The Academic
International Journal of Multidisciplinary Research

An Online Peer Reviewed / Refereed Journal
Volume 3 | Issue 4 | April 2025
ISSN: 2583-973X (Online)
Website: www.theacademic.in

# Optimized Feature Selection for Liver Cirrhosis Stage Prediction using combination of Explainable AI and Metaheuristic Algorithms

**Abhishek Dey**
Assistant Professor, Department of Computer Science, Bethune College, Kolkata
dey.abhishek7@gmail.com

**ARTICLE DETAILS**

**Research Paper**

**Accepted:** 16-04-2025

**Published:** 10-05-2025

**Keywords:**

*Liver cirrhosis, Explainable AI, Particle Swarm Optimization, Whale Optimization Algorithm, Grey Wolf Optimization, Firefly Algorithm*

**ABSTRACT**

Liver cirrhosis is a chronic liver disease marked by the gradual scarring of liver tissue, which ultimately results in liver dysfunction. Long-term liver damage from a variety of sources, including infections, alcoholism, and metabolic disorders, is the origin of this scarring. The progression of cirrhosis occurs in stages, moving from compensated, where the liver maintains relatively good function, to decompensated, characterized by severe complications. Therefore, for efficient clinical treatment, it is essential to accurately predict the stages of liver cirrhosis. This article introduces a hybrid approach that combines Explainable AI methods such as Local Interpretable Model-agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) to estimate feature importance, optimizing their contributions using four metaheuristic algorithms: Particle Swarm Optimization (PSO), Whale Optimization Algorithm (WOA), Grey Wolf Optimization (GWO), and Firefly Algorithm (FA). The intersection of top-ranked features across these methods is then determined, providing a robust selection of the most significant predictors. The findings highlight the efficacy of this hybrid approach in enhancing the reliability of feature selection for liver cirrhosis stage prediction.

**DOI : https://doi.org/10.5281/zenodo.15389884**

## Introduction

Cirrhosis, a persistent liver disease, results in significant and irreversible damage to the liver (Arroyo et al., 2016). This condition develops due to various underlying diseases, causing structural and functional deterioration of the liver. As cirrhosis progresses, healthy liver cells are gradually lost, leading to fibrosis (scarring), hardening, and shrinkage of the liver (Garcia-Martinez et al., 2013). The reduced blood flow to these hardened tissues forces the body to create new vascular pathways to compensate. However, this process further worsens liver function, ultimately leading to liver failure (Mozos, 2015). In the early stages of liver cirrhosis, symptoms are often mild and nonspecific but become more severe as liver damage progresses (Younossi & Henry, 2015). As the disease advances, patients may experience fluid retention (edema), abdominal swelling (ascites), muscle wasting, easy bruising, increased bleeding tendency, severe itching, and jaundice (Pinto, Schneider, & da Silveira, 2015).

An essential part of the body's metabolic functions is the liver. It creates vital proteins like albumin, which support blood vessel fluid equilibrium (van Zutphen et al., 2016). When liver function declines, albumin synthesis decreases, causing fluids to leak into tissues, leading to leg swelling and abdominal fluid accumulation (Guerci et al., 2019). Furthermore, liver impairment reduces the generation of blood clotting factors, which increases the risk of bleeding and makes bruising easier. Toxic compounds build up in the bloodstream as cirrhosis progresses, causing hepatic encephalopathy (liver failure-related brain dysfunction). Severe alcohol consumption, diabetes, obesity, gastrointestinal disorders, heart failure, and unprotected sexual activity are some of the variables that lead to cirrhosis (Ginès et al., 2021).

Traditionally, liver cirrhosis stage detection relies on biopsy, imaging tests (ultrasound, MRI), and blood tests, which can be invasive and time-consuming (Acharya et al., 2015). Artificial Intelligence (AI) algorithms can help in this issue by providing faster, and non-invasive diagnostic solutions. AI enables early detection, allowing timely interventions that slow disease progression and reduce complications (Nam et al., 2022).

In recent years, significant research has been conducted in the area of liver disease prediction. (Auxilia, 2018) proposed a predictive model for liver patient classification, employing Pearson Correlation for feature selection and comparing five classifiers: Decision Tree (DT), Naïve Bayes (NB), Artificial Neural Network (ANN), Support Vector Machine (SVM), and Random Forest (RF), with DT yielding the highest accuracy. Similarly, (Veena et al., 2018) applied five data mining algorithms to the Indian Liver Patient Dataset, achieving highest accuracy with the C5.0 boosted classifier. (Rahman et al., 2019) conducted a comparative study using supervised learning methods, including K-Nearest Neighbors

(KNN), Logistic Regression (LR), DT, SVM, NB, and RF, finding that LR performed best in terms of classification accuracy and F1-score. (Rabbi et al., 2020) also used Pearson Correlation for feature selection and compared LR, DT, RF, and Extra Trees (ET) for liver disease prediction, with ET yielding the best performance. (Shaheamlung & Kaur, 2021) proposed a voting-based hybrid classification method combining KNN, LR, and DT for liver disease prediction on the Indian Liver Patient Dataset. (Singh, Gourisaria, & Das, 2021) used the same dataset to evaluate the performance of several boosting algorithms for liver disease prediction, including XG Boost, CatBoost, and AdaBoost. (Utku, 2023) developed a Multilayer Perceptron (MLP) model for liver cirrhosis detection using the Cirrhosis Patient Survival Prediction dataset. (Topcu, Elbasi, & Alzoubi, 2024) also utilized this dataset to predict early-stage cirrhosis using various machine learning algorithms, including LR, KNN, RF, MLP, AdaBoost, and BernoulliNB, with RF achieving the best performance.

Compared to existing studies, the proposed method offers several key contributions in liver cirrhosis stage prediction. While prior research has primarily focused on traditional machine learning classifiers, this study integrates Explainable AI techniques, namely Local Interpretable Model-agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) to enhance the interpretability of feature selection. Most previous works have prioritized classification accuracy while overlooking the identification of the most critical features. Moreover, recognizing the limitations of relying on a single feature selection technique, as demonstrated in (Auxilia, 2018) and (Rabbi et al., 2020), this work adopts a hybrid optimization framework. The contributions of LIME and SHAP are combined and optimized using four metaheuristic algorithms: Particle Swarm Optimization (PSO), Whale Optimization Algorithm (WOA), Grey Wolf Optimization (GWO), and Firefly Algorithm (FA), ensuring a more reliable selection of relevant predictors. By identifying the common top-ranked features selected by these algorithms, the proposed approach enhances the robustness of feature selection, leading to more informed clinical decision-making, and better patient outcomes.

## 2. Materials and Methods

The aim of this study is to determine the most influential attributes for predicting liver cirrhosis stages. The proposed workflow comprises the following steps:

Data pre-processing: The raw data is pre-processed by handling missing values, encoding categorical variables, and normalizing features.

Feature importance estimation: A Random Forest classifier is trained on the dataset to serve as the base model for feature importance estimation and then feature importance is estimated using two Explainable AI (XAI) methods, namely Local Interpretable Model-agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP).

Optimization: Feature importance values obtained using LIME and SHAP are combined and optimized using metaheuristic algorithms such as, Particle Swarm Optimization (PSO), Whale Optimization Algorithm (WOA), Grey Wolf Optimization (GWO), and Firefly Algorithm (FA).

Feature selection: The most influential features are determined by identifying the common top-ranked features from each optimized model.

Details of the dataset and its pre-processing are described in Section 2.1, while feature importance estimation and optimization using metaheuristic algorithms are elaborated in Sections 2.2 and 2.3, respectively.

### 2.1. Dataset Description and Pre-processing

In this study, the Cirrhosis Patient Survival Prediction dataset is utilized, which comprises 20 attributes from 418 patients, encompassing demographic details, clinical observations, and laboratory test results. This dataset is accessible via the UCI Machine Learning Repository: https://archive.ics.uci.edu/dataset/878/cirrhosis+patient+survival+prediction+dataset-1.

A detailed description of the attributes is provided in Table 1. The 'Stage' attribute, which represents the histologic stage of liver cirrhosis, is designated as the target variable for prediction. However, the attributes 'ID', 'N_Days', 'Status', and 'Drug' were considered irrelevant for cirrhosis stage prediction and were therefore excluded from further analysis. 'ID' is simply a unique identifier for each patient and has no bearing on their liver condition or its stage. Keeping it in the dataset could introduce noise without contributing to model performance. 'N_Days' might be relevant for survival analysis but it does not directly reflect the current stage of cirrhosis. Since, this value is an outcome rather than a predictor, including it would lead to data leakage, where the model indirectly learns from future outcomes rather than true predictive patterns. The 'Status' attribute is related to patient survival rather than cirrhosis

staging, making it inappropriate for predicting the histologic stage of the disease. The 'Drug' attribute indicates whether a patient was given D-penicillamine or a placebo. While drug treatment may influence disease progression or survival, it does not directly reflect the histologic stage of cirrhosis at a given point in time. Additionally, including treatment as a predictor could bias the model, as the drug assignment is not a biological determinant of cirrhosis stage. Hence, removing these attributes prevents data leakage, eliminates non-informative variables, and ensures the model focuses on relevant predictors. Thus, the remaining 15 attributes serve as meaningful features for liver cirrhosis stage prediction.

**Table 1. Description of the Cirrhosis Patient Survival Prediction dataset**

| Column name | Description | Type |
|---|---|---|
| ID | Unique identifier of patients | Integer |
| N_Days | Duration from patient registration until the first occurrence of death, transplant, or study completion | Integer |
| Status | The patient's final status, categorized as C (censored), CL (because of liver transplantation), or D (death) | |
| Drug | Type of drug D-penicillamine or placebo | |
| Age | Age of the patients in the number of days | Integer |
| Sex | M (male) or F (female) | Categorical |
| Ascites | Indication of ascites: N (no) or Y (yes) | Categorical |
| Hepatomegaly | Indication of hepatomegaly: N (no) or Y (yes) | Categorical |
| Spiders | Indication of spiders: N (no) or Y (yes). | Categorical |
| Edema | Categorization of edema: N (absent, with no diuretic treatment), S (present without diuretics or resolved by diuretics), or Y (persistent despite diuretic use) | Categorical |
| Bilirubin | Concentration of bilirubin in serum (mg/dl) | Continuous |
| Cholesterol | Concentration of cholesterol in serum (mg/dl) | Integer |
| Albumin | Concentration of albumin in serum (gm/dl) | Continuous |
| Copper | Level of copper in urine (µg/day) | Integer |
| Alk_Phos | Alkaline phosphatase level (U/liter) | Continuous |
| SGOT | SGOT level (U/ml) | Continuous |

| | | |
|---|---|---|
| Tryglicerides | Tryglicerides level (mg/dl) | Integer |
| Platelets | Platelets value per cubic ml/1000 | Integer |
| Prothrombin | Prothrombin time in seconds | Continuous |
| Stage | Histologic stage of disease as 1, 2, 3, or 4 | Categorical |

Initially, the dataset contains missing values in some attributes, which are addressed using the MissForest imputation technique (Stekhoven & Bühlmann, 2021). MissForest, an iterative, non-parametric method based on Random Forest, is employed to impute missing values, ensuring the integrity of the dataset. This approach effectively preserves the relationships among features while estimating missing values (Stekhoven & Bühlmann, 2021).

Categorical features are identified and appropriately transformed into numerical representations to ensure compatibility with machine learning models. The categorical values of the 'Sex' attribute are changed to 1 for 'M' and 0 for 'F'. The 'Ascites', 'Hepatomegaly', and 'Spiders' attributes are encoded as 0 for 'N' and 1 for 'Y'. The 'Edema' attribute is transformed into 0 for 'N', 1 for 'Y', and 2 for 'S'. Continuous numerical features are normalized to standardize their values between 0 and 1, mitigating the impact of varying feature scales.

### 2.2. Feature Importance Estimation using LIME and SHAP

To identify the most significant features in predicting cirrhosis stages, two XAI techniques: Local Interpretable Model-Agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) are used in this article. These methods are widely used for interpreting machine learning models.

LIME operates by slightly altering the input data for a specific prediction and observing the resulting changes in the model's output (Zafar & Khan, 2021). Through the analysis of these perturbations, LIME identifies the features that most significantly impact the model's prediction for that particular instance. This provides a local explanation, highlighting which features contributed positively or negatively on the prediction for that specific data point (Zafar & Khan, 2021).

In contrast, SHAP leverages principles from cooperative game theory to offer a cohesive measure of feature importance. It computes Shapley values, which quantify each feature's contribution to the prediction (Nohara et al., 2022). While LIME focuses on local interpretability, SHAP aims to provide a global understanding of feature importance. Shapley values consider all possible feature combinations to

determine the impact of each feature on the model's output (Nohara et al., 2022). This approach ensures a fair distribution of importance among the features, providing a consistent and accurate representation of their influence on the model's predictions.

After pre-processing, as detailed in Section 2.1, the dataset is divided into training (80%) and testing (20%) sets using stratified splitting to maintain balanced class distributions. A Random Forest classifier is then trained on the training set to serve as the base model for the XAI techniques.

For analysis using SHAP, a TreeExplainer is employed to estimate Shapley values for each feature. TreeExplainer is a specific implementation of the SHAP, designed to efficiently calculate Shapley values specifically for tree-based machine learning models including Decision Tree, Random Forest, and Gradient Boosting Machines. These Shapley values offered insights into each variable's contribution to the model's predictions (Nohara et al., 2022). The mean absolute Shapley value across all test instances is computed to quantify feature importance.

For analysis using LIME, a LimeTabularExplainer is used to generate instance-wise feature explanations. LimeTabularExplainer is a key component of the LIME method, designed to explain individual predictions of machine learning models (Zafar & Khan, 2021). It is specifically tailored for tabular data, where samples (observations) are represented by rows and features by columns. Unlike global explainability techniques like SHAP, LIME provides local interpretability, explaining why a model made a particular prediction for a specific instance rather than for the entire dataset (Zafar & Khan, 2021). The model's prediction probabilities are perturbed, and feature contribution weights are aggregated across multiple instances. The average absolute LIME weights are calculated to determine each feature's importance.

### 2.3. Optimization using Metaheuristic Algorithms

As explained in Section 2.2, The SHAP and LIME importance scores for 15 features are independently computed. As a result, for each feature, there are two importance scores: one from LIME and one from SHAP. The goal is to combine the LIME and SHAP importance scores to get a more robust and reliable measure of feature importance. To achieve this, a weighted sum approach is used. A single objective function is defined to combine the individual feature importance values obtained using LIME and SHAP. This function is defined as,

$$F = \sigma.L + \eta.S;$$

where,

$L$ is a 1x15 column vector containing LIME importance scores.

$S$ is a 1x15 column vector containing SHAP importance scores.

$\sigma$ and $\eta$ are weights associated with LIME and SHAP scores, respectively, such that $\sigma + \eta = 1$ (This ensures a balanced distribution of importance values between the two techniques).

$F$ is a 1x15 column vector containing is the final combined importance scores.

Hence, using the above function, the final importance score for each feature $f_i$ is given by:

$\sigma.L_i + \eta.S_i$; where, $L_i$ and $S_i$ are LIME and SHAP importance values for $f_i$ respectively.

The key challenge is to find the optimal values for the weights $\sigma$ and $\eta$. For this purpose, four metaheuristic algorithms: PSO, WOA, GWO, and FA are used.

In each of the four optimization algorithms, a population-based search approach is employed. Each individual (particle, whale, wolf, or firefly) in the population represents a candidate solution, i.e., a potential value of $\sigma$. The corresponding value of $\eta$ is implicitly determined as $(1 - \sigma)$. The steps involved in the optimization process are as follows:

Initialization: The initial population of candidate solutions ($\sigma$ values) is generated in random manner within the range [0,1].

Fitness Evaluation: Each candidate $\sigma$ value is used to compute the combined LIME-SHAP importance scores for all features. A Random Forest model is trained using these scores. The classification error rate is computed as the fitness value.

Population Update: Based on the specific optimization algorithm, individuals adjust their positions to explore better solutions.

Convergence Check: The process continues for a predefined number of iterations or until a stopping criterion (such as minimal improvement in fitness) is met.

Best Solution Selection: The best-performing individual provides the optimized $\sigma$, and $\eta$ is computed as $(1 - \sigma)$. These final weights are used to compute the optimized feature importance scores.

Since, each method employs a distinct approach to update the population, the algorithm-specific update processes are described in the Sections 2.3.1, 2.3.2, 2.3.3, and 2.3.4.

### 2.3.1 Optimization using PSO

PSO draws inspiration from the collective behavior observed in bird flocks and fish schools, where individuals adjust their positions based on two key factors: the best solution an individual particle has found so far (pbest) and the best solution found by any particle in the swarm (gbest). Each particle in the PSO swarm represents a possible $\sigma$ value. The velocity of a particle determines how much it moves toward better solutions. During each iteration, particles update their positions ($\sigma$ values) based on their own past experiences and the swarm's collective knowledge (Wang, Tan, & Liu, 2018). The fitness of each particle is evaluated using the Random Forest model's error rate. If a particle finds a better $\sigma$, it updates its pbest. The global best gbest is updated if any particle outperforms the current best solution (Wang, Tan, & Liu, 2018). This iterative process continues until convergence, and the best-performing particle at the end provides the optimized $\sigma$ value.

### 2.3.2 Optimization using WOA

WOA mimics the bubble-net hunting strategy of humpback whales, balancing exploration, and exploitation through the following three mechanisms :

Encircling Prey (exploration and exploitation): Whales move closer to promising regions.

Bubble-Net Attacking (exploitation): Whales spiral inward toward the best solution.

Random Search (exploration): Whales move randomly to avoid local optima.

Each whale represents a candidate $\sigma$ value. During each iteration, the fitness of each whale is determined using the Random Forest error rate. Whales adjust their $\sigma$ values by either following the best whale (exploitation) or exploring new areas (exploration). Over multiple iterations, whales converge on the best $\sigma$ value, which minimizes classification error. The whale with the best fitness at the end provides the optimized $\sigma$, and the final feature importance scores are computed (Nadimi-Shahraki et al., 2023).

### 2.3.3 Optimization using GWO

GWO is modelled after the social hierarchy and hunting strategies of grey wolves. The wolf pack has a strict social hierarchy: alpha ($\alpha$), beta ($\beta$), delta ($\delta$), and omega ($\omega$). The $\alpha$ wolf is the leader, followed by $\beta$ and $\delta$ wolves. The $\omega$ wolves follow the other wolves. GWO simulates the hunting process, where the wolves encircle, chase, and attack prey (Makhadmeh et al., 2023). The algorithm uses this hierarchical

structure and hunting behavior to effectively search for the optimal solution. The grey wolves are ranked according to their fitness values. The wolf with the best fitness is designated as alpha ($\alpha$), the second-best as beta ($\beta$), and the third-best as delta ($\delta$). Each wolf represents a possible $\sigma$ value. The fitness of each wolf is calculated based on the Random Forest model's error rate. The $\alpha$, $\beta$, and $\delta$ wolves guide the remaining wolves toward the optimal solution. Wolves adjust their $\sigma$ values by moving toward the leaders while maintaining some randomness for exploration. After several iterations, the $\alpha$ wolf (best solution) provides the optimal $\sigma$ value (Makhadmeh et al., 2023).

### 2.3.4 Optimization using FA

FA is inspired by the flashing behavior of fireflies, where, brighter fireflies (better solutions) attract less bright ones. Movement towards brightness (exploitation) is balanced with random movement (exploration). The algorithm simulates the movement of fireflies towards brighter ones, representing better solutions in the search space (Kumar & Kumar, 2021). Each firefly represents a candidate $\sigma$ value. The process includes evaluating the fitness of each firefly based on the Random Forest error rate and moving toward brighter fireflies (better solutions). The movement also includes a random exploration component to avoid local optima, which helps in exploration. The position of each firefly is updated in each iteration based on the attraction and random movement. After multiple iterations, the brightest firefly (best solution) provides the optimized $\sigma$ value (Kumar & Kumar, 2021).

After applying these four optimization algorithms, the intersection of the top-ranked features identified by each method is determined, resulting in a reliable selection of the most important predictors. By focusing on the shared top-ranked features from these algorithms, the proposed approach strengthens the robustness of feature selection, contributing to more informed clinical decisions and improved patient outcomes.

## 3. Results and Discussion

As discussed in Section 2.2, the LIME and SHAP importance scores for 15 relevant features from the Cirrhosis Patient Survival Prediction dataset are independently computed. These individual importance values are presented in Table 2. As detailed in Section 2.3, the final combined importance score for each feature $f_i$ is calculated using the weighted sum approach: $\sigma.L_i + \eta.S_i$; where, $L_i$ and $S_i$ represent the

LIME and SHAP importance values for $f_i$, and $\sigma$ and $\eta$ are the weights assigned to the LIME and SHAP contributions, respectively, with the constraint $\sigma + \eta = 1$.

To determine the optimal values for $\sigma$ and $\eta$, four metaheuristic algorithms PSO, WOA, GWO, and FA are employed, as explained in Section 2.3. After optimization, the combined feature importance scores for each of the 15 features are calculated and shown in Table 3. The corresponding optimal values of $\sigma$ and $\eta$ for each algorithm are also indicated in Table 3, enclosed in parentheses next to the method name.

**Table 2. Feature importance scores obtained from the XAI techniques LIME and SHAP**

| Feature ($f_i$) | LIME importance score ($L_i$) | SHAP importance score ($S_i$) |
|---|---|---|
| Age | 0.3505 | 0.0210 |
| Albumin | 0.3993 | 0.0283 |
| Alk_Phos | 0.0833 | 0.0088 |
| Ascites | 0.0623 | 0.0896 |
| Bilirubin | 0.1738 | 0.2215 |
| Cholesterol | 0.1913 | 0.0300 |
| Copper | 0.2571 | 0.2918 |
| Edema | 0.0303 | 0.0527 |
| Hepatomegaly | 1.0000 | 1.0000 |
| Platelets | 0.3186 | 0.2698 |
| Prothrombin | 0.6237 | 0.2308 |
| Sex | 0.0000 | 0.0245 |
| SGOT | 0.0497 | 0.0000 |
| Spiders | 0.0666 | 0.0799 |
| Tryglicerides | 0.2326 | 0.1037 |

Furthermore, the sorted feature importance scores obtained from each of the four algorithms are visualized in the figures 1 to 4 using horizontal bar charts, where features are arranged from most to least important (from bottom to top). From Table 3 and the visualizations in the figures 1 to 4, it is evident that 'Hepatomegaly' consistently ranks as the most important feature across all four optimization techniques, with an importance score of 1.0 in each case. However, for other features,

importance rankings vary across the algorithms. 'Prothrombin', 'Platelets', 'Copper', and 'Albumin' consistently receive high importance values under PSO, WOA, and GWO. Another observation is that FA, which relies entirely on LIME for feature importance scoring, tends to assign lower importance to several features (such as 'Albumin' and 'Age') compared to the other three algorithms.

**Table 3. Optimized feature importance scores using metaheuristic algorithms**

| Feature | PSO ($\sigma$ = 0.6227, $\eta$ = 0.3773) | WOA ($\sigma$ = 0.0028, $\eta$ = 0.9972) | GWO ($\sigma$ = 0.1962, $\eta$ = 0.8038) | FA ($\sigma$ = 1, $\eta$ = 0) |
|---|---|---|---|---|
| Age | 0.1453 | 0.3496 | 0.2859 | 0.0210 |
| Albumin | 0.1683 | 0.3982 | 0.3265 | 0.0283 |
| Alk_Phos | 0.0369 | 0.0831 | 0.0687 | 0.0088 |
| Ascites | 0.0793 | 0.0624 | 0.0676 | 0.0896 |
| Bilirubin | 0.2035 | 0.1739 | 0.1831 | 0.2215 |
| Cholesterol | 0.0909 | 0.1908 | 0.1596 | 0.0300 |
| Copper | 0.2787 | 0.2572 | 0.2639 | 0.2918 |
| Edema | 0.0442 | 0.0304 | 0.0347 | 0.0527 |
| Hepatomegaly | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Platelets | 0.2882 | 0.3185 | 0.3090 | 0.2698 |
| Prothrombin | 0.3790 | 0.6226 | 0.5466 | 0.2308 |
| Sex | 0.0153 | 0.0001 | 0.0048 | 0.0245 |
| SGOT | 0.0188 | 0.0496 | 0.0400 | 0.0000 |
| Spiders | 0.0749 | 0.0666 | 0.0692 | 0.0799 |
| Tryglicerides | 0.1524 | 0.2322 | 0.2073 | 0.1037 |

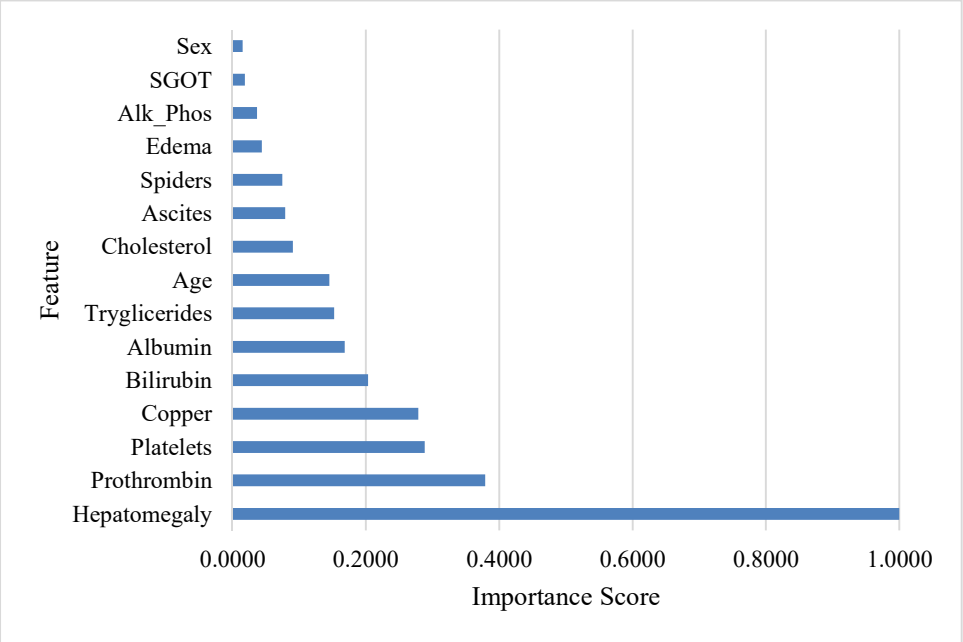


**Fig. 1: Graphical representation of feature importance scores obtained by applying PSO**

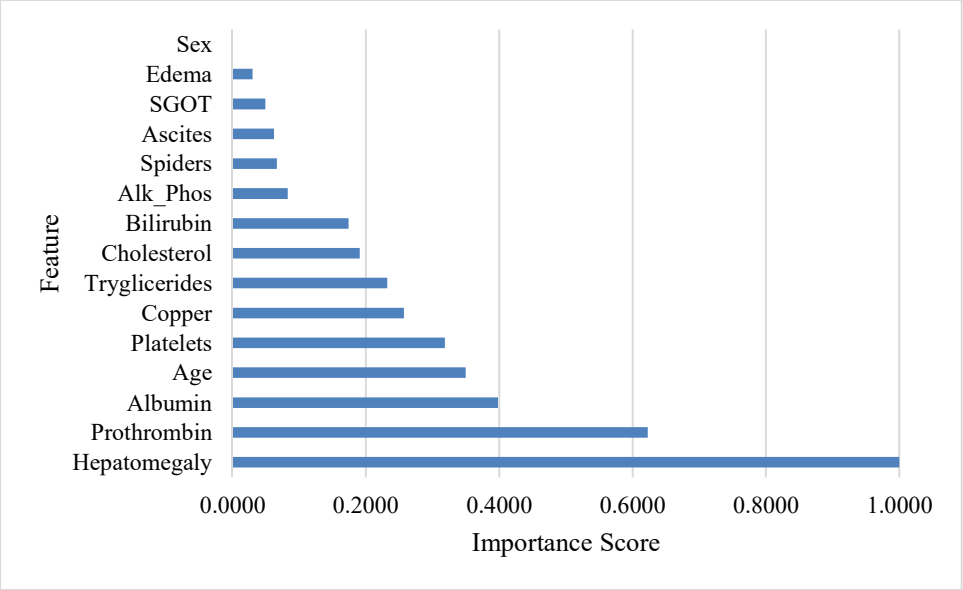


**Fig. 2: Graphical representation of feature importance scores obtained by applying WOA**

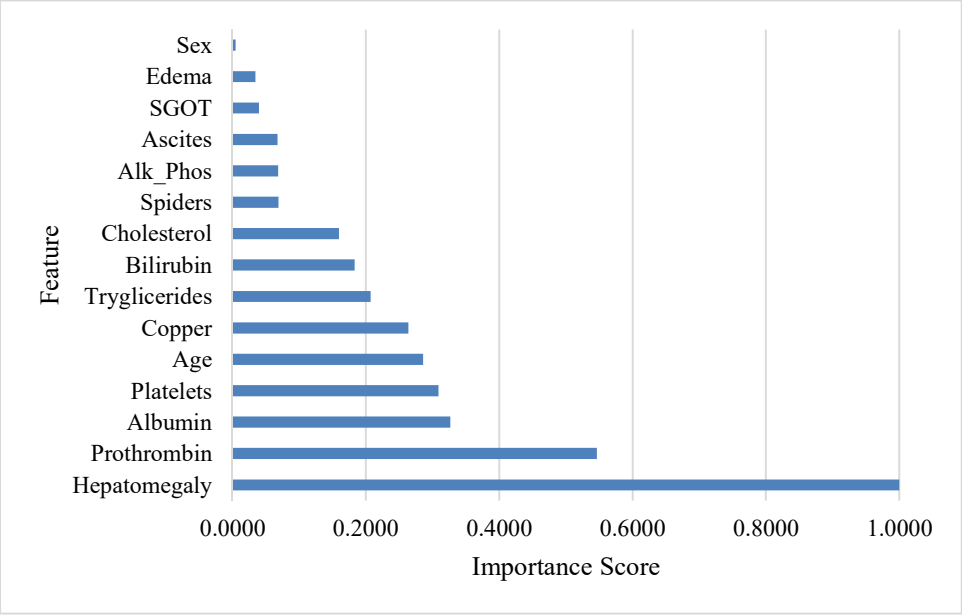


**Fig. 3: Graphical representation of feature importance scores obtained by applying GWO**

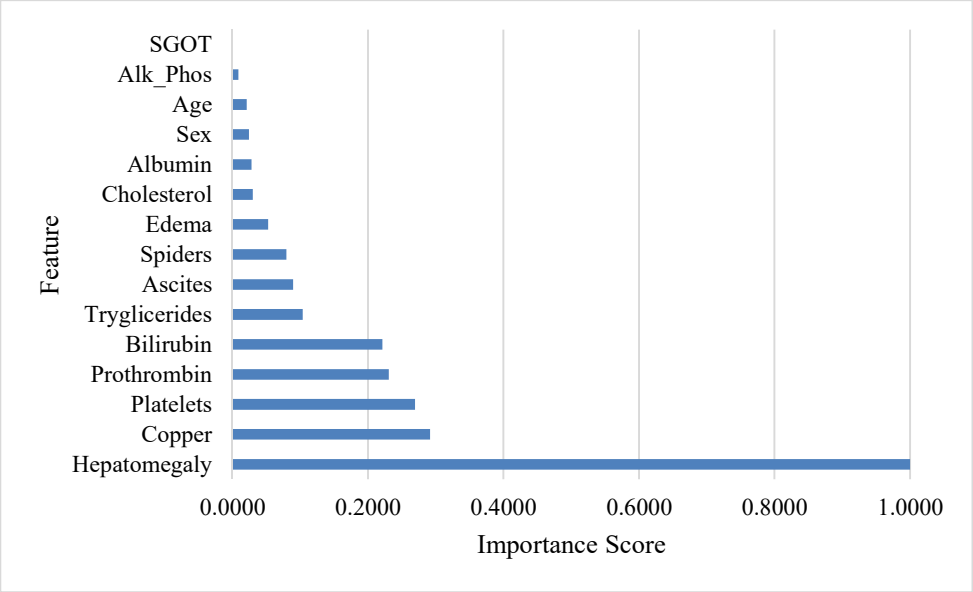


**Fig. 4: Graphical representation of feature importance scores obtained by applying FA**

From the previous discussions it is evident that, determining the intersection of feature rankings derived from the four metaheuristic algorithms is crucial for identifying the most robust set of important features for cirrhosis stage prediction. This intersection process strengthens the significance of consistently identified features and minimizes potential biases introduced by any single optimization algorithm. Table 4 presents the results of this intersection, demonstrating the consistency of feature selection across

the four algorithms. In this table, the first column INTER(n) denotes the intersection of the top n features from each method (where n ranges from 1 to 15), the second column displays the specific set of features (S) that are common to that intersection, and the third column indicates the total number of features (nf) present within each of the resulting intersected sets.

Table 4 illustrates how the intersection evolves as the number of top features considered from each method increases from 1 to 15. Notably, 'Hepatomegaly' is the only feature consistently ranked as the most important across all methods, as shown in the first three intersection levels INTER(1) to INTER(3). As the number of considered top features increases to five (i.e., nf = 5), additional four features 'Prothrombin', 'Platelets', 'Copper', and 'Triglycerides' are included in the set S (shown in row 7 of Table 4), indicating their high relevance, albeit with slightly more variability than 'Hepatomegaly'. By the time the top 10 features are considered, a broader and more diverse set of features including 'Bilirubin', 'Cholesterol', 'Albumin', 'Spiders' and, 'Ascites' appear in the intersection set, underscoring their collective importance. Since, Sex, and SGOT are assigned relatively low scores across all the four algorithms, they are included in S at the last intersection step i.e., INTER(15), indicating their limited relevance in predicting cirrhosis stages.

**Table 4. Intersection of top-ranked features identified by PSO, WOA, GWO, and FA**

| INTER(n) | Set of features (S) | nf |
|---|---|---|
| 1 | {'Hepatomegaly'} | 1 |
| 2 | {'Hepatomegaly'} | 1 |
| 3 | {'Hepatomegaly'} | 1 |
| 4 | {'Hepatomegaly', 'Prothrombin'} | 2 |
| 5 | {'Hepatomegaly', 'Platelets', 'Prothrombin'} | 3 |
| 6 | {'Copper', 'Hepatomegaly', 'Platelets', 'Prothrombin'} | 4 |
| 7 | {'Trygliceride s', 'Hepatomegaly', 'Platelets', 'Copper', 'Prothrombin'} | 5 |
| 8 | {'Trygliceride s', 'Hepatomegaly', 'Platelets', 'Copper', 'Prothrombin'} | 5 |
| 9 | {'Bilirubin', 'Trygliceride s', 'Hepatomegaly', 'Platelets', 'Copper', 'Prothrombin'} | 6 |
| 10 | {'Bilirubin', 'Trygliceride s', 'Hepatomegaly', 'Platelets', 'Cholesterol', 'Copper', 'Prothrombin'} | 7 |
| 11 | {'Bilirubin', 'Trygliceride s', 'Hepatomegaly', 'Platelets', 'Cholesterol', 'Albumin', 'Copper', 'Spiders', 'Prothrombin'} | 9 |

| | | |
|---|---|---|
| 12 | {'Bilirubin', 'Tryglicerides', 'Ascites', 'Hepatomegaly', 'Platelets', 'Cholesterol', 'Albumin', 'Copper', 'Spiders', 'Prothrombin'} | 10 |
| 13 | {'Bilirubin', 'Tryglicerides', 'Age', 'Ascites', 'Hepatomegaly', 'Platelets', 'Cholesterol', 'Albumin', 'Copper', 'Spiders', 'Prothrombin'} | 11 |
| 14 | {'Bilirubin', 'Tryglicerides', 'Age', 'Alk_Phos', 'Ascites', 'Edema', 'Hepatomegaly', 'Platelets', 'Cholesterol', 'Albumin', 'Copper', 'Spiders', 'Prothrombin'} | 13 |
| 15 | {'Bilirubin', 'Tryglicerides', 'Age', 'Alk_Phos', 'Ascites', 'Edema', 'Hepatomegaly', 'Platelets', 'Cholesterol', 'Albumin', 'Copper', 'Spiders', 'Sex', 'SGOT', 'Prothrombin'} | 15 |

## 4. Conclusion

In this study, a robust and interpretable feature selection framework is proposed for liver cirrhosis stage prediction by combining explainable AI techniques—SHAP and LIME—with metaheuristic optimization algorithms—PSO, WOA, GWO, and FA. The integration of LIME and SHAP allows for the computation of feature importance scores from both local and global perspectives, while the use of optimization techniques ensures the derivation of optimal weight combinations that maximize their complementary strengths. The comparative analysis of feature rankings generated by the four metaheuristic algorithms, followed by their intersection, enhances the stability and reliability of the selected feature set. This intersection-based feature selection approach effectively mitigates the bias introduced by any single optimization method and identifies a robust set of clinically significant predictors. These findings should provide actionable insights for clinicians, contributing to more accurate and interpretable decision-making in cirrhosis staging. Future work may involve validating the proposed methodology across diverse medical datasets and exploring other feature attribution techniques to further enrich the model's interpretability and generalizability.

## References


- Arroyo, V., Moreau, R., Kamath, P. S., Jalan, R., Ginès, P., Nevens, F., & Schnabl, B. (2016). Acute-on-chronic liver failure in cirrhosis. *Nature reviews Disease primers*, *2*(1), 1- 18.

- Garcia-Martinez, R., Caraceni, P., Bernardi, M., Gines, P., Arroyo, V., & Jalan, R. (2013). Albumin: pathophysiologic basis of its role in the treatment of cirrhosis and its complications. *Hepatology*, *58*(5), 1836-1846.
- Mozos, I. (2015). Arrhythmia risk in liver cirrhosis. *World journal of hepatology*, *7*(4).
- Younossi, Z., & Henry, L. (2015). Systematic review: patient-reported outcomes in chronic hepatitis C-the impact of liver disease and new treatment regimens. *Alimentary pharmacology & therapeutics*, *41*(6), 497-520.
- Pinto, R. B., Schneider, A. C. R., & da Silveira, T. R. (2015). Cirrhosis in children andadolescents: An overview. *World journal of hepatology*, *7*(3), 392-405.
- van Zutphen, T., Ciapaite, J., Bloks, V. W., Ackereley, C., Gerding, A., Jurdzinski, A., de Moraes, R. A., Zhang, L., Wolters, J. C., & Bischoff, R. (2016). Malnutrition-associated liver steatosis and ATP depletion is caused by peroxisomal and mitochondrial dysfunction. *Journal of Hepatology*, *65*(6), 1198-1208.
- Guerci, P., Ergin, B., Uz, Z., Ince, Y., Westphal, M., Heger, M., & Ince, C. (2019). Glycocalyxdegradation is independent of vascular barrier permeability increase in nontraumatichemorrhagic shock in rats. *Anesthesia & Analgesia*, *129*(2), 598-607.
- Ginès, P., Krag, A., Abraldes, J. G., Solà, E., Fabrellas, N., & Kamath, P. S. (2021). Livercirrhosis. *The Lancet*, *398*(10308), 1359-1376.
- Acharya, U. R., Faust, O., Molinari, F., Sree, S. V., Junnarkar, S. P., & Sudarshan, V. (2015). Ultrasound-based tissue characterization and classification of fatty liver disease: A screening and diagnostic paradigm. *Knowledge-Based Systems*, *75*, 66-77.
- Nam, D., Chapiro, J., Paradis, V., Seraphin, T. P., & Kather, J. N. (2022). Artificial intelligence in liver diseases: Improving diagnostics, prognostics and response prediction. *JHEP Reports*, *4*(4), 1-11.
- Veena, G., Sneha, D., Basavaraju, D., & Tanvi, T. (2018). Effective analysis and diagnosis of liver disorder. In *2018 International Conference on Communication and Signal Processing*, 86-90.
- Auxilia, L. A (2018). Accuracy prediction using machine learning techniques for Indian patient liver disease. In *2018 Second International Conference on Trends in Electronics and Informatics*, 45-50.

- Rahman, A. K. M., Shamrat, F. M., Tasnim, Z., Roy, J., & Hossain, S. (2019). A comparative study on liver disease prediction using supervised machine learning algorithms. *International Journal of Scientific & Technology Research*, *8*(4), 419-422.
- Rabbi, M. F., Hasan, S. M. M., Champa, A. I., Zaman, M. A., & Hasan, M. K. (2020). Prediction of liver disorders using machine learning algorithms: A comparative study. In *2020 2nd International Conference on Advanced Information and Communication Technology*, 111-116.
- Shaheamlung, G., & Kaur, H. (2021). The diagnosis of chronic liver disease using machine learning techniques. *IT in Industry*, *9*(2), 545-565.
- Singh, V., Gourisaria, M. K., & Das, H. (2021). Performance analysis of machine learning algorithms for prediction of liver disease. In *2021 6th International Conference on Inventive Computation Technologies*, 1-7.
- Utku, A. (2023). Deep learning based cirrhosis detection. *Operational Research in Engineering Sciences: Theory and Applications*, *6*(1), 95-114.
- Topcu, A. E., Elbasi, E., & Alzoubi, Y. I. (2024). Machine learning-based analysis and prediction of liver cirrhosis. In *2024 47th International Conference on Telecommunications and Signal Processing*, 191-194.
- Stekhoven, D. J., & Bühlmann, P. (2012). MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*, *28*(1), 112-118.
- Zafar, M. R., & Khan, N. (2021). Deterministic local interpretable model-agnostic explanations for stable explainability. *Machine Learning and Knowledge Extraction*, *3*(3), 525-541.
- Nohara, Y., Matsumoto, K., Soejima, H., & Nakashima, N. (2022). Explanation of machine learning models using shapley additive explanation and application for real data in hospital. *Computer Methods and Programs in Biomedicine*, *214*, 106584.
- Wang, D., Tan, D., & Liu, L. (2018). Particle swarm optimization algorithm: an overview. *Soft computing*, *22*(2), 387-408.
- Nadimi-Shahraki, M. H., Zamani, H., Asghari Varzaneh, Z., & Mirjalili, S. (2023). A systematic review of the whale optimization algorithm: theoretical foundation, improvements, and hybridizations. *Archives of Computational Methods in Engineering*, *30*(7), 4113-4159.
- Makhadmeh, S. N., Al-Betar, M. A., Doush, I. A., Awadallah, M. A., Kassaymeh, S., Mirjalili, S., & Zitar, R. A. (2023). Recent advances in Grey Wolf Optimizer, its versions and applications. *Ieee Access*, *12*, 22991-23028.

- Kumar, V., & Kumar, D. (2021). A systematic review on firefly algorithm: past, present, and future. *Archives of Computational Methods in Engineering*, *28*, 3269-3291.