Metaheuristic Optimization Review

♦ ISSN: 2692-4048 Vol: 01, No. 01, 2024 ♦ PP. 14–17

Article DOI: https://doi.org/10.54216/MOR.010104

REVIEW ARTICLE

Innovations in Machine Learning Models for Hepatitis Diagnosis and Disease Progression Prediction

Marwa M. Eid[1,*] **Wei Hong Lim**[2]

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

[2] Faculty of Engineering, Technology and Built Environment, UCSI University, Kuala Lumpur 56000, Malaysia

Emails: mmm@ieee.org · limwh@ucsiuniversity.edu.my

Received: August 22, 2023 Revised: December 09, 2023 Accepted: January 09, 2024 ⋆ Corresponding author

ABSTRACT

Chronic liver disease (CLD) is a group of conditions for which up to half of the global population remains at risk and causes serious complications, including liver cirrhosis and liver cancer. Therefore, early diagnosis and proper treatment enhance the prognosis of patients suffering from CLD. This review explores how machine learning (ML) techniques are used in practice to diagnose, prognose, and treat chronic liver diseases. Drawing on findings from several studies, broader implementation of these methods is expected to improve disease management and enable more accurate detection of liver disease. The review discusses supervised and unsupervised learning, neural networks, and ensemble learning methods applied to patient risk estimation, treatment guidance, and disease-progression assessment. Although ML technology in hepatology is progressing, issues such as model diversity, applicability, and ethical concerns still pose challenges. This paper highlights the importance of multidisciplinary work to develop suitable mechanisms for addressing these issues and for using ML adequately in clinics. Overall, the findings indicate that ML can transform chronic liver disease management and support innovative treatment methods and better patient care.

Keywords: Persistent hepatic illnesses ▪ Computational techniques ▪ Identification ▪ Prediction ▪ Management and evidence-based methods

1. INTRODUCTION

Chronic liver disease includes hepatitis B, hepatitis C, and hepatocellular carcinoma (HCC), conditions that affect millions of people worldwide. These diseases create a substantial burden in morbidity, mortality, and healthcare cost; consequently, improving diagnosis and treatment is essential. Although modern imaging, serological tests, and liver biopsy support clinical decision-making, they can be costly, invasive, or limited by resource availability. This has increased the demand for early intervention, effective diagnosis, treatment optimization, and efficient resource management [1].

Artificial intelligence in medicine, particularly machine learning (ML), has begun to reshape medical diagnostics and healthcare delivery. ML systems learn patterns from data and can reveal relationships that are difficult to detect using traditional analysis. In liver diseases, ML has been used to interpret serum biomarkers, clinical data, genetic material, medical imaging, and other patient records. These capabilities enable improved prediction of disease development, risk factors, and patient-specific treatment pathways [2, 3].

During the last decade, researchers have examined ML methods for determining causes, improving diagnosis, and managing liver diseases. Deep learning, support vector machines (SVM), decision trees, and ensemble strategies have been used to classify liver disease, even when imaging data are

© 2024 The Author(s). Published by ASPG. This article is licensed under CC BY 4.0.

unclear. ML models trained on imaging data can detect subtle liver changes that may precede HCC or fibrosis, while biomarker-based models can estimate risk by combining genetic markers with clinical variables [4, 5].

Despite encouraging outcomes, integrating ML into routine liver-disease practice remains challenging. Key barriers include data completeness, model interpretability, and validation across different settings. The changing nature of liver diseases, including differences in causes and progression rates, also requires flexible and robust models. Addressing these challenges requires collaboration among clinicians, data scientists, engineers, and ethicists to ensure that ML systems are accurate, explainable, and clinically useful [6, 7].

2. LITERATURE REVIEW

Several studies have demonstrated the potential of ML for liver disease diagnosis, prognosis, and treatment planning. Hashem et al. [8] compared ML approaches for predicting advanced liver fibrosis in chronic hepatitis C patients. Using serum biomarkers and clinical data, they evaluated decision trees, genetic algorithms, and particle swarm optimization with multi-linear regression. The study identified age, platelet count, aspartate aminotransferase (AST), and albumin levels as important predictors of advanced fibrosis.

Parisi et al. [9] proposed a hybrid algorithm for prognosis prediction in patients with hepatitis. Their method used 19 biomarkers collected from 155 patients and combined least-squares SVM (LSVM) with multilayer perceptron (MLP). The hybrid model achieved strong classification performance, demonstrating the value of combining multiple learning strategies for clinical prognosis.

Xu et al. [10] developed personalized predictive models for evaluating HCC risk in hepatitis B virus-related cirrhosis patients with low alpha-fetoprotein levels. They used clinical and pre-treatment laboratory information and compared Gaussian naive Bayes, XGBoost, random forest, and LASSO regression. XGBoost outperformed conventional approaches in estimating HCC risk.

Shousha et al. [11] examined the influence of IL28B gene polymorphism and biochemical markers in predicting advanced liver fibrosis in chronic hepatitis C. Their study used decision trees and Auto-WEKA and compared the results with APRI and FIB-4 scores. Auto-WEKA selected an MLP neural network as the best-performing model, outperforming traditional scoring systems.

Kukkar and Sandhu [12] proposed an automatic diagnosis system using electronic health records from 615 clients. The study considered categories such as blood donors, suspected donors, hepatitis patients, fibrosis patients, and cirrhosis patients. SVM, multivariate adaptive regression splines, Bayesian generalized linear models, random forests, and decision trees were tested, showing the usefulness of ML for early hepatitis C detection.

Fayaz et al. [13] investigated liver disease diagnosis using the Indian Liver Patients Dataset. Their preprocessing pipeline included sampling, normalization, and particle swarm optimization. The J48 algorithm performed better than other algorithms for distinguishing patients from non-patients, reinforcing the role of preprocessing and feature selection in liver disease diagnosis.

Sartakhti et al. [14] presented a hybrid hepatitis diagnosis approach combining SVM with simulated annealing. The UCI machine learning repository dataset was used, and classification accuracy was evaluated using 10-fold cross-validation. The method achieved 96.25% accuracy, indicating the feasibility of hybrid optimization-based classifiers for hepatitis diagnosis.

Jin et al. [15] developed a deep-learning radiomics model for predicting HCC occurrence in chronic hepatitis B patients over a five-year follow-up. Their model used two-dimensional shear wave elastography and B-mode ultrasound images, combined with clinical data. The HCC-R model achieved high area-under-the-curve values in training, validation, and testing cohorts and outperformed liver stiffness measurement and conventional clinical scores.

Lukáčová et al. [16] focused on improving hepatitis B and C diagnostics using questionnaire data from patients in Slovakia. Their work used exploratory data analysis and predictive data mining to identify relationships among attributes such as age, gender, and diagnosis, while also producing decision rules useful in clinical settings.

Kayvanjoo et al. [17] studied prediction of hepatitis C interferon/ribavirin therapy outcomes based on viral nucleotide attributes. Using Gini index, chi-squared analysis, and ML algorithms, they identified nucleotide features associated with treatment response and achieved useful predictive performance for distinguishing responders from non-responders.

Vijayalakshmi and Mohideen [18] proposed a survival prediction model for hepatitis B patients using supervised and unsupervised learning with lion optimization. Their study compared the efficiency of multiple algorithms for survival prediction and emphasized the importance of optimized models in patient-risk evaluation.

Qiu et al. [19] built a 28-day prognosis model for acute-on-chronic liver failure patients using clinical data, LASSO regression, and XGBoost. Their model outperformed conventional scoring systems for mortality prediction, demonstrating the value of ML in short-term clinical risk assessment.

Anyin et al. [20, 21] integrated bioinformatics and ML to analyze diagnostic biomarkers for HBV-induced HCC. Using the GEO database, weighted gene co-expression network analysis, LASSO, random forest, and SVM, they identified RACGAP1, ECT2, and NDC80 as essential genes associated with HBV-HCC and as potential diagnostic and therapeutic targets.

Babu et al. [22] developed classification and prediction models for hepatitis C using multimodal data sources. Decision trees, support vectors, and an unconventional consensus classifier were used to stratify patients according to disease severity and hepatitis C stage, improving predictive power through preprocessing and feature engineering.

Machine learning in hepatology is a valuable addition to early diagnosis, risk stratification, and treatment management. The reviewed studies show that ML can reveal patterns and risk factors that may be missed by conventional methods, ranging from non-invasive fibrosis prediction to hybrid prognosis algorithms for hepatitis and HCC. However, these advances are accompanied by challenges involving data quality, model in-

Table 1. Summary of Literature Review

Study	Objective	Methodology	Key Findings
[8]	Assess ML algorithms for advanced liver fibrosis prediction.	Serum biomarkers and clinical data with decision tree, genetic algorithm, and particle-swarm-optimized multi-linear regression.	Age, platelet count, AST, and albumin were significant predictors of advanced fibrosis.
[9]	Assess a hybrid AI classifier for chronic hepatitis prognosis.	Nineteen biomarkers from 155 patients with LSVM and MLP.	The hybrid method achieved excellent accuracy and AUROC.
[10]	Develop individualized predictive models of HCC risk.	Clinical and laboratory information with Gaussian naive Bayes, XGBoost, random forest, and LASSO regression.	XGBoost outperformed conventional methods in predicting HCC risk.
[11]	Assess IL28B polymorphism and compare ML with traditional scores.	Decision tree and Auto-WEKA compared with APRI, FIB-4, and MLP.	MLP outperformed APRI, FIB-4, and REP Tree in assessing advanced fibrosis.
[12]	Use ML to improve hepatitis C diagnosis and prognosis.	EHR data with SVM, MARS, BGLM, RF, and DT.	ML supported early hepatitis C detection.
[13]	Improve liver disease diagnosis accuracy.	ILPD dataset with J48 and other algorithms.	J48 gave better results for distinguishing patients from non-patients.
[14]	Diagnose hepatitis using hybrid SVM-SA.	Machine-learning repository data with SVM and simulated annealing.	The method achieved 96.25% diagnostic accuracy.
[15]	Predict HCC using ML.	2D-SWE and B-mode ultrasound with deep-learning radiomics.	HCC-R achieved high accuracy in predicting HCC occurrence.
[16]	Improve hepatitis B and C diagnostics.	Questionnaire data with exploratory analysis and predictive data mining.	The study identified useful diagnostic relations and decision rules.
[17]	Identify genetic determinants of HCV therapy outcomes.	Complete HCV nucleotide sequences with Gini index, chi-squared analysis, and ML.	Nucleotide features correlated with treatment outcomes.
[18]	Evaluate survival rate among hepatitis patients.	Serum hepatitis dataset with multiple ML algorithms.	The study compared algorithm efficiency for survival prediction.
[19]	Build a 28-day mortality model for ACLF patients.	Clinical data with LASSO regression and XGBoost.	XGBoost outperformed conventional scoring systems.
[20, 21]	Analyze HBV-induced HCC biomarkers.	GEO data with WGCNA, LASSO, random forest, and SVM.	RACGAP1, ECT2, and NDC80 were identified as diagnostic and therapeutic candidates.
[22]	Classify and predict liver disease severity.	Multimodal data with decision trees, support vectors, and unconventional consensus classifier.	Patients were stratified by severity and hepatitis C stage.

terpretability, and clinical validation. Future research should improve algorithms, increase database size, and develop flexible ML solutions that can be integrated into clinical practice.

3. DISCUSSION

The reviewed literature demonstrates that ML models can support liver disease care across several stages: early detection, classification, prognosis, and therapy-response prediction. Supervised learning algorithms such as SVM, decision trees, random forests, and XGBoost are commonly used when labeled clinical data are available. Neural networks and deep-learning radiomics are especially useful for high-dimensional medical imaging, while feature-selection techniques such as LASSO and bioinformatics workflows help identify compact diagnostic markers.

The general prediction task in many reviewed studies can be expressed as learning a mapping from patient features to a clinical outcome:

$$\hat{y} = f(\mathbf{x}; \theta), \tag{1}$$

where $\mathbf{x}$ denotes clinical, laboratory, genetic, or imaging features, θ represents model parameters, and $\hat{y}$ is the predicted diagnosis, prognosis, or risk category.

For regularized feature selection, several studies employ sparse modeling principles similar to LASSO:

$$\min_{\beta} \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda \|\beta\|_1, \tag{2}$$

where λ controls sparsity. This is clinically important because compact feature sets are easier to interpret and can reduce unnecessary measurements.

For classification models, performance is commonly evaluated using accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC). Accuracy is computed as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{3}$$

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively.

Although ML performance is promising, clinical deployment requires careful external validation. Models trained on limited or single-center datasets may not generalize to other populations. Ethical issues also remain important, including privacy, bias, transparency, and accountability. Consequently, ML systems should support clinicians rather than replace clinical judgment.

4. CONCLUSION

This review examined innovations in machine learning models for hepatitis diagnosis and disease progression prediction. The reviewed studies show that ML can improve early diagnosis, risk stratification, treatment planning, and prognosis assessment for chronic liver diseases. Methods such as SVM, decision trees, random forests, XGBoost, neural networks, LASSO, and deep-learning radiomics have achieved strong results across fibrosis prediction, hepatitis diagnosis, HCC risk estimation, and survival analysis.

Despite these benefits, successful clinical translation depends on high-quality data, model interpretability, external validation, and multidisciplinary collaboration. Future work should focus on larger and more diverse datasets, explainable algorithms, and clinically integrated systems that provide reliable decision support. Overall, machine learning has substantial potential to improve prevention, diagnosis, and management of chronic liver diseases at early stages and to enhance patient outcomes.

REFERENCES

[1] N. Mayr, N. Thoeringer, M. M. Zeitlinger, B. Koestenbauer, A. Kanitsar, and A. F. Stalzer, "A multimodal framework for computer-aided liver disease diagnosis," 2024.

[2] E.-S. M. El-Kenawy et al., "Machine learning for improved diagnosis and prognosis in healthcare," 2024.

[3] A. Abdulrahman, Z. M., A. M. Zaki, F. H. H. Rizk, M. M. Eid, and E.-S. M. El-Kenawy, "Exploring Optimization Algorithms: A Review of Methods and Applications," *Journal of Artificial Intelligence and Metaheuristics*, vol. 7, no. 2, pp. 08–17, 2024.

[4] M. Al-Hasani et al., "Machine learning improves early detection of liver fibrosis by quantitative ultrasound radiomics," *IEEE International Ultrasonics Symposium*, 2022.

[5] M. Li, X. Song, C. Xu, and C. Yu, "Response to the clinical significance and prognostic value of 3-mercaptopyruvate sulfurtransferase in hepatocellular carcinoma," *Liver International*, vol. 42, no. 7, pp. 1696–1696, 2022.

[6] G. Chongo and J. Soldera, "Use of machine learning models for the prognostication of liver transplantation: A systematic review," *World Journal of Transplantation*, vol. 14, no. 1, 2024.

[7] A. C. Filiberto, I. L. Leeds, and T. J. Loftus, "Editorial: Machine Learning in Clinical Decision-Making," *Frontiers in Digital Health*, vol. 3, 2021.

[8] S. Hashem et al., "Comparison of Machine Learning Approaches for Prediction of Advanced Liver Fibrosis in Chronic Hepatitis C Patients," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 15, no. 3, pp. 861–868, 2018.

[9] L. Parisi, N. Ravichandran, and M. L. Manaog, "A novel hybrid algorithm for aiding prediction of prognosis in patients with hepatitis," *Neural Computing and Applications*, vol. 32, no. 8, pp. 3839–3852, 2020.

[10] Y. Xu et al., "Development of machine learning-based personalized predictive models for risk evaluation of hepatocellular carcinoma in hepatitis B virus-related cirrhosis patients with low levels of serum alpha-fetoprotein," *Annals of Hepatology*, vol. 29, no. 6, 2024.

[11] H. I. Shousha, A. H. Awad, D. A. Omran, M. M. Elnegouly, and M. Mabrouk, "Data Mining and Machine Learning Algorithms Using IL28B Genotype and Biochemical Markers Best Predicted Advanced Liver Fibrosis in Chronic Hepatitis C," *Japanese Journal of Infectious Diseases*, vol. 71, no. 1, pp. 51–57, 2018.

[12] A. Kukkar and J. K. Sandhu, "An Automatic Diagnosis System Based on Machine Learning Models for Predicting Hepatitis C from Blood Samples," *Signals and Communication Technology*, pp. 91–108, 2024.

[13] S. T. Fayaz et al., "Prediction of Liver Patients Using Machine Learning Algorithms," *Lecture Notes in Electrical Engineering*, vol. 789, pp. 135–145, 2022.

[14] J. S. Sartakhti, M. H. Zangooei, and K. Mozafari, "Hepatitis disease diagnosis using a novel hybrid method based on support vector machine and simulated annealing," *Computer Methods and Programs in Biomedicine*, vol. 108, no. 2, pp. 570–579, 2012.

[15] J. Jin et al., "Deep learning radiomics model accurately predicts hepatocellular carcinoma occurrence in chronic hepatitis B patients: a five-year follow-up," *American Journal of Cancer Research*, vol. 11, no. 2, p. 576, 2021.

[16] A. Lukáčová, F. Babič, Z. Paraličová, and J. Paralič, "How to increase the effectiveness of the hepatitis diagnostics by means of appropriate machine learning methods," *Lecture Notes in Computer Science*, vol. 9267, pp. 81–94, 2015.

[17] A. H. Kayvanjoo, M. Ebrahimi, and G. Haqshenas, "Prediction of hepatitis C virus interferon/ribavirin therapy outcome based on viral nucleotide attributes using machine learning algorithms," *BMC Research Notes*, vol. 7, no. 1, pp. 1–11, 2014.

[18] C. Vijayalakshmi and S. P. Mohideen, "Survival prediction model with lion optimization for Hepatitis-B patients using supervised and unsupervised learning algorithm," *Measurement: Sensors*, vol. 32, 2024.

[19] S. Qiu et al., "Predicting the 28-day prognosis of acute-on-chronic liver failure patients based on machine learning," *Digestive and Liver Disease*, 2024.

[20] Y. Anyin, L. Jianping, L. Mengru, Z. Hong, Z. Xulei, and W. Lianping, "Integrating bioinformatics and machine learning methods to analyze diagnostic biomarkers for HBV-induced hepatocellular carcinoma," *Diagnostic Pathology*, vol. 19, no. 1, pp. 1–10, 2024.

[21] Y. Anyin, L. Jianping, L. Mengru, Z. Hong, Z. Xulei, and W. Lianping, "Integrating bioinformatics and machine learning methods to analyze diagnostic biomarkers for HBV-induced hepatocellular carcinoma," *Diagnostic Pathology*, vol. 19, no. 1, pp. 1–10, 2024.

[22] A. Babu, S. S. Krishna, K. S. Saran, and B. Aarthi, "Categorization and Forecasting of Hepatitis C Diagnosis via an Unconventional Consensus Classifier," *International Journal of Advance Research*, pp. 2454–132, 2024.