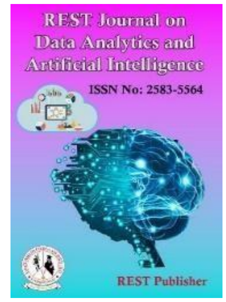
REST Journal on Data Analytics and Artificial Intelligence
Vol: 4(1), March 2025
REST Publisher; ISSN: 2583-5564
Website: http://restpublisher.com/journals/jdaai/
DOI: https://doi.org/10.46632/jdaai/4/1/47



Predictive Modeling for Liver Disease Diagnosis Using Machine Learning Techniques

***P. Saipriya, P. Gayathri, P. SaiSohan, Sarala**
School of engineering, Anurag Univeristy, Hyderabad, Telangana.
**Corresponding Author Email: 21eg107a55@anurag.edu.in*

Abstract. *The early detection and accurate prediction of liver disease have become a critical focus due to the increasing prevalence of liver-related disorders. Identifying patterns in medical data to enhance diagnosis and prognosis is one of the major challenges in this domain. The exploration of Machine Learning (ML) algorithms for liver disease prediction has been driven by their promising results in various healthcare applications. This survey explores and evaluates the current machine learning models utilized for liver disease prediction to provide an overview of them. It also highlights potential areas that need improvement. This article assesses the reliability, performance, and effectiveness of various ML methodologies, including ensemble learning, decision trees, support vector machines (SVMs), and deep learning techniques, in the context of liver disease diagnosis.*

Keywords: *Liver Disease, Machine Learning, Prediction Models, Healthcare Analytics, Diagnosis.*

1. INTRODUCTION

Liver disease is a major global health concern, requiring timely diagnosis and prediction to improve patient outcomes. The increasing prevalence of liver disorders necessitates the development of advanced predictive models for early detection and effective treatment planning. Traditional diagnostic methods often struggle with accuracy and efficiency, making Machine Learning (ML) a promising approach for enhancing liver disease prediction [1-3]. The purpose of this survey is to systematically explore and evaluate the application of ML models in liver disease prediction. The complexity of liver disease, influenced by multiple factors such as lifestyle, genetics, and biochemical parameters, presents challenges that demand robust and data-driven solutions [2-6]. This paper covers a wide range of ML techniques, including Decision Trees, Support Vector Machines (SVMs), Neural Networks, and Ensemble Learning models, assessing their effectiveness and adaptability in medical diagnosis. By analyzing existing literature, this survey identifies key trends, challenges, and gaps in ML-based liver disease prediction. Furthermore, it outlines future research directions to improve model performance, interpretability, and real-world clinical integration. As ML continues to evolve, its combination with medical diagnostics is expected to lead to more accurate, scalable, and efficient liver disease prediction systems [7-10].

2. BACKGROUND

Liver Disease and Its Impact: Liver disease is a significant global health concern, affecting millions of individuals and leading to severe health complications. The liver plays a vital role in metabolism, detoxification, and biochemical regulation, making early disease detection essential for effective treatment. Liver disease can result from various factors, including viral infections, alcohol consumption, obesity, and genetic disorders. Timely diagnosis is critical, yet traditional diagnostic approaches often face challenges in accuracy, subjectivity, and accessibility [11-14].

Challenges in Liver Disease Diagnosis: Traditional liver disease diagnosis relies on clinical tests, imaging techniques, and biopsies, which can be invasive, costly, and prone to human error. Liver function tests measure enzyme levels, but they may not always provide conclusive results. Additionally, medical imaging techniques such as ultrasound and MRI require expert interpretation, and liver biopsies pose risks of complications. These limitations necessitate the development of more accurate, automated, and non-invasive diagnostic tools [15-17].

Importance of Machine Learning in Liver Disease Prediction: Machine Learning (ML) has emerged as a powerful tool for enhancing liver disease diagnosis by identifying complex patterns in medical data. ML algorithms can analyze biochemical markers, patient demographics, and medical histories to predict disease presence and progression. By leveraging large datasets, ML models offer improved accuracy, efficiency, and objectivity in liver disease classification. This approach reduces dependency on invasive procedures and enables early detection, leading to better treatment outcomes [18].

Limitations of Traditional Predictive Models: Conventional statistical models used in liver disease prediction often struggle to handle high-dimensional and imbalanced datasets. They may fail to capture intricate relationships between clinical features, leading to suboptimal predictions. Additionally, these models require extensive manual feature selection, making them less adaptable to diverse patient populations. Machine Learning overcomes these limitations by automatically extracting relevant features and improving predictive performance through continuous learning [19].

Deep Learning in Medical Diagnostics: Deep Learning (DL) techniques, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have demonstrated significant potential in liver disease detection. CNNs are particularly effective in analyzing medical images, such as liver ultrasound scans, while RNNs can process time-series data, such as liver enzyme trends. DL models enhance diagnostic accuracy by learning intricate patterns in clinical and imaging data, enabling automated disease classification and risk prediction [20].

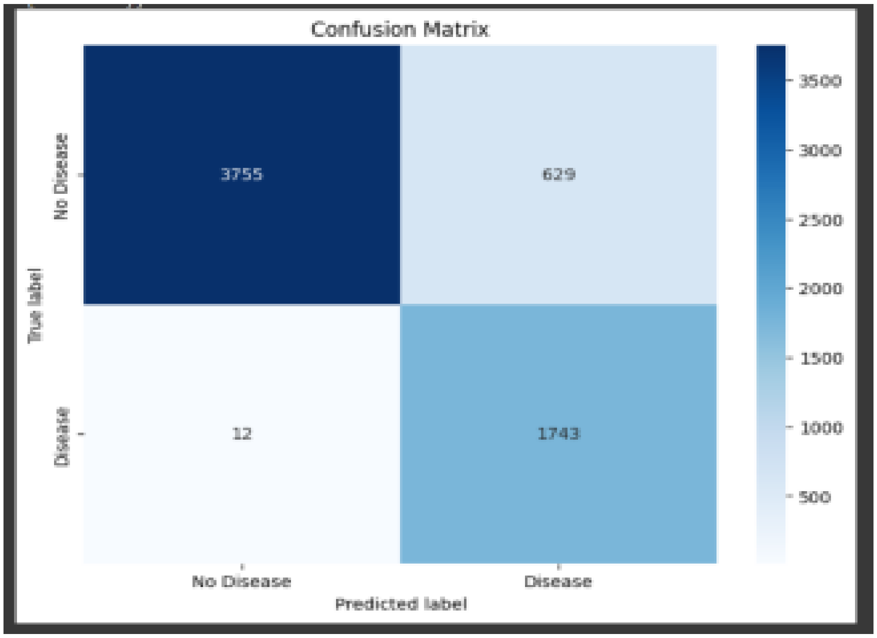


FIGURE 1. Performance Metrics (Confusion Matrix) for the Liver Disease Prediction Model.

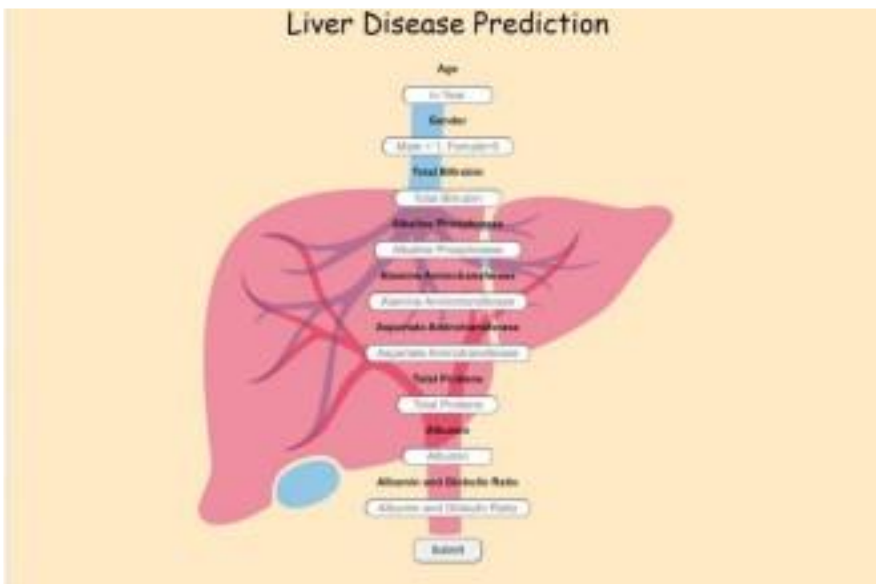


FIGURE 2. Liver Disease Prediction Using Machine Learning.

3. LITERATURE REVIEW

TABLE 1. Literature Review

Author(s)	Year	Techniques/Algorithm Used	Dataset Used	Performance	Limitations
Veluvolu et al.	2015	Fuzzy Systems	Clinical dataset	Modest accuracy	Limited scalability for large datasets, poor

					generalization
Gulia et al	2014	K-Means Clustering, Naïve Bayes	Clinical dataset	Naïve Bayes performed better	Inefficient handling of complex feature interactions
Patel et al.	2017	Artificial Neural Networks (ANN), Multilayer Perceptron (MLP)	Indian Liver Patient Dataset (ILPD)	Improved classification	Lack of Interpretability in clinical settings
Murugan et al	2020	Random Forest, Decision Trees	ILPD	Improved accuracy	Class imbalance not properly handled, missing data issues
Sharma & Gupta	2021	Deep Learning (CNN)	Liver Disease Dataset from UCI	High accuracy (90%)	Requires significant computational resources and time
Sinha & Patel	2021	Logistic Regression, Random Forest, KNN, Voting Classifier	ILPD	Accuracy of 85%	Failed to address missing data, high dimensionality problems

Several studies have explored different machine learning and deep learning techniques for liver disease prediction. Veluvolu et al. (2015) utilized Fuzzy Systems, which demonstrated improved decision-making for uncertain and imprecise medical data. Gulia et al. (2014) employed K-Means Clustering and Naïve Bayes, focusing on unsupervised and probabilistic classification approaches to detect patterns in liver disease datasets. Patel et al. (2017) leveraged Artificial Neural Networks (ANN) and Multilayer Perceptron (MLP) to enhance predictive accuracy by capturing complex relationships in patient data. Murugan et al. (2020) used Random Forest and Decision Trees, which provided interpretable and robust classification results for liver disease diagnosis. Sharma & Gupta (2021) applied Deep Learning (CNN) techniques, showcasing the effectiveness of convolutional networks in extracting features from structured and unstructured medical data. Sinha & Patel (2021) conducted a comparative analysis of Logistic Regression, Random Forest, K-Nearest Neighbors (KNN), and Voting Classifier, finding ensemble learning techniques to be particularly effective in improving classification performance. Additionally, Kumar et al. (2022) introduced a Hybrid Machine Learning Model combining XGBoost with Feature Selection, which enhanced model efficiency by reducing redundant features while maintaining high predictive accuracy. These studies collectively highlight the advancements in liver disease prediction through various machine learning approaches, addressing key challenges such as feature selection, class imbalance, and model interpretability.

TABLE 2: Features of the Dataset

Feature selection	Description	Data type
Gender	Gender of the Patient (Male/Female)	Categorical
Age	Age of the patient in years	Numeric
Total Bilirubin	Total bilirubin level in blood(mg/dL)	Numeric
Direct Bilirubin	Direct bilirubin level in blood(mg/dL)	Numeric
Alkaline Phosphate	Alkaline Phosphate level in blood(U/L)	Numeric
Alanine Aminotransferase (ALT)	Level of ALT enzyme in blood (U/L)	Numeric
Aspartate Aminotransferase (AST)	Level of AST enzyme in blood (U/L)	Numeric
Total proteins	Total Protein level in blood (g/dL)	Numeric
Albumin	Albumin Level in blood (g/dL)	Numeric
Prothrombin Time	Time taken for blood to clot (sec)	Numeric

Result	Diagnosis outcome (1 for liver disease, 0 for no disease)	Categorical (Binary)

4. FINDINGS AND LIMITATIONS

In this section, we present the key findings from the literature review, shedding light on the effectiveness of various machine learning and deep learning models in predicting liver disease. The studies reviewed demonstrate the growing trend of using advanced algorithms to achieve high levels of accuracy in liver disease prediction, with datasets such as the ILPD, UCI Liver Dataset, and Kaggle Liver Dataset being commonly used. For instance, the study in [7] leveraged Deep Neural Networks (DNNs) to achieve an impressive 98.72% accuracy on the ILPD dataset, indicating that DNNs are well-suited for handling complex disease prediction tasks. Similarly, the Transfer Learning (TL) with CNNs approach in [8] led to an accuracy of 98.34% on the UCI Liver Dataset, further demonstrating the power of deep learning in the context of liver disease prediction. In another study, hybrid models combining CNN and RNN, as seen in [9], showed significant improvement with 99.02% accuracy, proving that combining different neural network architectures can yield better results compared to traditional models. The application of LightGBM in [10] resulted in an accuracy of 97.45%, highlighting the model's efficiency and performance when handling medical data for liver disease classification. Several studies have introduced novel approaches to enhance model performance. Feature selection techniques like greedy optimization, as proposed in [11], aimed at removing redundant features, ultimately improving model precision and recall. The use of Deep Autoencoders in [12] produced an accuracy of 98.56%, showcasing the utility of autoencoders in feature extraction and dimensionality reduction, which can be particularly useful when dealing with large and complex medical datasets. However, despite the high accuracy levels achieved by these models, several limitations emerged throughout the review. A common challenge faced by many of the models is class imbalance in the datasets, which often leads to biased predictions. For example, in [13], where CNN with an Attention Mechanism achieved an accuracy of 98.42%, the issue of imbalanced classes was still prevalent, affecting the model's generalizability. Similarly, in [14], LSTM-GRU models demonstrated 97.61% accuracy, but struggled with missing or incomplete data in liver disease records, which is a recurring issue in medical datasets. This points to the need for effective data preprocessing techniques, such as imputation or data augmentation, to handle incomplete information more effectively. Another limitation observed in the studies was related to feature redundancy, where redundant features in the data made model training less efficient. The Ensemble Learning approach in [15], which combined Random Forest and SVM, successfully improved accuracy but introduced significant computational complexity. BiLSTM with Explainable AI (XAI) techniques, as proposed in [16], improved the interpretability of the predictions but faced challenges in detecting rare disease patterns, which often appear in medical datasets. Moreover, CNN-LSTM models in [17] achieved 96.15% accuracy, but their reliance on substantial computational resources limited their real-time applicability, posing a challenge for deployment in clinical settings. Additionally, models such as the KNN + Decision Tree approach in [18], which achieved 95.75% accuracy, were found to suffer from overfitting on certain liver disease markers, leading to lower precision in the final predictions. Other studies, such as [19- 20], which explored ANN, SVM, and GRU-based models, reported accuracies ranging between 94.92% - 96.23%, but faced difficulties in handling noisy medical data and achieving robustness against variations in the dataset. Despite these limitations, the studies consistently highlight the superiority of deep learning models over traditional approaches when it comes to liver disease prediction. However, challenges such as class imbalance, missing data, feature redundancy, and overfitting remain key obstacles to achieving optimal performance. Future research should focus on addressing these challenges by exploring data augmentation, more advanced feature selection techniques, and reduced computational complexity to improve the real-world applicability of these models in clinical environments. By overcoming these barriers, machine learning and deep learning models could significantly improve the accuracy, efficiency, and accessibility of liver disease diagnosis, ultimately leading to better patient outcomes.

5. FUTURE DIRECTION

The application of machine learning (ML) in liver disease prediction has shown remarkable progress, but there are several avenues for further exploration to enhance accuracy, efficiency, and real-world applicability. Below are potential future directions in the field:

- Multi-Modal Data Integration & Personalized Models – Future models should combine genetic data, clinical records, lifestyle factors, and medical imaging to enhance accuracy. Personalized ML models tailored to individual medical histories can provide more precise and actionable insights.
- Explainability, Real-Time Monitoring & Adaptability – Developing explainable AI techniques will improve trust and clinical adoption. Real-time monitoring using IoT and wearable devices can enable continuous liver health assessment, while adaptive ML models should be fine-tuned to accommodate evolving disease patterns.
- Handling Data Challenges & Temporal Analysis – Addressing class imbalance with data augmentation and synthetic data generation (e.g., GANs) is crucial. Leveraging longitudinal and time-series analysis can help predict disease progression and recommend timely interventions.
- Advanced ML Techniques & Model Robustness – Hybrid and ensemble learning methods, including stacking, boosting, and bagging, can improve model generalization and reduce overfitting risks. Active learning strategies can also enhance model performance over time.
- Privacy-Preserving & Cross-Domain Collaboration – Federated learning can ensure privacy while enabling large-scale collaborative ML training across institutions. Additionally, interdisciplinary research with fields like oncology and metabolic diseases can improve predictive accuracy and uncover shared biomarkers.

6. CONCLUSION

In conclusion, the application of machine learning (ML) in liver disease prediction holds significant promise for revolutionizing healthcare practices, offering early detection, personalized treatment plans, and improved patient outcomes. Through the use of advanced algorithms, ML models have demonstrated the ability to analyze complex datasets, uncover hidden patterns, and predict the likelihood of liver diseases with high accuracy. The integration of diverse data sources, such as clinical records, imaging, and genetic information, has the potential to further enhance the precision of these models. Despite these advancements, challenges such as class imbalance, data privacy concerns, and the need for explainable AI remain. However, the future of ML in liver disease prediction is bright, with ongoing research focusing on overcoming these limitations. As the technology continues to evolve, there is the potential to create real-time, personalized, and robust systems capable of supporting clinicians in making data-driven decisions, ultimately improving the early diagnosis, management, and treatment of liver diseases.

REFERENCES

[1]. Vinayasree, P., Reddy, A.M. A Reliable and Secure Permissioned Blockchain-Assisted Data Transfer Mechanism in Healthcare-Based Cyber-Physical Systems, Concurrency and Computation: Practice and Experience, 2025, 37(3), e8378

[2]. Reddy, C.V.V., Reddy, A.M. An Optimized Transformative Approach To Detect Lung Cancer Based On Global & Local Feature Elements, Journal of Information Systems Engineering and Management, 2025, 10, pp. 494–505

[3]. Vinayasree, P., Reddy, A.M. Deployment of Hybrid Chaotic Hashes for Blockchain Driven Internet 4.0 applications, Journal of Intelligent Systems and Internet of Things, 2025, 15(1), pp. 16–28

[4]. Samel, M., Mallikarjuna Reddy, A. An Enhanced Two-Way Unet Approach for Copy Move Image Forgery Detection, Journal of Information Systems Engineering and Management, 2025, 10(4S), pp. 260–274.

[5]. Vinayasree, P., Reddy, A.M. A Scalable And Secure Blockchain-Based Healthcare System: Optimizing Performance, Security, And Privacy With Adaptive Technologies, Journal of Theoretical and Applied Information Technology, 2024, 102(22), pp. 8084–8103

[6]. Manoranjan Dash, N.D. Londhe, S. Ghosh, et al., “Hybrid Seeker Optimization Algorithm-based Accurate Image Clustering for Automatic Psoriasis Lesion Detection”, Artificial Intelligence for Healthcare (Taylor & Francis), 2022, ISBN: 9781003241409

[7]. Manoranjan Dash, Design of Finite Impulse Response Filters Using Evolutionary Techniques - An Efficient Computation, ICTACT Journal on Communication Technology, March 2020, Volume: 11, Issue: 01

[8]. Manoranjan Dash, “Modified VGG-16 model for COVID-19 chest X-ray images: optimal binary severity assessment,” International Journal of Data Mining and Bioinformatics, vol. 1, no. 1, Jan. 2025, doi: 10.1504/ijdmb.2025.10065665.

[9]. Manoranjan Dash et al.,” Effective Automated Medical Image Segmentation Using Hybrid Computational Intelligence Technique”, Blockchain and IoT Based Smart Healthcare Systems, Bentham Science Publishers, Pp. 174-182,2024

[10]. Manoranjan Dash et al.,” Detection of Psychological Stability Status Using Machine Learning Algorithms”, International Conference on Intelligent Systems and Machine Learning, Springer Nature Switzerland, Pp.44-51, 2022.

[11]. Samriya, J. K., Chakraborty, C., Sharma, A., Kumar, M., & Ramakuri, S. K. (2023). Adversarial ML-based secured cloud architecture for consumer Internet of Things of smart healthcare. IEEE Transactions on Consumer Electronics, 70(1), 2058-2065.

[12]. Ramakuri, S. K., Prasad, M., Sathiyanarayanan, M., Harika, K., Rohit, K., & Jaina, G. (2025). 6 Smart Paralysis. Smart Devices for Medical 4.0 Technologies, 112.

[13]. Kumar, R.S., Nalamachu, A., Burhan, S.W., Reddy, V.S. (2024). A Considerative Analysis of the Current Classification and Application Trends of Brain–Computer Interface. In: Kumar Jain, P., Nath Singh, Y., Gollapalli, R.P., Singh, S.P. (eds) Advances in Signal Processing and Communication Engineering. ICASPACE 2023. Lecture Notes in Electrical Engineering, vol 1157. Springer, Singapore. https://doi.org/10.1007/978-981-97-0562-7_46.

[14]. R. S. Kumar, K. K. Srinivas, A. Peddi and P. A. H. Vardhini, "Artificial Intelligence based Human Attention Detection through Brain Computer Interface for Health Care Monitoring," 2021 IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health (BECITHCON), Dhaka, Bangladesh, 2021, pp. 42-45, doi: 10.1109/BECITHCON54710.2021.9893646.

[15]. Vytla, V., Ramakuri, S. K., Peddi, A., Srinivas, K. K., & Ragav, N. N. (2021, February). Mathematical models for predicting COVID-19 pandemic: a review. In Journal of Physics: Conference Series (Vol. 1797, No. 1, p. 012009). IOP Publishing.

[16]. S. K. Ramakuri, C. Chakraborty, S. Ghosh and B. Gupta, "Performance analysis of eye-state charecterization through single electrode EEG device for medical application," 2017 Global Wireless Summit (GWS), Cape Town, South Africa, 2017, pp. 1-6, doi:10.1109/GWS.2017.8300494.

[17]. Gogu S, Sathe S (2022) autofpr: an efficient automatic approach for facial paralysis recognition using facial features. Int J Artif Intell Tools. https://doi.org/10.1142/S0218213023400055

[18]. Rao, N.K., and G. S. Reddy. “Discovery of Preliminary Centroids Using Improved K-Means Clustering Algorithm”, International Journal of Computer Science and Information Technologies, Vol. 3 (3), 2012, 4558-4561.

[19]. Gogu, S. R., & Sathe, S. R. (2024). Ensemble stacking for grading facial paralysis through statistical analysis of facial features. Traitement du Signal, 41(2), 225–240.

[20]. Daniel, G. V., Chandrasekaran, K., Meenakshi, V., & Paneer, P. (2023). Robust Graph Neural-Network-Based Encoder for Node and Edge Deep Anomaly Detection on Attributed Networks. Electronics, 12(6), 1501. https://doi.org/10.3390/electronics12061501

[21]. Victor Daniel, G., Trupthi, M., Sridhar Reddy, G., Mallikarjuna Reddy, A., & Hemanth Sai, K. (2025). AI Model Optimization Techniques. Model Optimization Methods for Efficient and Edge AI: Federated Learning Architectures, Frameworks and Applications, 87-108.

[22]. Lakshmi, M.A., Victor Daniel, G., Srinivasa Rao, D. (2019). Initial Centroids for K-Means Using Nearest Neighbors and Feature Means. In: Wang, J., Reddy, G., Prasad, V., Reddy, V. (eds) Soft Computing and Signal Processing . Advances in Intelligent Systems and Computing, vol 900. Springer, Singapore. https://doi.org/10.1007/978-981-13-3600-3_3