Metaheuristic Optimization Review

♦ ISSN: 2692-4048 Vol: 04, No. 02, 2025 ♦ PP. 19–23

Article DOI: https://doi.org/10.54216/MOR.040204



REVIEW ARTICLE

Machine Learning-Based Liver Cirrhosis Prediction: A Review of Feature Selection, Optimization Algorithms, Clinical Applications, and Future Challenges

P. K. Dutta[1,*]

[1] School of Engineering and Technology, Amity University Kolkata, India

Emails: pkdutta@kol.amity.edu

Received: January 14, 2025 Revised: March 03, 2025 Accepted: May 06, 2025 ⋆ Corresponding author

ABSTRACT

Machine learning-based prediction has become an important research direction in the analysis of liver cirrhosis, where clinical, laboratory, and demographic data can be used to support early risk assessment, disease-stage identification, and outcome prediction. Liver cirrhosis represents a progressive and life-threatening condition associated with chronic hepatic injury, fibrosis, portal hypertension, liver-function deterioration, and several severe complications that may reduce patient survival and quality of life. Conventional diagnostic and prognostic approaches remain clinically valuable; however, they may be limited when dealing with complex, nonlinear, heterogeneous, and high-dimensional patient data. This review examines the role of machine learning in liver cirrhosis prediction by analyzing how supervised learning, ensemble models, deep learning, feature-selection strategies, and optimization-based approaches contribute to improving classification accuracy, reducing irrelevant clinical variables, and enhancing model reliability. The review highlights major predictive tasks, including cirrhosis diagnosis, disease severity classification, survival prediction, complication-risk estimation, and treatment-outcome assessment. It also discusses how machine learning can assist clinical decision-making by identifying hidden relationships among biochemical markers, patient characteristics, and disease progression patterns that may not be easily captured through traditional statistical methods. Despite these advantages, the implementation of machine learning models for liver cirrhosis prediction remains associated with several challenges, including limited dataset size, class imbalance, missing clinical values, insufficient external validation, model interpretability concerns, data heterogeneity, and the need for clinically transparent decision-support systems. Overall, this review shows that machine learning-based prediction can provide a valuable computational framework for improving liver cirrhosis assessment when it is developed using reliable datasets, robust validation procedures, explainable modeling strategies, and clinically meaningful performance evaluation.

Keywords: Machine learning ▪ Liver cirrhosis ▪ Disease prediction ▪ Feature selection ▪ Clinical decision support

1. INTRODUCTION

Liver cirrhosis is a progressive and clinically serious condition that develops as a consequence of sustained hepatic injury, long-term inflammation, and irreversible fibrotic remodeling of liver tissue. As the disease advances, normal hepatic architecture is gradually replaced by scar tissue and regenerative nodules, leading to impaired liver function, portal hypertension, metabolic dysregulation, and increased susceptibility to life-threatening complications. The clinical

© 2025 The Author(s). Published by ASPG. This article is licensed under CC BY 4.0.

importance of cirrhosis is not limited to its direct pathological burden, because delayed recognition may allow the disease to progress silently until decompensated manifestations emerge. Therefore, early prediction and structured risk assessment have become essential components of modern liver-disease management, particularly in settings where clinical decisions depend on combining biochemical indicators, demographic characteristics, imaging-derived findings, and patient-specific health records. In this context, machine learning has gained increasing attention because it can model complex relationships among clinical variables and support data-driven prediction frameworks that extend beyond conventional rule-based assessment [1].

The growing availability of healthcare datasets has created new opportunities for developing computational models that can assist in liver-disease diagnosis, severity estimation, and patient stratification. Unlike traditional statistical approaches that may require predefined assumptions about variable interactions, machine learning models can learn nonlinear patterns from structured and unstructured medical data, making them suitable for diseases characterized by multifactorial progression. Liver cirrhosis prediction is particularly relevant to this direction because the disease is influenced by heterogeneous factors such as viral hepatitis, alcohol-related injury, metabolic disorders, inflammatory activity, laboratory abnormalities, and imaging-visible tissue changes. These factors rarely act independently, and their combined effects may be difficult to interpret through isolated clinical measurements. Machine learning-based systems can therefore contribute to earlier identification of high-risk patients by discovering hidden dependencies among input variables and translating them into predictive outputs that may support clinical decision-making [2].

A major motivation for using machine learning in liver cirrhosis prediction is the need to improve the reliability of medical decision-support systems in environments where clinical data are often incomplete, imbalanced, noisy, or high-dimensional. Patient datasets may contain redundant attributes, missing values, overlapping disease patterns, and class distributions that do not equally represent all stages of illness. These issues can reduce the generalizability of predictive models if they are not handled through appropriate preprocessing, validation, and feature-selection procedures. Feature selection is therefore an important component of cirrhosis prediction because it helps identify the most informative clinical variables while reducing unnecessary complexity. Optimization algorithms, including swarm intelligence, evolutionary search, and hybrid metaheuristic approaches, have also been used to improve classifier performance, tune model parameters, and select compact feature subsets [3, 4].

Machine learning-based cirrhosis prediction can support several clinically relevant tasks. These include distinguishing cirrhotic from non-cirrhotic patients, predicting fibrosis or cirrhosis severity, estimating survival probability, detecting risk of complications, and supporting treatment-response assessment. In each task, the objective is not simply to maximize accuracy, but to generate predictions that are clinically interpretable, reproducible, and useful for decision-making. This requirement is particularly important in healthcare, where black-box predictions may face limited acceptance if clinicians cannot understand the reasoning behind model outputs. Consequently, explainable artificial intelligence, transparent feature ranking, and clinically meaningful evaluation metrics are increasingly important for the deployment of liver cirrhosis prediction models [5, 6].

This review examines machine learning-based liver cirrhosis prediction from the perspectives of data representation, feature selection, optimization algorithms, predictive modeling, clinical applications, and future challenges. It summarizes representative studies and discusses how computational methods can be developed into reliable decision-support frameworks when they are supported by robust datasets, careful validation, and explainable model design.

2. LITERATURE REVIEW

The literature review was organized around studies that addressed liver disease, cirrhosis prediction, feature selection, model optimization, medical classification, and decision-support applications. Studies were considered relevant when they investigated liver-related diagnosis or prognosis using clinical records, laboratory variables, medical images, gene-expression data, or real-world patient datasets, and when their methodological approach could support the main focus of the review. Articles that focused on feature selection, model optimization, classifier comparison, prediction performance, or clinical applicability were included because these elements are directly connected to the development of reliable liver cirrhosis prediction systems.

The literature on machine learning-based liver cirrhosis prediction has developed around the need to transform heterogeneous medical records into clinically useful predictive knowledge. Earlier computational studies in liver-disease analysis generally focused on constructing classification pipelines capable of separating diseased and non-diseased cases using measurable patient attributes, while more recent studies have moved toward richer frameworks that combine preprocessing, feature relevance estimation, model optimization, and interpretability-aware evaluation. This progression reflects a broader methodological shift in medical artificial intelligence, where the objective is no longer limited to producing a high numerical accuracy value, but also includes improving the clinical meaningfulness of the selected variables, reducing unnecessary computational burden, and strengthening the reliability of model decisions under real-world data constraints [7].

Figure 1 shows that viral hepatitis, alcohol-related liver disease, and MASLD/NAFLD represent the dominant etiological groups across most included studies. These causes appear with different proportions from one study to another, indicating that liver cirrhosis prediction models may be trained on datasets with different clinical profiles. Autoimmune, cholestatic, cryptogenic, and other etiologies appear as smaller but still relevant components. Their presence indicates that liver cirrhosis datasets may contain minority etiological subgroups that should not be ignored during model development. In machine learning-based prediction, such subgroup variation may affect feature importance, class representation, and model generalizability.

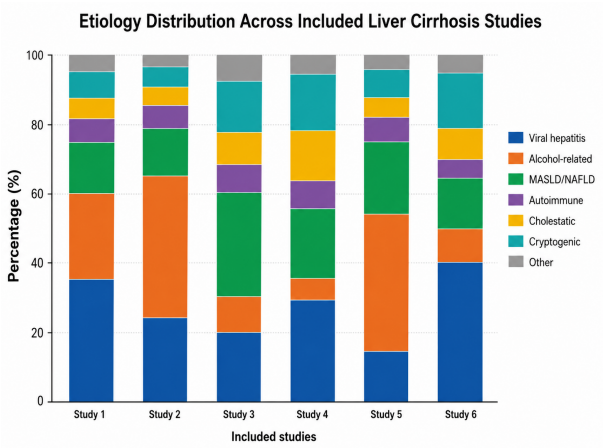


Figure 1. Etiology distribution across included liver cirrhosis studies.

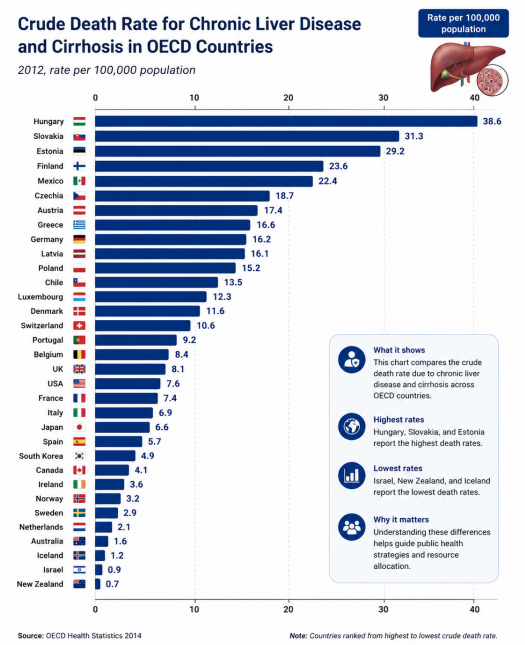


Figure 2. Crude death rate for chronic liver disease and cirrhosis in OECD countries.

Figure 2 presents a cross-national comparison of crude death rates for chronic liver disease and cirrhosis. Countries at the higher end of the chart show substantially larger mortality rates than those at the lower end. The middle group of countries shows moderate variation, suggesting that the burden of chronic liver disease and cirrhosis is not evenly distributed. This pattern may reflect differences in population health profiles, alcohol consumption patterns, viral hepatitis prevalence, obesity-related liver disease, healthcare access, prevention strategies, diagnostic practices, and public health interventions. The figure is descriptive, so these possible explanations should be interpreted as contextual factors rather than direct causal conclusions.

Table 1. Representative machine learning and optimization-based methodologies.

Study	Data/Task	Methodology	Main Contribution
Nasser et al. [1]	Liver disease prediction in a real-world case study	Machine learning framework with clinical variables	Demonstrated practical prediction value for liver-disease decision support.
Shaban [2]	Early liver disease diagnosis	Improved binary butterfly optimization with classifiers	Reduced feature redundancy and improved diagnostic performance.
Goswami et al. [3]	Gene selection and liver classification	Machine learning-based gene selection	Highlighted high-dimensional feature reduction in liver classification.
Rajaguru [4]	Ultrasound cirrhosis detection	Clustering-based features and classifiers	Supported image-based cirrhosis recognition from ultrasonic data.
Abinaya and Rajakumar [5]	Liver fibrosis detection	Hybrid MRMR–BiLSTM–CNN with optimization	Combined feature filtering and deep learning for fibrosis assessment.
Anisha et al. [6]	Liver lesion classification	Deep feature fusion and optimized ensemble classification	Improved lesion classification through optimized feature selection.
Selvaraj and Janakiraman [8]	Liver disease diagnosis	Particle swarm optimization feature selection	Used swarm search to select informative diagnostic features.
Malik et al. [9]	Noncommunicable disease detection	Hybrid metaheuristic optimization	Showed the relevance of hybrid optimization to medical diagnosis.

Table 1 presents a methodological synthesis of representative machine learning, deep learning, and optimization-based studies. These studies indicate that liver-related prediction has increasingly shifted from isolated classifier comparison toward integrated pipelines that combine data preprocessing, feature selection, model tuning, and clinical interpretation. Optimization-based feature selection is especially important because medical data frequently contain correlated variables, sparse measurements, and noisy observations.

2.1 Data Characteristics and Clinical Representation

Reliable liver cirrhosis prediction begins with appropriate representation of clinical information. Typical datasets may contain demographic variables, biochemical markers, hematological measurements, imaging-derived descriptors, disease-stage labels, and patient follow-up outcomes. These variables differ in scale, completeness, and clinical meaning. Missing values may appear because not all tests are performed for all patients, while class imbalance may occur when severe disease cases are less frequent than mild or non-diseased cases. If these issues are ignored, a model may learn biased patterns and perform poorly when deployed in different clinical populations.

Clinical representation also depends on whether the prediction task is based on tabular patient records, medical images, genomic information, or multimodal data. Tabular records are commonly used because they are easier to collect and interpret, whereas imaging and genomic datasets can provide richer information but require more complex preprocessing and higher computational resources. In all cases, the quality of the input representation directly affects the reliability of the model. Therefore, normalization, missing-value handling, class rebalancing, and feature engineering are central steps in the development of cirrhosis prediction systems.

2.2 Feature Selection and Model Optimization

Feature selection is a major theme in liver cirrhosis prediction because clinical datasets may contain redundant, irrel-

evant, or highly correlated attributes. Filter methods rank variables using statistical criteria, wrapper methods evaluate subsets through classifier performance, and embedded methods perform selection during model training. Metaheuristic optimization methods such as particle swarm optimization, firefly optimization, butterfly optimization, multi-verse optimization, and evolutionary algorithms have been used to search large feature spaces where exhaustive evaluation is impractical [8, 10, 11].

The feature-selection problem can be expressed as the optimization of two competing objectives:

$$\min F(S) = \alpha E(S) + (1-\alpha)\frac{|S|}{|D|}, \tag{1}$$

where S is the selected feature subset, D is the original feature set, $E(S)$ is the classification error obtained using the selected subset, and α controls the trade-off between predictive error and subset size. This formulation reflects the need to preserve predictive accuracy while reducing unnecessary clinical measurements.

For supervised classification, a model learns a mapping from patient features to disease labels:

$$\hat{y} = f(\mathbf{x};\theta), \tag{2}$$

where $\mathbf{x}$ denotes the patient feature vector, θ represents model parameters, and $\hat{y}$ is the predicted class or risk score. Model optimization seeks parameter values that minimize a loss function:

$$\theta^* = \arg\min_{\theta} \mathscr{L}(y, f(\mathbf{x};\theta)) + \lambda\Omega(\theta), \tag{3}$$

where $\mathscr{L}$ measures prediction error, Ω is a regularization term, and λ controls the regularization strength.

Common performance metrics include accuracy, precision, recall, specificity, F1-score, and area under the receiver operating characteristic curve. In imbalanced clinical datasets, accuracy alone may be misleading because a model can achieve high accuracy by favoring the majority class. Therefore, sensitivity, specificity, balanced accuracy, and calibration measures should be considered when evaluating cirrhosis prediction models.

2.3 Clinical Decision Support and Implementation Challenges

Machine learning can support clinical decision-making by identifying hidden relationships among biochemical markers, patient characteristics, and disease progression patterns that may not be easily captured through traditional statistical methods. A practical liver cirrhosis prediction system should provide risk estimates, highlight influential variables, and communicate model uncertainty in a way that clinicians can interpret. Explainability is therefore an essential requirement, particularly when predictions may influence diagnostic testing, referral decisions, monitoring frequency, or treatment planning [12].

Implementation remains challenging. Many studies rely on limited datasets, single-center records, or retrospective data. These settings may restrict external validity and make reported performance difficult to reproduce. Missing clinical values, inconsistent measurement protocols, and heterogeneous patient populations further complicate deployment. In addition, data privacy requirements and differences among hospital information systems can limit data sharing and model validation. To move toward clinical translation, future studies should emphasize multicenter validation, transparent reporting, prospective testing, and integration with existing clinical workflows.

3. DISCUSSION

The reviewed literature indicates that machine learning-based liver cirrhosis prediction is promising, but its value depends strongly on the quality of data, the appropriateness of feature representation, and the transparency of the modeling process. Supervised learning models can provide effective classification when the dataset contains informative variables and balanced representation of disease stages. Ensemble methods often improve robustness by combining multiple learners, while deep learning methods can extract complex patterns from images or high-dimensional data. However, increased model complexity does not automatically guarantee clinical usefulness, especially when interpretability, reproducibility, and validation are insufficient.

Optimization algorithms provide a useful mechanism for improving feature selection and parameter tuning. They are particularly valuable when the search space is large and when interactions among features are difficult to evaluate manually. Nevertheless, optimization-based studies should avoid overfitting to small datasets and should report validation procedures clearly. Cross-validation, independent test sets, and external validation are necessary to demonstrate that selected feature subsets and tuned parameters generalize beyond the development dataset.

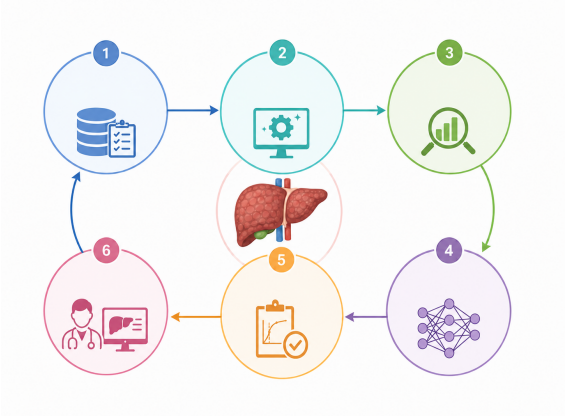


Figure 3. Step-based workflow for machine learning-based liver cirrhosis prediction.

Figure 3 summarizes a step-based workflow for machine learning-based liver cirrhosis prediction. The process begins with data collection and clinical variable acquisition, followed by preprocessing and model development. Feature selection, optimization, and classifier training are then used to generate predictive outputs. Finally, clinical interpretation and decision-support integration close the loop by translating computational findings into meaningful medical action. This workflow emphasizes that machine learning prediction should be treated as a complete pipeline rather than a single classifier-selection problem.

Future research should focus on clinically transparent systems that combine high predictive performance with interpretability. Explainable models can help clinicians understand why a patient is assigned to a high-risk category, while uncertainty estimation can support cautious decision-making when data are incomplete or ambiguous. Multimodal models that integrate laboratory tests, imaging features, genomics, and longitudinal patient records may improve prediction, but they also require careful harmonization and validation. Ethical considerations, including bias, fairness, privacy, and accountability, should be addressed before deployment in real healthcare environments.

4. CONCLUSION

This review presented a structured overview of machine learning-based liver cirrhosis prediction, with emphasis on feature selection, optimization algorithms, clinical applications, and future challenges. The reviewed studies show that machine learning can support diagnosis, severity estimation, survival prediction, complication-risk assessment, and treatment-outcome evaluation. Feature selection and optimization methods are important because they reduce irrelevant variables, improve model efficiency, and may increase interpretability when applied carefully.

Despite these advantages, several barriers remain. Limited dataset size, class imbalance, missing values, heterogeneous data sources, insufficient external validation, and interpretability concerns continue to restrict clinical implementation. Therefore, future systems should be developed using reliable datasets, robust validation procedures, transparent modeling strategies, and clinically meaningful evaluation metrics. When these requirements are met, machine learning-based prediction can provide a valuable computational framework for improving liver cirrhosis assessment and supporting evidence-based clinical decision-making.

REFERENCES

[1] I. M. Nasser *et al.*, "Liver cirrhosis prediction based on machine learning framework: a real-world case study in egypt," *Knowledge and Information Systems*, vol. 65, no. 6, pp. 2595–2617, 2023.

[2] 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, 2023.

[3] S. A. Goswami, K. C. Patel, and S. A. Dave, "Gene selection and liver classification using machine learning," in *Computational Intelligence for Genomics Data*. Elsevier, 2025, pp. 135–143.

[4] H. Rajaguru, "Enhancement of liver cirrhosis detection from ultrasonic images through clustering-based features and classifiers," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualisation*, vol. 13, no. 1, 2025.

[5] R. J. Abinaya and G. Rajakumar, "Accurate liver fibrosis detection through hybrid mrmr-bilstm-cnn architecture with histogram equalization and optimization," *Journal of Imaging Informatics in Medicine*, vol. 37, no. 3, pp. 1008–1022, 2024.

[6] A. Anisha, G. Jiji, and T. A. B. Raj, "Deep feature fusion and optimized feature selection based ensemble classification of liver lesions," *The Imaging Science Journal*, vol. 71, no. 6, pp. 518–536, 2023.

[7] Y. Liu, J.-S. Wang, J.-Y. Wen, Y.-T. Li, and P.-G. Yan, "A multi-objective evolutionary algorithm for solving the feature selection problem of high-dimensional sparse data and its application in admet classification," *Engineering Optimization*, vol. 58, no. 1, pp. 171–202, 2026.

[8] G. Selvaraj and S. Janakiraman, "Improved feature selection based on particle swarm optimization for liver disease diagnosis," in *Swarm, Evolutionary, and Memetic Computing*, vol. 8298. Springer International Publishing, 2013, pp. 214–225.

[9] S. Malik, S. G. K. Patro, C. Mahanty *et al.*, "Hybrid metaheuristic optimization for detecting and diagnosing noncommunicable diseases," *Scientific Reports*, vol. 15, no. 1, p. 7816, 2025.

[10] G. I. Sayed, A. Darwish, and A. E. Hassanien, "A new chaotic multi-verse optimization algorithm for solving engineering optimization problems," *Journal of Experimental and Theoretical Artificial Intelligence*, vol. 30, no. 2, pp. 293–317, 2018.

[11] A. A. Abd El-Mageed, A. A. Abohany, and K. M. Hosny, "Enhanced binary kepler optimization algorithm for effective feature selection of supervised learning classification," *Journal of Big Data*, vol. 12, no. 1, p. 93, 2025.

[12] J. R. Trillo, M. J. Del Moral, J. M. Tapia, J. García-Cabello, and F. J. Cabrerizo, "Explainable classifier with adaptive optimisation for medical data," *Applied Intelligence*, vol. 56, no. 3, p. 77, 2026.