features often present in biochemical datasets.

Despite these advances, relatively few studies include detailed exploratory visualization and statistical hypothesis testing as part of the machine learning workflow for cirrhosis classification. Many existing works view data preprocessing as a general step. They do not systematically analyze how irregularities in the data or outliers affect model performance. This gap shows the need to combine exploratory data analysis, statistical inference, and machine learning into one process. By using distributional evaluation, significance testing, and comparing models, our study tackles this methodological issue and offers a better approach to predicting cirrhosis.

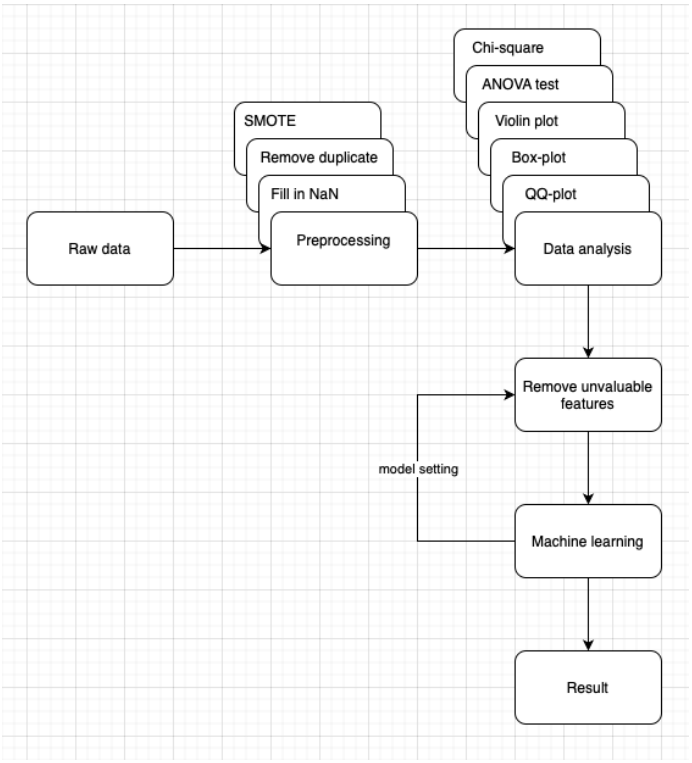

Fig. 1. Methodology pipeline

III. Methodology

A. Dataset Description

The cirrhosis dataset includes different biochemical and demographic indicators that show liver health, metabolic function, and structural damage. Each biomarker gives important diagnostic information for evaluating the presence and severity of cirrhosis. Their relevance is summarized as follows [1]–[3]:

- **ALT (Alanine Aminotransferase):** ALT is an enzyme mainly found in liver cells. High ALT levels suggest liver cell damage due to inflammation, cell death, or metabolic issues. In advanced cirrhosis, ALT levels may drop as the

number of liver cells decreases, which limits the release of the enzyme into the blood.

- **AST (Aspartate Aminotransferase):** AST is less liver-specific than ALT, but remains a critical marker of chronic hepatic injury. A sustained elevation in AST suggests prolonged hepatocellular stress. An $\mathrm{AST}/\mathrm{ALT} > 2$ ratio is strongly associated with alcohol-related cirrhosis.
- **Total Bilirubin (TB):** TB shows the total amount of conjugated and unconjugated bilirubin in the blood. A high TB level suggests problems with liver detoxification, cholestasis, or portal hypertension. In cirrhosis, an increased TB level is an important sign of reduced bilirubin removal.
- **Direct Bilirubin (DB):** DB reflects the conjugated fraction of bilirubin processed by the liver. Elevated DB indicates obstruction of bile flow or impaired excretion, conditions frequently observed in advanced or cholestatic types of cirrhosis.
- **ALP (Alkaline Phosphatase):** ALP is associated with biliary tract function. Increased ALP is characteristic of cholestasis or bile duct obstruction. Elevated ALP levels are commonly found in biliary cirrhosis and other cholestatic disorders.
- **Albumin (ALB):** Albumin is synthesized by the liver and serves as an important indicator of hepatic synthetic capacity. Reduced ALB concentration is a prominent feature of chronic liver failure and is a key predictor in severity scoring systems such as the Child–Pugh classification.
- **A/G Ratio (Albumin/Globulin Ratio):** The A/G ratio looks at the levels of albumin and globulin. Healthy individuals typically have ratios above 1.0, while patients with cirrhosis often show much lower ratios because of reduced albumin production and higher globulin levels due to ongoing inflammation.
- **Total Protein (TP):** TP encompasses both albumin and globulin components. Although TP may remain within normal ranges or mildly elevated, the internal redistribution between albumin and globulin provides supportive evidence of cirrhotic progression.
- **Age:** Age is an important demographic variable, as the likelihood of developing cirrhosis increases with cumulative exposure to hepatotoxic, metabolic, or infectious factors over time.
- **Gender:** Gender-based differences exist in the causes of cirrhosis. Alcohol-induced cirrhosis occurs more often in males, while autoimmune cirrhosis mainly affects females. As a result, gender is an important factor in predictive modeling.
- **Target Variable (Cirrhosis Status):** The binary target variable indicates whether the patient has been clinically diagnosed with cirrhosis (1) or is non-cirrhotic (0), serving as the supervisory label for classification models.

B. Data Preprocessing

Missing values were imputed using iterative chained equations. Numerical features were standardized via Z-score normalization. Gender was encoded using one-hot encoding. The dataset was split into training and testing sets, and class imbalance was handled using SMOTE to ensure fair model learning.

C. Exploratory Data Analysis

1) Distribution Inspection: Q-Q plots were employed to examine normality, revealing that most biochemical variables do not follow Gaussian distributions. Histogram–violin comparisons show skewness and wide variances between healthy and cirrhotic patients.

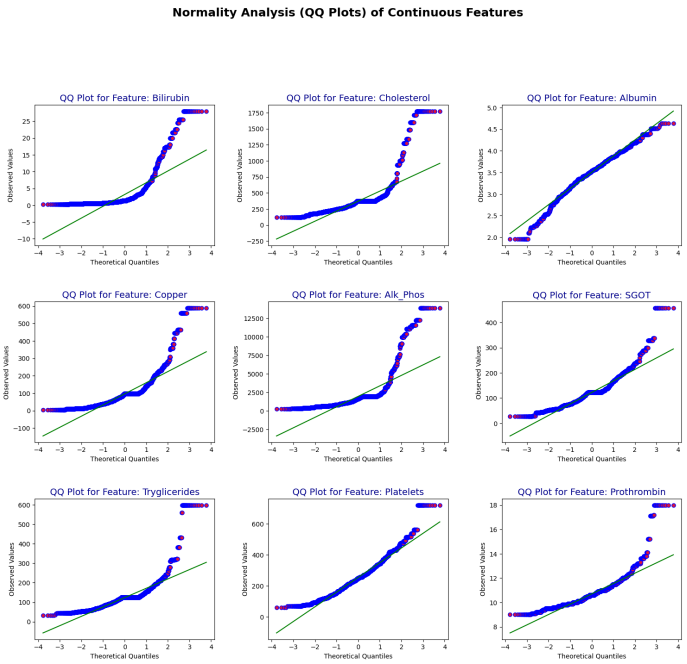

Fig. 2. Q-Q plot

2) Outlier Detection: Boxplots illustrate elevated and dispersed enzyme levels in cirrhotic patients, with notable outliers in AST, ALT, and Bilirubin. These outliers are clinically meaningful and are preserved rather than removed. The outlier distribution presented in Table VI provides clinically meaningful information on the progressive deterioration of hepatic function in the stages of cirrhosis. As the disease advances, biochemical markers shift further from their normal ranges. This reflects growing problems in liver cell integrity, bile flow, metabolism, and synthetic ability. To be specific, markers like Bilirubin, Copper, SGOT, and Alkaline Phosphatase show much higher outlier levels in Stage 3, which indicates severe bile buildup, liver cell death, and disruption of liver transport processes. In contrast, markers such as Albumin and Prothrombin, both dependent on liver synthetic function, show disproportionately high outlier rates in late-stage cirrhosis, consistent with impaired protein synthesis and coagulopathy.

The consistent rise in abnormal values at different stages highlights the changes in cirrhosis. As fibrosis advances, biochemical imbalances become clearer. These trends show that clinical biomarkers are useful not just for diagnosis but also

as measurable indicators of how serious the disease is. Outlier profiles act as important diagnostic signals. They give an early warning of liver failure and improve the effectiveness of machine learning models for staging and predicting cirrhosis.

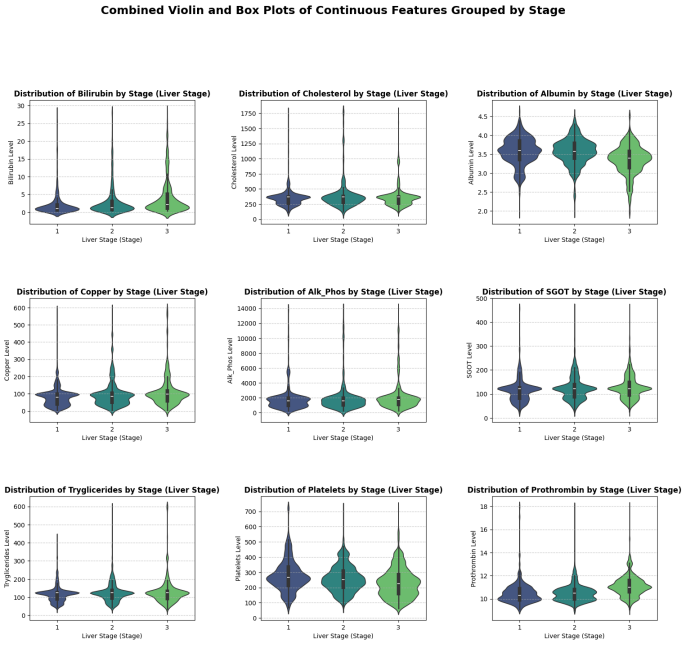

Fig. 3. Box plot and violin plot

TABLE I
OUTLIER DISTRIBUTION IN EACH CIRRHOSIS STAGE

Feature	Stage 1	Stage 2	Stage 3	Total Outliers
Bilirubin	16.98%	33.58%	49.43%	1060
Cholesterol	22.82%	45.38%	31.79%	758
Albumin	8.24%	19.61%	72.16%	255
Copper	14.40%	37.59%	48.01%	854
Alk_Phos	28.61%	33.38%	38.01%	755
SGOT	21.68%	43.81%	34.51%	452
Tryglicerides	22.49%	42.47%	35.13%	817
Platelets	62.59%	13.67%	23.74%	139
Prothrombin	21.74%	9.88%	68.38%	253
Total per stage	**1136**	**1912**	**2295**	**5343**

Note: Outliers detection based on theory of quantile detection with UB = $Q_3 + 1.5 * IQR and LB = Q_1 - 1.5 * IQR$.

3) Statistical Calculations: The statistical distributions shown in Table II show significant changes in clinical biomarkers across cirrhosis stages. These changes appear in the mean and standard deviation, along with other distribution characteristics like skewness and kurtosis. As the disease advances from Stage 1 to Stage 3, most biomarkers show rising mean concentrations. For example, Bilirubin, Copper, Alk_Phos, and SGOT levels increase. This indicates worsening metabolic dysfunction and reduced hepatocellular clearance. At the same time, reductions in markers like Albumin and Platelets show a decline in synthetic capacity and hematological issues common in advanced cirrhosis.

Moreover, the clear skewness and heavy-tailed kurtosis were seen in several features, particularly Bilirubin, Copper, Alk_Phos, and Triglycerides, suggest that the distributions in later stages include a higher proportion of extreme pathological values. This occurrence indicates that the biochemical imbalance intensifies as cirrhosis advances. Consequently, there are more asymmetric distributions that are heavy on outliers. On the other hand, certain markers, such as Albumin, exhibit a negative skew. This behavior corresponds with a consistent decline in the liver's production of proteins instead of irregular spikes.

The differences seen in these distribution measures suggest that the underlying data not only change in size but also vary in structure depending on the severity of the disease. These distributional discrepancies are expected to be statistically significant, and they motivate the subsequent application of inferential tests such as ANOVA (for continuous biomarkers) and Chi-square (for categorical clinical attributes). Such hypothesis-testing procedures let us confirm whether inter-stage variation is statistically significant. This helps us find biomarkers that have strong discriminative power for machine learning models later on.

4) Statistical Testing: One-way ANOVA tests confirm significant differences between healthy and cirrhotic groups across most numerical biomarkers [1], [3]. Chi-square analysis shows strong associations between categorical variables and disease stage. Overall, the features exhibit pronounced differences across stages, particularly for clinical measurements, suggesting that models are likely to learn effectively from those features with substantial inter-class separation.

D. Principal Component Analysis

PCA was performed to explore structure and dimensionality reduction. The first components explain most of the variance, and correlation matrices after PCA reveal transformed relationships among biomarkers.

- The PCA analysis reveals that the first principal component (PC1) captures a high variance among the features, indicating their strong contribution to the differentiation of the disease.
- Features with high positive loadings, such as Bilirubin (0.75), Copper (0.59), SGOT (0.53), and Tryglicerides (0.45), suggest that increases in these biomarkers are associated with a higher risk of cirrhosis.
- Cholesterol (0.44) and Alk_Phos (0.34) also contribute positively, indicating moderate associations with disease progression.
- Albumin (-0.49) shows a negative load, implying that lower levels of Albumin are associated with greater severity of the disease.
- Prothrombin (0.36) contributes moderately, reflecting its relevance in distinguishing stages of liver dysfunction.
- Overall, the variance captured by PC1 highlights that changes in these clinical features strongly influence the risk and stage of cirrhosis, supporting their importance for predictive modeling.

TABLE II
STATISTICAL DISTRIBUTION (MEAN, STD, KURTOSIS, SKEWNESS) ACROSS CIRRHOSIS STAGES

Feature	Stage	Mean	Std	Kurtosis	Skewness
Bilirubin	1	2.25	3.64	18.99	4.05
	2	3.07	4.43	9.81	2.98
	3	4.16	4.96	4.22	2.10
Cholesterol	1	355.01	162.19	27.96	4.36
	2	396.79	252.70	13.72	3.52
	3	362.29	168.43	15.70	3.30
Albumin	1	3.58	0.36	0.28	-0.40
	2	3.58	0.34	1.10	-0.47
	3	3.36	0.38	0.99	-0.59
Copper	1	81.84	54.44	12.80	2.29
	2	95.24	71.43	7.05	2.23
	3	110.05	85.71	10.45	2.78
Alk_Phos	1	1843.03	1678.36	18.84	3.85
	2	1955.38	1899.15	14.26	3.61
	3	2086.88	1889.31	10.76	3.11
SGOT	1	114.96	47.80	9.45	1.95
	2	124.02	49.00	4.33	1.36
	3	126.44	46.51	2.31	1.06
Triglycerides	1	116.56	43.30	7.54	1.81
	2	124.92	50.30	12.90	2.24
	3	128.42	68.98	21.97	3.96
Platelets	1	277.24	102.32	2.11	0.96
	2	258.86	85.36	0.32	0.53
	3	230.05	95.02	0.57	0.70
Prothrombin	1	10.49	0.95	22.60	3.61
	2	10.50	0.79	19.15	2.81
	3	11.10	0.86	3.46	1.03

Notes:
- Kurtosis > 3: heavy-tailed distribution (high probability of extreme values).
- Skewness > 0: right-skewed distribution, common in biochemical markers.
- Stage 3 includes combined Stage 3 and Stage 4.
- Values rounded to 2 decimal places for clarity.

TABLE III
CHI-SQUARE INDEPENDENCE TEST BETWEEN CATEGORICAL FEATURES AND CIRRHOSIS STAGE

Categorical Feature	vs Stage	χ^2 Statistic	p-value
Status	Stage	608.45	<0.001
Drug	Stage	11.19	<0.004
Sex	Stage	24.81	<0.001
Ascites	Stage	117.60	<0.001
Hepatomegaly	Stage	1007.35	<0.001
Spiders	Stage	234.71	<0.001
Edema	Stage	686.57	<0.001

IV. MACHINE LEARNING MODELS

We implement and evaluate:

- Softmax Regression
- Support Vector Machine (RBF kernel)
- K-Nearest Neighbors (K=5)
- Random Forest (200 trees)
- XGBoost (learning rate 0.1, depth 8)

TABLE IV
DIFFERENCE TESTS FOR NUMERICAL FEATURES ACROSS CIRRHOSIS STAGES

Numerical Feature	Test	Statistic	p-value
Bilirubin	ANOVA	137.79	<0.001
	Kruskal-Wallis	644.98	<0.001
Cholesterol	ANOVA	36.80	<0.001
	Kruskal-Wallis	16.96	<0.001
Albumin	riverside ANOVA	370.34	<0.001
	Kruskal-Wallis	702.88	<0.001
Copper	ANOVA	110.20	<0.001
	Kruskal-Wallis	207.89	<0.001
Alk_Phos	ANOVA	12.83	<0.001
	Kruskal-Wallis	57.08	<0.001
SGOT	ANOVA	45.55	<0.001
	Kruskal-Wallis	102.60	<0.001
Tryglicerides	ANOVA	34.14	<0.001
	Kruskal-Wallis	54.74	<0.001
Platelets	ANOVA	184.67	<0.001
	Kruskal-Wallis	370.19	<0.001
Prothrombin	ANOVA	485.00	<0.001
	Kruskal-Wallis	1304.17	<0.001

TABLE V
POST-HOC PAIRWISE COMPARISONS (DUNN TEST WITH BONFERRONI CORRECTION)

Feature	Group 1 vs Group 2	p-value	Significant
Bilirubin	Stage 1 vs 2	<0.001	Yes
	Stage 1 vs 3	<0.001	Yes
	Stage 2 vs 3	<0.001	Yes
Cholesterol	Stage 1 vs 2	<0.001	Yes
	Stage 1 vs 3	<1.000	No
	Stage 2 vs 3	<0.005	Yes
Albumin	Stage 1 vs 2	<1.000	No
	Stage 1 vs 3	<0.001	Yes
	Stage 2 vs 3	<0.001	Yes
Copper	Stage 1 vs 2	<0.001	Yes
	Stage 1 vs 3	<0.001	Yes
	Stage 2 vs 3	<0.001	Yes
Alk_Phos	Stage 1 vs 2	<0.115	No
	Stage 1 vs 3	<0.001	Yes
	Stage 2 vs 3	<0.001	Yes
SGOT	Stage 1 vs 2	<0.001	Yes
	Stage 1 vs 3	<0.001	Yes
	Stage 2 vs 3	<0.040	Yes
Tryglicerides	Stage 1 vs 2	<0.001	Yes
	Stage 1 vs 3	<0.001	Yes
	Stage 2 vs 3	<1.000	No
Platelets	Stage 1 vs 2	<0.001	Yes
	Stage 1 vs 3	<0.001	Yes
	Stage 2 vs 3	<0.001	Yes
Prothrombin	Stage 1 vs 2	<0.401	No
	Stage 1 vs 3	<0.001	Yes
	Stage 2 vs 3	<0.001	Yes

Model performance is evaluated using accuracy, precision, recall, F1-score, and AUC. Hyperparameters are tuned using 5-fold cross-validation.

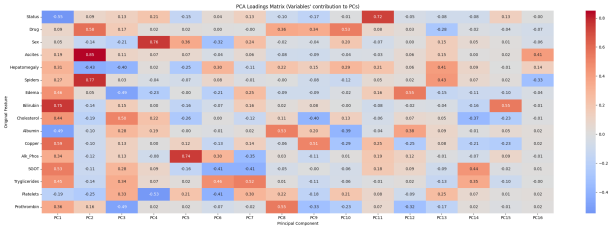

Fig. 4. PCA loading matrix

TABLE VI
MODELS EVALUATION

Model	Accuracy	F1 score	Precision	Recall
Softmax Regression	56.14%	55.35%	55.61%	56.14%
Support Vector Machine	55.51%	54.45%	54.81%	55.51%
K Nearest neighbors	72.05%	71.97%	71.85%	72.05%
Random Forest	**84.66%**	**84.59%**	**84.63%**	**84.66%**
XGBoost	**88.47%**	**88.44%**	**88.43%**	**88.47%**

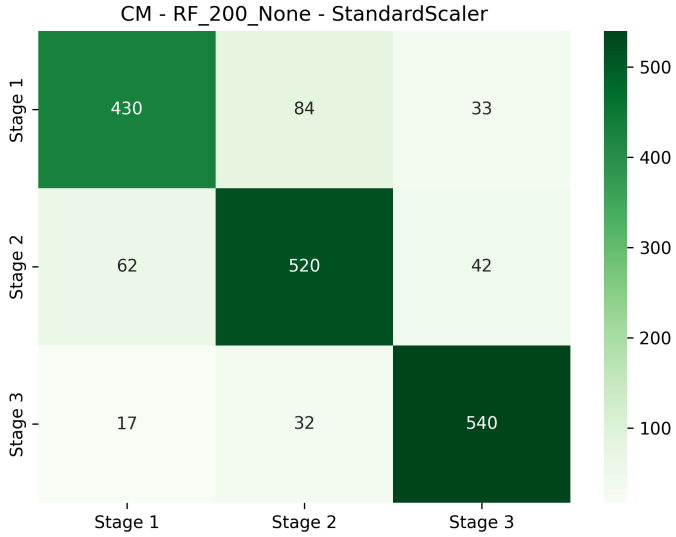

Fig. 5. Random Forest confusion matrix

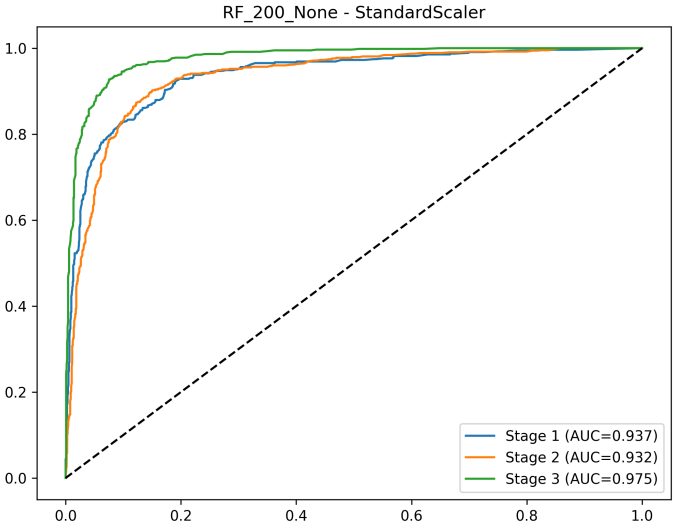

Fig. 6. Random Forest Roc curve

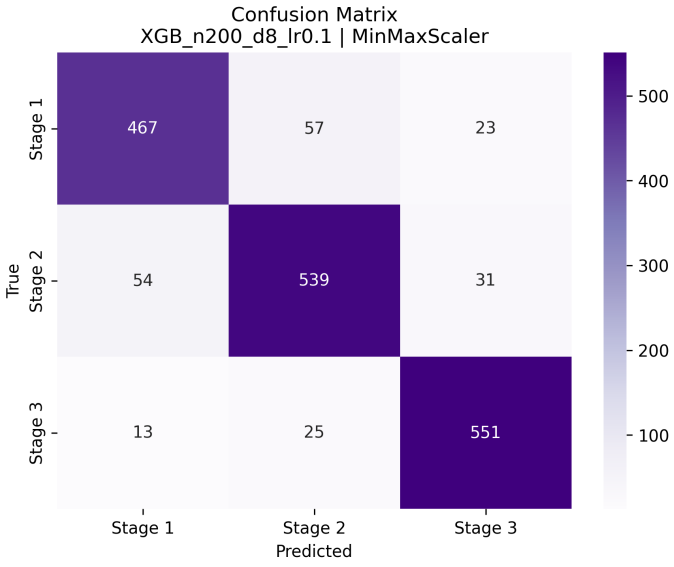

Fig. 7. XGBoost confusion matrix

V. RESULTS

Table VI summarizes the performance of different classification models in predicting cirrhosis stages. Among the models evaluated, Softmax Regression and Support Vector Machine (SVM) achieved relatively low performance, with accuracies of 56.14% and 55.51%, respectively, indicating limited capability in capturing complex patterns in the dataset. K Nearest Neighbors (KNN) showed improved performance with an accuracy of 72.05%, suggesting that local neighborhood information contributes to better classification.

Ensemble-based methods demonstrated superior predictive ability. Random Forest achieved an accuracy of 84.66%, while XGBoost outperformed all other models with the highest accuracy of 88.47%, along with corresponding improvements in F1 score, precision, and recall. These results show that tree-based ensemble methods are very good at learning from clinical features and understanding the nonlinear relationships involved in distinguishing different stages of cirrhosis. Ensemble-based models perform better than linear and distance-based methods. XGBoost reaches the highest accuracy and F1-score, with Random Forest slightly lower. Moreover, ROC curves demonstrate that both models achieve strong AUC values. In contrast, Softmax Regression and KNN are less effective because they rely on linearity and are sensitive to the scales of the features.

Statistical analysis aligns with model findings: Bilirubin levels, ALP, AST, and the A/G ratio emerge as key predictive features.

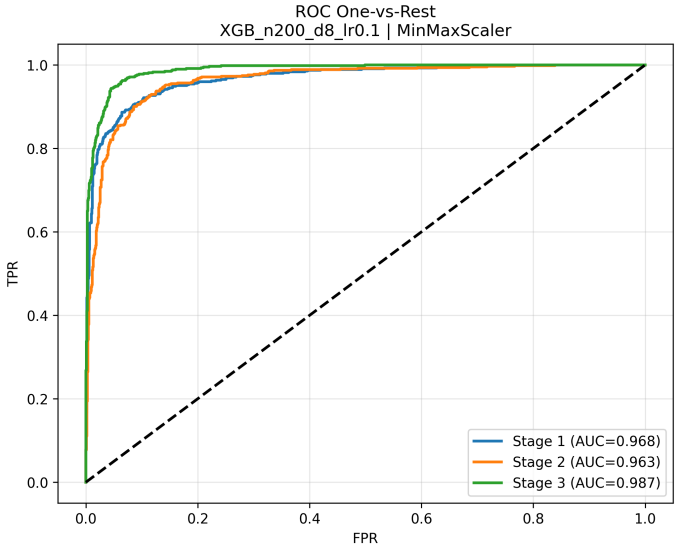

Fig. 8. XGBoost Roc curve

VI. Discussion

A. The Significance of Integrated Statistical and Visual Analysis

Our study began with an integrated approach to Exploratory Data Analysis (EDA) and Statistical Inference, which proved essential for characterizing the complexity of liver cirrhosis biomarkers [1], [3], [4]. The initial visualization, using Q-Q plots (Fig. 2), boxplots, and violin plots (Fig. 3), confirmed that most clinical biomarkers, such as Bilirubin, Copper, and Alkaline Phosphatase (Alk_Phos), exhibit non-Gaussian, right-skewed distributions and contain significant outliers. This is not merely a statistical artifact; it is a pathophysiological signal reflecting the extreme biochemical disruption seen in advanced liver disease.

The systematic distribution of outliers across disease stages, as detailed in Table III, validates this. Markers of impaired hepatic synthesis, notably Albumin and Prothrombin, show the highest outlier proportions in Stage 3 (72.16% and 68.38%, respectively), underscoring the severe loss of functional liver mass at the late stage. Conversely, Bilirubin and Copper, reflecting cholestasis and metabolic accumulation, show a progressive increase in outliers from Stage 1 to Stage 3. This distributional understanding guided the necessary robust preprocessing techniques, confirming that retaining these clinically meaningful outliers was crucial for the predictive task.

B. Discriminative Power of Clinical Biomarkers

The inferential statistical tests formally quantified the discriminative potential of the features. The ANOVA and Kruskal-Wallis tests (Table V) showed that all numerical biomarkers exhibited a statistically significant difference across cirrhosis stages ($p<0.001$). This confirms that biochemical changes are not random but systematically track the severity of the disease.

The Post-hoc pairwise comparisons (Dunn test with Bonferroni correction) in Table VI provided finer clinical insight. Key markers like Bilirubin, Copper, and Platelets showed significant differences across all stage pairs (Stage 1 vs 2, 1 vs 3, and 2 vs 3), making them highly effective features for a machine learning classifier. Markers of synthetic function, such as Albumin and Prothrombin, showed significant drops only from Stage 2 to Stage 3, suggesting they are more effective indicators of the transition to late-stage (decompensated) cirrhosis, rather than early-stage detection.

Furthermore, the Principal Component Analysis (PCA) (Fig. 4 and associated text) revealed that the first principal component is strongly loaded by the most clinically relevant markers (Bilirubin, Copper, Albumin, and SGOT). This highlights that a combination of synthetic, excretory, and metabolic markers collectively accounts for the largest variance in distinguishing patient outcomes, directly supporting the feature selection underlying our machine learning pipeline.

C. Comparative Model Performance and Clinical Utility

The comparison of machine learning models (Table 7) showed that ensemble methods are better than traditional linear and instance-based classifiers. Softmax Regression and SVM had difficulty reaching consistent accuracy, performing just a bit better than random chance. This lack of performance validates the initial hypothesis from the EDA: the relationships between clinical biomarkers and cirrhosis status are highly non-linear and complex, making simple linear decision boundaries ineffective.

In contrast, Random Forest and XGBoost achieved the highest overall performance, with XGBoost reaching an accuracy of 88.47% and a high F1-score (88.44%). The ROC curves (Fig. 7, 9) confirm that these models maintain high predictive stability across various thresholds. The superior performance of XGBoost stems from its ability to: (1) handle the complex, non-linear interactions among biomarkers, (2) effectively manage the skewness and outliers that characterize the dataset (as revealed by the EDA), and (3) iteratively refine predictions, concentrating on hard-to-classify samples.

This finding has significant clinical implications. An XGBoost model trained on routine, non-invasive lab tests provides an accuracy that approaches or even exceeds the diagnostic confidence derived from traditional screening methods. The model offers a low-cost, rapid, and non-invasive tool that can be integrated into clinical settings to:

- Prioritize high-risk patients for immediate imaging or biopsy.
- Reduce diagnostic overload on clinicians by providing a robust second-opinion risk assessment.
- Support large-scale, early screening efforts in primary care, potentially leading to earlier intervention and improved patient outcomes.

VII. Conclusion

This work shows that machine learning methods used on common clinical biomarkers can accurately classify liver cir-

rhosis. In particular, XGBoost and Random Forest provide the best predictive performance and clinical insight. Furthermore, statistical visualization and hypothesis testing help us understand disease-related patterns better. Finally, they also support feature selection based on biological

VIII. Future Work

Future extensions include:

- Expanding to larger or multi-center datasets
- Integrating imaging features (CT, MRI)
- Applying deep learning models such as TabNet or DNNs
- Using explainable AI (SHAP, LIME) for interpretability
- Developing clinical decision-support applications

References

[1] A. Ismaiel, K. Evrard, D.-C. Leucuta, S.-L. Popa, C. S. Catana, D. L. Dumitrascu and T. Surdea-Blaga, "The Impact of Non-Invasive Scores and Hemogram-Derived Ratios in Differentiating Chronic Liver Disease from Cirrhosis," *J. Clin. Med.*, vol. 14, no. 9, 3072, 2025. :contentReference[oaicite:0]index=0

[2] S. Zeng, Z. Liu, B. Ke, et al., "The non-invasive serum biomarkers contributes to indicate liver fibrosis staging and evaluate the progress of chronic hepatitis B," *BMC Infect. Dis.*, vol. 24, 638, 2024. :contentReference[oaicite:1]index=1

[3] P. Zhen, "A Statistical Analysis of Chronic Liver Disease Diagnosis with Noninvasive Biomarkers," in *Proceedings of the 2022 International Conference on Biomedical Engineering and Bioinformatics (BIOINFORMATICS/BIOSTEC)*, 2022. :contentReference[oaicite:2]index=2

[4] T. Xu, Y. Fang, A. Rong and J. Wang, "Flexible combination of multiple diagnostic biomarkers to improve diagnostic accuracy," *BMC Med. Res. Methodol.*, vol. 15, 94, 2015. :contentReference[oaicite:3]index=3

[5] Scikit-learn Developers, "Hyper-parameter Tuning (GridSearch cross-validation)," *scikit-learn.org* — official documentation.

[6] A recent review: "Non-invasive Biomarkers and Tests for Diagnosis and Monitoring of Chronic Liver Diseases," summarizing serum panels, elastography, and other non-invasive modalities. :contentReference[oaicite:4]index=4