Predicting Histological Progression in Primary Biliary Cirrhosis Using Advanced Machine Learning Techniques

Laberiano Andrade-Arenas[1], Cesar Yactayo-Arias[2], Inoc Rubio Paucar[3]
Facultad de Ciencias e Ingeniería, Universidad de Ciencias y Humanidades, Lima, Perú[1]
Departamento de Estudios Generales, Universidad Continental, Lima, Perú[2]
Facultad de Ingeniería y Negocios, Universidad Privada Norbert Wiener, Lima, Perú[3]

***Abstract*—Cirrhosis is considered one of the most serious liver diseases worldwide, closely related to excessive alcohol consumption and inadequate eating habits, factors that progressively deteriorate people's health. In this context, the present research aimed to develop and validate a predictive model based on Machine Learning (ML), specifically using the Random Forest (RF) algorithm, to determine the histological stage (from 1 to 4) in patients with primary biliary cirrhosis. For the development of the study, the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology was applied, which includes the stages of business understanding, data understanding, data preparation, modeling, and evaluation. The results obtained showed that the model presents better performance in the more advanced stages of the disease. The area under the curve (AUC) increased from 0.612 in Stage 1 to 0.874 in Stage 4, reflecting a notable improvement in its discriminative capacity. Similarly, metrics such as sensitivity, precision, and F1-score showed an upward trend, reaching their highest values in Stage 4. In this sense, the proposed model represents a complementary diagnostic support tool, since it allows estimating the histological stage through the analysis of clinical data, contributing to medical decision-making without relying exclusively on invasive procedures.**

***Keywords*—Clinical decision support; histological stage prediction; primary biliary cirrhosis; random forest**

I. Introduction

Currently, hepatic cirrhosis is a chronic disease that continues to represent a serious public health problem worldwide according to the World Health Organization (WHO) [1], due to its high rate of complications and mortality. It is characterized by progressive and irreversible liver damage, which significantly affects the quality of life of those who suffer from it [2]. Various factors, such as viral infections, prolonged alcohol consumption, and metabolic disorders, contribute to its development. In this scenario, early diagnosis and proper risk assessment are essential to prevent disease progression and reduce the occurrence of severe complications, such as liver failure or liver cancer [3]. Thus, the clinical and laboratory information generated during medical care plays a key role in understanding the evolution of cirrhosis and supporting timely medical decision-making.

However, the clinical management of hepatic cirrhosis still presents significant difficulties. On the one hand, the disease does not evolve in the same way in all patients, which means that traditional assessment tools are not always sufficient to accurately predict its progression. Thus, commonly used clinical models, such as the Child-Pugh or MELD scales, although useful, may be limited when applied in isolation [4],[5]. In addition, there is the constant growth of available medical information, coming from electronic medical records and laboratory results, whose interpretation can be complex and demanding for healthcare professionals. In this sense, the lack of automated tools that allow comprehensive analysis of clinical and laboratory information can delay diagnosis [6], hinder the early identification of patients at higher risk, and affect the planning of appropriate and personalized treatments for each case. This situation is especially concerning considering that cirrhosis is a disease whose progression can occur silently, without the person being fully aware of its advancement, until more severe complications appear. As a consequence, many patients are diagnosed at advanced stages, when therapeutic options are more limited and the impact on quality of life is considerable. Therefore, cirrhosis should be considered a disease with accelerated progression in the absence of timely control, highlighting the importance of early identification of its main causes and risk factors.

Facing these limitations, ML emerges as an alternative capable of providing practical solutions for the analysis of complex medical data. By using algorithms that learn from data, it is possible to identify patterns and relationships that are not always evident in conventional analysis [7],[8]. In the context of hepatic cirrhosis, these techniques allow the simultaneous analysis of multiple clinical and laboratory variables, facilitating the prediction of disease evolution and the risk of developing complications. Furthermore, ML models can be continuously updated as new data are incorporated, making them flexible and adaptable tools that effectively support clinical decision-making and contribute to more personalized medical care [9]. Thus, ML is presented as a viable and promising option to support decision-making related to personalized treatments and medical assistance. By jointly analyzing large volumes of clinical data, laboratory results, and patient history, these techniques allow the identification of relevant patterns that facilitate a more accurate evaluation of each case [10]. This not only contributes to selecting more appropriate therapeutic strategies according to individual patient characteristics but also improves disease monitoring, optimizes clinical resources, and strengthens evidence-based

medical decision-making.

The objective of this research is to develop and validate a predictive model based on the RF ML algorithm for determining the histological stage (1 to 4) in patients with primary biliary cirrhosis, using exclusively clinical variables and non-invasive biochemical parameters. This approach seeks to reduce dependence on liver biopsy as a diagnostic method, proposing an alternative based on statistical and computational analysis of serum biomarkers. In this way, it is intended to estimate the degree of progression of liver diseases, particularly cirrhosis, through ML techniques that allow a precise, safe, and reproducible evaluation without resorting to invasive procedures.

II. Literature Review

In this section, the literature review focuses on fundamental elements, such as the theoretical foundations, which include the study variables related to cirrhosis and the RF algorithm. Likewise, related works are considered, which comprise research conducted by various authors who have contributed to scientific knowledge and social development through the application of ML techniques in the study and analysis of cirrhosis.

A. Theoretical Bases

1) Cirrhosis: The topic presented about cirrhosis is known as the result of chronic liver injury, which can be caused by various conditions such as viral infection, excessive alcohol consumption, non-alcoholic fatty liver disease (NAFLD/EHGNA), or autoimmune processes. Thus, this prolonged injury induces diffuse fibrosis and the formation of regenerative nodules, altering both hepatic blood flow and normal liver function [11],[12]. On the other hand, in its initial stages, compensated cirrhosis may not present symptoms; however, when it progresses to a decompensated phase, complications such as ascites, variceal bleeding, hepatic encephalopathy, and organ failure are frequently observed. It is important to consider that fibrosis progression is mainly mediated by the activation of hepatic stellate cells and extracellular matrix deposition, processes regulated by transforming growth factor beta (TGF-β) [13]. Thus, cirrhosis is not only a liver disease: it is a process that, over time, can generate various complications that affect the daily life of those who suffer from it, as shown in Table I, ranging from fluid accumulation in the abdomen (ascites) to bleeding, mental confusion (encephalopathy), or serious problems in other organs.

2) Random forest: Within the research, RF is considered a viable option to help reduce disease progression. Subsequently, the RF model is a supervised learning algorithm based on the principle of ensemble learning using multiple independent decision trees [14],[15]. One characteristic is that each tree is constructed from a random subset of the training data using bootstrap sampling, and is trained by randomly selecting a subset of variables at each node to reduce correlation between trees [16]. Consequently, the individual predictions of each tree are combined through majority voting in classification problems or averaging in regression problems [17], allowing the development of a robust model capable of minimizing overfitting, reducing variance, and improving overall accuracy. RF is based on statistical concepts of model aggregation and diversity, constituting a solid theoretical tool for predictive analysis in different data contexts due to its conceptual structure and components mentioned in Table II. It is important to consider that this algorithmic model belongs to the ensemble method family and is built from a set of decision trees trained randomly, allowing it to combine multiple predictions to improve the accuracy and robustness of the model.

B. Related Work

Several researchers have focused their efforts on optimizing the diagnosis of hepatic cirrhosis, and some authors [18],[19] developed a study aimed at improving the detection of this disease in patients with chronic hepatitis B. In this sense, the methodology consisted of dividing the sample into training and internal validation sets, on which ML algorithms were applied, including Logistic Regression (LR), K-Nearest Neighbors (KNN), RF, Artificial Neural Network (ANN), Support Vector Machine (SVM), and Extreme Gradient Boosting (XGBoost). Therefore, model performance was evaluated using metrics such as area under the receiver operating characteristic curve (ROC), area under the curve AUC, calibration analysis, clinical decision-making, and result interpretation using Shapley Additive Explanations (SHAP). Consequently, in a cohort of 1609 patients, of whom 470 had cirrhosis, the RF model showed accuracy and AUC values higher than 0.80, demonstrating adequate calibration and clinically relevant benefit, highlighting the potential of ML algorithms to improve early detection and clinical decision-making in hepatic cirrhosis. On the other hand, in other research, the authors [20] showed that the prediction of liver diseases, such as cirrhosis or liver cancer, represents a valuable source of information for evaluating and comparing the performance of ML models in large datasets. In this context, three algorithms were analyzed: KNN, Gaussian Naive Bayes (GNB), and RF, evaluating their performance using key metrics such as precision, accuracy, sensitivity, and F1-score [21]. The results indicated that RF outperformed the other models, achieving a precision of 97.3%, accuracy of 97%, sensitivity of 96%, and F1-score of 95%, demonstrating its ability to accurately classify patients with liver diseases and highlighting its potential as a reliable tool for early detection and clinical decision-making. In this context, another relevant study conducted by the authors [22],[23] developed two ML-based algorithmic models for predicting nosocomial infections and mortality risk in cirrhotic patients. To this end, various algorithms were evaluated, selecting the best-performing ones through bootstrapping and external validation, and comparing their accuracy using key metrics such as sensitivity, specificity, predictive values, and likelihood ratios. Among the analyzed models, RF stood out for its superior performance [24]. Specifically, for nosocomial infection prediction, the PIPC-NI model achieved an AUC of 0.784 (95% CI 0.741–0.826), with sensitivity of 0.712 and specificity of 0.702. Regarding hospital mortality, the PIPC-Mortality model showed an AUC of 0.793 (95% CI 0.749–0.836), sensitivity of 0.769, and specificity of 0.701, demonstrating its ability to reliably classify patients and highlighting its usefulness as a clinical decision support tool and in early risk management in hospitalized cirrhotic patients. Another important aspect concerns hepatocellular

TABLE I. MAIN COMPOUNDS AND CHEMICAL PROCESSES ASSOCIATED WITH CIRRHOSIS

Process / Factor	Chemical compound / Biomarker	Clinical relevance
Hepatic fibrosis	Type I and III collagen, fibronectin, laminin	Accumulation in the extracellular matrix that contributes to liver stiffness and the formation of regenerative nodules.
Oxidative stress	Free radicals (ROS), glutathione, lipid peroxides	Promotes hepatocellular damage, activation of stellate cells, and progression of fibrosis.
Chronic inflammation	Pro-inflammatory cytokines (TNF-α, IL-6), transforming growth factor beta (TGF-β)	Mediated by immune response, stimulates fibrogenesis, and alters hepatic homeostasis.
Altered lipid metabolism	Triglycerides, free fatty acids, lipoproteins	Contributes to steatosis and liver damage in cirrhosis associated with NAFLD/EHGNA.
Hepatic detoxification	CYP450 enzymes, bilirubin, ammonia	Reduced activity in cirrhosis affects drug metabolism and detoxification of nitrogenous compounds.

TABLE II. CONCEPTUAL STRUCTURE AND COMPONENTS OF RF

Component	Key element	Function	Highlighted property
Decision tree	Root node	Initial data splitting	Decision hierarchy
	Leaves	Final output	Categorical or continuous prediction
Bagging	Bootstrap sampling	Generates training subsets	Variance reduction
Random feature selection	Variables per node	Increases tree diversity	Improves generalization
Ensemble	Set of trees	Combination of predictions	Robustness and stability
Aggregation	Voting / Averaging	Final model prediction	Bias-variance balance

carcinoma (HCC), which is a high-risk malignant liver tumor whose early detection is crucial. Therefore, in this study, 550 patients were analyzed, divided into training and validation cohorts, and LR, SVM, RF, and LASSO Regression models were developed, evaluating their performance using ROC curves, AUC, calibration, and decision curve analysis (DCA), comparing them with the ASAP model [25]. Consequently, the RF algorithm showed the best discrimination, with an AUC of 0.969 in training and 0.858 in validation, outperforming the other models and demonstrating robust calibration and clinical validity, highlighting its usefulness for early HCC diagnosis and decision-making in patients at risk of liver cancer.

In accordance with the above, [26] proposes a paradigm shift in cirrhosis identification by processing unstructured clinical language from electronic health records (EHR). Using Convolutional Neural Networks (CNN), their methodology achieved a diagnostic accuracy of 0.965, demonstrating that artificial intelligence can phenotype the disease more effectively than traditional ICD codes. Unlike this text-based data approach, [27],[28] explore non-invasive detection through physical biomarkers, employing deep learning networks to analyze volatile molecular patterns in exhaled breath. This system not only achieved 90% accuracy but also guaranteed 100% sensitivity, positioning it as a viable alternative to intrusive diagnostic tests in hospital settings. Nevertheless, facing the complexity of deep learning models, [29] defend the usefulness of low-cost and high-accessibility tools for community risk stratification. Their research determined that routine scores such as APRI and FIB-4 possess robust capacity to predict cirrhosis complications over 10 years, additionally concluding that the integration of genetic data provides a marginal improvement of less than 5% in predictive performance [30]. This emphasis on clinical feasibility is further discussed by Mane and Bhosale, who argue that applying data mining techniques to liver function tests (LFT) allows objective identification of hidden patterns. According to [31][32], algorithms such as SVM and RF facilitate early diagnosis by transforming standard biochemical results into precise medical decisions. From a critical perspective, however, [33], [34] warn that the high technical performance of these models does not guarantee healthcare equity. When analyzing liver prediction algorithms, they found significant gender bias where women experienced much higher false-negative rates than men, suggesting that AI systems could digitize and perpetuate pre-existing health inequalities if not properly audited [35]. In contrast to this ethical analysis, [36],[37] provide a technical view of structural limitations, noting that although RF and SVM algorithms are the most stable for automatic cirrhosis detection, their generalization is seriously compromised by the small size of publicly available datasets, which are usually smaller than one thousand samples. Expanding diagnostic scope toward therapeutic prognosis, [38],[39] argue that ML is superior to traditional statistics in predicting outcomes after complex treatments such as surgical resection or transplantation in patients with hepatocellular carcinoma (HCC). Their review highlights that AI's ability to process nonlinear data enables personalized therapeutic strategies with reliability unattainable by conventional regression models. Complementarily, [40],[41] postulate that the future of precision medicine in hepatology lies in multi-omics. Unlike models based on simple clinical variables, these authors integrate transcriptomic and metabolomic data using deep learning to unravel the molecular heterogeneity of end-stage liver disease. On the other hand, [42] reinforces the importance of early cirrhosis detection to prevent progression toward liver cancer, emphasizing that ML is the key tool for managing the dynamic nature of fibrosis. Their technical review stresses that early diagnosis using various algorithms is essential to reduce the global mortality burden associated with this chronic disease. Finally, and in a perspective divergent from studies focused purely on algorithms, [43],[44] underline that successful integration of AI into clinical practice depends on standardized technical infrastructure. The authors emphasize that maintaining PACS systems and data ontology uniformity are indispensable requirements for predictive liver disease models to be interoperable and useful in a digitalized healthcare environment.

In the field of assisted diagnosis, [45], [46] argue that integrating multiple algorithms through a Voting Classifier

increases diagnostic accuracy to 80%, surpassing the variability of individual models. Complementarily, validate the robustness of RF as the most reliable predictor when applying complex physiological parameters, consolidating the transition toward early detection methods that reduce risks associated with human intervention. Nevertheless, while these authors focus on disease identification,[47] specialize this methodology in the oncological context [48], demonstrating that Decision Trees can diagnose cirrhosis in patients with hepatocellular carcinoma with an AUC of 0.85, which is critical for surgical planning. Regarding mortality prediction, there is a clear divergence between traditional statistical methods and deep learning approaches. [49] argue that Deep Neural Networks (DNN) offer superior capacity compared to the clinical MELD-Na standard for long-term survival prognosis based on large-scale electronic record data. This finding is reinforced by [50], who through the CiMM mixed model achieve a more transparent and actionable score than conventional systems, allowing more personalized risk management in cirrhotic patients. However, while [51],[52] use routine clinical data, [53],[54] increase prognostic complexity through multi-omics analysis, demonstrating that combining five specific metabolites can predict mortality in acute liver failure with 93.3% accuracy, surpassing any isolated biomarker. Finally, the biological and genetic component introduces an additional dimension in the literature. Authors [55],[56] propose the FibroGENE model, which integrates IFNL genotype with clinical data to stage fibrosis, marking progress toward precision genetic medicine. Conversely, as mentioned in previous research, authors such as [57],[58] suggest that for mass community practice, genetic data offers marginal benefits compared to simple biochemical scoring systems. This contradiction suggests that the use of AI and genetics should be stratified: while complex models are vital in acute or specialized stages, ML applied to low-cost tests remains the most viable strategy for population-level cirrhosis screening.

III. Methodology

A. Definition of the CRISP-DM Methodology

The construction of the predictive model was based on the CRISP-DM methodology, chosen for its flexible and systematic approach to structuring complex data analysis projects [59]. This methodology allowed proper integration between the processing, modeling, and evaluation stages, ensuring the robustness of the results. In the health field, CRISP-DM organizes the lifecycle of a ML project into interrelated phases that include understanding the clinical context, data understanding, data preparation, modeling, and evaluation, facilitating the transformation of heterogeneous clinical information into useful and clinically relevant knowledge, as shown in Fig. 1. For this study, a clinical database obtained from the Kaggle platform was used [60], selected for its scientific relevance in ML applications in healthcare [61],[62]. To facilitate data understanding, an architecture based on information flow was developed, as shown in Fig. 2. This architecture summarizes each stage of the process in a precise manner, from data acquisition to data analysis. The methodological process was developed from problem understanding to result interpretation, starting with a careful selection of the most important variables for model training, ensuring the reliability of the analysis. Finally, Fig. 3 illustrates the functioning of the RF model, which combines multiple decision trees constructed from preprocessed clinical and biochemical data to generate more robust and reliable predictions.

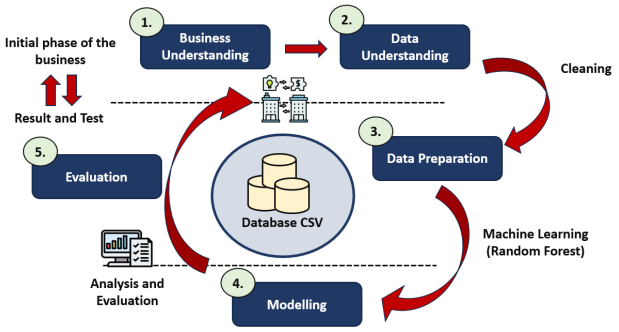


Fig. 1. CRISP-DM methodology.

1) Business understanding: Within this first section, the business understanding stage was initiated, where the problems that motivated the development of the ML model based on primary biliary cirrhosis were analyzed [63]. For this purpose, a descriptive table was presented that clearly organizes the fundamental criteria of the study, including: the justification of the problem, the business problem, the general objective and specific objectives, the expected impact, the business questions, the scope of the study, and the limitations projected and referenced in Table III. This approach allows establishing a solid foundation that guides the development of the model, ensuring that all subsequent stages are aligned with the proposed objectives and facilitate obtaining precise, reliable, and planned results according to the methodological proposal [64],[65]. It is important to highlight that this starting point was considered fundamental, since it allows defining and consolidating the relevant criteria that will be developed and applied in the subsequent stages of the study, ensuring that the analysis and construction of the model are carried out in an orderly manner and aligned with the established objectives.

2) Data understanding: Data understanding is a key stage of the process, as it allows, first, the collection and exploration of the available information within the proposed model. In this way, evaluating the obtained data facilitates the identification of relevant patterns and, at the same time, allows recognizing inconsistencies present in the database. Likewise, it is important to highlight that in this phase no data cleaning or transformation processes are performed, since its main purpose is to understand the state and characteristics of the information, which is essential for making informed decisions in subsequent stages of the analysis [66],[67]. Thus, Table IV presents the dimensionality of the dataset, detailing the variables associated with each dimension, their data type, and a brief description. This allows a visual and structured understanding of the information that will be used in the model, as well as its main characteristics. Consequently, data understanding provides clearer knowledge about data behavior, which is essential for proper interpretation of the information and for improving the performance of the predictive model.

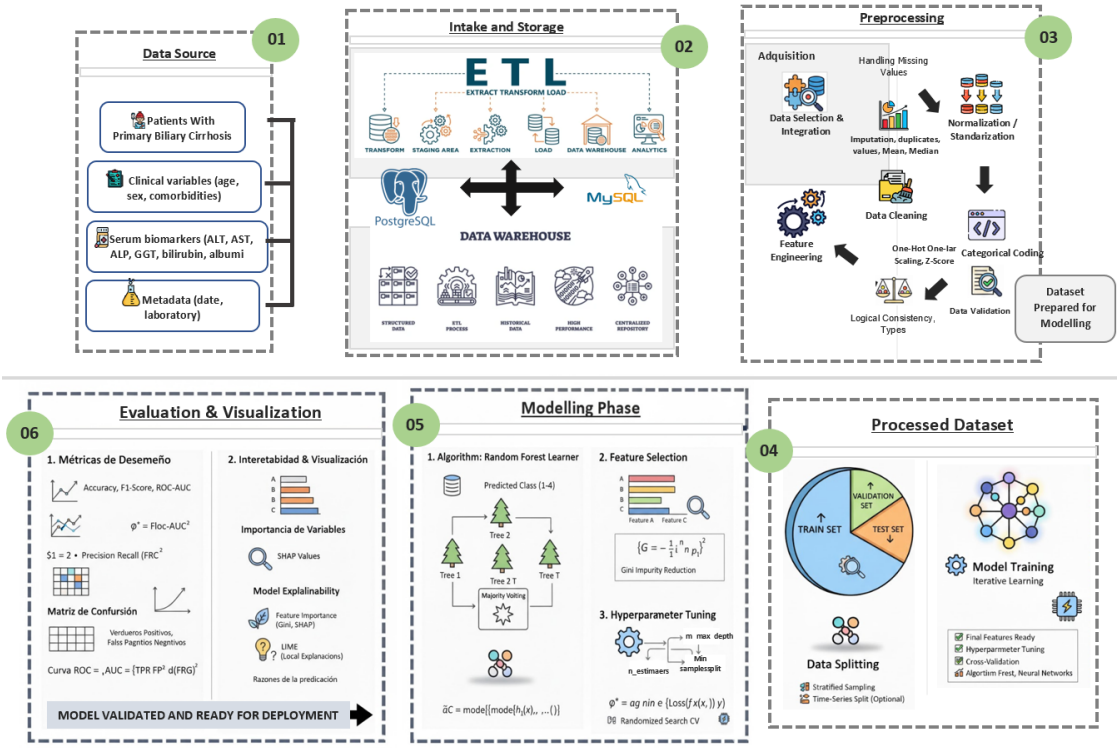


Fig. 2. Data flow architecture.

TABLE III. FIRST PHASE - BUSINESS UNDERSTANDING

Element	Description
Business problem	The determination of the histological stage depends on liver biopsies, an invasive, costly procedure with risk of complications.
General objective	Develop and validate a predictive model based on RF that estimates the histological stage (1–4) using non-invasive clinical and biochemical variables.
Specific objectives	1. Identify the most relevant clinical and biochemical variables. 2. Train and optimize a RF model. 3. Evaluate model accuracy and reliability. 4. Propose a prototype applicable in clinical practice.
Expected impact	- Reduction of invasive procedures. - Faster and safer diagnosis. - Reduction of hospital costs. - Scientific contribution to the use of ML in medicine.
Business questions	- Can histological stage be predicted without biopsy? - Which variables are most relevant for prediction? - What is the accuracy of the RF model? - Is the model reproducible in new patients?
Study scope	The model applies only to patients with primary biliary cirrhosis, using non-invasive clinical and biochemical variables. It does not include other types of liver diseases or invasive evaluation methods.

3) Data preparation: The data preparation phase constitutes an essential component within the analytical process, as it integrates systematic procedures of cleaning, transformation, and structuring aimed at ensuring the quality, coherence, and consistency of the dataset before its incorporation into the predictive model [68]. This stage not only contributes to optimizing model performance but also helps mitigate potential biases associated with outliers, missing values, or inappropriate distributions. In this context, Fig. 4 presents a descriptive analysis that allows characterizing variable behavior through measures of central tendency, such as mean, median, and mode, providing a preliminary overview of their distribution and supporting decision-making during data preparation. Additionally, the set of Tables V summarizes the main processes applied in this phase. Table Va details basic descriptive metrics, including mean, standard deviation, missing values, and their respective percentage per variable [69], [70]. Table Vb summarizes outlier detection, indicating their frequency, percentage, and interquartile range, allowing the evaluation of dispersion and statistical behavior of numerical variables. Finally, Table Vc describes the distribution of patients according to clinical stage, showing the number of cases and their percentage proportion, key information for understanding the dataset structure and its impact on predictive analysis.

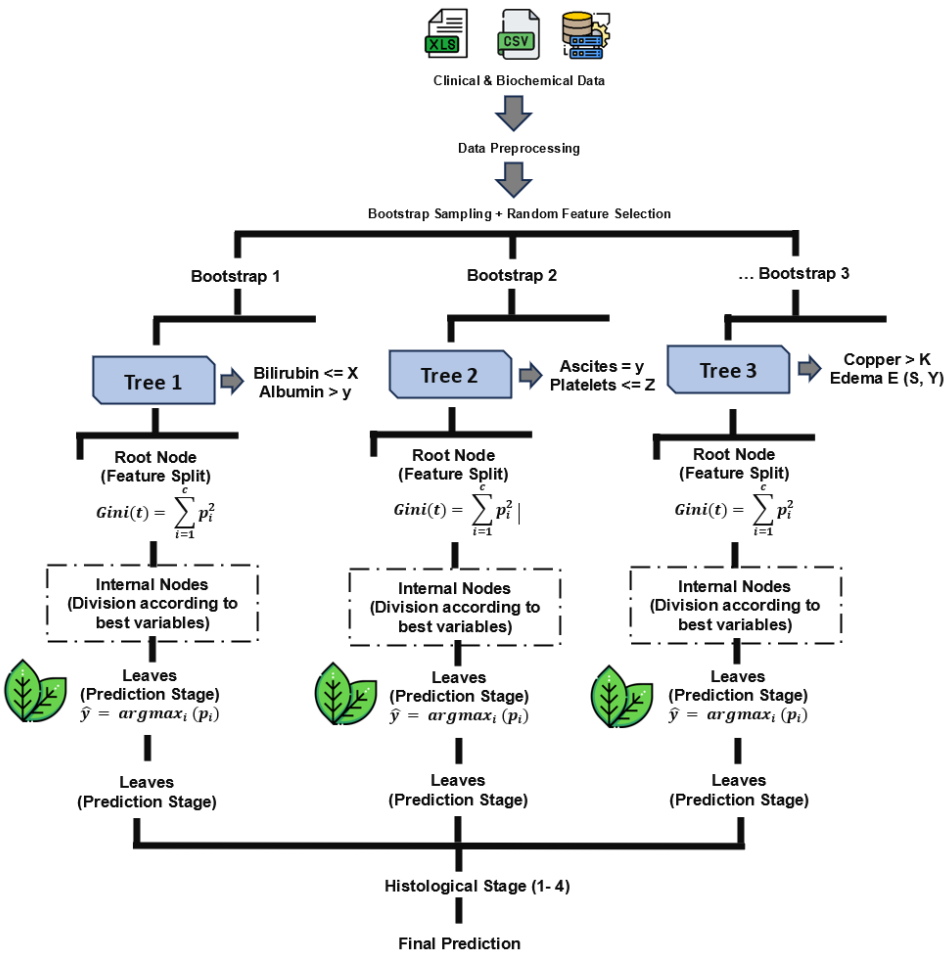


Fig. 3. Random forest model architecture.

IV. Mathematical Formulas for Data Preparation

This section presents the equations used to calculate missing values, descriptive statistics, outliers, and patient distribution, corresponding to Tables V, Va, Vb, and Vc. The summary is as follows [71]:

Percentage of missing values: For each variable X_i with n observations and m_i missing values, the percentage of missing data is calculated as [see Eq. (1)]:

$$\text{Missing}\%_i = \frac{m_i}{n} \times 100 \tag{1}$$

where, m_i is the number of missing values for variable i and n is the total number of observations.

Mean and standard deviation: The mean $\bar{X}_i$ of a numeric variable X_i is defined as [see Eq. (2)], [see Eq. (3)]:

$$\bar{X}_i = \frac{1}{n}\sum_{j=1}^{n} x_{ij} \tag{2}$$

and the standard deviation σ_i as:

$$\sigma_i = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(x_{ij} - \bar{X}_i)^2} \tag{3}$$

where, x_{ij} represents the j-th observation of variable i.

Outlier identification using IQR: Let Q_1 be the first quartile and Q_3 the third quartile of variable X_i. The interquartile range is [see Eq. (4)], [see Eq. (5)], [see Eq. (6)]:

$$\text{IQR}_i = Q_3 - Q_1 \tag{4}$$

Values are considered outliers if:

$$x_{ij} < Q_1 - 1.5 \cdot \text{IQR}_i \quad \text{or} \quad x_{ij} > Q_3 + 1.5 \cdot \text{IQR}_i \tag{5}$$

and the outlier percentage is calculated as:

$$\text{Outlier}\%_i = \frac{\#\text{ of outliers in } X_i}{n} \times 100 \tag{6}$$

Patient distribution by category: For a categorical variable S with k categories (e.g., histological stage) [see Eq. (7)]:

$$\text{Percentage}_k = \frac{n_k}{n} \times 100 \tag{7}$$

where, n_k is the number of patients in category k and n is the total number of patients.

TABLE IV. CLINICAL AND BIOCHEMICAL VARIABLES GROUPED BY DIMENSIONS

Dimension	Variable	Data type	Description
General clinical	Status	Categorical	Patient vital status (C: censored, D: deceased, CL: liver transplant)
	Drug	Categorical	Assigned treatment (D-penicillamine or Placebo)
Demographic	Age	Numeric (integer)	Patient age expressed in days
	Sex	Categorical	Patient sex (M/F)
Clinical signs	Ascites	Categorical	Presence of ascites
	Hepatomegaly	Categorical	Presence of hepatomegaly
	Spiders	Categorical	Presence of spider angiomas
	Edema	Ordinal categorical	Degree of edema (N, S, Y)
Hepatic biochemistry	Bilirubin	Continuous numeric	Serum bilirubin level
	Albumin	Continuous numeric	Serum albumin concentration
	Copper	Continuous numeric	Serum copper level
	Alk_Phos	Continuous numeric	Alkaline phosphatase
	SGOT	Continuous numeric	Aspartate aminotransferase (AST)
	Prothrombin	Continuous numeric	Prothrombin time
Metabolic / Hematological	Cholesterol	Continuous numeric	Serum cholesterol level
	Triglycerides	Continuous numeric	Serum triglyceride level
	Platelets	Continuous numeric	Platelet count
Target variable	Stage	Ordinal categorical	Histological stage of primary biliary cirrhosis (1–4)

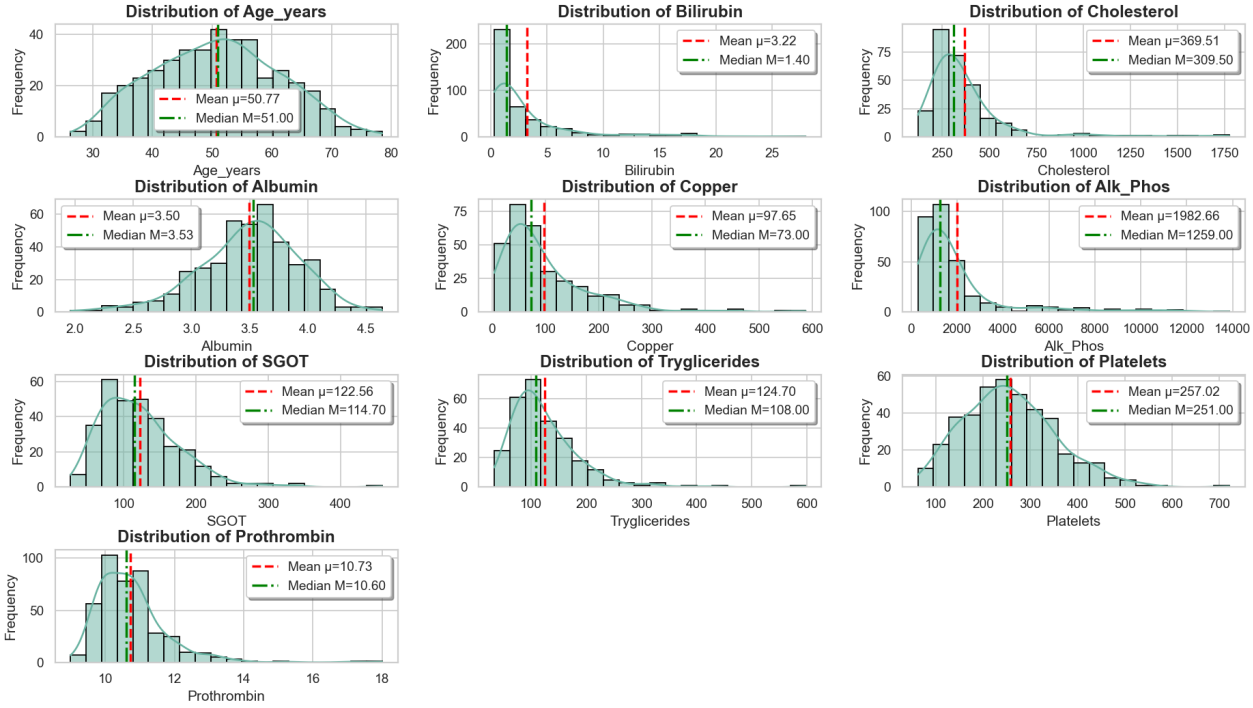


Fig. 4. Descriptive statistics data preparation.

Overall variable quality index (optional): A clean data index can be defined as [see Eq. (8)]:

$$\text{Data Quality Index}_i = 100 - \text{Missing}\%_i - \text{Outlier}\%_i \quad (8)$$

This index reflects the percentage of usable data without missing values or outliers.

1) Modelling: In the Modeling phase, Fig. 5 shows the behavior of the RF model from two complementary perspectives: variable importance and generalization capacity. On one hand, in Fig. 5a, the cumulative importance curve shows that a reduced subset of features concentrates a significant proportion of the total contribution (approximately 33.8% in the three most important variables), indicating that the model identifies variables with high explanatory power and hierarchically structures their influence [72], [73]. On the other hand, the learning curve reveals that training accuracy remains close to 1.0, while validation accuracy is considerably lower and presents fluctuations, evidencing a gap that suggests possible overfitting as represented in Fig. 5b. Overall, these results indicate that although the model successfully captures relevant patterns in the data, it is necessary to optimize its configuration to improve its generalization capacity and ensure greater predictive robustness.

V. MATHEMATICAL FOUNDATIONS OF RF

A. Prediction of a Single Decision Tree

Let T_m be a decision tree trained on a bootstrap sample D_m. It predicts the outcome for a given feature vector $\mathbf{x}$. Each tree uses a different bootstrap sample to increase diversity [74] [see Eq. (9)].

TABLE V. EXTENDED SUMMARY OF DATA PREPARATION PHASE PROCESSES

(A) SUMMARY OF MISSING VALUES

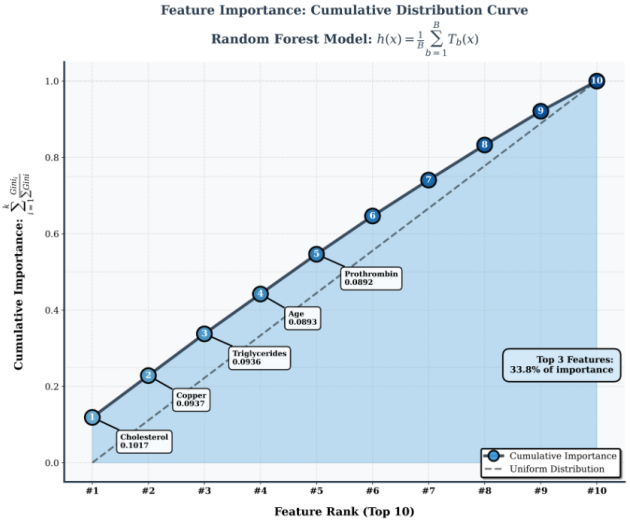
Variable	Type	Mean	Std Deviation	Missing Values	Percentage (%)
Age_years	Numeric	50.74	10.45	0	0.0
Bilirubin	Numeric	3.22	4.41	5	2.3
Cholesterol	Numeric	369.51	231.94	12	5.5
Albumin	Numeric	3.50	0.42	0	0.0
Copper	Numeric	97.65	85.61	7	3.2
Alk_Phos	Numeric	1982.66	2140.39	0	0.0
SGOT	Numeric	122.56	56.70	3	1.4
Tryglicerides	Numeric	124.70	65.15	8	3.7
Platelets	Numeric	257.02	98.33	2	0.9
Prothrombin	Numeric	10.73	1.02	4	1.8
Stage	Ordinal	3.02	0.88	0	0.0

(B) SUMMARY OF VARIABLE OUTLIERS

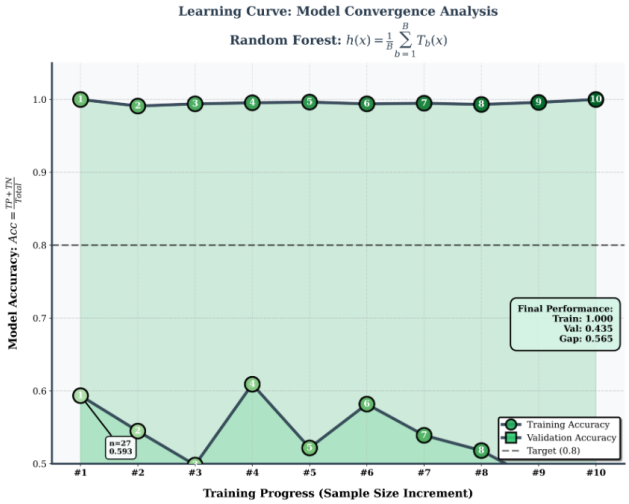
Variable	Outlier Values	Percentage (%)	Interquartile Range
Bilirubin	15	6.9	0.8 - 4.0
Cholesterol	10	4.6	220 - 450
Copper	12	5.5	50 - 110
Alk_Phos	20	9.2	500 - 2500
SGOT	8	3.7	60 - 150
Tryglicerides	7	3.2	70 - 200

(C) DISTRIBUTION OF PATIENTS BY HISTOLOGICAL STAGE

Stage	Number of Patients	Percentage (%)
1	45	30.0
2	60	40.0
3	30	20.0
4	15	10.0
Total	150	100



(a) General data flow architecture.



(b) Detailed modelling process.

Fig. 5. Data flow architecture of the proposed modelling framework.

$$\hat{y}_m(\mathbf{x}) = T_m(\mathbf{x}) \tag{9}$$

Key points:

- $\hat{y}_m(\mathbf{x})$ is the predicted class for classification.
- $\hat{y}_m(\mathbf{x})$ is a numeric estimate for regression.
- Bootstrap sampling increases diversity among trees.

B. RF Prediction

The RF aggregates predictions from M decision trees. It uses majority voting for classification. For regression, it averages the predictions to reduce variance [see Eq. (10)].

$$\hat{y}_{RF}(\mathbf{x}) = \begin{cases} \text{mode}\{\hat{y}_1(\mathbf{x}), \ldots, \hat{y}_M(\mathbf{x})\}, & \text{classification} \\ \frac{1}{M}\sum_{m=1}^{M} \hat{y}_m(\mathbf{x}), & \text{regression} \end{cases} \tag{10}$$

Key points:

- Majority voting determines classification outcomes.
- Averaging reduces variance for regression.
- Combines multiple trees to improve prediction stability.

C. Feature Selection at Each Node

At each node, a random subset of features $F_m \subseteq F$ with size k is selected. The best split j^* is chosen according to an

impurity criterion. Random selection increases tree diversity and robustness [see Eq. (11)], [see Eq. (12)].

Key points:

- Optimal split minimizes impurity.
- Impurity is Gini (classification) or variance (regression).
- Random selection improves generalization.

1) Gini Index (Classification): The Gini Index measures node impurity. It quantifies the likelihood of misclassifying a randomly chosen sample. It reflects the class distribution within the node.

$$Gini(D_j) = 1 - \sum_{c=1}^{C} p_{jc}^2 \tag{11}$$

$$j^* = \arg\min_{j \in F_m} \sum_{t \in \{\text{left, right}\}} \frac{|D_t|}{|D_j|} Gini(D_t) \tag{12}$$

Key points:

- p_{jc} is the proportion of class c in node j.
- Weighted Gini guides splits.
- Minimizing Gini improves node purity.

2) Variance Reduction (Regression): Variance reduction measures how much a split decreases the variability of the target variable. It identifies splits that create more homogeneous child nodes. The best split maximizes this reduction [see Eq. (13)].

$$\Delta Var = Var(D_j) - \left(\frac{|D_L|}{|D_j|} Var(D_L) + \frac{|D_R|}{|D_j|} Var(D_R) \right) \tag{13}$$

Key points:

- Measures reduction in target variance.
- Split maximizing ΔVar is selected.
- Ensures homogeneous child nodes for regression.

D. Feature Importance

Feature importance quantifies the contribution of each feature X_k to the model prediction. It is calculated as the average decrease in impurity across all trees, defined as $VI(X_k) = \frac{1}{M} \sum_{m=1}^{M} \sum_{t \in T_m : v(t) = X_k} \Delta i(t)$. Features with higher $VI(X_k)$ have greater influence on the predictions.

Key points:

- $\Delta i(t)$ is the decrease in impurity at node t.
- Averaging across all trees gives global importance.
- Higher values indicate more influential features.

E. Out-of-Bag (OOB) Error

The Out-of-Bag (OOB) error estimates the model's generalization performance. It is calculated as $\text{Error}_{OOB} = \frac{1}{N} \sum_{i=1}^{N} L(y_i, \hat{y}_{OOB}(\mathbf{x}_i))$, using only samples not included in each tree's bootstrap training. This provides an unbiased estimate without requiring a separate validation set.

Key points:

- Uses only samples excluded from bootstrap training.
- L is 0-1 loss (classification) or MSE (regression).
- Provides unbiased error estimation without cross-validation.

VI. Result

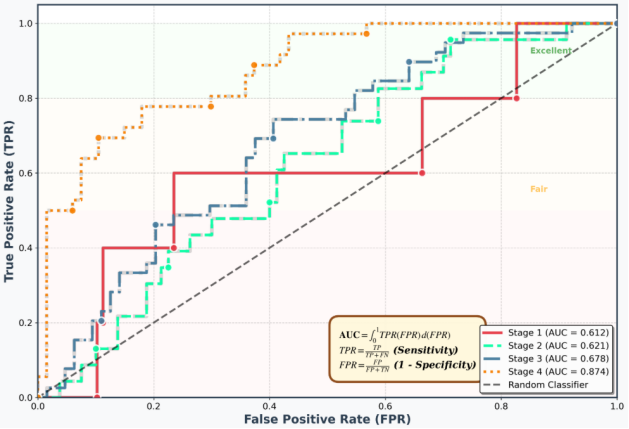
The results obtained through the implementation of the RF model are presented, performing a detailed comparison of the evaluation metrics in order to analyze its performance. Likewise, a comparative review of different methodologies is carried out, with the purpose of identifying the one that best adapts to the characteristics of the proposed model and the context of the study. It is important to highlight that the results obtained not only allow validating the predictive capacity of the model but also constitute a solid basis for decision-making and the proposal of improvements aimed at optimizing its performance.

A. Evaluation of Result

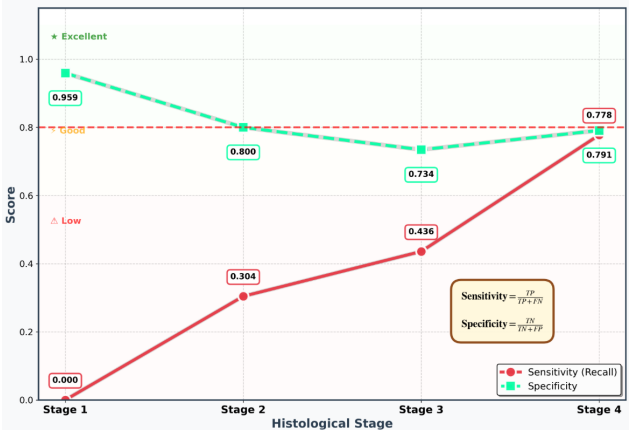
Fig. 6 presents an exhaustive evaluation of the multiclase classification model performance through six key metrics. The panel in Fig. 6a shows the ROC curves, where the area under the curve AUC progressively improves from a borderline value in Stage 1 (0.612) to a robust performance in Stage 4 (0.874). This trend is confirmed in Fig. 6b, where sensitivity increases from 0.000 to 0.778, contrasting with an initially excellent specificity (0.959) that gradually declines toward advanced stages. Similarly, Fig. 6c and 6d show that both precision and F1-score exhibit linear growth, reaching values of 0.667 and 0.718, respectively, in Stage 4, confirming minimal predictive capability in early phases that becomes stronger in later stages. Fig. 6e reveals that class-wise accuracy fluctuates between a maximum of 0.913 (Stage 1) and a minimum of 0.621 (Stage 3), compared to a global accuracy of 0.505 and an average class accuracy of 0.752. Finally, Fig. 6f analyzes predictive values, highlighting that the Negative Predictive Value (NPV) consistently remains higher than the Positive Predictive Value (PPV) across all levels, suggesting that the model has greater diagnostic reliability for excluding stages than for confirming them specifically.

B. Comparison of Methodologies

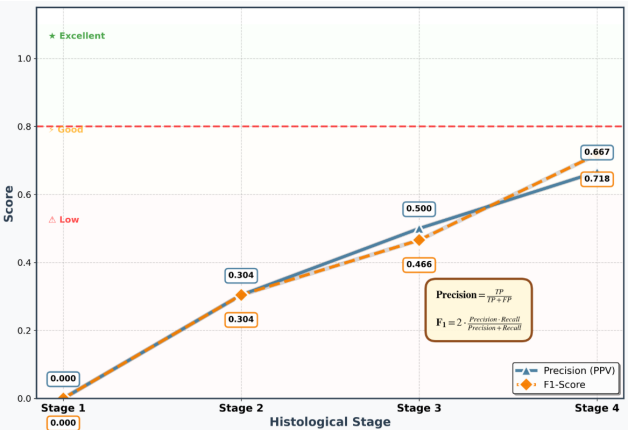
Table VI presents an integrated analysis of three widely used data mining methodologies: KDD, SEMMA, and CRISP-DM. Table VIa provides a qualitative comparison based on key criteria such as main focus, methodological structure, flexibility, scope, advantages, and limitations, highlighting the conceptual differences and strengths of each approach. Table VIb complements this analysis with a quantitative evaluation by assigning realistic percentage scores to each criterion, allowing a clear appreciation of the relative performance



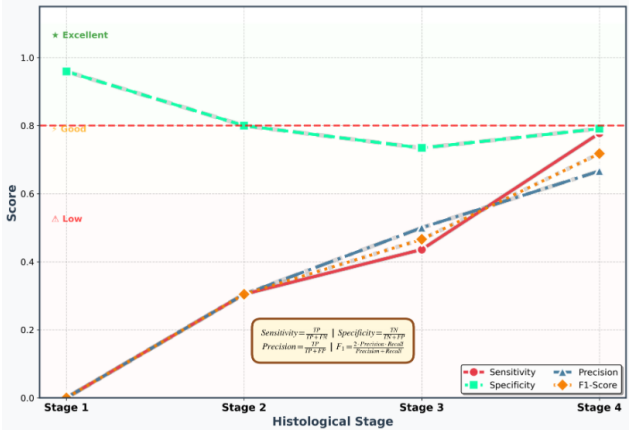
(a) Roc curves - multiclass clasisification.



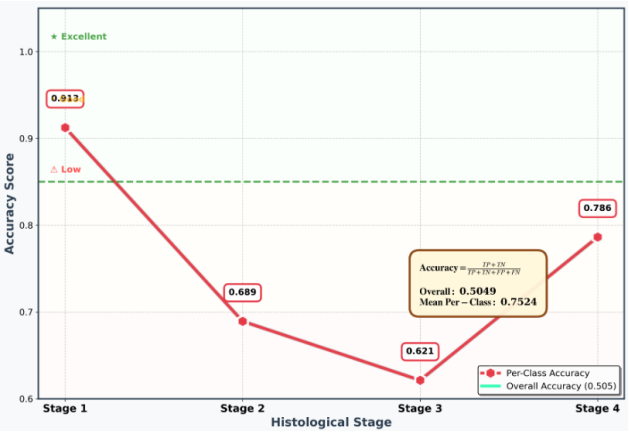
(b) Sensitivity and specificity per stage.



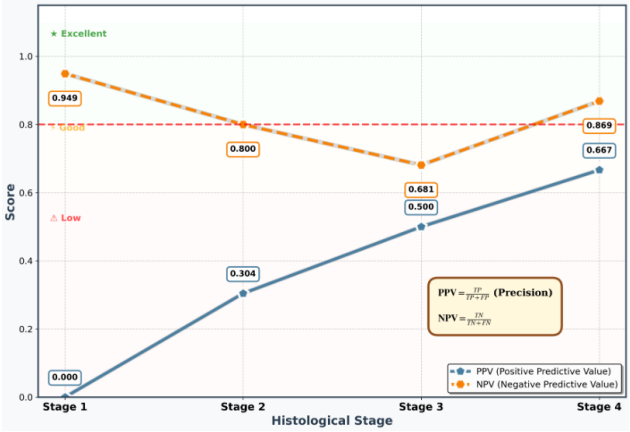
(c) Presicion and F1 score per stage.



(d) Comprehensive performance metrics comparison.



(e) Accuracy metrics per stage.



(f) Predictive values per stage.

Fig. 6. Comparison of ML models based on performance metrics.

of each methodology. The results show that CRISP-DM consistently obtains the highest scores across most criteria, standing out for its adaptability, comprehensive coverage, and alignment with business objectives, thus consolidating itself as the most recommended methodology. This joint presentation allows the reader to simultaneously consider both qualitative and quantitative aspects within a coherent and complete perspective.

VII. Discussion

In the discussion phase, the controversy between the results obtained in our research and those reported by other authors who have addressed similar topics is analyzed. In this section, the coincidences and discrepancies identified are critically examined, contrasting our findings with the existing scientific literature. To this end, the description of our study shows a progressive behavior of the multiclass classification model in predicting the different stages of hepatic cirrhosis, with clear differences in performance metrics as the disease advances. The area under the curve AUC increases steadily, from 0.612 in Stage 1 to 0.874 in Stage 4, reflecting a gradual improvement in the model's ability to distinguish between the different stages. This trend is consistent with previous studies, where algorithms such as RF (RF) and RF (BA) demonstrated good performance in stratifying patients with cirrhosis and other liver diseases [18, 19, 20, 21]. Therefore, the analysis of sensitivity and specificity reveals a complementary pattern: sensitivity increases from 0.000 to 0.778 in advanced stages, while specificity, although initially high (0.959), slightly decreases as the disease progresses. This suggests that the model is more reliable for ruling out early stages than for confirming them, a finding that is consistent with previous observations regarding models trained with limited samples, which tend to underestimate early cases but maintain good accuracy for excluding negatives [29, 30].

On the other hand, precision and the F1-score show a steady increase until reaching 0.667 and 0.718 in Stage 4, indicating that the model consolidates its predictive capacity in advanced phases of the disease. This behavior is consistent with reports by [24] and [22], who also highlight that RF and BA offer strong performance in patients with advanced cirrhosis, especially when evaluating balanced precision through metrics such as the F1-score. Similarly, class-wise accuracy, however, shows some variability across stages: it reaches a maximum of 0.913 in Stage 1 and a minimum of 0.621 in Stage 3, while global accuracy remains at 0.505 and mean class accuracy at 0.752. This dispersion reflects the clinical and biochemical heterogeneity inherent to cirrhosis, a phenomenon also noted in previous studies that emphasize how patient variability and biomarker complexity can limit the generalization of predictive models, particularly in intermediate stages [42, 38].

Finally, the predictive values show that the Negative Predictive Value (NPV) remains higher than the Positive Predictive Value (PPV), indicating that the model is more reliable for ruling out stages than for confirming them. This is consistent with observations from studies that use low-cost biochemical scores such as APRI and FIB-4, which perform well in identifying patients without significant fibrosis progression but have limitations in confirming advanced stages [29, 57, 58]. Compared to the literature on AI-assisted diagnosis, our results complement findings that highlight the usefulness of RF and SVM for early prediction of cirrhosis, hepatocellular carcinoma, and hospital mortality [45, 46, 37]. Although deep learning models or convolutional neural networks (CNN) achieve very high accuracy values [26, 27], our approach has the advantage of being more interpretable and more easily applicable in clinical contexts where multi-omic or textual data are not always available. This is consistent with the recommendations of [43, 44], who emphasize the need for standardized infrastructure so that predictive models can be truly useful and transferable to clinical practice. In summary, these results show that multiclass classification models, although with limitations in early stages, are valuable tools for cirrhosis prediction and monitoring. The consistency with previous studies indicates that evaluating multiple metrics ROC, AUC, precision, F1-score, and predictive values — allows a more comprehensive view of model performance, which supports its potential implementation in clinical practice for informed decision-making and early intervention planning.

VIII. Conclusion

Cirrhosis, along with related diseases, has progressively increased and has shown serious health problems, especially in people with a history of excessive alcohol consumption, viral hepatitis, or fatty liver disease, among other risk factors. In this sense, the proposed research aimed to develop a ML model using the RF algorithm as the main focus. Likewise, for better understanding, the methodology used was CRISP-DM, which includes five stages in the development of the project.

The first phase of the methodology corresponds to business understanding, in which the research problem is defined, the general objectives are established, and the scope of the study is delimited, among other relevant aspects. The second phase focuses on data understanding, allowing the collection, exploration, and preliminary analysis of the obtained information. The third phase is oriented toward data preparation, a process that includes cleaning, transformation, and structuring of the data to ensure its quality and consistency. The fourth phase consists of evaluating the behavior of the RF model, both in identifying explanatory variables and in training the dataset. Finally, the fifth phase presents the comparison of evaluation metrics to analyze the performance of the RF algorithm in the prediction task.

The research evaluated the multiclass classification model applied to histological study and demonstrated a performance that progressively improves as the disease stages advance. The area under the curve AUC increased from 0.612 in Stage 1 to 0.874 in Stage 4, evidencing greater discriminative capacity in the more advanced phases. Consistently, sensitivity, precision, and F1-score also showed an upward trend, confirming that the model more effectively identifies cases corresponding to late stages. Although overall accuracy was moderate, the results indicate that the model has stronger predictive capability in advanced stages than in early phases. Likewise, the fact that the Negative Predictive Value is higher than the Positive Predictive Value indicates that the system is more reliable for ruling out stages than for confirming them specifically. Overall, these findings support the usefulness of the model as a complementary diagnostic support tool,

TABLE VI. COMPARISON AND EVALUATION OF DATA MINING METHODOLOGIES

(A) QUALITATIVE COMPARISON

Criterion	KDD	SEMMA	CRISP-DM
Main Focus	Extraction of implicit patterns and relationships in data	Structured approach to building predictive models	Alignment with business goals and decision-making
Methodological Structure	Conceptual framework with flexible interpretation	Sequential workflow emphasizing data transformation and modeling	Defined stages with iterative feedback loops and documentation
Flexibility	Moderate adaptability depending on domain expertise	Restricted by tool-specific implementation (SAS)	Highly adaptable across industries and project types
Scope	Emphasis on discovering novel insights from raw data	Primarily focused on model optimization and validation	Comprehensive coverage from problem definition to deployment
Advantages	Strong foundation for academic research and innovation	Efficient for rapid prototyping in controlled environments	Widely accepted framework with extensive community support
Limitations	Limited operational guidance for real-world applications	Tool dependency reduces general applicability	Requires more effort in planning and documentation stages

(B) QUANTITATIVE EVALUATION WITH REALISTIC SCORES

Criterion	KDD (%)	SEMMA (%)	CRISP-DM (%)
Main Focus	65	75	90
Methodological Structure	70	80	95
Flexibility	60	65	90
Scope	75	70	95
Advantages	70	80	95
Limitations	50	60	85
Total / Average Score	65.0	71.7	91.7
Winning Methodology	**CRISP-DM**		

especially in the identification and differentiation of advanced stages, contributing to better clinical decision-making.

For future research, it would be appropriate to expand the study by incorporating larger and more heterogeneous clinical datasets from different medical institutions, in order to strengthen the model's generalization capacity in predicting the histological stage of primary biliary cirrhosis. Likewise, it would be possible to explore the comparison of the RF model with other ML algorithms and more advanced ensemble techniques, evaluating potential improvements in predictive performance. It would also be valuable to integrate additional variables, such as specific biomarkers or longitudinal patient information, allowing the analysis of disease progression over time. Finally, the incorporation of model interpretability methods would help facilitate understanding by clinical personnel and promote its application as a decision-support tool in medical practice. Despite the obtained results, the study presents some important limitations. The model showed difficulties in correctly identifying cases in Stage 1, since it registered low sensitivity and AUC values in this early phase. Additionally, the overall accuracy achieved indicates that there is still room to improve general performance. On the other hand, since this is a multiclass model, imbalances may exist among the different stages, which may influence metric outcomes.

ACKNOWLEDGMENT

Thanks to the University of Sciences and Humanities, which made this project a reality.

REFERENCES

[1] 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," *European Journal of Gastroenterology & Hepatology*, vol. 34, no. 10, pp. 1070–1077, 2022, doi: 10.1097/MEG.0000000000002424.

[2] H. Yoshiji, S. Nagoshi, T. Akahane, Y. Asaoka, Y. Ueno, K. Ogawa, T. Kawaguchi, M. Kurosaki, I. Sakaida, M. Shimizu, M. Taniai, S. Terai, H. Nishikawa, Y. Hiasa, H. Hidaka, H. Miwa, K. Chayama, N. Enomoto, T. Shimosegawa, T. Takehara, and K. Koike, "Evidence-based clinical practice guidelines for liver cirrhosis 2020," *Journal of Gastroenterology*, vol. 56, no. 7, pp. 593–619, 2021, doi: 10.1007/s00535-021-01792-0.

[3] X. Xu, H. Ding, W. Li, J. Xu, Y. Han, J. Jia, L. Wei, Z. Duan, E.-Q. Ling-Hu, and Z. Hui, "Chinese guidelines on the management of liver cirrhosis (abbreviated version)," *World Journal of Gastroenterology*, vol. 26, no. 45, pp. 7079–7104, 2020, doi: 10.3748/wjg.v26.i45.7079.

[4] P. Ginès, A. Krag, J. G. Abraldes, E. Solà, N. Fabrellas, and P. S. Kamath, "Liver cirrhosis," *The Lancet*, vol. 398, no. 10308, pp. 1359–1376, 2021, doi: 10.1016/S0140-6736(21)01374-7.

[5] M. Sun, S. Liu, J. Min, L. Zhong, J. Zhang, and Z. Du, "Predicting in-hospital mortality in patients with alcoholic cirrhosis complicated by severe acute kidney injury: development and validation of an explainable machine learning model," *Frontiers in Medicine*, vol. 12, p. 1570928, 2025, doi: 10.3389/fmed.2025.1570928.

[6] L. T. Nalasari, S. Anam, and N. Shofianah, "Liver cirrhosis classification using extreme gradient boosting classifier and harris hawk optimization as hyperparameter tuning," *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 7, no. 2, p. 730, 2025, doi: 10.35882/jeeemi.v7i2.730.

[7] N. R. Mazumder, A. Kazen, A. Carek, M. Etemadi, and J. Levitsky, "The answer at our fingertips: Volume status in cirrhosis determined by machine learning and pulse oximeter waveform," *Physiological Reports*, vol. 10, no. 5, p. e15223, 2022, doi: 10.14814/phy2.15223.

[8] Z. Wang, A. Zhang, Y. Yin, J. Tian *et al.*, "Clinical prediction of HBV-associated cirrhosis using machine learning based on platelet and bile acids," *Clinica Chimica Acta*, vol. 551, p. 117589, 2023, doi: 10.1016/j.cca.2023.117589.

[9] M. Mardewi, "Prediski Dini Liver Cirrhosis Untuk Kesehatan Hati Menggunaka Metode Machine Learning ," *Advances in Computer System Innovation Journal*, vol. 1, no. 1, pp. 1–10, 2023, doi: 10.51577/acsijournal.v1i1.474.

[10] I. Hanif and M. M. Khan, "Liver cirrhosis prediction using machine learning approaches," in *2022 IEEE 13th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON)*. IEEE, 2022, doi: 10.1109/UEMCON54665.2022.9965718, pp. 0534–0540.

[11] J. Zhang, H. Shen, J. Xu, L. Liu, J. Tan, M.-H. Li, N. Xu, S. Luo, J. Wang, F. Yang, J. Tang, Q. Li, Y. Wang, L. Yu, and Z. Yan, "Liver-targeted sirna lipid nanoparticles treat hepatic cirrhosis by dual anti-fibrotic and anti-inflammatory activities," *ACS Nano*, vol. 14, no. 5, pp. 5850–5869, 2020, doi: 10.1021/acsnano.9b09334.

[12] F. Meyer, K. Bannert, M. Wiese, S. Esau, L. Sautter, L. Ehlers, A. Aghdassi, C. Metges, L. A. Garbe, R. Jaster, M. M. Lerch, G. Lamprecht, and L. Valentini, "Molecular mechanism contributing to malnutrition and sarcopenia in patients with liver cirrhosis," *International Journal of Molecular Sciences*, vol. 21, no. 15, p. 5352, 2020, doi: 10.3390/ijms21155352.

[13] J. Traub, L. Reiss, B. Aliwa, and V. Stadlbauer, "Malnutrition in patients with liver cirrhosis," *Nutrients*, vol. 13, no. 2, p. 540, 2021, doi: 10.3390/nu13020540.

[14] S. Li, Y. Zhang, Y. Lin, L. Zheng, K. Fang, and J. Wu, "Development and validation of prediction models for nosocomial infection and prognosis in hospitalized patients with cirrhosis," *Antimicrobial Resistance & Infection Control*, vol. 13, no. 1, p. 103, aug 2024, doi: 10.1186/s13756-024-01444-y.

[15] Y. Y. Wang, W. X. Yang, Q. J. Du, Z. H. Liu, M. H. Lu, and C. G. You, "Construction and evaluation of a liver cancer risk prediction model based on machine learning," *World Journal of Gastrointestinal Oncology*, vol. 16, no. 9, pp. 3839–3850, sep 2024, doi: 10.4251/wjgo.v16.i9.3839.

[16] E. Audureau, F. Carrat, R. Layese, C. Cagnot, T. Asselah *et al.*, "Personalized surveillance for hepatocellular carcinoma in cirrhosis – using machine learning adapted to HCV status," *Journal of Hepatology*, vol. 73, no. 6, pp. 1374–1388, 2020, doi: 10.1016/j.jhep.2020.05.052.

[17] S. T. Y. Yau, E. Y. M. Leung, C. T. Hung, M. C. S. Wong, K. C. Chong, A. Lee, and E. K. Yeoh, "Scoring system for predicting the risk of liver cancer among diabetes patients: A random survival forest-guided approach," *Cancers*, vol. 16, no. 13, p. 2310, jun 2024, doi: 10.3390/cancers16132310.

[18] X. Bai, C. Pu, W. Zhen, Y. Huang, Q. Zhang, Z. Li, Y. Zhang, R. Xu, Z. Yao, W. Wu, M. Sun, and X. Li, "Identifying liver cirrhosis in patients with chronic hepatitis b: an interpretable machine learning algorithm based on LSM," *Annals of Medicine*, vol. 57, no. 1, p. 2477294, 2025, doi: 10.1080/07853890.2025.2477294.

[19] O. M. Khaled, A. Z. Elsherif, A. Salama, M. Herajy, and E. Elsedimy, "Evaluating machine learning models for predictive analytics of liver disease detection using healthcare big data," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 15, no. 1, pp. 1162–1174, 2025, doi: 10.11591/ijece.v15i1.pp1162-1174.

[20] S. E. Müller, M. Casper, C. Ripoll, A. Zipprich, P. Horn, M. Krawczyk, F. Lammert, and M. C. Reichert, "Machine learning models predicting decompensation in cirrhosis," *Journal of Gastrointestinal and Liver Diseases*, vol. 34, no. 1, pp. 45–52, mar 2025, doi: 10.15403/jgld-5876.

[21] F. Yang, C. Li, W. Yang, Y. He, L. G. Wu, K. Jiang, and C. Sun, "Development and validation of an explainable machine learning model for predicting multidimensional frailty in hospitalized patients with cirrhosis," *Briefings in Bioinformatics*, vol. 25, no. 6, p. bbae491, nov 2024, doi: 10.1093/bib/bbae491.

[22] F. Yang, J. Li, Z. Yang, L. Wu, H. Wang, and C. Sun, "Construction and validation of a machine learning-based model predicting early readmission in patients with decompensated cirrhosis: a prospective two-center cohort study," *BioData Mining*, vol. 18, no. 1, p. 5, jan 2025, doi: 10.1186/s13040-025-00479-0.

[23] P. Jain and P. Saxena, "A liver cirrhosis segmentation and detection using modified deep learning model," *International Journal of Scientific Research in Science and Technology*, vol. 11, no. 4, pp. 141–152, 2024, doi: 10.32628/ijsrst2411414.

[24] D. S. Ali and M. A. Aljabery, "Predicting liver cirrhosis stages using extra trees, random forest, and SVM with data mining techniques," *Informatica (Slovenia)*, vol. 48, no. 21, p. 6752, 2024, doi: 10.31449/inf.v48i21.6752.

[25] H. Wei, "A random forest-based prediction for liver cirrhosis," *Applied and Computational Engineering*, vol. 67, no. 1, pp. 42–49, 2024, doi: 10.54254/2755-2721/78/20240642.

[26] F. Rui, L. Xu, Y. H. Yeo, Y. Xu, W. Ni, Y. Tan, Q. Zheng, X. Tian, Q. L. Zeng, Z. He, Y. Qiu, C. Zhu, W. Ding, J. Wang, R. Huang, Q. Xue, X. Wang, Y. Chen, J. Fan, Z. Fan, E. Ogawa, M. S. Kwak, X. Qi, J. Shi, V. W. S. Wong, C. Wu, and J. Li, "Machine learning–based models for advanced fibrosis and cirrhosis diagnosis in chronic hepatitis b patients with hepatic steatosis," *Clinical Gastroenterology and Hepatology*, vol. 22, no. 11, pp. 2282–2292, nov 2024, doi: 10.1016/j.cgh.2024.06.014.

[27] P. Jain and P. Saxena, "Machine learning based liver cirrhosis detection using different algorithm: A review," *International Journal of Scientific Research in Science and Technology*, vol. 11, no. 4, pp. 69–78, 2024, doi: 10.32628/IJSRST2411353.

[28] M. Wieczorek *et al.*, "A deep learning approach for detecting liver cirrhosis from volatolomic analysis of exhaled breath," *Frontiers in Medicine*, vol. 9, p. 992703, 2022, doi: 10.3389/fmed.2022.992703.

[29] H. Innes *et al.*, "Performance of routine risk scores for predicting cirrhosis-related morbidity in the community," *Journal of Hepatology*, vol. 77, no. 2, pp. 365–376, 2022, doi: 10.1016/j.jhep.2022.02.022.

[30] I. Straw and H. Wu, "Investigating for bias in healthcare algorithms: a sex-stratified analysis of supervised machine learning models in liver disease prediction," *BMJ Health Care Informatics*, vol. 29, no. 1, p. e100457, 2022, doi: 10.1136/bmjhci-2021-100457.

[31] D. S. Ali and M. A. Aljabery, "A study on cirrhosis prediction based on machine learning techniques," *Journal of College of Education for Pure Sciences*, vol. 14, no. 4, pp. 15–24, 2024, doi: 10.32792/jeps.v14i4.461.

[32] Z.-M. Zou, D.-H. Chang, H. Liu, and Y.-D. Xiao, "Current updates in machine learning in the prediction of therapeutic outcome of hepatocellular carcinoma: what should we know?" *Insights into Imaging*, vol. 12, no. 1, p. 31, 2021, doi: 10.1186/s13244-021-00977-9.

[33] S. Ghosh *et al.*, "Artificial intelligence applied to 'omics data in liver disease: towards a personalised approach for diagnosis, prognosis and treatment," *Gut*, vol. 73, no. 8, pp. 1364–1376, 2024, doi: 10.1136/gutjnl-2023-331740.

[34] M. Field *et al.*, "Machine learning applications in radiation oncology," *Physics and Imaging in Radiation Oncology*, vol. 19, pp. 13–24, 2021, doi: 10.1016/j.phro.2021.05.007.

[35] S. Malik, L. J. Frey, and K. Qureshi, "Evaluating the predictive power of machine learning in cirrhosis mortality: a systematic review," *Journal of Medical Artificial Intelligence*, vol. 8, pp. 1–12, 2025, doi: 10.21037/jmai-24-205.

[36] Y. S. Choi and E. Oh, "Investigating of machine learning based algorithms for liver cirrhosis prediction," *Advances in Engineering and Intelligence Systems*, vol. 3, no. 1, pp. 158–170, 2024, doi: 10.22034/aeis.2024.446087.1177.

[37] F. Kanwal, T. J. Taylor, J. R. Kramer, Y. Cao, D. Smith, A. L. Gifford, H. B. El-Serag, A. D. Naik, and S. M. Asch, "Development, validation, and evaluation of a simple machine learning model to predict cirrhosis mortality," *JAMA Network Open*, vol. 3, no. 11, p. e2023780, 2020, doi: 10.1001/jamanetworkopen.2020.23780.

[38] C. Hu, V. Anjur, K. Saboo *et al.*, "Low predictability of readmissions and death using machine learning in cirrhosis," *The American Journal of Gastroenterology*, vol. 116, no. 2, pp. 336–346, 2021, doi: 10.14309/ajg.0000000000000971.

[39] G. Ma and Y. Li, "Liver cirrhosis prediction: The employment of the machine learning-based approaches," *Intelligent Systems with Applications*, vol. 28, p. 200573, 2025, doi: 10.1016/j.iswa.2025.200573.

[40] X. Liu, D. Liu, C. Tan, and W. Feng, "Gut microbiome-based machine learning for diagnostic prediction of liver fibrosis and cirrhosis: a systematic review and meta-analysis," *BMC Medical Informatics and Decision Making*, vol. 23, no. 1, p. 240, 2023, doi: 10.1186/s12911-023-02402-1.

[41] K. A. Mohamud, S. A. E. Eltahir, H. A. A. Alhardalo *et al.*, "The role of machine learning models in predicting cirrhosis mortality: A systematic review," *Cureus*, vol. 17, no. 1, p. e78155, 2025, doi:

10.7759/cureus.78155.

[42] J. Song, Z. Gao, L. Lai, J. Zhang, B. Liu *et al.*, "Machine learning-based plasma metabolomics for improved cirrhosis risk stratification," *BMC Gastroenterology*, vol. 25, no. 1, p. 3655, 2025, doi: 10.1186/s12876-025-03655-y.

[43] X. F. Gu, X. J. Liang, and J. L. Dong, "Construction of risk prediction models for nosocomial infection in patients with liver cirrhosis based on machine learning," *World Chinese Journal of Digestology*, vol. 33, no. 3, pp. 199–208, 2025, doi: 10.11569/wcjd.v33.i3.199.

[44] Q. Y. Liang, J. Wang, Y. F. Yang, K. Zhao, R. L. Luo, Y. Tian, and F. X. Li, "Machine learning to identify potential biomarkers for sarcopenia in liver cirrhosis," *World Journal of Hepatology*, vol. 17, no. 6, p. 105332, 2025, doi: 10.4254/wjh.v17.i6.105332.

[45] E. Dritsas and M. Trigka, "Supervised machine learning models for liver disease risk prediction," *Computers*, vol. 12, no. 1, p. 19, 2023, doi: 10.3390/computers12010019.

[46] A. Guo *et al.*, "Predicting mortality among patients with liver cirrhosis in electronic health records with machine learning," *PLOS ONE*, vol. 16, no. 8, p. e0256428, 2021, doi: 10.1371/journal.pone.0256428.

[47] Z. Zhou *et al.*, "A decision tree model to predict liver cirrhosis in hepatocellular carcinoma patients: a retrospective study," *PeerJ*, vol. 11, p. e15950, 2023, doi: 10.7717/peerj.15950.

[48] M. Eslam *et al.*, "FibroGENE: a gene-based model for staging liver fibrosis," *Journal of Hepatology*, vol. 64, no. 2, pp. 390–398, 2016, doi: 10.1016/j.jhep.2015.09.012.

[49] N. Sharma *et al.*, "Biomolecular map of albumin identifies signatures of severity and early mortality in acute liver failure," *Journal of Hepatology*, vol. 79, no. 3, pp. 677–691, 2023, doi: 10.1016/j.jhep.2023.04.029.

[50] K. R. Makkena and K. Natarajan, "Enhancing liver cirrhosis diagnosis using machine learning with explainable AI and cross-validated hyperparameter tuning techniques," *IEEE Access*, vol. 13, p. 3568918, 2025, doi: 10.1109/ACCESS.2025.3568918.

[51] Z. J. Wang, F. Y. Li, J. J. Cai, Z. T. Xue, Y. Zhou, and Z. Wang, "Construction and validation of a machine learning-based prediction model for short-term mortality in critically ill patients with liver cirrhosis," *Clinics and Research in Hepatology and Gastroenterology*, vol. 49, no. 1, p. 102507, 2025, doi: 10.1016/j.clinre.2024.102507.

[52] N. Solanki, J. Thobhani, and P. Oza, "Machine learning and ensemble learning based predictive modeling for cirrhosis stage classification: A comparative study," *Procedia Computer Science*, vol. 259, pp. 100–109, 2025, doi: 10.1016/j.procs.2025.04.100.

[53] S. M. K. Hosseini-Asl, S. J. Masoumi, G. Rashidizadeh *et al.*, "Sina score as a new machine learning-derived online prediction model of mortality for cirrhotic patients awaiting liver transplantation," *Journal of Clinical Medicine*, vol. 14, no. 13, p. 4559, 2025, doi: 10.3390/JCM14134559.

[54] G. Sparacia, G. Colelli, G. Parla, G. Mamone *et al.*, "Brain magnetic resonance imaging radiomic signature and machine learning model prediction of hepatic encephalopathy in adult cirrhotic patients," *Life*, vol. 15, no. 3, p. 346, 2025, doi: 10.3390/life15030346.

[55] N. Rengaraj, K. M. Vamsi, and G. Vijayaraghavan, "Evaluation of machine learning approaches for cirrhosis survival prediction," p. 10915362, 2025, doi: 10.1109/IITCEE64140.2025.10915362.

[56] J. Blessy R and N. A. H, "Liver cirrhosis stage classification using ensemble machine learning techniques," *International Journal of Scientific Research in Engineering and Management (IJSREM)*, vol. 9, no. 4, p. 44060, 2025, doi: 10.55041/ijsrem44060.

[57] M. Luo, D. Yan, X. Wang, Y. Wang, H. Li *et al.*, "Construction of a machine learning prognostic prediction model based on psoas muscle index for patients with decompensated liver cirrhosis," *Chinese Journal of Hepatology*, vol. 33, no. 7, 2025, doi: 10.3760/cma.j.cn501113-20231123-00222.

[58] P. K. Mishra, B. K. Chaurasia, and M. M. Shukla, "XAIHO: explainable AI leveraging hybrid optimized framework for liver cirrhosis detection," *Discover Artificial Intelligence*, vol. 5, no. 1, p. 470, 2025, doi: 10.1007/s44163-025-00470-y.

[59] D. Kurniawan and M. Yasir, "Optimization sentimen analysis using CRISP-DM and Naive Bayes methods implemented on social media," *Cyberspace: Jurnal Pendidikan Teknologi Informasi*, vol. 6, no. 2, pp. 127–135, 2022, doi: 10.22373/cj.v6i2.12793.

[60] Kaggle, "Primary biliary cirrhosis (pbc) dataset," 2023, accessed: 05 Ago. 2024.

[61] J. C. G. Mejia, F. A. V. Agudelo, D. A. M. Samboní, and M. F. C. Ortega, "Method for the application of BI in software development," *Salud, Ciencia y Tecnología*, vol. 5, p. 1424, 2025, doi: 10.56294/saludcyt20251424.

[62] U. Roman-Concha, A. López, K. Ruiz-Carrasco, C. Chavez-Herrera *et al.*, "Propensity model using decision trees (LightGBM) for the management of the effective credit product in a financial entity," *Advances in Artificial Intelligence and Machine Learning*, vol. 5, no. 1, p. 51184, 2025, doi: 10.54364/AAIML.2025.51184.

[63] D. A. Hamidah, R. Salkiawati, and R. Sari, "Analisis sentimen ulasan customer kopi TMLST menggunakan algoritma Naïve Bayes," *Journal of Students' Research in Computer Science*, vol. 5, no. 1, pp. 89–97, 2024, doi: 10.31599/mrm89y71.

[64] J. Adiputra and D. Mahdiana, "Analisis Sentimen Degan Algoritma Support Vector Machine Terhadap Penyakit Hepatitis Akut Misterius," *IDEALIS: InDonEsiA journaL Information System*, vol. 6, no. 1, p. 2985, 2023, doi: 10.36080/idealis.v6i1.2985.

[65] J. de Pedro-Carracedo, J. Clemente, D. Fuentes-Jimenez, M. F. Cabrera-Umpiérrez, and A. P. Gonzalez-Marcos, "Photoplethysmographic signal-diffusive dynamics as a mental-stress physiological indicator using convolutional neural networks," *Applied Sciences*, vol. 13, no. 15, p. 8902, 2023, doi: 10.3390/app13158902.

[66] A. Fajriansyah, D. Yusup, and K. Prihandini, " Prediski Kelulusan Nilai Kalkulus Menggunakan Naive Bayes," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 2, p. 13343, 2025, doi: 10.36040/jati.v9i2.13343.

[67] R. M. Apsari, "Penerapan metode Naïve bayes dalam memprediksi prestasi siswa," *Jurnal Pustaka AI (Pusat Akses Kajian Teknologi Artificial Intelligence)*, vol. 4, no. 2, pp. 76–85, 2024, doi: 10.55382/jurnalpustakaai.v3i3.760.

[68] S. K. Kamath, S. K. Pendekanti, and D. Rao, "Livmarx: An optimized low-cost predictive model using biomarkers for interpretable liver cirrhosis stage classification," *IEEE Access*, vol. 12, pp. 98 712–98 725, 2024, doi: 10.1109/ACCESS.2024.3422451.

[69] X. Jiang, R. Zhou, F. Jiang, Y. Yan, Z. Zhang, and J. Wang, "Construction of diagnostic models for the progression of hepatocellular carcinoma using machine learning," *Frontiers in Oncology*, vol. 14, p. 1401496, jun 2024, doi: 10.3389/fonc.2024.1401496.

[70] M. Tabatabai, D. Wilus, C. K. Chen, K. P. Singh, and T. L. Wallace, "Taba binary, multinomial, and ordinal regression models: New machine learning methods for classification," *Bioengineering*, vol. 12, no. 1, p. 2, jan 2025, doi: 10.3390/bioengineering12010002.

[71] S. Vithal, S. Shah, V. C. Kulkarni, and S. Terdal, "Liver cirrhosis prediction using random forest," *International Journal of Latest Technology in Engineering, Management & Applied Science (IJLTEMAS)*, vol. 14, no. 5, pp. 115–122, 2025, doi: 10.51583/ijltemas.2025.140500115.

[72] B. Da, H. Chen, W. Wu, W. Guo, A. Zhou, Q. Yin, J. Gao, J. Chen, J. Xiao, L. Wang, M. Zhang, Y. Zhuge, and F. Zhang, "Development and validation of a machine learning-based model to predict survival in patients with cirrhosis after transjugular intrahepatic portosystemic shunt," *eClinicalMedicine*, vol. 79, p. 103001, jan 2025, doi: 10.1016/j.eclinm.2024.103001.

[73] F. Yao, J. Luo, Q. Zhou, L. Wang, and Z. He, "Development and validation of a machine learning-based prediction model for hepatorenal syndrome in liver cirrhosis patients using MIMIC-IV and eICU databases," *Scientific Reports*, vol. 15, no. 1, p. 3412, feb 2025, doi: 10.1038/s41598-025-86674-9.

[74] A. M. Al Alawi, H. Al Kaabi, Z. Al Falahi, Z. Al-Naamani, and S. Al Busafi, "Machine learning-powered 28-day mortality prediction model for hospitalized patients with acute decompensation of liver cirrhosis," *Oman Medical Journal*, vol. 39, no. 3, p. e624, may 2024, doi: 10.5001/omj.2024.79.