International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online) Volume 14, Issue 8, August 2025)

Review of Liver Diseases Detection using Machine Learning Techniques

[1]Shreya Anurag, [2]Dr. Kamlesh Namdev
[1]M.Tech Scholar, Department of CSE, Lakshmi Narain College of Technology, Bhopal, India
[2]Professor, Department of CSE, Lakshmi Narain College of Technology, Bhopal, India

***Abstract*— This review explores the application of machine learning techniques in the detection and diagnosis of liver diseases, highlighting how various algorithms improve accuracy and efficiency in identifying conditions such as hepatitis, cirrhosis, and liver cancer. By analyzing recent studies, it discusses the use of classification methods like support vector machines, decision trees, neural networks, and ensemble models on medical imaging and clinical data. The review emphasizes the potential of machine learning to support early diagnosis, enhance treatment planning, and reduce the burden on healthcare systems, while also addressing challenges such as data quality, model interpretability, and the need for large, diverse datasets.**

Keywords*— Liver, Diseases, Machine Learning, Cancer, Prediction model, Accuracy, Deep learning.

I. INTRODUCTION

Liver diseases represent a significant global health challenge due to their high prevalence and potential to cause severe complications, including liver failure and cancer. The liver, being a vital organ responsible for multiple essential functions such as detoxification, metabolism, and protein synthesis, is susceptible to a variety of diseases caused by infections, lifestyle factors, genetics, and environmental exposures. Common liver conditions include hepatitis (inflammation of the liver), fatty liver disease, cirrhosis (scarring of liver tissue), and hepatocellular carcinoma (liver cancer). Early and accurate prediction of these diseases is crucial for effective treatment, improving patient outcomes, and reducing mortality rates.

Traditional diagnostic methods for liver diseases often involve invasive procedures like liver biopsies, alongside imaging techniques such as ultrasound, CT scans, and MRI. While these approaches provide valuable information, they may be expensive, time-consuming, or carry risks for patients. Additionally, clinical diagnosis based solely on symptoms and standard blood tests may not always detect liver conditions in their early stages, when interventions could be most beneficial. This gap has led researchers and healthcare professionals to explore advanced computational techniques to improve liver disease detection and prediction.

Machine learning (ML), a subset of artificial intelligence, has emerged as a powerful tool in medical diagnosis due to its ability to learn patterns from complex and large datasets. By leveraging historical clinical data, laboratory test results, and medical images, machine learning algorithms can identify subtle patterns that might be missed by traditional analysis. In the context of liver diseases, ML techniques have been applied to classify different types of liver conditions, predict disease progression, and assess the risk of complications. Models such as decision trees, support vector machines, random forests, and deep learning networks have demonstrated promising results in predicting liver disease outcomes with high accuracy and sensitivity.

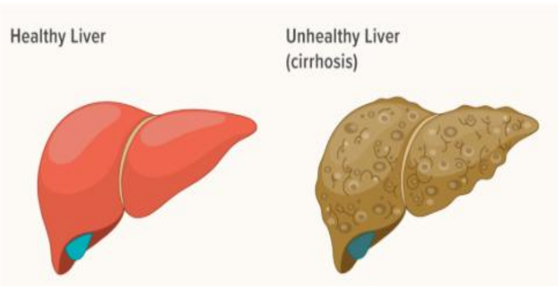

Figure 1: healthy and unhealthy liver

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online) Volume 14, Issue 8, August 2025)

The prediction of liver diseases using machine learning involves several important steps including data collection, preprocessing, feature selection, model training, and evaluation. Clinical data such as liver function tests, patient demographics, lifestyle habits, and imaging findings are gathered and cleaned to prepare for analysis. Selecting relevant features that contribute most significantly to prediction accuracy is critical to enhance model performance and reduce computational complexity. Trained models can then be tested on new patient data to provide predictive insights that aid clinicians in decision-making.

Despite the considerable advancements, there remain challenges in applying machine learning for liver disease prediction. These include the availability of large, diverse, and high-quality datasets, handling imbalanced data where some disease categories have fewer samples, and ensuring model interpretability so that healthcare providers can trust and understand algorithmic recommendations. Furthermore, integrating machine learning models into clinical workflows requires careful validation, regulatory approval, and user-friendly interfaces.

Advancements in machine learning have brought a variety of algorithms into the spotlight for liver disease prediction. Supervised learning techniques, such as support vector machines (SVM), random forests, and logistic regression, have been extensively used to classify liver disease conditions based on labeled clinical data. These algorithms excel at learning decision boundaries that separate healthy patients from those with disease, or distinguishing between different types of liver disorders. More recently, deep learning methods, particularly convolutional neural networks (CNNs), have shown remarkable success in analyzing medical imaging data like ultrasound or MRI scans to detect liver abnormalities. These models automatically extract relevant features from raw images, reducing the need for manual intervention and enabling more accurate and faster diagnosis. Ensemble methods, which combine multiple machine learning models, also provide robust predictions by balancing the strengths of individual algorithms, thus improving overall performance in complex datasets.

The integration of machine learning in liver disease prediction is not only revolutionizing diagnostics but also paving the way for personalized medicine. By analyzing patient-specific data, including genetic information, lifestyle factors, and response to previous treatments, ML models can predict disease progression and treatment outcomes tailored to individual patients. This helps clinicians devise optimized treatment plans and monitor patients more effectively. However, successful implementation requires overcoming challenges related to data privacy, interoperability of healthcare systems, and ensuring equitable access to these technologies. Moreover, interdisciplinary collaboration among data scientists, medical professionals, and regulatory bodies is essential to translate machine learning research into practical clinical tools. As machine learning models continue to evolve, their role in enhancing liver disease prediction and management is expected to grow, offering hope for better healthcare solutions worldwide.

II. LITERATURE SURVEY

A Oguntimilehin et al. Liver disease prediction has garnered significant attention in recent years due to the rising incidence of liver-related ailments worldwide. The work by Nithyashri, Goel, and Hada (2024) explores intelligent classification of liver diseases through ensemble machine learning techniques, highlighting how combining multiple classifiers can improve accuracy and robustness. Their study demonstrated that ensemble methods like bagging and boosting outperform single models by effectively handling variability and complexity in clinical data. The ensemble approach utilized in their research addressed issues of overfitting and improved generalization, crucial for reliable clinical applications. Their methodology incorporated several base classifiers, including decision trees and support vector machines, whose outputs were aggregated to enhance predictive performance. This work underscores the importance of ensemble learning as a superior strategy for liver disease classification compared to traditional standalone algorithms.

In a similar vein, Chicco and Jurman (2021) presented an ensemble learning framework specifically designed for enhanced classification of hepatitis and cirrhosis patients. Their approach integrated multiple machine

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online) Volume 14, Issue 8, August 2025)

learning models such as random forests, gradient boosting machines, and neural networks to capture diverse patterns in patient data. The ensemble system showed improved sensitivity and specificity compared to individual classifiers, which is critical in medical diagnosis where false negatives can have severe consequences. Notably, their work emphasized the use of feature selection techniques that helped identify the most relevant clinical parameters contributing to disease progression. By focusing on hepatitis and cirrhosis, this study provided valuable insights into managing chronic liver diseases through predictive analytics, which could facilitate timely intervention and personalized treatment strategies.

Tamilarasi et al. (2022) conducted predictive analysis of hepatitis and cirrhosis liver diseases using multiple machine learning algorithms, including logistic regression, support vector machines, and k-nearest neighbors. Their comparative study highlighted the strengths and weaknesses of each algorithm with respect to accuracy, precision, and computational efficiency. The authors also investigated the impact of data preprocessing steps such as normalization and missing value imputation on model performance. Their findings revealed that while traditional classifiers could yield reasonable accuracy, ensemble models consistently outperformed them, especially in handling imbalanced datasets typical in medical records. This study reinforced the importance of preprocessing and algorithm selection in building reliable liver disease prediction systems.

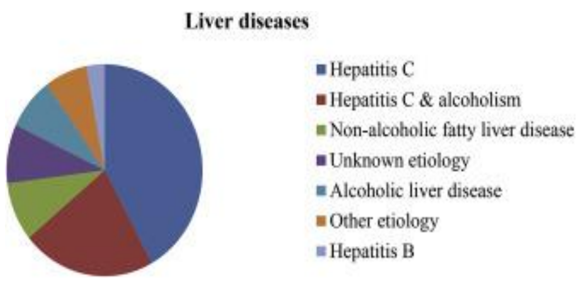

Figure 2: Liver diseases

The work of Lalithesh and Raghavendra (2023) focused on the analysis and prediction of liver cirrhosis through machine learning models trained on patient biochemical and demographic data. They explored decision trees, random forests, and extreme gradient boosting (XGBoost) techniques to predict cirrhosis with high accuracy. Their research highlighted the significance of incorporating domain knowledge into feature engineering, improving interpretability and clinical relevance of the models. Moreover, they addressed challenges related to class imbalance by employing oversampling methods like SMOTE (Synthetic Minority Over-sampling Technique), which enhanced the model's sensitivity towards the minority class. This study contributes valuable perspectives on tailoring machine learning techniques to clinical datasets that often suffer from imbalance and noise.

Hanif and Khan (2022) presented a study specifically on liver cirrhosis prediction using various machine learning approaches. Their research compared support vector machines, logistic regression, and ensemble classifiers in terms of performance metrics such as accuracy, F1-score, and area under the ROC curve (AUC). The authors underscored the critical role of feature selection and hyperparameter tuning in optimizing model performance. They also experimented with cross-validation techniques to ensure robustness and avoid overfitting. Their results indicated that ensemble classifiers, particularly those combining decision trees with boosting algorithms, delivered superior predictive power for cirrhosis detection. This work adds to the growing body of evidence favoring ensemble learning in liver disease prognosis.

Geetha and Maruthuperumal (2022) also focused on liver cirrhosis prediction using ensemble machine learning algorithms, presenting a framework that combined bagging and boosting methods to improve accuracy. Their approach incorporated multiple base learners, such as random forests and AdaBoost, which were trained on clinical and biochemical data. The authors demonstrated that ensemble models not only improved classification accuracy but also enhanced model stability across different patient cohorts. They

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online) Volume 14, Issue 8, August 2025)

emphasized the potential for such models to be integrated into decision support systems in healthcare, aiding clinicians in early diagnosis and monitoring of liver cirrhosis. This study reinforces the critical role of ensemble techniques in enhancing liver disease detection reliability.

In an early detection context, Naseem et al. (2021) conducted a performance assessment of various classification algorithms on liver syndrome diagnosis. Their research compared traditional machine learning models like decision trees, naive Bayes, and logistic regression on a dataset comprising patient clinical variables. The study found that tree-based models, especially random forests, delivered the best trade-off between accuracy and interpretability. They also highlighted the importance of feature scaling and data normalization to improve classifier performance. Their work underlines the necessity of adapting preprocessing techniques according to dataset characteristics to maximize the predictive capabilities of machine learning algorithms in liver disease detection.

Tokala et al. (2023) provided a comprehensive analysis of liver disease prediction and classification using machine learning techniques. Their study explored an array of algorithms including support vector machines, random forests, and artificial neural networks on a diverse dataset combining clinical and imaging data. They employed feature extraction and dimensionality reduction methods such as principal component analysis (PCA) to streamline input features, reducing computational load while preserving relevant information. Their results emphasized the effectiveness of neural networks in capturing nonlinear relationships in the data, leading to improved diagnostic accuracy. This work highlights the advantage of integrating clinical data with imaging features for more holistic liver disease prediction models.

Spann et al. (2020) conducted a comprehensive review on the application of machine learning in liver disease and transplantation, offering valuable insights into the current state and future directions of the field. They discussed how predictive models are used not only for early diagnosis but also for predicting transplantation outcomes and patient survival rates. The review stressed challenges such as data heterogeneity, model interpretability, and ethical considerations in deploying AI in clinical settings. Their analysis showed that while machine learning holds promise, successful clinical translation requires multidisciplinary collaboration and rigorous validation on diverse patient populations. This review provides a broad context for understanding the impact of AI-driven liver disease prediction.

Rajput and Kaur (2019) carried out a comparative study of machine learning algorithms for liver disease diagnosis, evaluating models such as naive Bayes, decision trees, and k-nearest neighbors on a publicly available dataset. Their findings indicated that decision trees and random forests achieved higher accuracy and better handling of missing data compared to simpler classifiers. They also examined the effect of feature selection techniques on model performance, concluding that appropriate feature selection is vital to reduce overfitting and enhance generalizability. Their study laid foundational knowledge for applying machine learning in liver disease detection, guiding researchers in algorithm choice and data preparation strategies.

Shobana and Umamaheswari (2021) explored liver disease prediction using gradient boosting machine learning techniques combined with feature scaling to improve classification accuracy. They demonstrated that scaling features such as liver enzyme levels and bilirubin concentrations before model training significantly enhanced the performance of gradient boosting classifiers. The study also emphasized the importance of hyperparameter tuning to optimize tree depth, learning rate, and number of estimators. Their results highlighted gradient boosting as a powerful technique capable of capturing complex interactions among clinical variables, making it well-suited for liver disease prognosis.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online) Volume 14, Issue 8, August 2025)

Zhao et al. (2021) proposed a novel liver disease prediction model using a hybrid algorithm called W-L R-XGB, which integrates wavelet transforms with light gradient boosting machine (LightGBM) and extreme gradient boosting (XGBoost). This hybrid model aimed to improve feature extraction and enhance classification performance by combining the strengths of different boosting algorithms. Their approach showed significant improvements in sensitivity and specificity over traditional single models when tested on a real-world liver disease dataset. By applying wavelet transforms, they extracted multi-resolution features that better captured the underlying patterns in the data. This study represents an innovative direction in liver disease prediction, leveraging advanced signal processing with machine learning to boost diagnostic accuracy.

III. CHALLENGES

Challenges in Liver Disease Prediction Using Machine Learning:

1. **Data Quality and Availability:** Medical datasets often contain missing, noisy, or inconsistent data, which can degrade model performance and lead to inaccurate predictions.
2. **Class Imbalance:** There is usually a significant imbalance between healthy and diseased patient samples, causing models to be biased toward the majority class and perform poorly on minority classes.
3. **Feature Selection and Engineering:** Selecting relevant features from complex, high-dimensional, and heterogeneous clinical data is difficult, and poor feature choices can result in overfitting and reduced interpretability.
4. **Model Interpretability and Explainability:** Many advanced ML models act as black boxes, providing limited insight into how predictions are made, which reduces clinician trust and hinders adoption in practice.
5. **Generalizability and External Validation:** Models trained on data from specific populations or healthcare settings may not generalize well to others due to demographic and environmental differences.
6. **Computational Complexity and Resource Requirements:** Some algorithms require high computational resources and technical expertise, limiting their usability in resource-constrained or low-income healthcare environments.
7. **Lack of Standardized Protocols and Benchmarks:** Variability in datasets, preprocessing methods, evaluation metrics, and reporting standards makes it difficult to compare studies and reproduce results.
8. **Integration of Multimodal Data:** Combining diverse data types like clinical records, biochemical tests, and medical imaging into a single predictive model remains a complex task.
9. **Handling Imbalanced and Noisy Clinical Data:** Medical records often have incomplete or erroneous entries, which complicate training reliable models.
10. **Timely and Accurate Early Diagnosis:** Achieving both high sensitivity and specificity for early-stage liver disease detection is challenging but critical for improving patient outcomes.

IV. PROPOSED PLAN

Proposed Plan for Liver Disease Prediction Using Machine Learning

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online) Volume 14, Issue 8, August 2025)

1. **Data Collection and Preprocessing:** Collect diverse, high-quality datasets from multiple sources and clean the data by handling missing values and inconsistencies to ensure reliability.
2. **Addressing Class Imbalance:** Apply data balancing techniques such as oversampling, undersampling, or synthetic data generation to ensure fair and accurate model training.
3. **Feature Selection and Engineering:** Identify and select the most relevant clinical and biochemical features with expert collaboration to improve model accuracy and interpretability.
4. **Model Development with Explainability:** Develop robust machine learning models, emphasizing interpretable algorithms or integrating explainability methods to foster clinician trust.
5. **Model Validation and Generalization:** Validate models using cross-validation and external datasets to ensure they perform well across different patient populations and settings.

V. CONCLUSION

Liver disease prediction using machine learning holds great promise for enhancing early diagnosis, improving treatment planning, and ultimately saving lives. Despite notable advances in algorithm development and data analysis techniques, several challenges remain that must be addressed to fully realize this potential. Issues related to data quality, model interpretability, class imbalance, and generalizability currently limit the clinical applicability of many predictive systems. Overcoming these hurdles requires continued research focused on improving data preprocessing, developing transparent and robust models, and ensuring ethical use of patient information. With collaborative efforts between clinicians, data scientists, and policymakers, machine learning-based liver disease prediction can become an integral part of personalized and precision medicine, leading to better healthcare outcomes worldwide.

REFERENCES

1. Nithyashri, H. Goel and M. S. Hada, "Intelligent Classification of Liver Diseases using Ensemble Machine Learning Techniques," *2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI)*, Coimbatore, India, 2024, pp. 1183-1188, doi: 10.1109/ICoICI62503.2024.10696789.
2. Davide Chicco ; Giuseppe Jurman - "An Ensemble Learning Approach for Enhanced Classification of Patients With Hepatitis and Cirrhosis " - "IEEE Access, Vol 9, Page(s): 24485–24498 " - 2021.
3. Tamilarasi A ; Chitra K ; Swetha J ; Nihila R - "Predictive Analysis for Hepatitis and Cirrhosis Liver Disease using Machine Learning Algorithms " - "Third International Conference on Electronics and Sustainable Communication Systems " - 2022.
4. Lalithesh D Sawant ; Raghavendra Ritti - "Analysis and Prediction of Liver Cirrhosis Using Machine Learning Algorithms " - "3rd International Conference on Intelligent Technologies (CONIT) " - 2023.
5. Ishtiaqe Hanif ; Mohammad Monirujjaman Khan - "Liver Cirrhosis Prediction using Machine Learning Approaches " - "13th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON) "- 2022.
6. C. Geetha ; S. Maruthuperumal - "Prediction of Liver Cirrhosis using Ensemble Machine Learning Algorithms " - "International Conference on Computer, Power and Communications (ICCPC) " - 2022.
7. Rashid Naseem, Bilal Khan, Muhammad Arif Shah, Karzan Wakil, Atif Khan, Wael Alosaimi, M. Irfan Uddin, Badar Alouffi,"Performance assessment of

classification algorithms on early detection of Liver syndrome ", Journal of Healthcare Engineering, Vol 2020, Issue 1, 2021.

8. Srilatha Tokala, Koduru Hajarathaiah, Sai Ram Praneeth Gunda, Srinivasrao Botla, Lakshmikanth Nalluri, Pathipati Nagamanohar, Satish Anamalamudi, Murali Krishna Enduri, "Liver Disease Prediction and classification using Machine Learning techniques, International Journal of Advanced Computer Science and Applications, Vol. 14 (2), pp: 871–878, 2023.
9. A. Spann, A. Yasodhara, J. Kang, K. Watt, B. Wang, A. Goldenberg, and M. Bhat, "Applying machine learning in liver disease and trans plantation: a comprehensive review," Hepatology, vol. 71, no. 3, pp. 1093–1105, 2020.
10. A. S. Rajput and R. Kaur - "A Comparative Study of Machine Learning Algorithms for Liver Disease Diagnosis " - "International Journal of Computer Applications, Vol 179, Issue 18, Pages: 1–8 "- 2019.
11. G. Shobana, K. Umamaheswari - "Prediction of Liver Disease using Gradient Boost Machine Learning Techniques with Feature Scaling "-"5th International Conference on Computing Methodologies and Communication (ICCMC) "-
12. Ruizhe Zhao, Xiumei Wen, Hui Pang, and Zhaohui Ma - "Liver disease prediction using W-L R-XGB Algorithm " - "International Conference on Computer, Blockchain and Financial Development (CBFD) " - 2021.