ISSN: 2581-3404 (Online) IF: 5.86 (SJIF)
IJIRTM, Volume-10, Issue-5, September-2026

Machine Learning Algorithms for Liver Disease Classification: A Review

Iliyas Akhtar[1]
Research Scholar, Computer Science & Engineering Department
All Saint's College of Technology, Bhopal[1,2]
Sarwesh Site[2]
Assistant Professor, Computer Science & Engineering Department[2]

Abstract. *Liver disease is a major healthcare concern because its progression may remain undetected until significant liver damage has occurred. Early and accurate identification of liver disease can support timely clinical assessment and improve the management of affected patients. Recent developments in artificial intelligence (AI), machine learning (ML), and deep learning (DL) have created opportunities to develop predictive models that analyze demographic, clinical, biochemical, and imaging-related information. This study presents a literature-based investigation of machine learning and deep learning techniques for liver disease prediction, drawing on research involving the Indian Liver Patient Dataset (ILPD), larger healthcare datasets, metabolic dysfunction-associated steatotic liver disease (MASLD), liver cirrhosis, and nonalcoholic fatty liver disease (NAFLD). The reviewed approaches include Logistic Regression, Decision Tree, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest, Gradient Boosting, XGBoost, Artificial Neural Networks, convolutional neural networks, and hybrid learning frameworks. Particular attention is given to data preprocessing, missing-value treatment, class-imbalance handling, feature selection, hyper-parameter optimization, cross-validation, and explainable AI. Existing studies demonstrate the potential of ensemble classifiers, including Random Forest and XGBoost, while other investigations indicate that multimodal learning and quantum–classical approaches may provide additional opportunities for predictive modeling. However, differences in datasets, target definitions, validation procedures, and evaluation metrics make direct comparisons difficult. The literature therefore highlights the need for a reproducible and carefully validated prediction framework that evaluates predictive performance, minority-class detection, model interpretability, and generalization. Such a framework could support research into reliable computer-assisted liver disease screening, while clinical diagnosis and treatment decisions remain the responsibility of qualified healthcare professionals.*

Keywords: Liver disease prediction, machine learning, deep learning, ILPD, Random Forest.

Introduction

Liver disease is a major health concern because it can affect the normal functioning of the human body and may lead to severe complications when not detected at an early stage. The liver performs several essential functions, including metabolism, detoxification, nutrient storage, and the production of important proteins. Various factors, such as viral infections, excessive alcohol consumption, obesity, unhealthy dietary habits, and metabolic disorders, can contribute to liver-related diseases. Early

identification of liver disease is important for timely medical intervention, appropriate treatment, and improved patient outcomes. However, conventional diagnosis may require multiple clinical investigations and expert interpretation of patient data. Therefore, data-driven approaches can support healthcare professionals by identifying patterns in clinical measurements and estimating the likelihood of liver disease [1].

Machine learning (ML) has emerged as a useful approach for analyzing healthcare data and developing predictive models for disease detection. It enables computer systems to learn patterns and relationships from historical patient records and use these patterns to classify new cases. In liver disease prediction, ML algorithms can analyze clinical attributes such as age, gender, total bilirubin, direct bilirubin, alkaline phosphatase [2], alanine aminotransferase (SGPT), aspartate aminotransferase (SGOT), total proteins, albumin, and the albumin-to-globulin ratio. These attributes may contain relevant information about liver function and can help a trained model distinguish between patients with and without indicators of liver disease. The effectiveness of a prediction model depends on the quality of the dataset, the selection of relevant features, appropriate preprocessing, and the choice of classification algorithm [3].

Several machine learning algorithms have been investigated for liver disease prediction, including Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Naive Bayes, Gradient Boosting, and Extreme Gradient Boosting (XGBoost). Logistic Regression provides a baseline classification approach, while Decision Tree models generate predictions through rule-based decisions. Random Forest combines multiple decision trees to improve predictive performance and reduce overfitting [4, 5]. SVM identifies decision boundaries between classes, whereas KNN classifies patients according to the similarity of their clinical attributes. Gradient Boosting and XGBoost construct ensembles of predictive models sequentially to improve classification performance. Previous studies have investigated these algorithms using liver disease datasets and reported differences in their predictive effectiveness depending on the dataset, preprocessing strategy, and evaluation methodology.

One of the commonly used datasets for this research is the Indian Liver Patient Dataset (ILPD), which contains clinical attributes of Indian patients and is used to investigate the classification of liver disease. However, developing a reliable prediction model using this dataset presents several challenges. These include missing or inconsistent values, differences in feature scales, potentially irrelevant attributes, and unequal representation of patient classes [6]. Class imbalance is particularly important because a model may achieve apparently satisfactory accuracy while incorrectly classifying a substantial proportion of patients belonging to the minority class. Therefore, data preprocessing, feature selection, and suitable strategies for handling class imbalance are important components of an effective machine learning pipeline. Previous research has explored feature selection, hybrid classification approaches, and resampling methods to address these challenges [7].

The performance of liver disease prediction models should be evaluated using multiple classification metrics rather than accuracy alone. Accuracy measures the overall proportion of correct predictions, whereas precision indicates the proportion of positive predictions that are correct. Recall measures the proportion of actual positive cases correctly identified, and the F1-score balances precision and recall. A confusion matrix provides a detailed representation of correct and incorrect classifications, while the area under the receiver operating characteristic curve (ROC-AUC) evaluates the model's ability to distinguish between classes across different classification thresholds [8, 9]. Cross-validation can further support a more reliable comparison by evaluating model performance across multiple training and validation splits.

These measures are particularly relevant when the dataset contains an unequal distribution of liver disease and non-liver-disease cases.

Although existing studies have reported promising results using individual classifiers, ensemble learning, feature selection, and class-balancing techniques, the performance of a machine learning model can vary according to the dataset and experimental setup [10, 11]. Consequently, a systematic comparison of multiple classifiers under consistent experimental conditions is necessary to identify their relative strengths and limitations. The present study focuses exclusively on machine learning-based liver disease prediction, using clinical data from the selected liver disease dataset. It aims to preprocess the data, train and evaluate selected machine learning classifiers, compare their performance using accuracy, precision, recall, F1-score, ROC-AUC, and confusion matrices, and identify the classifier that provides the most suitable predictive performance for the experimental dataset [12, 13]. The results can contribute to the development of data-driven decision-support methods for liver disease screening, while clinical diagnosis and treatment decisions remain the responsibility of qualified healthcare professionals.

Literature Review

Gupta et al. [1] investigated liver disease prediction using the UCI Liver Patient Dataset and evaluated multiple machine learning algorithms, including Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbors, Gradient Boosting, XGBoost, and LightGBM. Their work examined the use of feature selection to improve classification performance. The study illustrates the importance of comparing multiple algorithms rather than relying on a single predictive model. However, the results depend on the selected features, preprocessing methods, and experimental configuration, making standardized evaluation important for subsequent research.

Khaled et al. [2] evaluated machine learning models for liver disease detection using healthcare big data containing 32,000 records. The study compared KNN, Gaussian Naive Bayes, and Random Forest, reporting 97.3% accuracy for Random Forest under its experimental conditions. The findings demonstrate the potential of ensemble learning for clinical prediction. Nevertheless, performance obtained from a larger healthcare dataset should not be directly compared with results from smaller benchmark datasets such as ILPD without accounting for differences in population, features, and validation procedures.

El Atifi et al. [3] investigated ensemble learning for liver disease prediction using clinical data and compared Random Forest, AdaBoost, and Gradient Boosting with hyperparameter optimization. Their study reported that Random Forest achieved the best performance among the evaluated models, with approximately 85.17% accuracy. The research emphasizes the importance of model optimization and appropriate evaluation when working with clinical datasets. Its findings also support further investigation into ensemble classifiers and their performance under different data-balancing strategies.

Kate et al. [4] proposed a hybrid framework for liver disease prediction using the Indian Liver Patient Dataset. The methodology integrated data preprocessing, Synthetic Minority Over-sampling Technique (SMOTE), Mutual Information-based feature selection, Recursive Feature Elimination (RFE), and a genetically optimized XGBoost classifier. The study reported 85.03% test accuracy, 90.67% recall, and 91.09% ROC-AUC. These findings highlight the potential benefits of combining feature selection with hyperparameter optimization. However, the performance of the framework should be evaluated under independent testing and leakage-free validation to determine its reliability beyond the experimental dataset.

Rani et al. [5] examined class-imbalance handling for liver disease prediction using the Indian Liver Patient Dataset and the BUPA Liver Disorders Dataset. The researchers compared upsampling, ADASYN, SMOTE, SMOTE-ENN, and SMOTE-Tomek, and proposed hybrid approaches that combine balancing techniques with classification algorithms. Their hybrid ensemble framework reported 93.2% accuracy on ILPD and 95.4% on BUPA, with Brier scores of 0.032 and 0.031, respectively. The study indicates that class-balancing methods and ensemble learning can contribute to improved prediction. Nevertheless, the results should be interpreted in relation to the particular train-test design and evaluation procedures used.

Other research summarized in the uploaded literature has investigated feature-selection techniques such as correlation-based feature selection and Greedy Stepwise search. One reported ILPD study selected Total Bilirubin, Direct Bilirubin, Alkaline Phosphatase, Alanine Aminotransferase, and Aspartate Aminotransferase as relevant attributes. Logistic Regression achieved 74.36% accuracy in that study, while Naive Bayes achieved 55.9%. These results suggest that reducing the feature space may simplify a prediction model, although the effectiveness of feature selection depends on the algorithm, dataset, and evaluation procedure.

Makkena and Natarajan [6] investigated liver cirrhosis diagnosis using machine learning, explainable AI, and cross-validated hyperparameter tuning. Their approach combined predictive modeling with techniques for understanding model behavior. This direction is important because clinical prediction systems should provide more than a disease label; they should also help researchers investigate which patient characteristics contribute to the model's predictions. However, explanations of model behavior do not establish causal relationships, and clinical validity must be assessed independently.

Yu et al. [7] developed an explainable machine learning model for predicting metabolic dysfunction-associated steatotic liver disease (MASLD) using data from the National Health and Nutrition Examination Survey (NHANES). Their study included 13,436 participants and evaluated six machine learning algorithms. Random Forest achieved the strongest discrimination among the evaluated models, and the selected ten-feature model achieved AUC values of 0.928 and 0.918 in internal and external validation, respectively. Recursive Feature Elimination was used for feature selection, while SHAP was applied to interpret predictions. The study demonstrates the value of feature reduction, model interpretability, and external validation. However, its MASLD prediction task differs from the binary classification problem commonly studied with ILPD.

Hong et al. [8] investigated machine learning prediction of metabolic dysfunction-associated fatty liver disease using body-composition measurements and explainable AI. The researchers compared six algorithms and used the Boruta algorithm for feature selection. Gradient Boosting achieved AUC values of 0.875 in training and 0.879 in validation. SHAP analysis identified visceral adipose tissue, body mass index, and subcutaneous adipose tissue as influential predictors. Their findings demonstrate the value of combining clinical and body-composition variables with model interpretability. Nevertheless, the model's findings are specific to its target condition and population and should not be assumed to apply directly to ILPD-based prediction.

Feng et al. [9] conducted a bibliometric analysis of AI-based computer vision methods for liver diseases. Their review covered 3,629 documents published between January 1990 and July 2023 and identified substantial growth in research activity after 2017. Convolutional Neural Networks and U-Net architectures were prominent in liver image classification and segmentation research. This work illustrates how liver disease prediction has expanded beyond tabular clinical data to medical imaging.

However, image-based prediction typically requires different datasets, preprocessing methods, architectures, and evaluation strategies from conventional clinical-data classification.

Chen et al. [10] proposed DeepFLD, a multimodal learning framework for nonalcoholic fatty liver disease prediction. Their approach integrated clinical metadata with facial images and evaluated a dataset containing more than 6,000 participants. The reported findings indicated that multimodal learning performed better than approaches using metadata alone, while facial-image-only prediction also showed competitive potential in the study. This research introduces a promising direction for noninvasive prediction. Nevertheless, multimodal systems require appropriate image datasets, careful validation, and consideration of privacy, fairness, and generalizability.

Donaire et al. [11] examined a hybrid quantum–classical approach to liver disease detection and reviewed classical and quantum machine learning studies using ILPD. This work broadens the methodological landscape beyond conventional classifiers and highlights the growing interest in quantum machine learning for healthcare. However, emerging approaches must be evaluated against established algorithms using consistent datasets, suitable baselines, and reproducible metrics before claims of superiority can be made.

Akther et al. [12] proposed a Streamlit-based AI system for multi-disease prediction involving diabetes, heart disease, breast cancer, and liver disease. Their framework discussed classical machine learning, deep learning, ensemble methods, and explainable AI tools such as LIME and SHAP. The study emphasizes accessibility and the potential integration of predictive models into an interactive application. However, a unified interface does not by itself establish the clinical reliability of its individual disease models; each model requires appropriate independent evaluation and validation.

Kankal and Kasar [13] conducted a systematic review of machine learning techniques for liver disease detection. Their review examined the application of machine learning and deep learning to clinical prediction and discussed the contribution of AI to medical decision support. The review provides a broad methodological foundation for understanding the development of liver disease prediction systems. It also reinforces the need to consider data quality, model selection, evaluation procedures, and the practical limitations of AI-based diagnosis.

Overall, the reviewed studies demonstrate progress in conventional machine learning, ensemble classification, feature selection, class-imbalance handling, explainable AI, medical imaging, and multimodal prediction. However, their results vary according to dataset characteristics, target definitions, and experimental methodology. This motivates a controlled comparison of selected classifiers using the same dataset, leakage-free preprocessing, stratified validation, and comprehensive evaluation metrics. For an ILPD-based study, the most useful contribution would be a reproducible framework that assesses predictive performance and interpretability without assuming that the highest reported accuracy in another study will generalize to the selected dataset.

Problem Statement

Research on liver disease prediction emphasizes early identification to facilitate timely medical evaluation. While machine learning approaches, including ensemble classifiers and deep learning models, show promise, their effectiveness is contingent on data quality, class distribution, feature relevance, model configuration, and validation methods. The Indian Liver Patient Dataset serves as a benchmark but presents challenges due to its limited size and class imbalance. Previous studies have examined various classifiers and feature-selection strategies, yet inconsistencies in experimental design hinder

comparability. This research aims to create a systematic framework for liver disease prediction that integrates data preprocessing, class imbalance handling, relevant feature selection, and hyperparameter optimization. The framework will compare classification algorithms under standardized conditions using multiple metrics, and apply explainable AI techniques to enhance model transparency, ultimately determining the most suitable predictive model while supporting research and screening efforts.

Research Gaps

Research on machine learning and deep learning for liver disease prediction reveals several gaps. First, a lack of standardized evaluation means comparisons across studies are difficult due to differing datasets and methods. Second, class-imbalance issues persist, necessitating systematic comparisons of balancing techniques. Third, effective feature selection is crucial, and further research is needed on its impact on predictive performance. Fourth, complex models often lack interpretability, which needs addressing through explainable methods. Fifth, generalization across populations is essential, as models may not perform well on different datasets. Sixth, reliance solely on accuracy is insufficient; various clinically relevant metrics must be evaluated. Lastly, balancing predictive performance with computational costs is vital, especially with emerging complex methods. Addressing these issues would enhance machine learning's effectiveness in liver disease prediction.

Conclusion

Machine learning is an effective method for predicting liver disease by analyzing clinical data and identifying health patterns. This study examines the implementation of various classification algorithms, including Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Gradient Boosting, and XGBoost, using patient attributes such as age, bilirubin levels, and enzyme measurements. The success of these models hinges on data quality, preprocessing, feature selection, and dataset characteristics.

Key to establishing a dependable prediction model is the systematic assessment of classifiers under consistent conditions, utilizing performance metrics like accuracy, precision, recall, F1-score, ROC-AUC, and confusion matrix. Proper management of missing values, feature scaling, and class imbalance can enhance prediction reliability. The Indian Liver Patient Dataset (ILPD) serves as a basis for evaluating algorithm effectiveness. In conclusion, machine learning can facilitate early screening and data-informed clinical decisions for liver disease. A careful comparison of classifiers reveals the most fitting algorithm for specific datasets. Future advancements should concentrate on enhancing generalization via larger datasets, rigorous validation, and effective feature selection, viewing the developed system as a decision-support mechanism rather than a substitute for medical evaluation.

References

[1] K. Gupta, N. Jiwani, N. Afreen, and D. Divyarani, “Liver disease prediction using machine learning classification techniques,” in Proc. 2022 IEEE 11th Int. Conf. Communication Systems and Network Technologies (CSNT), 2022, pp. 221–226, doi: 10.1109/CSNT54456.2022.9787574.

[2] O. M. Khaled, A. Z. Elsherif, A. Salama, M. Herajy, and E. Elsedimy, “Evaluating machine learning models for predictive analytics of liver disease detection using healthcare big data,” International Journal

of Electrical and Computer Engineering, vol. 15, no. 1, pp. 1162–1174, Feb. 2025, doi: 10.11591/ijece.v15i1.pp1162-1174.

[3] C. Kate, Rajkumari, A. B. C., G. S. Chandra, P. Sudheer, and W. M. Stalin, "A hybrid feature selection and hyperparameter-tuned XGBoost approach for liver disease prediction," in Proc. 5th Int. Conf. Soft Computing for Security Applications (ICSCSA), 2025.

[4] W. El Atifi, O. El Rhazouani, F. M. Khan, and H. Sekkat, "Optimizing ensemble machine learning models for accurate liver disease prediction in healthcare," PLOS ONE, vol. 20, no. 8, Art. no. e0330899, Aug. 2025, doi: 10.1371/journal.pone.0330899.

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

[6] S. S. Kankal and S. Kasar, "A systematic review on machine learning techniques for liver disease detection," Biomedical Materials & Devices, 2025, doi: 10.1007/s44174-025-00341-1.

[7] R. Rani, G. Jaiswal, Nancy, Lipika, S. Bhushan, F. Ullah, P. Singh, and M. Diwakar, "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, Art. no. 1502749, May 2025, doi: 10.3389/fmed.2025.1502749.

[8] G. Zhu et al., "Machine learning models for predicting metabolic dysfunction-associated steatotic liver disease prevalence using basic demographic and clinical characteristics," Journal of Translational Medicine, vol. 23, Art. no. 381, 2025, doi: 10.1186/s12967-025-06387-5.

[9] L. M. Donaire, G. Ortega, F. Orts, E. Martín Garzón, and E. Filatovas, "A hybrid quantum-classical approach for liver disease detection using quantum machine learning," Engineering Applications of Artificial Intelligence, vol. 164, Art. no. 113240, 2026.

[10] Y. Chen et al., "Multimodal learning-based prediction for nonalcoholic fatty liver disease," Machine Intelligence Research, vol. 22, no. 5, pp. 871–887, Oct. 2025, doi: 10.1007/s11633-024-1506-4.

[11] Y. Feng et al., "Application of artificial intelligence-based computer vision methods in liver diseases: A bibliometric analysis," Intelligent Medicine, vol. 5, pp. 111–122, 2025.

[12] Y. Yu, Y. Yang, Q. Li, J. Yuan, and Y. Zha, "Predicting metabolic dysfunction associated steatotic liver disease using explainable machine learning methods," Scientific Reports, vol. 15, Art. no. 12382, 2025, doi: 10.1038/s41598-025-96478-6.
[13] Y. Hong, X. Chen, L. Wang, F. Zhang, Z. Zeng, and W. Xie, "Machine learning prediction of metabolic dysfunction-associated fatty liver disease risk in American adults using body composition:

Explainable analysis based on SHapley Additive exPlanations," Frontiers in Nutrition, vol. 12, Art. no. 1616229, Jun. 2025, doi: 10.3389/fnut.2025.1616229.

[14] F. Akther, M. Begum, T. Mahmud, A. Boltayev, A. Hanip, and M. S. Hossain, "Streamlit-based AI for multi-disease prediction." Publication venue, year, and DOI should be verified from the original PDF before final submission.