A Human-Inspired Metaheuristic Optimization Framework for Accurate Liver Disease Prediction Using Clinical Laboratory Data

Benyamin Abdollahzadeh[1,*], Marwa M. Eid[2,3]

[1]Department of Mathematics, Faculty of Science, University of Hradec Králové, 500 03, Hradec Králové, Czech Republic

[2]Faculty of Artificial Intelligence, Delta University for Science and Technology, Mansoura 11152, Egypt

[3]Jadara Research Center, Jadara University, Irbid 21110, Jordan

Emails: benyamin.abdolahzade@gmail.com; mmm@ieee.org

Abstract

The rapid increase in liver disease prevalence worldwide, particularly in developing regions, necessitates accurate and reliable diagnostic systems capable of supporting early clinical decision-making based on routine laboratory data. Traditional diagnostic approaches and unoptimized machine learning models often struggle to fully capture the complex, nonlinear relationships among biochemical liver indicators, leading to suboptimal predictive reliability. Motivated by these challenges, this study proposes a human-inspired metaheuristic optimization framework that integrates the iHow Optimization Algorithm (iHOW) with the Extreme Gradient Boosting model (XGBoost) to enhance liver disease prediction performance. The main contribution of this work lies in the development of an optimized diagnostic pipeline that systematically tunes XGBoost hyperparameters using iHOW and rigorously benchmarks its effectiveness against established metaheuristic optimizers, including Genetic Algorithm (GA), Particle Swarm Optimizer (PSO), Grey Wolf Optimizer (GWO), and Greylag Goose Optimization (GGO). Experimental evaluation is conducted on a clinically sourced liver disease dataset using multiple diagnostic metrics. In the baseline stage, the unoptimized XGBoost model achieves an accuracy of 0.921875, sensitivity of 0.920245399, specificity of 0.923566879, and F-Score of 0.923076923. After hyperparameter optimization, the proposed iHOW+XGBoost framework demonstrates substantial performance enhancement, attaining an accuracy of 0.983696458, sensitivity of 0.983391608, specificity of 0.984012066, and F-Score of 0.983965015, outperforming GA+XGBoost, PSO+XGBoost, GWO+XGBoost, and GGO+XGBoost across all evaluated metrics. These results confirm the effectiveness of human-inspired optimization in navigating complex hyperparameter search spaces and improving diagnostic robustness. The findings of this study highlight the practical implications of integrating advanced metaheuristic optimization with ensemble learning models, offering a highly accurate, reliable, and scalable decision-support framework that can be leveraged for early liver disease screening and extended to other medical diagnostic and predictive healthcare applications.

Keywords: Liver disease prediction; Machine learning in healthcare; Metaheuristic optimization; Hyperparameter tuning; Clinical decision support systems

1 Introduction

Liver diseases represent one of the most critical and rapidly escalating global public health challenges, imposing a substantial burden on healthcare systems, national economies, and affected populations

DOI: https://doi.org/10.54216/JAIM.110104
Received: September 26, 2025 Revised: November 29, 2025 Accepted: January 06, 2026

worldwide [1], [2], [3]. Chronic hepatic disorders, including hepatitis, cirrhosis, fatty liver disease, and progressive liver failure, continue to rank among the leading causes of morbidity and mortality, particularly in developing and low-income regions. In such settings, limited access to advanced diagnostic infrastructure, shortages of specialized medical professionals, and constrained healthcare resources frequently result in delayed diagnosis and insufficient monitoring. Consequently, a significant proportion of patients are diagnosed only at advanced stages of disease progression, when therapeutic interventions become less effective and long-term clinical outcomes are markedly compromised [4], [5].

Beyond their direct clinical impact, liver diseases exert profound socioeconomic consequences that extend to patients, families, and healthcare systems. Individuals suffering from chronic liver dysfunction often experience diminished quality of life, reduced work capacity, and long-term disability, while healthcare providers face increased financial pressure due to prolonged hospitalization, repeated laboratory testing, and the need for complex and costly treatment regimens [6]. Delayed or inaccurate diagnosis further exacerbates these challenges by accelerating disease progression, increasing the likelihood of irreversible liver damage, and limiting the effectiveness of available therapeutic options. Collectively, these factors highlight the urgent need for reliable, efficient, and early diagnostic methodologies capable of supporting timely clinical decision-making and mitigating the growing burden of liver disease [7].

Early detection plays a pivotal role in improving patient prognosis and enabling timely medical intervention [8]. Clinical assessment of liver function primarily relies on biochemical laboratory tests, including measurements of bilirubin levels, liver enzymes, total proteins, albumin concentration, and protein ratios. These biomarkers provide essential insight into hepatic metabolic activity, detoxification capability, and synthetic function [9], [10]. However, interpreting such indicators remains a challenging task due to their complex nonlinear relationships, strong interdependencies, and sensitivity to demographic factors such as age and gender. Conventional diagnostic approaches are largely dependent on manual interpretation, predefined threshold values, and physician expertise, which may vary across clinical settings and practitioners. As a result, traditional methods often struggle to capture subtle multivariate patterns within high-dimensional clinical data, leading to diagnostic variability and limited scalability in large patient populations.

In recent years, machine learning has emerged as a powerful paradigm for addressing these limitations in medical diagnosis [11], [12]. By leveraging data-driven learning mechanisms, machine learning models can automatically identify complex nonlinear interactions among clinical biomarkers that are difficult to model using traditional statistical techniques. Unlike conventional approaches that rely on simplifying assumptions regarding data distribution, machine learning algorithms can adaptively learn discriminative patterns from high-dimensional feature spaces, enabling more accurate, consistent, and reproducible predictions. Furthermore, machine learning-based diagnostic systems facilitate automation and objective decision support, reducing reliance on subjective clinical judgment and improving diagnostic consistency. A growing body of literature has demonstrated the superiority of machine learning-based approaches over classical statistical models in a wide range of disease prediction tasks, thereby motivating their application to liver disease detection and screening [13].

Despite these advantages, the application of machine learning to liver disease prediction presents several inherent challenges that complicate model development and deployment. One of the most prominent challenges arises from the high dimensionality of biochemical feature spaces. Liver-related clinical datasets typically comprise numerous laboratory indicators that reflect overlapping physiological processes. While these features collectively encode valuable diagnostic information, their simultaneous inclusion increases computational complexity and exacerbates the curse of dimensionality, potentially impairing learning efficiency and model generalization.

Closely associated with high dimensionality is the issue of feature redundancy and irrelevance. Many biochemical indicators, such as liver enzymes and protein ratios, exhibit strong correlations, while certain features contribute limited discriminatory power for disease classification. The presence of redundant or weakly informative variables can obscure meaningful patterns, inflate model variance, and degrade predictive performance. Consequently, identifying and selecting the most informative subset of clinical biomarkers is essential for constructing efficient, interpretable, and robust diagnostic models.

DOI: https://doi.org/10.54216/JAIM.110104

Received: September 26, 2025 Revised: November 29, 2025 Accepted: January 06, 2026

Another significant challenge stems from class imbalance and demographic bias within liver disease datasets. Clinical records often exhibit unequal distributions between liver patients and non-liver individuals, reflecting real-world screening and referral practices. Such imbalance can bias learning algorithms toward the majority class, reducing sensitivity in detecting affected patients. Additionally, demographic characteristics, including male-dominant patient populations and age-capping practices for elderly individuals, introduce further complexity by embedding population-specific biases into the data. If not adequately addressed, these factors may limit the fairness, robustness, and generalizability of predictive models.

Moreover, the performance of advanced machine learning models is highly sensitive to hyperparameter configurations. Ensemble-based learning methods, particularly boosting architectures, require careful tuning to achieve an optimal balance between bias and variance. Inadequate hyperparameter selection may lead to underfitting, overfitting, or unstable learning behavior, especially when applied to noisy and limited clinical datasets. Manual tuning of hyperparameters is both time-consuming and suboptimal, underscoring the necessity for automated and intelligent optimization strategies [14].

Ensuring robust generalization remains an additional challenge in medical machine learning applications. Clinical datasets are frequently constrained in size, heterogeneous in quality, and susceptible to noise arising from laboratory variability and patient-specific factors. These characteristics increase the risk of overfitting, whereby models exhibit strong apparent performance on training data but fail to maintain reliability when applied to unseen patient cases. Addressing this issue is essential for developing clinically trustworthy liver disease prediction systems capable of real-world deployment.

Against this backdrop, the primary objective of this study is to develop a robust and intelligent machine learning framework for liver disease prediction using routine clinical and biochemical data. Initially, multiple machine learning classifiers are developed and systematically evaluated to establish baseline predictive performance. This comparative assessment provides insight into the relative suitability of different learning paradigms for modeling liver-related clinical features.

Building upon the baseline analysis, the study aims to enhance predictive performance through metaheuristic-based feature selection and hyperparameter optimization. By intelligently identifying the most informative subset of clinical biomarkers and optimizing model configurations, the proposed approach seeks to improve classification accuracy while reducing computational complexity. Particular emphasis is placed on achieving balanced predictive behavior that is clinically meaningful and resilient to data variability.

In addition to performance enhancement, the proposed framework prioritizes robustness and generalization. The optimization strategy is designed to mitigate overfitting, stabilize learning dynamics, and support consistent predictive behavior across diverse patient profiles. These objectives align with the broader goal of advancing intelligent decision support systems for early and accurate liver disease diagnosis.

To this end, this study proposes a hybrid metaheuristic–machine learning framework tailored specifically for liver disease prediction using biochemical and demographic clinical data. The methodology integrates an improved metaheuristic optimization strategy with a powerful ensemble learning model, enabling simultaneous feature selection and hyperparameter tuning within a unified optimization framework. This integration provides a systematic solution to the intertwined challenges of dimensionality, redundancy, and model sensitivity commonly encountered in clinical prediction tasks.

Furthermore, the study establishes a comprehensive evaluation framework to assess predictive performance before and after optimization, facilitating a transparent analysis of the impact of the proposed approach. By emphasizing clinically relevant evaluation criteria, the framework highlights the role of intelligent optimization in enhancing diagnostic reliability without increasing model complexity.

The proposed framework also contributes to clinical interpretability by reducing reliance on redundant biomarkers while preserving essential diagnostic information. This aspect is particularly important in healthcare applications, where transparency, efficiency, and trust are critical prerequisites for real-world adoption.

DOI: https://doi.org/10.54216/JAIM.110104

Received: September 26, 2025 Revised: November 29, 2025 Accepted: January 06, 2026

The main contributions of this study can be summarized as follows:

- A comprehensive and intelligent framework is proposed for liver disease prediction based on routine clinical and biochemical laboratory data, addressing key challenges associated with nonlinear feature interactions, model sensitivity, and diagnostic reliability.
- A human-inspired metaheuristic optimization strategy is integrated with an ensemble machine learning model to perform automated hyperparameter tuning, enabling adaptive exploration and exploitation of the search space and improving model robustness.
- An extensive comparative evaluation is conducted in which the proposed methodology is systematically benchmarked against multiple state-of-the-art machine learning models and well-established metaheuristic optimization algorithms under a unified experimental protocol.
- The study provides a rigorous multi-metric assessment framework tailored to medical diagnosis, ensuring balanced evaluation of disease detection capability, non-disease recognition, and predictive reliability.
- The proposed approach emphasizes clinical interpretability and practical applicability by relying exclusively on routinely collected laboratory indicators, making it suitable for real-world deployment in early liver disease screening and decision support systems.
- The findings of this work contribute to the growing body of research on intelligent optimization in healthcare by demonstrating the effectiveness of human-inspired metaheuristics in enhancing advanced machine learning models for clinical prediction tasks.

The remainder of this paper is organized as follows. Section 2 describes the dataset, preprocessing procedures, machine learning models, and optimization strategies employed in the study. Section 3 outlines the experimental setup and evaluation methodology. Section 4 presents the empirical analysis and comparative evaluation of the proposed framework. Section 5 discusses the findings in relation to existing literature, highlighting implications and limitations. Finally, Section 6 concludes the paper and outlines directions for future research.

2 Literature Review

The global prevalence of liver diseases has increased substantially due to environmental pollution, unhealthy dietary habits, metabolic disorders, and excessive alcohol consumption. This growing burden has motivated extensive research into intelligent diagnostic systems capable of improving early detection, prediction accuracy, and clinical decision-making.

A laboratory-data-driven diagnostic framework, known as the Liver Patients Detection Strategy (LPDS), was proposed to overcome the limitations of conventional diagnostic approaches [15]. The framework integrates preprocessing, feature selection, and classification into a unified pipeline. A key contribution of this work is the Improved Binary Butterfly Optimization Algorithm (IB2OA), which employs Information Gain for initial feature reduction and opposition-based learning for improved initialization. Experimental results using multiple classifiers demonstrated superior diagnostic performance, with KNN achieving the highest accuracy.

A comprehensive review of machine learning techniques for early-stage chronic liver disease detection was presented in [16]. The study critically analyzed existing research with respect to precision, recall, generalization capability, and transferability. It highlighted that although ML models demonstrate strong predictive potential, their performance is highly sensitive to sample size, feature selection strategies, and model complexity.

Beyond diagnosis, machine learning has been applied to prognostic risk stratification in metabolic dysfunction-associated steatotic liver disease (MASLD). A survival-based CART model was developed

DOI: https://doi.org/10.54216/JAIM.110104

Received: September 26, 2025 Revised: November 29, 2025 Accepted: January 06, 2026

to predict major adverse cardiac events in MASLD patients [17]. The study identified age and cumulative metabolic risk factors as strong predictors, offering a practical ML-based tool for long-term cardiovascular risk assessment.

Ensemble learning combined with hyperparameter optimization has shown notable success in liver disease prediction. In [18], multiple tree-based models were integrated with the Tree-Structured Parzen Estimator (TPE) for automated tuning. After addressing class imbalance and feature scaling, the optimized Extra Tree Classifier achieved a high diagnostic accuracy, demonstrating the effectiveness of probabilistic hyperparameter search.

Model interpretability was addressed through the integration of Polynomial features and SHapley Additive exPlanations in [19]. The study reported consistent improvements across multiple performance metrics and emphasized the importance of interpretability for clinical trust, particularly in capturing nonlinear feature interactions relevant to liver disease diagnosis.

Machine learning has also been employed in population-level nutritional studies. A large-scale analysis using random forest classifiers investigated the association between dietary nutrients and hepatic steatosis [20]. The findings revealed that higher manganese intake was associated with reduced steatosis risk, illustrating the potential of ML in nutritional epidemiology and disease prevention.

Risk prediction for hepatocellular carcinoma (HCC) in MASLD patients was explored using machine learning models trained on routine clinical and laboratory parameters [21]. The study demonstrated high predictive accuracy, with fibrosis-related scores emerging as the strongest predictors, supporting the role of ML in personalized cancer surveillance.

Image-based machine learning approaches have also contributed to liver disease staging. A quantitative vessel analysis system was developed to estimate arterial density in liver biopsies using a decision tree algorithm [22]. The results showed strong correlations between arterial density, collagen proportion, and fibrosis stage, suggesting a novel ML-assisted histological staging approach.

At the molecular level, integrative machine learning frameworks have enabled biomarker discovery. In [23], differential expression analysis, co-expression networks, and ML-based feature selection identified five inflammation-related hub genes associated with MASLD. These biomarkers demonstrated strong diagnostic potential and biological relevance.

Genetic studies have further benefited from machine learning. A trans-ancestral analysis leveraged ML-based phenotype prediction to enhance the detection of rare coding variants associated with MASLD [24]. The study highlighted the effectiveness of ML in overcoming phenotyping challenges and improving discovery power in large biobank datasets.

The shared molecular mechanisms between coronary artery disease and non-alcoholic fatty liver disease were investigated using integrated bioinformatics and machine learning approaches [25]. Several hub genes were identified as potential diagnostic biomarkers, with immune and inflammatory pathways emerging as key shared mechanisms.

The impact of data imbalance and ensemble learning on chronic liver disease diagnosis was examined in [26]. Hybrid models combining SMOTE-ENN, recursive feature elimination, and ensemble classifiers significantly improved predictive accuracy on benchmark datasets, underscoring the importance of balanced data representation.

Advanced feature selection and ensemble fusion strategies were proposed in [27], where LASSO-ANOVA-based recursive selection identified key biochemical and demographic features. The ensemble fusion of Random Forest, AdaBoost, and CatBoost achieved high diagnostic accuracy, demonstrating the effectiveness of hybrid ML pipelines.

Explainable artificial intelligence was further emphasized in [28], where SHAP and LIME were integrated with boosting-based classifiers. The approach successfully balanced predictive performance with transparency, identifying clinically meaningful predictors such as liver function tests, alcohol consumption, and age.

DOI: https://doi.org/10.54216/JAIM.110104

Received: September 26, 2025 Revised: November 29, 2025 Accepted: January 06, 2026

Finally, deep learning approaches were evaluated for liver disease classification using the Indian Liver Patient Dataset [29]. Convolutional Neural Networks outperformed traditional machine learning models by effectively capturing complex nonlinear patterns, highlighting the growing role of deep learning in automated liver disease diagnosis.

Overall, the reviewed studies demonstrate that machine learning and deep learning techniques significantly enhance liver disease diagnosis, prognosis, and biomarker discovery. Despite these advances, challenges related to interpretability, data heterogeneity, and clinical integration remain critical directions for future research.

To provide a structured and comparative overview of the existing literature, Table 1 summarizes the fifteen reviewed studies. The table highlights the primary research focus, adopted methodologies, datasets or data sources, and the key findings and contributions of each work. This comparative presentation facilitates a clearer understanding of how different machine learning, deep learning, and data-driven strategies have been applied to liver disease diagnosis, prognosis, risk stratification, and biomarker discovery, while also revealing methodological trends and research gaps across the literature.

3 Materials and Methods

3.1 Dataset Description

The dataset employed in this study was collected from clinical test samples obtained in the North East region of Andhra Pradesh, India, and is designed to support the automated detection of liver disease using routine laboratory measurements. The dataset reflects a real-world clinical context in which biochemical liver function indicators are routinely recorded as part of diagnostic and screening procedures. Its structure and composition make it suitable for investigating data-driven approaches to liver disease prediction while accounting for demographic and clinical variability commonly observed in hospital-based records.

The dataset comprises a total of 583 patient records, including individuals diagnosed with liver disease as well as non-liver patients. Each record corresponds to a unique patient and contains both demographic information and biochemical test results. The class distribution distinguishes between liver patients and non-liver patients using a predefined binary class label, enabling supervised learning for disease classification. The dataset exhibits demographic diversity, with records spanning a wide age range and including both male and female patients. To ensure consistency in age representation and to address sparsity at extreme values, all patients whose age exceeded 89 years were grouped under a single age category labeled as 90. This preprocessing decision reflects common practices in clinical data handling and supports model stability.

Gender distribution within the dataset is imbalanced, with a higher proportion of male patients compared to female patients. Such imbalance reflects observed epidemiological trends in liver disease prevalence but also introduces challenges related to demographic bias, which are considered during model development and evaluation. The dataset structure enables analysis of how demographic factors interact with biochemical indicators in liver disease prediction.

The feature set consists of both demographic and biochemical attributes. Demographic features include patient age and gender, which are known to influence liver function and disease susceptibility. Biochemical features correspond to standard liver function tests and related clinical indicators, including total bilirubin, direct bilirubin, liver enzymes, total proteins, albumin levels, and albumin-to-globulin ratio. These features collectively capture multiple aspects of hepatic metabolism, enzymatic activity, and protein synthesis, providing a comprehensive representation of liver health. The inclusion of both enzyme-based and protein-based indicators allows the learning models to exploit complementary diagnostic information embedded within routine laboratory measurements.

The target variable is defined as a binary class label denoted as `is_patient`, which indicates whether a given record corresponds to a liver disease patient or a non-liver individual. This formulation

DOI: https://doi.org/10.54216/JAIM.110104

Received: September 26, 2025 Revised: November 29, 2025 Accepted: January 06, 2026

Table 1: Summary of related studies on liver disease diagnosis, prediction, and analysis

Reference	Focus Area	Methodology	Key Findings and Contributions
[15]	Liver disease diagnosis	IB2OA-based feature selection with SVM, KNN, NB, DT, RF	Proposed LPDS framework; IB2OA reduced feature redundancy and improved accuracy, with KNN achieving the best performance.
[16]	Early chronic liver disease detection	Literature review of ML models	Identified strengths and limitations of ML methods; emphasized challenges related to sample size, feature selection, and generalization.
[17]	Cardiovascular risk in MASLD	Survival-based CART model	Demonstrated that age and cumulative metabolic risk factors effectively stratify MACE risk in MASLD patients.
[18]	Liver disease prediction	Ensemble ML with TPE hyperparameter tuning	Achieved high diagnostic accuracy; highlighted the importance of automated hyperparameter optimization.
[19]	Model performance and interpretability	Polynomial features with SHAP analysis	Improved predictive metrics and interpretability; enabled insight into nonlinear feature interactions.
[20]	Nutritional factors and steatosis	Random forest feature importance analysis	Identified manganese intake as inversely associated with hepatic steatosis in large cohorts.
[21]	HCC risk prediction in MASLD	ML models using clinical and lab data	Achieved high accuracy in predicting HCC; fibrosis scores emerged as dominant predictors.
[22]	Liver fibrosis staging	Decision tree-based image analysis (qVessel)	Quantified arterial density correlated with fibrosis stage, proposing a novel ML-assisted staging approach.
[23]	Biomarker discovery for MASLD	WGCNA with SVM-RFE, LASSO, RF	Identified five hub inflammatory genes with strong diagnostic potential.
[24]	Genetic analysis of MASLD	ML-based phenotype prediction with GWAS	Discovered rare and ancestry-specific variants; demonstrated ML utility in large biobank studies.
[25]	CAD and NAFLD association	Integrated bioinformatics and ML	Identified shared hub genes and immune pathways; proposed biomarkers for dual diagnosis.
[26]	Chronic liver disease diagnosis	SMOTE-ENN with ensemble ML	Improved performance on imbalanced datasets; highlighted benefits of hybrid data balancing.
[27]	Feature selection and ensemble fusion	LASSO-ANOVA with RF, AdaBoost, CatBoost	Selected key biochemical features; achieved high diagnostic accuracy and robustness.
[28]	Explainable liver disease prediction	Boosting classifiers with SHAP and LIME	Balanced high accuracy with interpretability; aligned model insights with clinical knowledge.
[29]	Deep learning-based diagnosis	CNN compared with traditional ML	CNN outperformed ML models by capturing complex nonlinear patterns in liver disease data.

DOI: https://doi.org/10.54216/JAIM.110104
Received: September 26, 2025 Revised: November 29, 2025 Accepted: January 06, 2026

frames the problem as a binary classification task, suitable for evaluating a wide range of machine learning models and optimization strategies. The binary structure of the target variable aligns with clinical decision-making scenarios, where the primary objective is to distinguish between affected and non-affected individuals.

For model development and evaluation, the dataset is partitioned into training, validation, and testing subsets. This data splitting strategy is designed to ensure unbiased assessment of model performance and to support robust hyperparameter tuning and feature selection. The training subset is used to learn model parameters, the validation subset supports optimization and model selection, and the testing subset provides an independent basis for evaluating generalization capability. Such a structured partitioning approach is essential for mitigating overfitting and ensuring that predictive models remain reliable when applied to unseen clinical data.

To examine the underlying distribution and variability of the input features used in the classification process, Fig. 1 presents overlaid histograms for all dataset attributes, including demographic, biochemical, and clinical indicators. This visualization provides an initial exploratory assessment of feature dispersion, overlap, and potential skewness across variables, which is essential for understanding data characteristics prior to preprocessing and model training. As shown in Fig. 1, differences in scale, concentration, and distribution among features highlight the necessity of normalization and feature selection strategies to ensure balanced model learning and robust predictive performance.

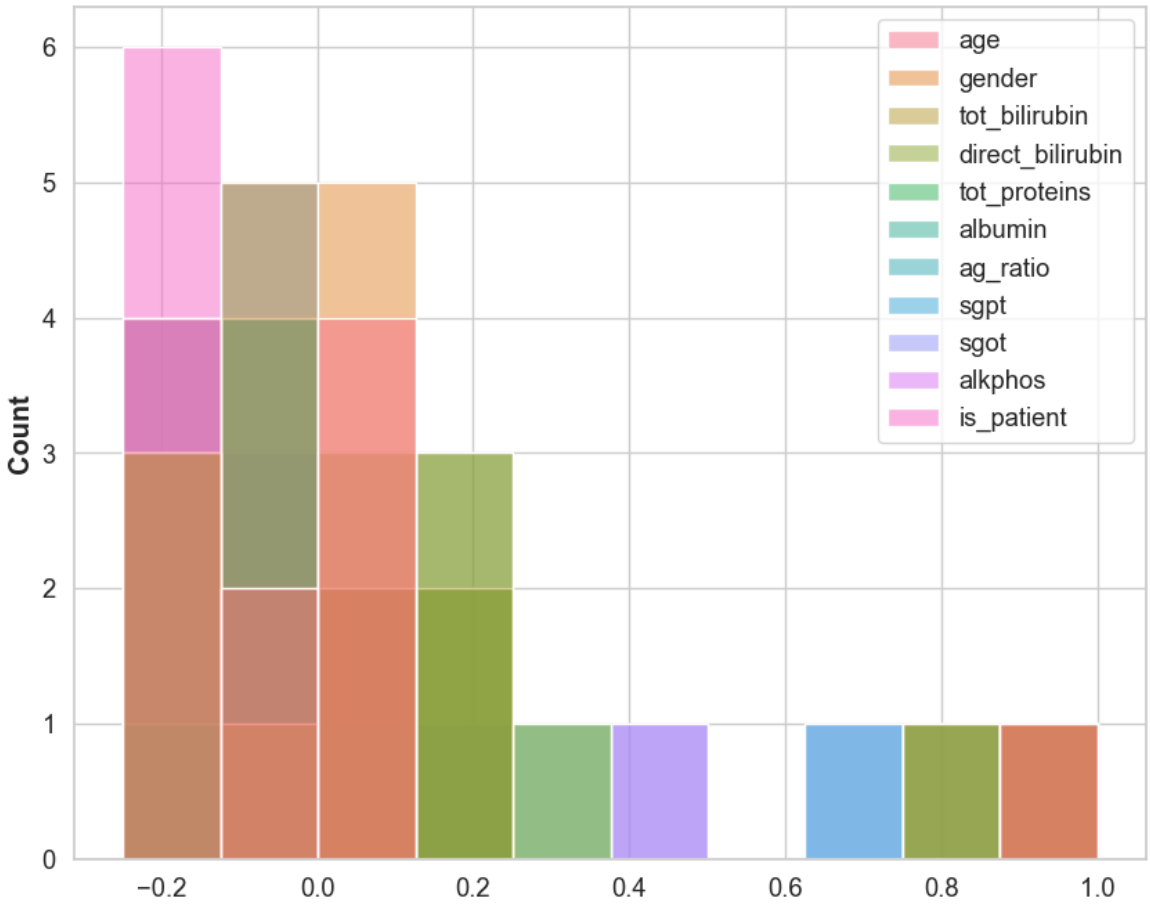

Figure 1: Overlaid histograms illustrating the distribution of dataset features used in the classification process.

To analyze the statistical behavior and comparative distribution of key liver enzyme biomarkers, Fig. 2 illustrates the kernel density estimates of serum glutamic pyruvic transaminase (SGPT) and serum glutamic oxaloacetic transaminase (SGOT) levels. This visualization enables a direct comparison of the central tendency, spread, and overlap between the two enzyme distributions, which are clinically

DOI: https://doi.org/10.54216/JAIM.110104

Received: September 26, 2025 Revised: November 29, 2025 Accepted: January 06, 2026

relevant indicators of hepatic function. As shown in Fig. 2, the distinct distributional patterns provide insight into their individual diagnostic contributions and justify their inclusion as discriminative features in liver disease classification models.

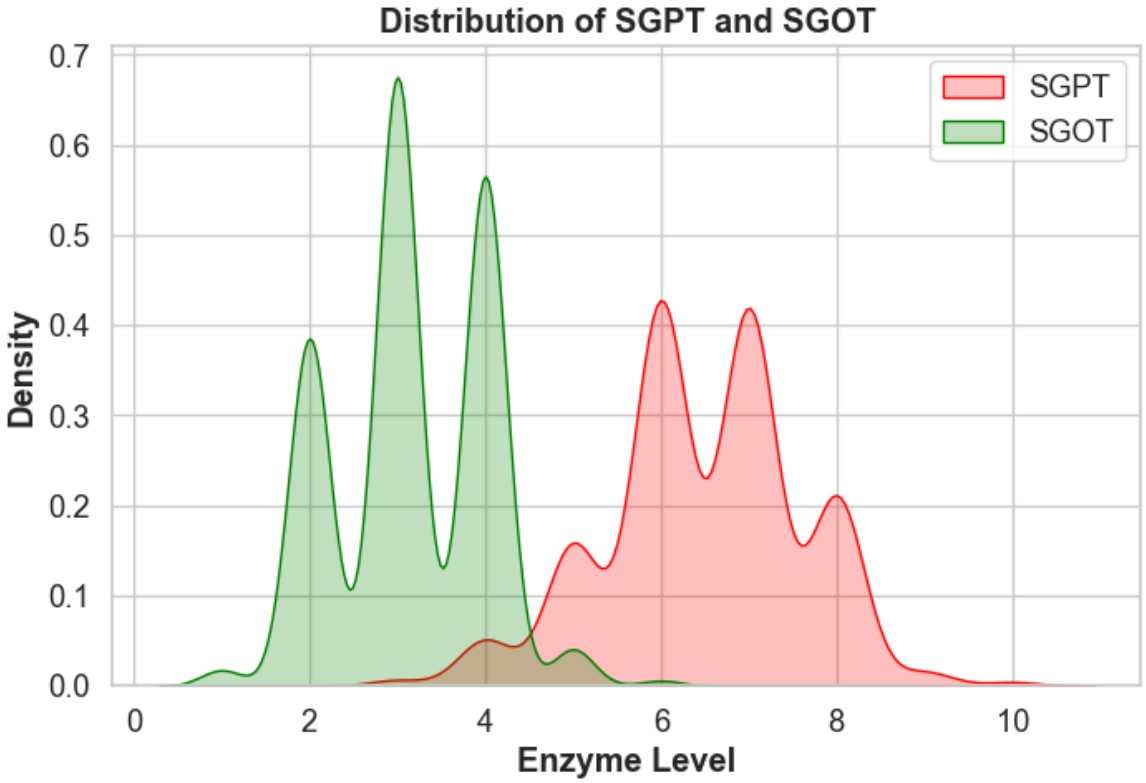

Figure 2: Kernel density distributions of SGPT and SGOT enzyme levels.

To further explore the distributional characteristics and variability of key liver enzyme biomarkers, Fig. 3 presents violin plots for serum glutamic pyruvic transaminase (SGPT), serum glutamic oxaloacetic transaminase (SGOT), and alkaline phosphatase (ALP). Unlike summary statistics alone, violin plots combine kernel density estimation with measures of central tendency, enabling a detailed visualization of data dispersion, multimodality, and skewness. As illustrated in Fig. 3, the distinct distributional profiles of these enzymes highlight their differing biochemical behaviors and underscore their relevance as informative features for liver disease classification.

DOI: https://doi.org/10.54216/JAIM.110104

Received: September 26, 2025 Revised: November 29, 2025 Accepted: January 06, 2026

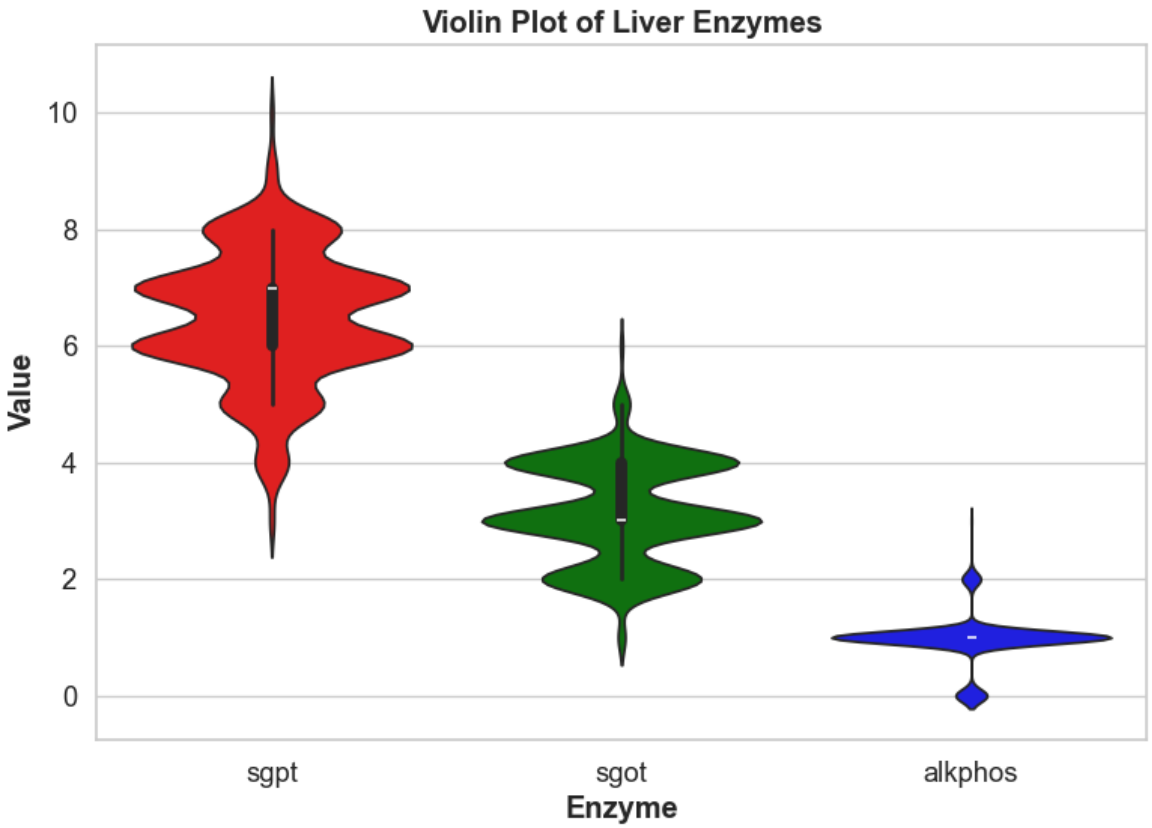

Figure 3: Violin plots illustrating the distribution of SGPT, SGOT, and alkaline phosphatase enzyme levels.

3.2 Data Preprocessing

Data preprocessing constitutes a critical stage in the development of reliable machine learning models, particularly in medical applications where clinical data may be affected by noise, variability, and inconsistencies. In this study, a systematic preprocessing pipeline is employed to enhance data quality, improve learning stability, and ensure that subsequent modeling stages operate on well-conditioned inputs.

Clinical datasets often contain missing or noisy values arising from measurement errors, laboratory variability, or incomplete patient records. Such imperfections can adversely affect the learning process if left unaddressed. Therefore, prior to model training, the dataset is examined to identify missing entries and anomalous values within the clinical features. Appropriate strategies are applied to handle these issues in a manner consistent with medical data integrity, ensuring that the preprocessing steps do not distort the underlying clinical meaning of the variables. This process contributes to reducing noise-induced bias and improving the robustness of downstream learning models.

Categorical variables require special treatment to be compatible with machine learning algorithms that operate on numerical representations. In the considered dataset, gender is the only categorical feature and is encoded using a numerical encoding scheme that preserves its binary nature while enabling effective model processing. This encoding approach avoids the introduction of artificial ordinal relationships and ensures that gender information is incorporated into the learning process without inflating feature dimensionality.

Feature scaling and normalization are applied to address differences in magnitude and measurement units across biochemical indicators. Liver function features, such as enzyme levels and protein concentrations, naturally span different numerical ranges, which can bias distance-based and

DOI: https://doi.org/10.54216/JAIM.110104

Received: September 26, 2025 Revised: November 29, 2025 Accepted: January 06, 2026

gradient-based learning algorithms if not properly scaled. To mitigate this issue, normalization techniques are employed to rescale numerical features to a comparable range. This step improves numerical stability, accelerates model convergence, and prevents features with large numeric values from dominating the learning process.

Outlier and distribution analysis is performed to examine the statistical characteristics of each feature and to identify extreme values that may unduly influence model training. Clinical outliers may arise from rare pathological conditions, measurement artifacts, or data entry errors. Rather than indiscriminately removing such values, the analysis is conducted carefully to distinguish between clinically meaningful extremes and spurious anomalies. This ensures that important diagnostic information is retained while minimizing the adverse effects of noise and irregularities.

Finally, correlation and redundancy analysis is conducted to investigate interdependencies among the biochemical and demographic features. Liver function indicators are often physiologically related, leading to strong correlations that can introduce redundancy into the feature space. Highly correlated variables may increase model complexity without providing additional predictive value. Identifying such relationships supports informed feature selection and optimization strategies by highlighting redundant attributes and enabling the construction of more compact and efficient predictive models. This analysis forms a foundational step for subsequent feature selection and optimization procedures applied in the study.

3.3 Machine Learning Models

The selection of appropriate machine learning models is a fundamental step in the development of reliable clinical decision support systems, particularly for disease prediction tasks involving heterogeneous biochemical and demographic data. In this study, multiple machine learning classifiers are employed to capture diverse learning behaviors and to explore different inductive biases when modeling liver-related clinical features. The chosen models span probabilistic, instance-based, tree-based, and ensemble learning paradigms, enabling a comprehensive methodological investigation of their suitability for liver disease classification.

Decision Tree models are adopted as one of the baseline classifiers due to their inherent interpretability and structured learning mechanism. Decision Trees operate by recursively partitioning the feature space into homogeneous regions based on decision rules derived from feature thresholds. This hierarchical structure allows the model to capture nonlinear relationships and feature interactions without requiring explicit assumptions about data distributions. In the context of clinical applications, Decision Trees are particularly valuable because their decision paths can be traced and interpreted, facilitating clinical insight into how demographic and biochemical features contribute to diagnostic outcomes. However, their susceptibility to overfitting, especially in high-dimensional and noisy datasets, necessitates careful consideration of model complexity and motivates comparative evaluation with more robust learning techniques.

Naïve Bayes classifiers are included as representative probabilistic models grounded in Bayes' theorem. These models estimate posterior class probabilities under the assumption of conditional independence among features given the class label. While this assumption is often violated in medical datasets—where biochemical indicators are physiologically correlated—Naïve Bayes classifiers have demonstrated practical effectiveness in various biomedical classification tasks due to their simplicity, robustness to small sample sizes, and low computational cost. Their probabilistic formulation provides an explicit measure of uncertainty, which is advantageous in clinical decision-making scenarios. Moreover, the inclusion of Naïve Bayes serves as a reference point for assessing the benefits of more complex nonlinear and ensemble-based models.

The K-Nearest Neighbors model is incorporated as a non-parametric, instance-based learning approach that relies on similarity measures in the feature space. Rather than learning an explicit parametric model during training, K-Nearest Neighbors classifies unseen instances based on the class labels of their nearest neighbors. This local learning behavior enables the model to adapt to complex and irregular decision boundaries that may arise in clinical data. However, its performance is highly

DOI: https://doi.org/10.54216/JAIM.110104

Received: September 26, 2025 Revised: November 29, 2025 Accepted: January 06, 2026

sensitive to feature scaling, noise, and the dimensionality of the input space, making it particularly dependent on effective preprocessing and feature selection. Including K-Nearest Neighbors in the comparative framework provides insight into how distance-based methods behave when applied to biochemical liver function indicators.

Gradient Boosting is employed as a powerful ensemble learning technique that constructs a predictive model through the sequential aggregation of weak learners, typically decision trees. Each new learner is trained to correct the errors made by the preceding ensemble, resulting in a model that progressively refines its predictions. This additive learning strategy enables Gradient Boosting to capture complex nonlinear patterns and higher-order feature interactions that may not be effectively modeled by single learners. Its flexibility and strong theoretical foundations have led to widespread adoption in medical data analysis, where relationships among features are often intricate and nonlinear. However, Gradient Boosting models are sensitive to hyperparameter settings, which can significantly influence learning dynamics and generalization behavior.

XGBoost is selected as an advanced ensemble learning framework that extends traditional gradient boosting through algorithmic and computational enhancements. These enhancements include regularization mechanisms to control model complexity, efficient tree construction strategies, and support for parallel computation. Such characteristics make XGBoost particularly well-suited for high-dimensional tabular data, where balancing model expressiveness and generalization is critical. In clinical datasets containing correlated biochemical indicators and demographic variables, XGBoost provides a flexible architecture capable of modeling complex interactions while incorporating mechanisms to mitigate overfitting.

Due to its modular design and compatibility with automated optimization strategies, XGBoost is chosen as the primary model for subsequent feature selection and hyperparameter optimization. Its structure allows seamless integration with metaheuristic optimization techniques, enabling systematic exploration of feature subsets and parameter configurations within a unified framework. This methodological choice supports a focused investigation of how intelligent optimization can enhance model efficiency, stability, and suitability for liver disease prediction, while maintaining alignment with the broader comparative evaluation of machine learning approaches employed in this study.

3.4 Metaheuristic Optimization

Metaheuristic optimization plays a central role in enhancing the effectiveness and robustness of machine learning models, particularly in complex clinical prediction tasks characterized by high dimensionality, nonlinear feature interactions, and heterogeneous data distributions. Unlike conventional optimization techniques that rely on gradient information or exhaustive search, metaheuristic algorithms employ population-based and stochastic search strategies inspired by natural, biological, or physical processes. These characteristics enable them to explore large and complex search spaces efficiently, making them well suited for optimizing machine learning models applied to medical datasets.

In the context of liver disease prediction, model performance is strongly influenced by both feature representation and internal configuration parameters. The interaction between biochemical indicators, demographic attributes, and learning model structure creates a highly nonconvex optimization landscape in which traditional tuning approaches may become trapped in suboptimal configurations. Metaheuristic optimization provides a flexible and adaptive mechanism for navigating this landscape by balancing global exploration and local exploitation. This balance is particularly important when optimizing advanced ensemble learning models, where multiple parameters jointly determine learning behavior, model capacity, and generalization ability.

The optimization framework adopted in this study is designed to systematically adjust model-related parameters in order to improve learning stability, reduce overfitting risk, and enhance the model's ability to capture clinically meaningful patterns. By embedding the learning model within a metaheuristic search process, the optimization procedure operates as a wrapper mechanism that evaluates candidate configurations based on predictive fitness while preserving the integrity of the clinical data.

DOI: https://doi.org/10.54216/JAIM.110104

Received: September 26, 2025 Revised: November 29, 2025 Accepted: January 06, 2026

3.5 Hyperparameter Tuning

Hyperparameter tuning constitutes a critical component of the proposed metaheuristic optimization framework. Hyperparameters govern the learning dynamics, structural complexity, and regularization behavior of machine learning models, and their selection has a direct impact on model generalization and robustness. In ensemble-based learning frameworks such as XGBoost, hyperparameters control how individual learners are constructed, how successive learners interact, and how model complexity is regulated during training.

The optimization process focuses on identifying suitable values for key hyperparameters that influence learning rate, model capacity, and structural configuration. Parameters controlling the learning rate determine how rapidly the model adapts to training data and how strongly new information influences the ensemble. Structural parameters regulate the complexity of the underlying learning components, shaping how feature interactions are represented and how expressive the model becomes. Additionally, parameters related to model depth and internal representation play an important role in balancing underfitting and overfitting, particularly when working with clinical datasets containing correlated biochemical features and limited sample sizes.

Manual selection of such hyperparameters is both time-consuming and prone to suboptimal choices, especially when interactions among parameters are nonlinear and data-dependent. Therefore, the hyperparameter tuning process is formulated as an optimization problem, where the objective is to identify parameter configurations that promote stable learning behavior and effective representation of clinical patterns. Metaheuristic algorithms are employed to automate this process by iteratively generating, evaluating, and refining candidate hyperparameter sets based on a defined fitness criterion.

By integrating hyperparameter tuning within the metaheuristic optimization framework, the study ensures that model configuration is not treated as an isolated preprocessing step, but rather as an adaptive and data-driven process. This approach supports the development of machine learning models that are better aligned with the underlying characteristics of liver disease data and more suitable for clinical decision support applications.

3.5.1 The iHow Optimization Algorithm (iHOW)

The iHow Optimization Algorithm (iHOW) is a recently developed human-inspired metaheuristic designed to solve complex optimization problems by emulating the hierarchical cognitive processes observed in human learning and decision-making. Unlike nature-inspired or physics-based optimizers, iHOW is grounded in an abstract representation of human intellectual behavior, progressing from raw data acquisition to expert-level reasoning. This hierarchical structure enables the algorithm to gradually refine candidate solutions through a structured transition from exploration to exploitation.

The conceptual architecture of iHOW is organized into five sequential cognitive components: Data Gathering (DS), Learning State (LS), Information Structuring (IS), Knowledge State (KS), and Expert Strategy (ES). Each component contributes a distinct influence to the evolution of candidate solutions, collectively guiding the search process toward high-quality regions of the solution space.

Let $X_i^t \in \mathbb{R}^d$ denote the position of the i-th search agent at iteration t. The general position update rule of iHOW is formulated as:

$$X_i^{t+1} = X_i^t + r_1 \cdot DS_{t+1} + r_2 \cdot LS_{t+1} + r_3 \cdot IS_{t+1} + r_4 \cdot KS_{t+1} + r_5 \cdot ES_{t+1}, \tag{1}$$

where $r_1, r_2, r_3, r_4, r_5 \in (0, 1)$ are adaptive stochastic coefficients controlling the relative contribution of each cognitive component.

DOI: https://doi.org/10.54216/JAIM.110104

Received: September 26, 2025 Revised: November 29, 2025 Accepted: January 06, 2026

Data Gathering and Exploration During the early stages of optimization, exploration is emphasized through the Data Gathering mechanism, which promotes population diversity and broad coverage of the search space. The exploration component is computed as:

$$DS_{t+1} = r_1 \cdot DS_1 + r_1 r_2 \cdot DS_2 + r_1 r_2 r_3 \cdot DS_3, \tag{2}$$

where DS_1, DS_2, and DS_3 represent exploratory displacement vectors derived from randomly selected agents and search directions.

Learning and Information Structuring The Learning State consolidates experience accumulated from previous iterations and transforms exploratory information into structured knowledge. It is defined as:

$$LS_{t+1} = r_1 \cdot LS_1 + r_1 r_2 \cdot LS_2 + r_1 r_2 r_3 \cdot LS_3, \tag{3}$$

where the LS components encode historical solution feedback. The Information Structuring state refines this learned information to improve orientation within the search landscape:

$$IS_{t+1} = \alpha \cdot LS_{t+1} + (1 - \alpha) \cdot DS_{t+1}, \tag{4}$$

with $\alpha \in (0, 1)$ controlling the balance between experience and exploration.

Knowledge Formation The Knowledge State represents accumulated high-level understanding derived from exploration and learning. It is updated as:

$$KS_{t+1} = DS_{t+1} + LS_{t+1} + K_t, \tag{5}$$

where K_t is a time-varying knowledge factor governing the exploration–exploitation transition:

$$K_t = 2 - 2\left(\frac{t}{T_{\max}}\right), \tag{6}$$

with $T_{\max}$ denoting the maximum number of iterations. This formulation ensures a gradual reduction of exploration pressure as the optimization progresses.

Expert Strategy and Exploitation In the exploitation phase, each agent generates multiple candidate moves based on expert-level reasoning:

$$X_{i,1}^{t+1} = X_i^t + r \cdot (KS_{t+1} + DS_{t+1}), \tag{7}$$

$$X_{i,2}^{t+1} = X_i^t + r_3 r_4 \cdot (KS_{t+1} + LS_{t+1}), \tag{8}$$

$$X_{i,3}^{t+1} = X_i^t + r_3 r_4 r_5 \cdot (KS_{t+1} + IS_{t+1}), \tag{9}$$

where r is a uniformly distributed random number in $(0, 1)$.

Selection of the Best Candidate Among the generated candidates, the best solution is selected according to:

$$X_i^{t+1} = \arg \min_{X \in \{X_{i,1}^{t+1}, X_{i,2}^{t+1}, X_{i,3}^{t+1}\}} f(X), \tag{10}$$

where $f(\cdot)$ denotes the objective function to be minimized. Algorithm 1 summarizes the complete procedural flow of the iHOW optimization process [30].

DOI: https://doi.org/10.54216/JAIM.110104

Received: September 26, 2025 Revised: November 29, 2025 Accepted: January 06, 2026

Algorithm 1 Proposed iHOW Optimization Algorithm

Input: population size N, control parameters $(r_1, r_2, r_3, r_4, r_5)$, maximum iterations T
Output: best solution x_{best}
Initialize population $X = \{x_1, x_2, \ldots, x_N\}$ randomly within bounds
Initialize states DS_i, LS_i, and KS_i for $i = 1, \ldots, N$
Evaluate fitness for all individuals and set x_{best}
Phase 1: Data Collection
for $i = 1$ to N **do**
    Collect data state DS_i
end for
Phase 2: Learning
for $i = 1$ to N **do**
    Update learning state LS_i using (r_1, r_2, r_3)
end for
Phase 3: Information Processing
for $i = 1$ to N **do**
    Process LS_i to generate structured information
end for
Phase 4: Knowledge Acquisition
for $i = 1$ to N **do**
    Fuse information and experience to update KS_i
end for
Phase 5: Optimization
for $t = 1$ to T **do**
    Update knowledge factor $K_t = 2 - 2 \times \left(\frac{t}{T}\right)$
    for $i = 1$ to N **do**
        Generate candidate solution using DS_i, LS_i, and KS_i
        Evaluate candidate fitness
        if candidate is better than x_i **then**
            $x_i \leftarrow$ candidate
        end if
        if x_i is better than x_{best} **then**
            $x_{\text{best}} \leftarrow x_i$
        end if
    end for
end for
return x_{best}

The hierarchical cognitive structure of iHOW enables a smooth and adaptive transition from global exploration to focused exploitation. By integrating human-inspired learning and knowledge accumulation mechanisms, iHOW provides a flexible and effective optimization framework suitable for complex model configuration tasks, such as machine learning hyperparameter optimization.

To benchmark the proposed iHOW optimizer under a consistent and methodologically sound experimental protocol, four well-established metaheuristic algorithms are employed as baseline optimizers, namely Genetic Algorithm (GA), Particle Swarm Optimizer (PSO), Grey Wolf Optimizer (GWO), and Greylag Goose Optimization (GGO). The selection of these benchmark algorithms is deliberately designed to ensure diversity in search strategies, population update mechanisms, and exploration–exploitation trade-offs. By incorporating optimizers drawn from distinct metaheuristic families, the benchmarking framework provides a comprehensive reference landscape against which the behavior of the proposed iHOW algorithm can be examined in a controlled and reproducible manner.

Genetic Algorithm is included as a canonical evolutionary optimization technique rooted in the principles of natural selection and biological evolution. GA operates on a population of candidate solutions that evolve over successive generations through selection, recombination, and mutation operators. The selection mechanism favors fitter individuals, while crossover and mutation introduce genetic diversity and stochastic variation into the population. This evolutionary process enables GA

DOI: https://doi.org/10.54216/JAIM.110104

Received: September 26, 2025 Revised: November 29, 2025 Accepted: January 06, 2026

to perform extensive global exploration of the search space and reduces the likelihood of premature convergence, particularly in complex optimization problems characterized by multimodal and nonlinear fitness landscapes. As one of the most widely studied and applied evolutionary algorithms, GA serves as a fundamental baseline for evaluating the effectiveness of alternative optimization strategies.

Particle Swarm Optimizer is adopted as a representative swarm-intelligence-based algorithm inspired by the collective behavior observed in social organisms. In PSO, candidate solutions are modeled as particles that move through the search space by updating their velocities and positions based on both individual experience and shared social knowledge. Each particle adjusts its trajectory by considering its own historical best position as well as the best position found by the swarm. This cooperative information-sharing mechanism allows PSO to achieve efficient convergence while maintaining a balance between exploration and exploitation. PSO is particularly relevant as a benchmark due to its simplicity, fast convergence characteristics, and widespread use in hyperparameter tuning and model optimization tasks.

Grey Wolf Optimizer is incorporated as a nature-inspired metaheuristic that simulates the leadership hierarchy and cooperative hunting behavior of grey wolves. GWO structures the population according to hierarchical roles, typically modeling dominant and subordinate agents that guide the search process. The algorithm alternates between exploration and exploitation by encircling prey, updating candidate positions relative to leading solutions, and gradually refining the search around promising regions of the solution space. This hierarchical control mechanism provides a structured search dynamic that differs fundamentally from both evolutionary and swarm-based approaches, making GWO a valuable benchmark for assessing the adaptability and convergence behavior of the proposed optimization framework.

Greylag Goose Optimization is included as a biologically inspired optimizer based on the migratory and social coordination behaviors of greylag geese. GGO models collective movement patterns, adaptive leadership, and coordinated navigation strategies to guide the population through the search space. By leveraging collective alignment and adaptive coordination among candidate solutions, GGO promotes effective exploration while enabling gradual refinement toward high-quality regions. Its inclusion adds further diversity to the benchmarking set and allows examination of optimization behavior influenced by migration-inspired dynamics.

All benchmark optimizers are referenced using the abbreviations and corresponding expansions provided in the optimizer abbreviation list adopted throughout this study, ensuring terminological consistency and reproducibility. Collectively, the inclusion of iHOW as the proposed optimization method alongside GA, PSO, GWO, and GGO as benchmark algorithms establishes a rigorous and balanced optimization framework for model configuration. This framework enables a controlled investigation of how differing population-based search mechanisms, information-sharing strategies, and exploration–exploitation dynamics influence the stability and suitability of optimized configurations within the same experimental setting.

3.6 Evaluation Metrics

The evaluation of predictive models in medical diagnosis requires more than a single aggregate performance indicator, as different types of classification errors may lead to significantly different clinical consequences. In the context of liver disease prediction, false negatives may delay treatment and increase the risk of disease progression, while false positives may result in unnecessary follow-up tests, patient anxiety, and increased healthcare costs. Therefore, a comprehensive set of classification performance metrics is employed to ensure that the proposed framework is assessed from multiple diagnostic perspectives.

In this study, model performance is evaluated using six widely accepted classification metrics, namely Accuracy, Sensitivity (True Positive Rate), Specificity (True Negative Rate), Positive Predictive Value, Negative Predictive Value, and F-Score. These metrics jointly assess overall correctness, disease detection capability, non-disease recognition, predictive reliability, and the balance between precision

DOI: https://doi.org/10.54216/JAIM.110104

Received: September 26, 2025 Revised: November 29, 2025 Accepted: January 06, 2026

and recall. Such a multi-metric evaluation strategy is particularly important when working with clinical datasets that may exhibit class imbalance or heterogeneous patient characteristics.

Let TP, TN, FP, and FN denote the number of true positives, true negatives, false positives, and false negatives, respectively. The mathematical definitions of the evaluation metrics used in this study are summarized in Table 2. These formulations are derived from the confusion matrix and provide a standardized basis for comparative analysis across different models and optimization strategies.

Table 2: Evaluation metrics used for model performance assessment

Metric	**Mathematical Definition**
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$
Sensitivity (TPR)	$\frac{TP}{TP + FN}$
Specificity (TNR)	$\frac{TN}{TN + FP}$
Positive Predictive Value (PPV)	$\frac{TP}{TP + FP}$
Negative Predictive Value (NPV)	$\frac{TN}{TN + FN}$
F-Score	$\frac{2 \times TP}{2 \times TP + FP + FN}$

Accuracy represents the proportion of correctly classified instances across the entire dataset and provides an overall indication of model effectiveness. While accuracy is useful for general assessment, it may be misleading in medical datasets where class distributions are imbalanced. Consequently, accuracy is interpreted in conjunction with other metrics rather than in isolation.

Sensitivity, also referred to as the true positive rate, quantifies the ability of the model to correctly identify patients with liver disease. High sensitivity is essential in screening and early diagnostic applications, as it minimizes the number of false negatives and reduces the risk of undetected disease cases. In contrast, specificity measures the model's ability to correctly identify non-liver individuals. High specificity is important for avoiding unnecessary diagnostic procedures and reducing the burden on healthcare systems.

Positive predictive value reflects the probability that a patient classified as having liver disease truly belongs to the diseased class, thereby indicating the reliability of positive predictions. Negative predictive value measures the likelihood that a patient predicted as non-liver is indeed healthy, which is critical for clinical confidence in ruling out disease. These predictive values are particularly relevant in real-world clinical settings, as they directly influence decision-making and patient management.

The F-Score provides a harmonic balance between sensitivity and precision, offering a single metric that captures the trade-off between false positives and false negatives. This measure is especially useful when evaluating models on datasets with uneven class distributions, as it emphasizes balanced diagnostic performance rather than dominance in a single metric. By jointly considering all six evaluation metrics, the proposed framework ensures a rigorous, clinically meaningful, and statistically robust assessment of liver disease prediction performance.

DOI: https://doi.org/10.54216/JAIM.110104

Received: September 26, 2025 Revised: November 29, 2025 Accepted: January 06, 2026

4 Experimental Results

4.1 Baseline Model Performance

This subsection provides an in-depth quantitative analysis of the baseline predictive behavior of the evaluated machine learning models prior to the application of any feature selection or metaheuristic-based optimization strategies. The primary purpose of this analysis is to establish a robust reference benchmark that reflects the inherent learning capability of each model when trained using the full set of preprocessed clinical features and standard, unoptimized configurations. Such a baseline assessment is essential for objectively evaluating the contribution of subsequent optimization procedures. The baseline performance results are summarized in Table 3 using a comprehensive set of clinically relevant evaluation metrics.

As reported in Table 3, the XGBoost model demonstrates the strongest baseline performance among all evaluated classifiers. It achieves an accuracy of 0.9218, indicating that more than ninety-two percent of the instances are correctly classified without any optimization. This strong overall performance is further supported by a sensitivity of 0.9202 and a specificity of 0.9235, reflecting a balanced capability in correctly identifying both liver disease patients and non-liver individuals. In addition, the model attains a positive predictive value of 0.9259 and a negative predictive value of 0.9177, highlighting reliable confidence in both positive and negative diagnostic predictions. The corresponding F-Score of 0.9230 confirms the model's balanced trade-off between precision and recall at the baseline stage.

Gradient Boosting ranks second in baseline performance, achieving an accuracy of 0.9120. Its sensitivity and specificity values of 0.9090 and 0.9150, respectively, indicate stable and consistent classification behavior, albeit with slightly reduced discriminative power compared to XGBoost. The identical values observed for specificity and positive predictive value (0.9150) suggest coherent decision boundaries across outcome classes. The F-Score of 0.9120 further reflects a well-balanced predictive profile, positioning Gradient Boosting as a competitive baseline learner.

The K-Nearest Neighbors model exhibits moderate baseline performance, with an accuracy of 0.9013. Its sensitivity of 0.8965 and specificity of 0.9060 reveal a modest imbalance between disease detection and non-disease recognition. While the model records a positive predictive value of 0.9027 and a negative predictive value of 0.9000, its F-Score of 0.8996 indicates comparatively weaker balance between precision and recall. These results suggest that although K-Nearest Neighbors can capture local structures in the data, its effectiveness is influenced by feature dimensionality and distance-based sensitivity.

Naive Bayes demonstrates a baseline accuracy of 0.8916, which is lower than that of ensemble-based and instance-based models. The sensitivity and specificity values of 0.8865 and 0.8965, respectively, indicate reduced effectiveness in distinguishing between liver and non-liver cases. The corresponding positive and negative predictive values of 0.8928 and 0.8904 further suggest narrower confidence margins in diagnostic decisions. The resulting F-Score of 0.8896 reflects the limitations associated with the conditional independence assumption when applied to correlated biochemical liver indicators.

Decision Tree exhibits the lowest baseline performance among the evaluated models, achieving an accuracy of 0.8812. Its sensitivity of 0.8759 and specificity of 0.8865 indicate reduced capability in both disease detection and non-disease recognition. Similarly, the positive predictive value of 0.8823 and negative predictive value of 0.8802 highlight comparatively weaker predictive reliability. The F-Score of 0.8791 confirms an imbalanced trade-off between precision and recall, which can be attributed to the model's susceptibility to overfitting and sensitivity to data variability under unoptimized conditions.

Overall, the baseline results presented in Table 3 reveal a clear performance hierarchy across machine learning models. Ensemble-based approaches consistently demonstrate superior baseline predictive capability, while probabilistic and tree-based models exhibit comparatively lower effectiveness. These observations provide strong quantitative motivation for the subsequent application of metaheuristic optimization techniques aimed at enhancing predictive accuracy, improving metric balance, and strengthening model robustness.

DOI: https://doi.org/10.54216/JAIM.110104

Received: September 26, 2025 Revised: November 29, 2025 Accepted: January 06, 2026

Table 3: Baseline performance comparison of unoptimized machine learning models

Model	Accuracy	Sensitivity (TPR)	Specificity (TNR)	PPV	NPV	F-Score
XGBoost	0.9218	0.9202	0.9235	0.9259	0.9177	0.9230
Gradient Boosting	0.9120	0.9090	0.9150	0.9150	0.9090	0.9120
K-Nearest Neighbors	0.9013	0.8965	0.9060	0.9027	0.9000	0.8996
Naive Bayes	0.8916	0.8865	0.8965	0.8928	0.8904	0.8896
Decision Tree	0.8812	0.8759	0.8865	0.8823	0.8802	0.8791

To provide a comprehensive and systematic comparison of the classification models employed in this study, multiple evaluation metrics are visually summarized in Fig. 4. These metrics include accuracy, sensitivity (true positive rate), specificity (true negative rate), positive predictive value, negative predictive value, and F-score, each of which captures a distinct aspect of predictive performance. The joint analysis of these measures enables a balanced assessment of overall correctness, class discrimination capability, and robustness against false predictions. Such a multidimensional evaluation is particularly important in medical decision-support applications, where both false positives and false negatives can have significant clinical implications. As illustrated in Fig. 4, the comparative visualization facilitates a clear understanding of how different learning algorithms perform across complementary diagnostic criteria.

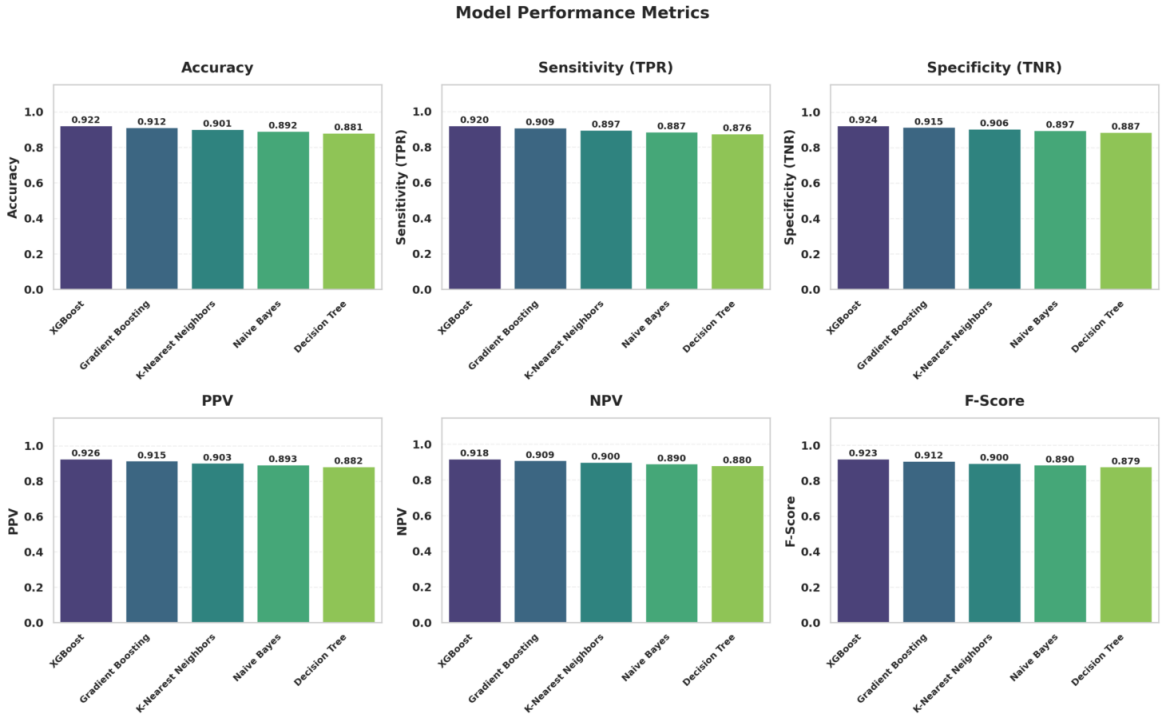

Figure 4: Comparative performance evaluation of different classification models using multiple metrics, including accuracy, sensitivity (TPR), specificity (TNR), positive predictive value (PPV), negative predictive value (NPV), and F-score.

To further analyze the relationship between balanced predictive performance and the ability to correctly identify positive instances, Fig. 5 presents a combined visualization of the F-score and sensitivity values across the evaluated classification models. In this dual representation, the bar plot illustrates the F-score, capturing the trade-off between precision and recall, while the overlaid line plot depicts sensitivity, emphasizing the models' effectiveness in detecting true positive cases. This integrated visualization facilitates a clearer comparative interpretation of how each classifier balances overall robustness with diagnostic sensitivity, which is particularly important in classification tasks involving medical or risk-sensitive decision-making.

DOI: https://doi.org/10.54216/JAIM.110104

Received: September 26, 2025 Revised: November 29, 2025 Accepted: January 06, 2026

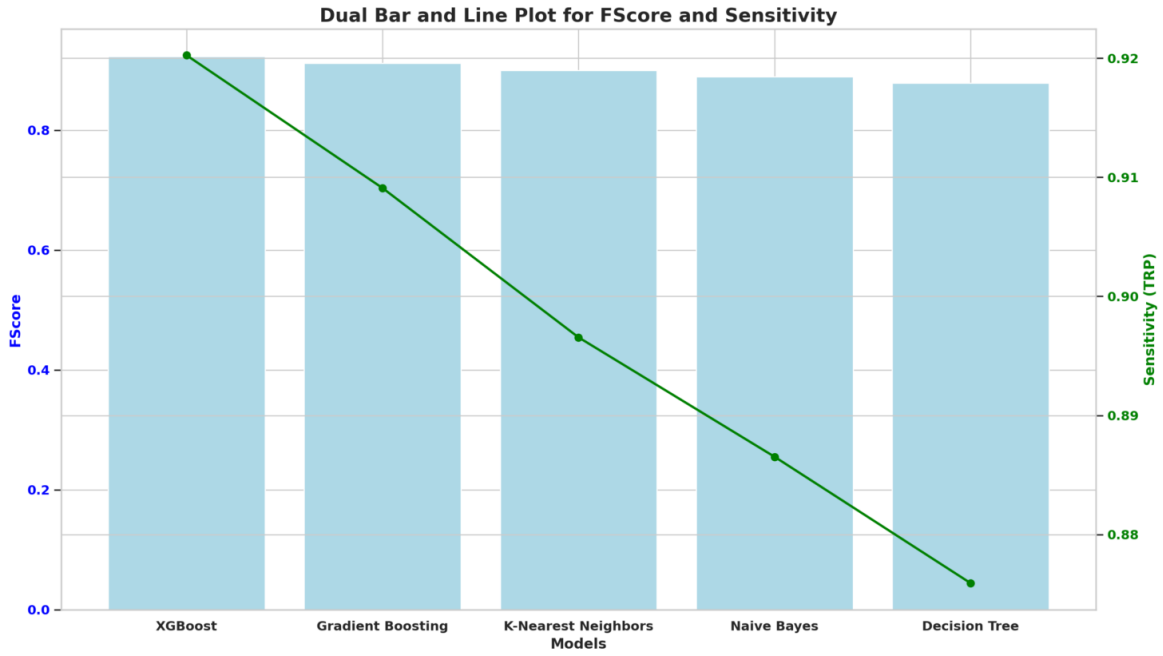

Figure 5: Dual bar and line plot illustrating the F-score and sensitivity (TPR) performance of the evaluated classification models.

To examine the distributional characteristics and central tendency of the evaluation metrics across different classification models, Fig. 6 presents a series of box plots for accuracy, sensitivity (TPR), specificity (TNR), positive predictive value, negative predictive value, and F-score. In addition to illustrating the spread and relative variability of each metric, the figure explicitly highlights both the mean and median values, enabling a direct comparison between average performance and robust central estimates. This visualization provides deeper insight into performance stability and consistency across models, thereby complementing the aggregate metric comparisons and supporting a more statistically grounded interpretation of classifier behavior.

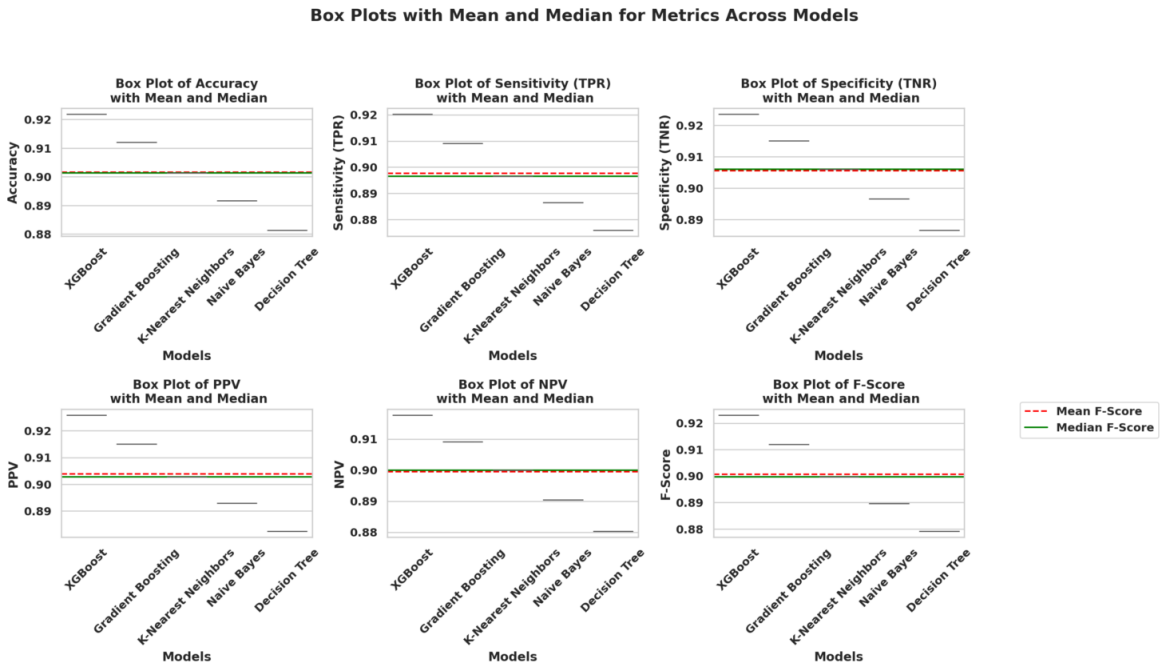

Figure 6: Box plots illustrating the distribution of performance metrics across classification models, with mean and median values indicated for each metric.

DOI: https://doi.org/10.54216/JAIM.110104

Received: September 26, 2025 Revised: November 29, 2025 Accepted: January 06, 2026

4.2 Hyperparameter Optimization Results

This subsection presents the experimental results obtained after applying metaheuristic-based hyperparameter optimization to the XGBoost model. The objective of this analysis is to evaluate the effectiveness of the proposed iHOW optimization strategy and to compare its performance against several well-established benchmark optimizers under an identical experimental protocol. The comparative results are summarized in Table 4, using the same evaluation metrics employed in the baseline analysis to ensure consistency and fairness.

Table 4: Performance comparison of XGBoost optimized using different metaheuristic algorithms

Model	Accuracy	Sensitivity (TPR)	Specificity (TNR)	PPV	NPV	F-Score
iHOW + XGBoost	0.9836	0.9833	0.9840	0.9845	0.9828	0.9839
GA + XGBoost	0.9727	0.9712	0.9743	0.9743	0.9712	0.9727
PSO + XGBoost	0.9686	0.9671	0.9700	0.9686	0.9685	0.9679
GWO + XGBoost	0.9648	0.9623	0.9671	0.9640	0.9656	0.9631
GGO + XGBoost	0.9598	0.9569	0.9623	0.9583	0.9611	0.9576

As shown in Table 4, the application of hyperparameter optimization leads to substantial performance enhancement across all evaluated metrics when compared to the baseline XGBoost configuration. Among the optimized models, iHOW+XGBoost achieves the highest overall performance, recording an accuracy of 0.9836. This result indicates that approximately 98.37% of instances are correctly classified after optimization. The sensitivity value of 0.9833 demonstrates a strong capability in correctly identifying liver disease cases, while the specificity of 0.9840 confirms reliable discrimination of non-liver individuals. Furthermore, the positive predictive value of 0.9845 and negative predictive value of 0.9828 reflect high confidence in both positive and negative diagnostic predictions. The resulting F-Score of 0.9839 highlights the balanced and robust classification behavior achieved through iHOW-based optimization.

The GA-optimized XGBoost model attains an accuracy of 0.9727, accompanied by a sensitivity of 0.9712 and a specificity of 0.9743. These values indicate notable improvements over the unoptimized baseline, confirming the effectiveness of evolutionary search mechanisms in tuning model hyperparameters. The symmetry between sensitivity, specificity, and predictive value metrics contributes to an F-Score of 0.9727, reflecting stable classification performance, albeit at a lower level than that achieved by iHOW.

The PSO-optimized XGBoost model yields an accuracy of 0.9686, with sensitivity and specificity values of 0.9671 and 0.9700, respectively. The corresponding PPV and NPV values of 0.9686 and 0.9685 suggest consistent predictive confidence across classes. The F-Score of 0.9679 confirms the effectiveness of swarm-based optimization; however, the performance remains inferior to that obtained using iHOW and GA.

Optimization using GWO results in an accuracy of 0.9648, with sensitivity and specificity values of 0.9623 and 0.9671, respectively. While the model exhibits improved discriminative capability compared to the baseline configuration, its predictive metrics, including an F-Score of 0.9631, indicate comparatively reduced effectiveness relative to iHOW-, GA-, and PSO-based optimization.

The GGO-optimized XGBoost model achieves an accuracy of 0.9598, which, although representing a clear improvement over the unoptimized model, is the lowest among the evaluated optimizers. Its sensitivity of 0.9569 and specificity of 0.9623 suggest modest gains in both disease detection and non-disease recognition. The corresponding F-Score of 0.95767 reflects the cumulative impact of these improvements, positioning GGO as a competitive but less effective optimization strategy in this context.

Overall, the results in Table 4 demonstrate that metaheuristic-based hyperparameter optimization significantly enhances the predictive performance of XGBoost. Among the evaluated strategies, iHOW consistently outperforms benchmark optimizers across all evaluation metrics, indicating its superior ability to identify high-quality hyperparameter configurations. These findings highlight the

DOI: https://doi.org/10.54216/JAIM.110104
Received: September 26, 2025 Revised: November 29, 2025 Accepted: January 06, 2026

effectiveness of human-inspired optimization mechanisms in improving model robustness and diagnostic reliability for liver disease prediction.

As reported in Table 3, XGBoost demonstrates the strongest baseline performance among all evaluated models, achieving an accuracy of 0.9218. This indicates that approximately 92.19% of patient instances are correctly classified in the absence of optimization. The model exhibits a sensitivity of 0.9202, reflecting a high ability to correctly identify liver disease cases, alongside a specificity of 0.9235, which confirms reliable discrimination of non-liver patients. In addition, XGBoost records a positive predictive value of 0.9259 and a negative predictive value of 0.9177, highlighting consistent confidence in both positive and negative predictions. The corresponding F-Score of 0.9230 further confirms a well-balanced trade-off between precision and recall under baseline conditions.

Gradient Boosting ranks second in baseline performance, attaining an accuracy of 0.91205. Its sensitivity and specificity values of 0.9090 and 0.915032680, respectively, indicate stable classification behavior with a slight reduction in discriminative power compared to XGBoost. The model maintains identical values for specificity and positive predictive value (0.9150), suggesting consistent decision boundaries across outcome classes. The achieved F-Score of 0.91205 reinforces the model's balanced performance profile, positioning Gradient Boosting as a competitive yet slightly less effective baseline learner.

The K-Nearest Neighbors model achieves a baseline accuracy of 0.9013. Its sensitivity of 0.8965 and specificity of 0.9060 reveal a moderate imbalance between disease detection and non-disease recognition. While the model records a positive predictive value of 0.9027 and a negative predictive value of 0.9000, its F-Score of 0.8996 indicates a comparatively weaker balance between precision and sensitivity. These numerical results suggest that although K-Nearest Neighbors can effectively capture local patterns in the data, its baseline performance is influenced by feature scaling and the high dimensionality of clinical attributes.

Naive Bayes exhibits a baseline accuracy of 0.8916, which is lower than that of ensemble-based and instance-based models. The sensitivity and specificity values of 0.8865 and 0.8965, respectively, indicate reduced effectiveness in correctly identifying both liver and non-liver cases. Correspondingly, the positive predictive value of 0.8928 and negative predictive value of 0.8904 reflect narrower confidence margins. The resulting F-Score of 0.8896 highlights the limitations imposed by the conditional independence assumption when applied to correlated biochemical liver indicators.

Decision Tree yields the lowest baseline performance among the evaluated models, with an accuracy of 0.8812. Its sensitivity of 0.875912409 and specificity of 0.8865 demonstrate reduced capability in detecting liver disease cases and distinguishing healthy individuals. This trend is further reflected in the positive predictive value of 0.8823 and negative predictive value of 0.8802. The F-Score of 0.8791 indicates a weaker balance between precision and recall, which can be attributed to the model's sensitivity to data variability and tendency toward overfitting under unoptimized conditions.

Overall, the baseline results in Table 3 reveal a clear performance hierarchy across machine learning models. Ensemble-based approaches consistently outperform probabilistic and tree-based methods across all evaluation metrics, while simpler classifiers exhibit comparatively lower baseline effectiveness. These numerically supported observations provide a strong empirical foundation for the subsequent application of metaheuristic optimization techniques, which aim to enhance predictive accuracy, improve metric balance, and strengthen model robustness in the following experimental stages.

To further investigate the performance trends of the hybrid optimization–based classification models, Fig. 7 illustrates time-series style plots for multiple evaluation metrics, including accuracy, sensitivity (TPR), specificity (TNR), positive predictive value, negative predictive value, and F-score. Although the horizontal axis represents different hybrid model configurations rather than temporal progression, the sequential visualization enables a clear assessment of relative performance variation and degradation across models. This representation facilitates an intuitive comparison of how each hybrid approach influences predictive stability and overall diagnostic effectiveness across complementary evaluation criteria.

DOI: https://doi.org/10.54216/JAIM.110104

Received: September 26, 2025 Revised: November 29, 2025 Accepted: January 06, 2026

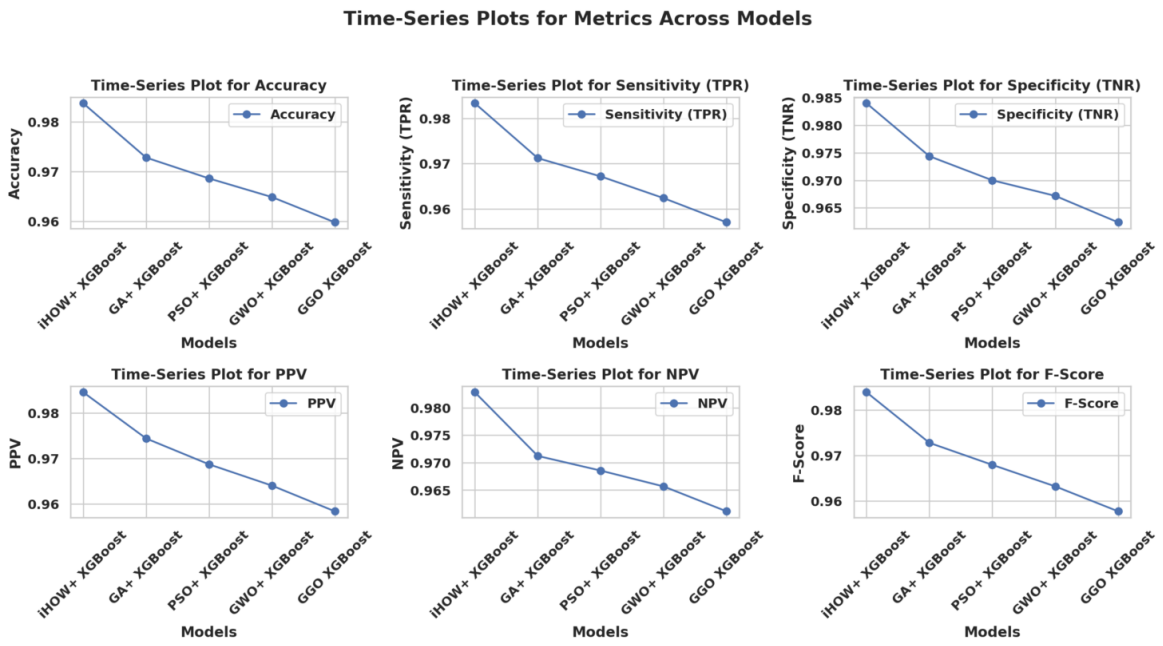

Figure 7: Time-series style plots illustrating the performance trends of hybrid optimization-based classification models across multiple evaluation metrics.

To summarize the central tendency and variability of the overall model performance, Fig. 8 presents a heatmap-based visualization of the mean, median, and standard deviation values across all evaluation metrics, including accuracy, sensitivity (TPR), specificity (TNR), positive predictive value, negative predictive value, and F-score. By jointly reporting measures of central tendency and dispersion, this figure provides a concise statistical overview of performance consistency. The close agreement between the mean and median values, together with the low standard deviation observed across metrics, indicates stable and reliable predictive behavior, thereby reinforcing the robustness of the evaluated classification framework.

DOI: https://doi.org/10.54216/JAIM.110104

Received: September 26, 2025 Revised: November 29, 2025 Accepted: January 06, 2026

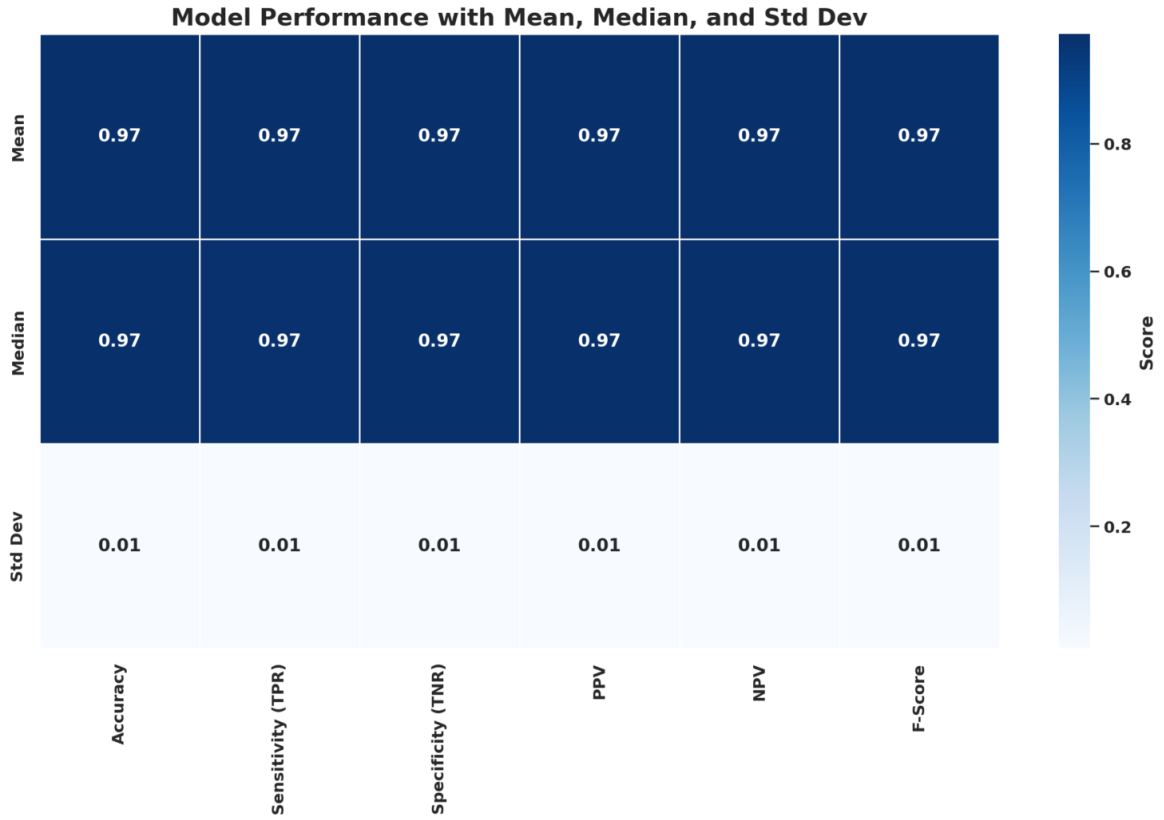

Figure 8: Heatmap illustrating the mean, median, and standard deviation of performance metrics across the evaluated classification models.

To provide a deeper distributional analysis of the classification performance metrics, Fig. 9 presents empirical cumulative distribution function (ECDF) plots for accuracy, sensitivity (TPR), specificity (TNR), positive predictive value, negative predictive value, and F-score across the evaluated models. In addition to the ECDF curves, the figure highlights the corresponding mean, median, and standard deviation boundaries, enabling a joint assessment of central tendency and dispersion. This visualization facilitates an intuitive understanding of performance concentration, variability, and robustness, thereby complementing aggregate and box-plot-based analyses with a more comprehensive statistical perspective.

DOI: https://doi.org/10.54216/JAIM.110104

Received: September 26, 2025 Revised: November 29, 2025 Accepted: January 06, 2026

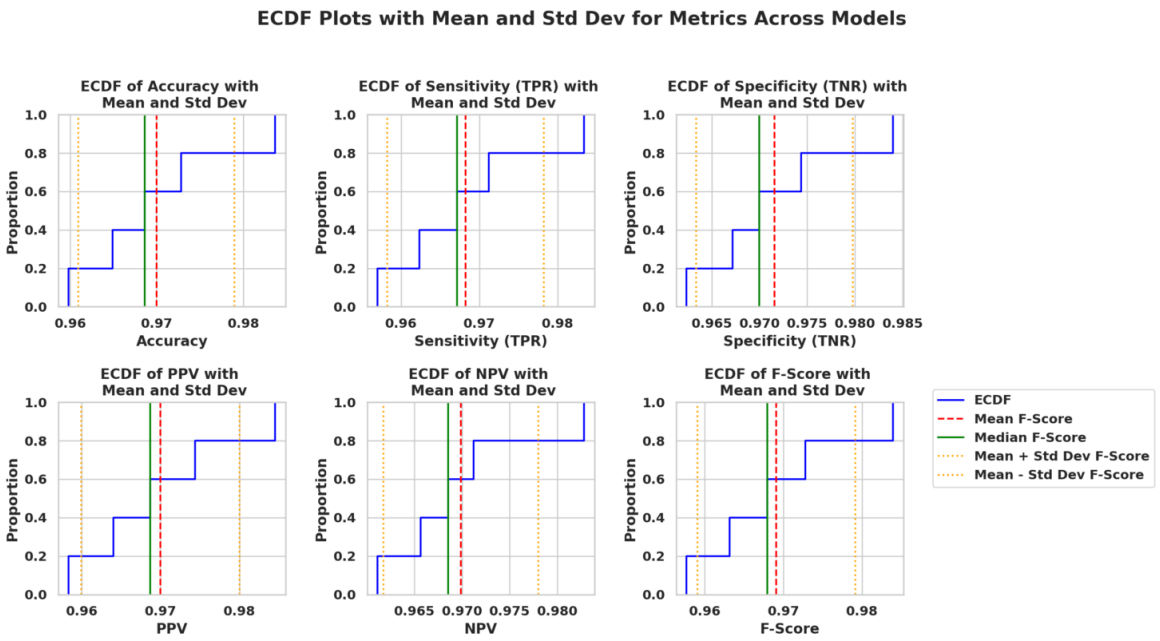

Figure 9: ECDF plots illustrating the distribution of performance metrics across classification models, with mean, median, and standard deviation indicators.

5 Discussion

This section provides an in-depth interpretation of the experimental findings, situates the results within the context of existing liver disease prediction studies, and discusses the clinical relevance and limitations of the proposed framework. The discussion aims to clarify why the optimized learning strategy achieves superior performance and to outline the practical considerations associated with its deployment.

The experimental results demonstrate that the application of metaheuristic-driven hyperparameter optimization leads to substantial and consistent improvements in liver disease prediction performance. In particular, the optimized XGBoost model exhibits marked gains across all evaluated metrics when compared to its baseline configuration. These improvements indicate that model performance in clinical classification tasks is not solely dependent on the choice of learning algorithm, but is strongly influenced by the quality of hyperparameter configuration.

The superior performance of the optimized XGBoost model can be attributed to several interacting factors. First, XGBoost inherently benefits from its ensemble-based architecture, which enables effective modeling of nonlinear relationships and complex feature interactions commonly observed in biochemical liver indicators. However, without optimization, the model's learning capacity may remain underutilized due to suboptimal parameter settings. Metaheuristic optimization enables systematic exploration of the hyperparameter space, allowing the model to better balance learning rate, structural complexity, and regularization effects. This balance enhances generalization capability and reduces susceptibility to overfitting, particularly in datasets characterized by correlated features and limited sample sizes.

Furthermore, the human-inspired search dynamics of the proposed optimization strategy facilitate an adaptive transition from global exploration to focused exploitation. This behavior increases the likelihood of identifying high-quality parameter configurations that conventional tuning strategies may overlook. The observed improvements in sensitivity and specificity indicate that the optimized model achieves a more balanced diagnostic behavior, improving both disease detection and non-disease recognition. Collectively, these findings confirm that intelligent hyperparameter optimization is a critical component in maximizing the diagnostic potential of advanced ensemble learning models.

DOI: https://doi.org/10.54216/JAIM.110104

Received: September 26, 2025 Revised: November 29, 2025 Accepted: January 06, 2026

When compared with existing studies on liver disease prediction, the findings of this work are consistent with the growing consensus that ensemble learning methods outperform traditional classifiers in medical diagnosis tasks. Previous research has reported that decision trees, Naive Bayes, and instance-based learners often exhibit limited discriminative capability when applied to complex clinical datasets due to simplifying assumptions or sensitivity to data distribution. The baseline results observed in this study align with these observations, as simpler models demonstrate lower predictive effectiveness relative to ensemble-based approaches.

More recent studies have highlighted the benefits of combining machine learning models with optimization techniques, particularly metaheuristic algorithms, to enhance predictive accuracy. The results reported here reinforce these conclusions by demonstrating that hyperparameter optimization yields substantial performance gains even for already strong baseline models. Moreover, the use of a human-inspired optimization strategy introduces a novel perspective into liver disease prediction research, complementing existing bio-inspired and swarm-based optimization approaches. The consistent outperformance of benchmark optimizers suggests that incorporating adaptive learning and knowledge accumulation mechanisms can further enhance model configuration quality.

From a clinical perspective, the optimized predictive framework offers several important advantages. The high sensitivity achieved by the optimized model supports reliable identification of liver disease cases, which is critical for early screening and timely intervention. At the same time, strong specificity reduces the likelihood of false positives, minimizing unnecessary follow-up testing and patient anxiety. This balanced diagnostic behavior is essential for practical deployment in real-world healthcare settings.

The reliance on routinely collected laboratory measurements enhances the feasibility of clinical adoption, as no additional invasive procedures or specialized tests are required. The automated and data-driven nature of the framework can support clinicians by providing consistent and objective diagnostic recommendations, thereby reducing inter-observer variability and supporting evidence-based decision-making. Additionally, the robustness demonstrated across multiple evaluation metrics suggests that the optimized model can maintain stable performance under varying clinical conditions.

Despite the promising results, several limitations should be acknowledged. First, the dataset used in this study is of moderate size and originates from a single geographical region, which may limit the generalizability of the findings. Although the results indicate strong predictive performance, validation on larger and more diverse patient populations is necessary to confirm robustness across different demographic and clinical settings.

Second, the dataset exhibits demographic imbalance, particularly with respect to gender distribution, which may introduce bias into the learned model. While this imbalance reflects real-world prevalence patterns to some extent, future studies should explore strategies for mitigating demographic bias to ensure equitable diagnostic performance.

Finally, the current study focuses on binary liver disease classification and does not differentiate between disease subtypes or severity levels. Although this formulation is suitable for screening applications, extending the framework to multi-class classification would enhance its clinical utility. Addressing these limitations in future research will be essential for advancing the proposed approach toward broader clinical adoption.

6 Conclusion and Future Work

This study presented a comprehensive machine learning and metaheuristic optimization framework for liver disease prediction based on routine clinical and biochemical data. A systematic experimental analysis was conducted to evaluate the baseline performance of multiple machine learning models and to assess the impact of hyperparameter optimization on predictive accuracy and robustness. The baseline evaluation demonstrated that ensemble-based learning models exhibit strong inherent predictive capability; however, their performance remains constrained by suboptimal hyperparameter configurations when deployed without optimization.

DOI: https://doi.org/10.54216/JAIM.110104

Received: September 26, 2025 Revised: November 29, 2025 Accepted: January 06, 2026

To address this limitation, a human-inspired metaheuristic optimization strategy was introduced through the integration of the iHow Optimization Algorithm with the XGBoost learning model. The optimization process resulted in consistent and substantial improvements across all evaluated diagnostic metrics, including accuracy, sensitivity, specificity, predictive values, and F-Score. Comparative analysis against established metaheuristic optimizers confirmed that the proposed optimization framework provides superior model configuration capability, effectively enhancing learning stability and diagnostic reliability. These findings collectively validate the effectiveness of metaheuristic-driven hyperparameter tuning in improving machine learning-based medical diagnosis systems.

The outcomes of this research have important practical and clinical implications for the development of intelligent decision support systems in healthcare. The proposed optimized framework demonstrates strong potential for integration into early liver disease screening pipelines, where rapid and accurate interpretation of laboratory test results is critical. By leveraging routinely collected clinical indicators, the framework can assist clinicians in identifying high-risk patients at early stages, thereby facilitating timely intervention and reducing the likelihood of disease progression.

Furthermore, the robustness and scalability of the optimized model make it suitable for deployment in resource-constrained clinical environments, where access to specialized diagnostic expertise may be limited. The automated and data-driven nature of the proposed approach can support consistent decision-making, reduce diagnostic variability, and enhance overall clinical efficiency. As such, the framework represents a viable foundation for computer-aided diagnostic tools aimed at improving patient outcomes and optimizing healthcare resource utilization.

Although the proposed framework demonstrates strong performance for binary liver disease classification, several avenues for future research remain open. One important direction involves extending the current binary classification formulation to a multi-class setting, enabling differentiation among multiple liver disease subtypes and severity levels. Such an extension would enhance the clinical utility of the system by providing more granular diagnostic insights.

Another promising direction is the integration of the optimized framework with real-time hospital decision support systems. This would involve embedding the model within clinical information systems to support continuous monitoring, real-time risk assessment, and dynamic decision-making based on incoming patient data. Addressing challenges related to system interoperability, data latency, and real-time model updating will be critical in this context.

Finally, future work should focus on validating the proposed approach using larger and multi-center clinical datasets. Expanding the evaluation to diverse patient populations and healthcare settings will improve generalizability, reduce potential demographic bias, and strengthen the clinical credibility of the framework. Such large-scale validation studies are essential steps toward the safe and effective translation of intelligent diagnostic models into real-world clinical practice.

Data Availability

The dataset used in this study is publicly available on Kaggle at https://www.kaggle.com/datasets/jeevannagaraj/indian-liver-patient-dataset.

Declarations

- **Acknowledgments**
  Not applicable.

- **Conflict of interest/Competing interests**
  The authors declare that they have no conflicts of interest to report regarding the present study.

- **Ethics approval and consent to participate**
  Not applicable.

- **Consent for publication**
  Not applicable.
- **Funding**
  No Fund

References

[1] C. Gan et al., "Liver diseases: Epidemiology, causes, trends and predictions," *Signal Transduction and Targeted Therapy*, vol. 10, no. 1, p. 33, 2025. DOI: https://doi.org/10.1038/s41392-024-02072-z.

[2] Z. M. Younossi, G. Wong, Q. M. Anstee, and L. Henry, "The global burden of liver disease," *Clinical Gastroenterology and Hepatology*, vol. 21, no. 8, pp. 1978–1991, 2023, Updates in Clinical Hepatology, ISSN: 1542-3565. DOI: https://doi.org/10.1016/j.cgh.2023.04.015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1542356523003154.

[3] V. K. Pathak and C. Lahariya, "Preventing liver diseases: A framework for service delivery across levels of the health system," in *Preventive Hepatology*, CRC Press, 2025, pp. 159–170.

[4] Z. M. Younossi et al., "Addressing the high and rising global burden of metabolic dysfunction-associated steatotic liver disease (masld) and metabolic dysfunction-associated steatohepatitis (mash): From the growing prevalence to payors' perspective," *Alimentary pharmacology & therapeutics*, vol. 61, no. 9, pp. 1467–1478, 2025. DOI: https://doi.org/10.1111/apt.70020.

[5] X. Luo, Y. He, Z. Jiang, and J. Liao, "Global burden of liver cirrhosis: Trends from 1990–2021 and projections to 2060," *Journal of Health, Population and Nutrition*, vol. 44, no. 1, pp. 1–20, 2025. DOI: https://doi.org/10.1186/s41043-025-01109-5.

[6] J. L. Allen and R. K. Sterling, "Palliative care & management of symptoms in advanced liver disease: An expert review," *Expert Review of Gastroenterology & Hepatology*, vol. 19, no. 5, pp. 515–526, 2025. DOI: https://doi.org/10.1080/17474124.2025.2491529.

[7] T.-Y. Wang et al., "Association of metabolic dysfunction-associated fatty liver disease with kidney disease," *Nature Reviews Nephrology*, vol. 18, no. 4, pp. 259–268, 2022. DOI: https://doi.org/10.1038/s41581-021-00519-y.

[8] P. Kumar and P. Dwivedi, "Chapter 6-1 - diagnosis of liver disorder," in *Computational Intelligence for Genomics Data*, ser. Advances in Biomedical Informatics, B. Pandey, S. L. Tripathi, V. Emilia Balas, D. K. Pandey, and M. Mahmud, Eds., Academic Press, 2025, pp. 199–224, ISBN: 978-0-443-30080-6. DOI: https://doi.org/10.1016/B978-0-443-30080-6.00001-8. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780443300806000018.

[9] S. Malik, R. Das, T. Thongtan, K. Thompson, and N. Dbouk, "Ai in hepatology: Revolutionizing the diagnosis and management of liver disease," *Journal of Clinical Medicine*, vol. 13, no. 24, 2024, ISSN: 2077-0383. DOI: 10.3390/jcm13247833. [Online]. Available: https://www.mdpi.com/2077-0383/13/24/7833.

[10] R. Pryke and I. N. Guha, "Time to focus on chronic liver diseases in the community: A review of primary care hepatology tools, pathways of care and reimbursement mechanisms," *Journal of Hepatology*, vol. 78, no. 3, pp. 663–671, 2023, ISSN: 0168-8278. DOI: https://doi.org/10.1016/j.jhep.2022.10.010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168827822031464.

[11] S. S. Kankal and S. Kasar, "A systematic review on machine learning techniques for liver disease detection," *Biomedical Materials & Devices*, pp. 1–26, 2025. DOI: https://doi.org/10.1007/s44174-025-00341-1.

[12] M. Bhat, M. Rabindranath, B. S. Chara, and D. A. Simonetto, "Artificial intelligence, machine learning, and deep learning in liver transplantation," *Journal of Hepatology*, vol. 78, no. 6, pp. 1216–1233, 2023, Liver Transplantation, ISSN: 0168-8278. DOI: https://doi.org/10.1016/j.jhep.2023.01.006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016882782300017X.

[13] S. Grover and S. Gupta, "Automated diagnosis and classification of liver cancers using deep learning techniques: A systematic review," *Discover Applied Sciences*, vol. 6, no. 10, p. 508, 2024.

[14] S. Rani et al., "Machine learning-powered smart healthcare systems in the era of big data: Applications, diagnostic insights, challenges, and ethical implications," *Diagnostics*, vol. 15, no. 15, 2025, ISSN: 2075-4418. DOI: 10.3390/diagnostics15151914. [Online]. Available: https://www.mdpi.com/2075-4418/15/15/1914.

[15] W. M. Shaban, "Early diagnosis of liver disease using improved binary butterfly optimization and machine learning algorithms," *Multimedia Tools and Applications*, vol. 83, no. 10, pp. 30 867–30 895, 2024. DOI: https://doi.org/10.1007/s11042-023-16686-y.

[16] J. Allenki and H. Kumar Soni, "Analysis of chronic liver disease detection by using machine learning techniques," in *2024 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*, 2024, pp. 1–8. DOI: 10.1109/SCEECS61402.2024.10481680.

[17] N. Shibata et al., "A machine learning algorithm for stratification of risk of cardiovascular disease in metabolic dysfunction-associated steatotic liver disease," *European Journal of Internal Medicine*, vol. 129, pp. 62–70, 2024, ISSN: 0953-6205. DOI: https://doi.org/10.1016/j.ejim.2024.07.005. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0953620524002887.

[18] R. Islam, A. Sultana, and M. N. Tuhin, "A comparative analysis of machine learning algorithms with tree-structured parzen estimator for liver disease prediction," *Healthcare Analytics*, vol. 6, p. 100 358, 2024, ISSN: 2772-4425. DOI: https://doi.org/10.1016/j.health.2024.100358. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2772442524000601.

[19] C. J. Ejiyi et al., "Polynomial-shap analysis of liver disease markers for capturing of complex feature interactions in machine learning models," *Computers in Biology and Medicine*, vol. 182, p. 109 168, 2024, ISSN: 0010-4825. DOI: https://doi.org/10.1016/j.compbiomed.2024.109168. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482524012538.

[20] S. Schophaus et al., "Machine learning uncovers manganese as a key nutrient associated with reduced risk of steatotic liver disease," *Liver International*, vol. 44, no. 10, pp. 2807–2821, 2024. DOI: https://doi.org/10.1111/liv.16055.

[21] S. Sarkar et al., "A machine learning model to predict risk for hepatocellular carcinoma in patients with metabolic dysfunction-associated steatotic liver disease," *Gastro Hep Advances*, vol. 3, no. 4, pp. 498–505, 2024, ISSN: 2772-5723. DOI: https://doi.org/10.1016/j.gastha.2024.01.007. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2772572324000074.

[22] Z. Li et al., "A machine learning based algorithm accurately stages liver disease by quantification of arteries," *Scientific Reports*, vol. 15, no. 1, p. 3143, 2025. DOI: https://doi.org/10.1038/s41598-025-87427-4.

[23] W. Zhang, W. Lu, Y. Jiao, T. Li, H. Wang, and C. Wan, "Identifying disease progression biomarkers in metabolic associated steatotic liver disease (masld) through weighted gene co-expression network analysis and machine learning," *Journal of Translational Medicine*, vol. 23, no. 1, p. 472, 2025. DOI: https://doi.org/10.1186/s12967-025-06490-7.

[24] R. Chen et al., "Trans-ancestral rare variant association study with machine learning-based phenotyping for metabolic dysfunction-associated steatotic liver disease," *Genome Biology*, vol. 26, no. 1, p. 50, 2025. DOI: https://doi.org/10.1186/s13059-025-03518-5.

[25] C. Lu, M. Han, Q. Ma, L. Ying, and Y. Zhang, "Identification of biomarkers associated with coronary artery disease and non-alcoholic fatty liver disease by bioinformatics analysis and machine learning," *Scientific Reports*, vol. 15, no. 1, p. 3557, 2025. DOI: https://doi.org/10.1038/s41598-025-87923-7.

[26] R. Rani et al., "Enhancing liver disease diagnosis with hybrid smote-enn balanced machine learning models—an empirical analysis of indian patient liver disease datasets," *Frontiers in Medicine*, vol. 12, p. 1 502 749, 2025. DOI: https://doi.org/10.3389/fmed.2025.1502749.

Received: September 26, 2025 Revised: November 29, 2025 Accepted: January 06, 2026

[27] P. Pavani, P Sravanthi, C. Suneetha, M. Azhar, L. H. supriya Reddy, and T. Maddineni, "Advanced machine learning approaches for liver disease prediction: A focus on feature selection and hybrid ensemble model," in *2025 6th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI)*, 2025, pp. 1003–1010. DOI: 10.1109/ICMCSI64620.2025.10883633.

[28] N. A. Khan, M. F. Bin Hafiz, M. A. Pramanik, S. Hossain, S. Barman, and N. Hossain, "Machine learning and explainable ai for liver disease prediction: An integrated interpretability framework," *Biomedical Materials & Devices*, pp. 1–14, 2025.

[29] A. A. Syed, A. Banu Shaik, S. Konatam, M. V. Phani Kumar, and N. Gupta, "Cnn-based analytical approach for liver disease classification and prediction," in *2025 3rd International Conference on Advancement in Computation & Computer Technologies (InCACCT)*, 2025, pp. 682–685. DOI: 10.1109/InCACCT65424.2025.11011283.

[30] E.-S. M. El-Kenawy et al., "Ihow optimization algorithm: A human-inspired metaheuristic approach for complex problem solving and feature selection," *Journal of Artificial Intelligence in Engineering Practice*, vol. 1, no. 2, pp. 36–53, 2024.

DOI: https://doi.org/10.54216/JAIM.110104

Received: September 26, 2025 Revised: November 29, 2025 Accepted: January 06, 2026