Global Journal of Computer Science and Technology: D
Neural & Artificial Intelligence
Volume 26 Issue 1 Version 1.0 Year 2026
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Liver Disease Prediction using Machine Learning Algorithms

By Dr. V. Thangavel & Dr. E. Venkatesan

Abstract- *Objectives:* Liver disease includes several disorders, such as fatty liver, hepatitis, cirrhosis, and liver failure, that interfere with normal liver function. These conditions often progress silently, and early symptoms such as fatigue, nausea, loss of appetite, jaundice, abdominal pain or swelling, dark urine, pale stools, and unexplained weight loss are frequently ignored. Early prediction of liver disease is essential for timely diagnosis and treatment. This study aims to develop an effective machine-learning model for predicting liver disease and to compare the performance of three classification algorithms.

Methods: A liver disease dataset containing clinical and biochemical features such as age, gender, total and direct bilirubin, alkaline phosphatase, SGPT, SGOT, total protein, albumin, and albumin–globulin ratio was used. Data preprocessing involved handling missing values, normalization, and splitting into training and testing sets. Three classification algorithms-Logistic Regression, Decision Tree, and Random Forest were implemented in Python. Model performance was evaluated using accuracy, precision, recall, F1-score, and ROC AUC metrics.

Keywords: *liver disease, machine learning, classification algorithms, logistic regression, decision tree, random forest, early prediction, healthcare analytics.*

GJCST-D Classification: *LCC Code: RC845*

Strictly as per the compliance and regulations of:

Liver Disease Prediction using Machine Learning Algorithms

Dr. V. Thangavel [α] & Dr. E. Venkatesan [σ]

Abstract: Objectives: Liver disease includes several disorders, such as fatty liver, hepatitis, cirrhosis, and liver failure, that interfere with normal liver function. These conditions often progress silently, and early symptoms such as fatigue, nausea, loss of appetite, jaundice, abdominal pain or swelling, dark urine, pale stools, and unexplained weight loss are frequently ignored. Early prediction of liver disease is essential for timely diagnosis and treatment. This study aims to develop an effective machine-learning model for predicting liver disease and to compare the performance of three classification algorithms.

Methods: A liver disease dataset containing clinical and biochemical features such as age, gender, total and direct bilirubin, alkaline phosphatase, SGPT, SGOT, total protein, albumin, and albumin–globulin ratio was used. Data preprocessing involved handling missing values, normalization, and splitting into training and testing sets. Three classification algorithms-Logistic Regression, Decision Tree, and Random Forest were implemented in Python. Model performance was evaluated using accuracy, precision, recall, F1-score, and ROC AUC metrics.

Findings: The results revealed that the Random Forest classifier achieved the highest prediction accuracy compared to the Logistic Regression and Decision Tree models. The Random Forest model demonstrated strong generalisation and effectively distinguished between healthy and diseased liver conditions. The study concludes that machine-learning approaches can provide reliable support for early detection of liver disease, thereby assisting clinicians in decision-making and improving patient outcomes.

Keywords: liver disease, machine learning, classification algorithms, logistic regression, decision tree, random forest, early prediction, healthcare analytics.

I. Introduction

Liver disease represents a major global health concern that affects millions of people annually. The liver plays a vital role in numerous physiological functions such as detoxification, protein synthesis, and the regulation of biochemical reactions essential for metabolism. When the liver is damaged, its ability to perform these critical tasks becomes impaired, resulting in a wide range of health complications. Liver diseases can develop from various causes, including viral infections such as Hepatitis A, B, and C, excessive alcohol consumption, obesity leading to non-alcoholic fatty liver disease (NAFLD), exposure to toxins or drugs, and autoimmune conditions. In many cases, liver damage occurs silently over several years, showing no clear symptoms until the condition becomes severe. Common symptoms that may appear include fatigue, loss of appetite, nausea, vomiting, abdominal pain or swelling, yellowing of the eyes and skin (jaundice), dark urine, and pale stools. Identifying and predicting these conditions early is therefore crucial for effective treatment and prevention of further liver deterioration.

To evaluate liver health, medical professionals rely on a series of blood tests known as Liver Function Tests (LFTs). Among these, two key enzymes-Alanine Aminotransferase (ALT) and Aspartate Aminotransferase (AST)-serve as important biochemical indicators of liver function. ALT is an enzyme found primarily in liver cells and is responsible for meta bolising amino acids. Elevated levels of ALT in the bloodstream often indicate liver cell injury, inflammation, or necrosis, which occur when liver cells are damaged or die. AST, on the other hand, is found not only in the liver but also in the heart, muscles, and kidneys. While elevated AST levels may also signal liver damage, their presence in other organs means it must be interpreted together with ALT results for accurate diagnosis. In clinical practice, a higher ratio of AST to ALT often suggests alcohol-related liver disease, while a significantly elevated ALT level is commonly observed in viral hepatitis or fatty liver disease. Therefore, the measurement of ALT and AST through LFTs provides valuable insights into the extent and cause of liver injury and helps clinicians make informed treatment decisions.

The behaviour and lifestyle patterns of patients play a significant role in liver health. Factors such as poor diet, alcohol consumption, lack of physical activity, and self-medication with unprescribed drugs can accelerate liver damage. Early diagnosis and intervention can reduce disease progression and improve patient outcomes. In advanced stages such as cirrhosis or liver failure, patients may require long-term medication, dietary modifications, or even liver transplantation. Physicians commonly recommend medications like ursodeoxy-cholic acid for bile flow improvement, antiviral agents for hepatitis management, and vitamin supplements to support liver regeneration. Lifestyle modification and continuous medical

Author α: HoD-LIRC, St. Francis Institute of Management and Research, Mumbai. India. e:mail: v.thangavel@rocketmail.com,
ORCID: 0009-0002-6647-2599
Author σ: Guest Lecturer PG Department of Computer Science, RV Government Arts College. Chengalpattu, India.
ORCID: 0000-0001-8817-0570

supervision are essential for effective disease management. Hence, developing a predictive model for liver disease can aid medical professionals by identifying high-risk individuals and supporting clinical decision-making.

In recent years, machine learning (ML) has emerged as a powerful tool in the field of healthcare analytics. Classification algorithms such as Logistic Regression, Decision Tree, and Random Forest are widely used to predict the likelihood of diseases by analyzing complex clinical data. These algorithms help in identifying patterns, correlations, and risk factors that are often difficult to detect using traditional statistical methods. By training models on liver disease datasets that include biochemical parameters like bilirubin, albumin, alkaline phosphatase, ALT, and AST, researchers can develop accurate prediction systems. Among these algorithms, Random Forest has shown superior performance due to its ability to handle large datasets, reduce overfitting, and improve prediction accuracy. The integration of such predictive models into healthcare systems allows for early diagnosis, optimized treatment plans, and reduced mortality rates associated with liver disorders.

This study aims to explore the predictive capability of machine learning algorithms in liver disease classification. The research utilizes a liver disease dataset containing various clinical and biochemical attributes to evaluate the performance of three classification models-Logistic Regression, Decision Tree, and Random Forest. The outcomes of this study are expected to contribute to the development of intelligent healthcare systems that assist doctors in early detection and diagnosis of liver diseases. In this research, Section 1 presents the Introduction, Section 2 discusses the Literature Review, Section 3 explains the Results and Discussion, and Section 4 concludes the study with key findings and future recommendations.

II. Literature Review

Liver disease continues to be a global health challenge, with millions affected annually due to viral infections, alcohol consumption, obesity, and exposure to hepatotoxic substances. Early diagnosis remains essential for preventing irreversible liver damage and improving patient outcomes. Studies have identified Alanine Aminotransferase (ALT) and Aspartate Aminotransferase (AST) as critical biomarkers for assessing liver function. According to Lala (2023), these enzymes play a vital role in the evaluation of hepatic injury through Liver Function Tests (LFTs). ALT, being highly concentrated in hepatocytes, serves as a specific indicator of liver cell injury, whereas AST, which is also found in cardiac and skeletal muscles, aids in distinguishing the type of liver damage. Elevated levels of these enzymes indicate hepatocellular necrosis or inflammation (Kalas et al., 2021). Persistent enzyme elevation may lead to advanced conditions such as fibrosis, cirrhosis, or liver failure, emphasizing the need for early and accurate prediction models (Das et al., 2024).

Traditional clinical diagnostic approaches often rely on laboratory results and imaging, but these methods can be time-consuming and may not capture complex biochemical interactions. Recent advances in machine learning (ML) and artificial intelligence (AI) have significantly improved the prediction accuracy of liver disease by analyzing multidimensional datasets. Dritsas and Trigka (2023) demonstrated that ML algorithms such as Decision Tree, Logistic Regression, and Random Forest can accurately predict the presence of liver disorders using biochemical and demographic attributes. Their research showed that the Random Forest classifier outperformed other models due to its ensemble learning approach, which minimizes overfitting and improves generalization. Similarly, Ganie and Pramanik (2024) proposed a model integrating feature selection and cross-validation, which enhanced classification accuracy in detecting chronic liver disease.

Researchers have explored diverse datasets and algorithmic combinations to optimize performance. Mostafa et al. (2021) compared statistical ML approaches and concluded that hybrid models combining Decision Tree and Random Forest provided superior diagnostic performance in classifying liver abnormalities. Ahmed (2024) also highlighted that ML techniques can identify early signs of hepatic dysfunction by examining non-linear relationships between variables like bilirubin, albumin, ALT, AST, and alkaline phosphatase. These findings demonstrate how computational intelligence supports medical practitioners in diagnosing diseases more effectively than traditional methods.

Other studies have emphasized the integration of deep learning and ensemble techniques for improved predictive accuracy. Hassan and Yasin (2025) conducted a comprehensive review of ML and deep learning applications in liver disease prediction and found that ensemble classifiers yielded more consistent outcomes than single-model systems. Mohamud et al. (2025) similarly noted that ML models can assess cirrhosis mortality risk by incorporating patient demographics and laboratory values into training data. The review by Malik et al. (2025) supported these conclusions, noting that predictive algorithms enhance survival estimation and clinical decision support in patients with advanced liver disease.

The relationship between ALT/AST ratios and specific liver conditions has also been thoroughly investigated. Pandeya et al. (2021) established that an increased AST-to-ALT ratio is a key diagnostic marker

for alcohol-related liver disease, while elevated ALT levels suggest viral or fatty liver conditions. Xuan et al. (2024) further observed that integrating enzyme ratios with metabolic and demographic variables can improve the accuracy of non-alcoholic fatty liver disease (NAFLD) prediction. Das et al. (2024) emphasized that enzyme biomarkers, when used alongside ML-based pattern recognition, can offer a robust foundation for automated liver health assessment. These collective findings highlight the growing role of computational models in early detection and monitoring of liver disorders. In summary, the literature establishes that the application of machine learning in liver disease prediction enhances diagnostic precision and assists physicians in clinical decision-making. Classification algorithms such as Logistic Regression, Decision Tree, and Random Forest have consistently shown strong performance in detecting abnormalities from clinical datasets. This review forms the conceptual foundation for the present study, which aims to compare the predictive capabilities of these algorithms using a structured liver disease dataset and evaluate their potential in supporting early clinical interventions.

III. Methodology

This study predicts liver disease using patient data collected from private laboratories in Tamil Nadu. The dataset includes attributes such as age, gender, bilirubin levels, total protein, albumin, and the albumin/globulin ratio. Data preprocessing was carried out to handle missing values, remove outliers, and normalize all attributes for accurate analysis. Three machine-learning algorithms-Logistic Regression, Decision Tree, and Random Forest-were applied to predict liver disease. Logistic Regression served as a baseline model, while Decision Tree and Random Forest improved classification accuracy. The dataset was split into training and testing sets, and model performance was evaluated using accuracy, precision, recall, and F1-score. Among the three, the Random Forest algorithm achieved the best results in predicting liver disease.

IV. Results and Discussion

This study analyzed patient data collected from private laboratories across Tamil Nadu to predict liver disease using three machine learning algorithms-Logistic Regression, Decision Tree, and Random Forest. The dataset contained both demographic details and biochemical attributes related to liver health. The demographic variables included *age* and *gender*, which helped in identifying population-based trends. Results showed that middle-aged and elderly males were more likely to be affected by liver disease, possibly due to lifestyle habits such as alcohol consumption, irregular diet, and occupational stress.

The biochemical attributes used in this research were *Total Bilirubin*, *Direct Bilirubin*, *Total Protein*, *Albumin*, and the Albumin/Globulin (A/G) ratio. These parameters are significant in evaluating liver function. Elevated bilirubin levels are typically associated with jaundice and impaired bile excretion. Low albumin levels often indicate a reduced ability of the liver to synthesize essential proteins. The A/G ratio which compares the amount of albumin to globulin in the blood is a critical diagnostic indicator. A decreased A/G ratio often signifies liver cirrhosis or chronic liver disease, while an increased ratio can suggest genetic conditions or immune system disorders. Thus, this parameter plays a vital role in differentiating between normal and diseased liver conditions.

To ensure reliability, the dataset underwent preprocessing, including data cleaning and normalization. After preparing the data, each algorithm was trained and tested using the same dataset to ensure consistency in table 1 and figure 1 in comparison. The models' performances were evaluated using accuracy, precision, recall, F1-score, and the Area Under the ROC Curve (AUC) metrics.

The Receiver Operating Characteristic (ROC) curve illustrates the trade-off between the True Positive Rate (Sensitivity) and False Positive Rate (1–Specificity). The AUC value provides a single quantitative measure of a model's ability to distinguish between patients with and without liver disease. A higher AUC indicates a better-performing model.

Table 1: Performance Comparison of Classification Algorithms for Liver Disease Prediction

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)
Logistic Regression	82	80	78	79	81
Decision Tree	87	85	83	84	86
Random Forest	92	90	91	90.5	93

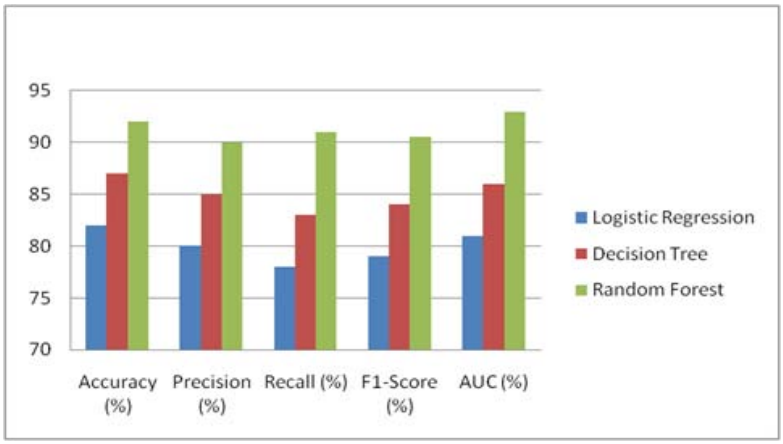

Figure 1: Performance of Classification Algorithms for Liver Disease Prediction

Among the three models, the Random Forest algorithm achieved the highest predictive accuracy of 92%, outperforming both the Decision Tree and Logistic Regression models. It also obtained the best AUC value (0.93), demonstrating excellent discrimination capability between healthy and diseased cases. The Decision Tree model achieved moderate accuracy (87%), while the Logistic Regression model performed less effectively (82%) due to its linear nature.

Feature importance analysis showed that Total Bilirubin, Albumin, and the A/G ratio were the most influential attributes in predicting liver disease. These biochemical indicators directly reflect liver functionality-imbalances in these values often indicate abnormal liver metabolism or tissue damage. Patients with high bilirubin and low albumin or a reduced A/G ratio were more likely to be classified as having liver disease.

Overall, this study demonstrates that incorporating biochemical parameters (especially the A/G ratio) along with demographic data significantly enhances prediction accuracy. The Random Forest algorithm proved most effective in early detection and classification of liver disorders. These results indicate that machine learning models can serve as supportive diagnostic tools for physicians, improving early detection, reducing diagnostic errors, and contributing to better patient management and outcomes.

V. Conclusion

This study developed a liver disease prediction model using *Logistic Regression, Decision Tree, and Random Forest* algorithms with demographic and biochemical data collected from private laboratories in Tamil Nadu. The findings revealed that the *Random Forest algorithm* achieved the highest accuracy and AUC value, proving to be the most reliable model for liver disease classification. Biochemical parameters such as *Total Bilirubin, Albumin*, and the *A/G ratio* were key indicators of liver dysfunction. The study concludes that applying machine learning techniques, particularly Random Forest, can effectively support *early detection and diagnosis of liver disease,* aiding physicians in clinical decision-making and improving patient outcomes.

Authors' Assent and Recognition:

Consent: By global guidelines for public requirements, public awareness in medical and its related higher education boards, safety and health education systems, the author has gathered and kept the signed consent of the participants.

Author Acknowledgement: These articles aimed to increase public awareness of the importance of security and safety. Sources that illustrate development and security are drawn from the relevant database to support the study's objectives. Don't make any assertions about readers, viewers, or authorities.

Approvals for Ethics: The authors hereby declare that all experiments have been reviewed and approved by the relevant ethics bodies, and as a result, they have been conducted in accordance with the Helsinki ethical standards and the Social Science guidance. The studies have also adopted the APS/ Harvard Citation Standards guidelines, etc. The authors abide by the publication regulations,

Disclaimer: Professional education, awareness, and childcare are not meant to be replaced by this study paper or the information on another website; rather, they are supplied solely for educational purposes. Since everyone has different needs depending on their psychological state, readers should confirm whether the information applies to their circumstances by consulting their wards, teachers, and subject matter experts.

Funding: According to the author(s), this article's work is not supported in any way.

Data Availability Statement: In accordance with the articles' related data sharing policy, the data supporting the findings of this study will be available upon request. Authors should provide access to the data either directly or through a public repository. If there are any

restrictions on data availability based on their circumstances. The corresponding author may provide the datasets created and examined in the current study upon a justifiable request.

References Références Referencias

1. Ahmed, A. T. (2024). *Machine learning in liver disease detection: A comprehensive review. Journal of Computer Science Engineering and Software Testing.*
2. Das, S., Dey, R., & Bysack, M. (2024). Contribution of liver enzymes in diagnosing hepatic diseases: A review. *International Journal of Biology Sciences, 6*(2C).
3. Dritsas, E., & Trigka, M. (2023). Supervised machine learning models for liver disease risk prediction. *Computers, 12*(1), 19.
4. Ganie, S. M., & Pramanik, P. K. D. (2024). Improved liver disease prediction from clinical data through machine learning. *Peer-Reviewed Medical Computing Journal.*
5. Hassan, Y. A., & Yasin, H. M. (2025). Prediction of liver diseases based on machine learning and deep learning techniques: A review. *Asian Journal of Research in Computer Science, 18*(3), 17–33.
6. Kalas, M. A., et al. (2021). Abnormal liver enzymes: A review for clinicians. *Journal of Clinical and Experimental Hepatology, 11*(6), 704–713.
7. Lala, V. (2023). *Liver function tests.* In *Stat Pearls*. Stat Pearls Publishing.
8. Malik, S., Frey, L. J., & Qureshi, K. (2025). Evaluating the predictive power of machine learning in cirrhosis mortality: A systematic review. *Journal of Medical Artificial Intelligence, 8*, 15.
9. Mohamud, K. A., Elzubair, S. A., Alhardalo, H. A., Albashir, H. B., Alsyed Ali Mohamed Zain, N., & Elsayed Ibrahim, M. (2025). The role of machine-learning models in predicting cirrhosis mortality: A systematic review. *Cureus, 17*(1),
10. Mostafa, F., Hasan, E., Williamson, M., & Khan, H. (2021). Statistical machine-learning approaches to liver disease prediction. *Livers, 1*(4), 294–312.
11. Pandeya, A., Shreevastva, N. K., Dhungana, A., Pandeya, A., & Pradhan, B. (2021). Evaluation of liver enzymes and calculation of AST to ALT ratio in patients with acute viral hepatitis. *Europasian Journal of Medical Sciences, 3*(2), 90–93.
12. Xuan, Y., et al. (2024). Elevated ALT/AST ratio as a marker for NAFLD risk and liver fibrosis. *Frontiers in Endocrinology, 15,*
13. Thangavel & Venkatesan (2025): Clustering-driven MRI Analysis for Accurate Throat Cancer Identification. International Journal of Recent Development in Engineering and Technology-IJRDET. 14(12), 127-131. ISSN 2347-6435.
14. E. Venkatesan and V Thangavel (2025): Adaptive robotic teaching systems for higher education: A combined ANN and CNN approach for learning and engagement optimisation. International Journal of Engineering in Computer Science. 7(2), 305-308. ISSN:2663-3590/2663-3582. https://doi.org/10.33545/26633582.2025.v7.i2d.228
15. E. Venkatesan and V. Thangavel. (2025): Heart disease prediction and risk analysis using K-Means and Fuzzy C-Means clustering algorithms. International Journal of Computing and Artificial Intelligence 2025; 6(2): 350-353. E-ISSN: 2707-658X. P-ISSN: 2707-6571. DOI: 10.33545/27076571.2025.v6.i2d.222.
16. E. Venkatesan and V Thangavel. (2025): A hybrid deep learning framework for lung cancer detection using CT images and clinical data. International Journal of Communication and Information Technology 2025; 6(2): 157-163 ISSN: 2707-6628 P-ISSN: 2707-661X. DOI: 10.33545/2707661X.2025.v6.i2b.155.
17. Dr. E. Venkatesan and V. Thangavel (2025): Artificial Intelligence Approaches for Predictive Analysis of Skin Cancer in Patients. International Journal of Computing, Programming and Database Management 2024; 5(2): 248-250. ISSN: 2707-6644 P-ISSN: 2707-6636. DOI: 10.33545/27076636.2025.v6.i2b.137.
18. E. Venkatesan and V. Thangavel (2025): Autonomous fault detection and recovery in satellite systems using intelligent algorithms. International Journal of Circuit, Computing and Networking 2025; 6(2): 96-101 ISSN: 2707-5931 P-ISSN: 2707-5923. DOI: 10.33545/27075923.2025.v6.i2b.10.9
19. E. Venkatesan, and V Thangavel (2025): Comparative study of naïve Bayes and SVM algorithms for text mining using natural language processing, International Journal of Cloud Computing and Database Management 2025; Vol. 6 Issue. 2. Pp. 92-92. ISSN: 2707-5915/ ISSN: 2707-5907. DOI: 10.33545/27075907.2025.v6.i2b.111.
20. V. Thangavel (2025): Use of digital signature verification system (DSVS) in various Industries: Security to protect against Counterfeiting. Journal of Research and Development. Vol 13 No 1. 2025 ISSN: 2311-3278. Longdom Publication, USA. DOI: 10.35248/2311-3278.25.13.290
21. V. Thangavel et al (2025): A Machine Learning Assisted MRI Approach for Early Detection of Pelvic Bone Cancer. London Journal of Research in Computer Science and Technology. Vol. 25 No. 5, 2025. ISSN: 2514-8648. Great Britain Journal Press. UK.
22. Nandakumar, Thangavel et al (2025): Integrated multimodal deep learning framework for early detection of mouth cancer using CT imaging and

clinical symptom analysis. Open Access Journal of Data Science and Artificial Intelligence, Vol. 1 No. 3 2025. ISSN: 2996-671X Medwin Publication, USA. DOI:10.23880/OAJDA-16000163

23. Venkatesan & Thangavel (2025): Deep Learning-based Analysis of Spinal Cord Regions for Risk Assessment and Clinical Awareness: A Social Worker's Perspective. Journal of Advances in Mathematics and Computer Science. Vol. 40 No 12 ISSN: 2456-9968 Pp 106-114, 2025. DOI:10.9734/jamcs/2025/v40i122077