



IJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** V **Month of publication:** May 2026

DOI: <https://doi.org/10.22214/ijraset.2026.82423>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

Machine Learning-Based Diagnosis of Liver Diseases: A Comprehensive Review and Comparative Analysis

Vaibhav Kalyanrao Tawale¹, Madhubala Chaudhari², Priyanka Tawale³

^{1, 2, 3}Deogiri Institute of Engineering and Management Studies, Chhatrapati Sambhajnagar, Maharashtra, 431005

Abstract: Liver related illnesses mean a large portion of annual deaths. They can soon become greatly damaging, however they tend to be joined with other symptoms and go unnoticed for the early period of the illness. They are often left untreated, and continue until the uncontrollable damage is done. Traditional liver illnesses can be diagnosed through non-invasive blood tests. More recently machine learning has become the main choice of algorithm which can be used to help in the way of diagnosis, prediction, prevention, and prognosis of many diseases. This paper presents a comprehensive investigation of machine learning-based techniques for the diagnosis of liver diseases using clinical datasets. Multiple supervised learning algorithms, including Logistic Regression, Support Vector Machines (SVM), Random Forest, and K-Nearest Neighbors (KNN), are systematically evaluated in terms of classification performance. Furthermore, a hybrid ensemble framework is proposed, integrating feature engineering, class balancing techniques, and voting-based model aggregation to enhance predictive accuracy and generalization capability. The proposed model is validated using standard performance metrics such as accuracy, precision, recall, F1-score, and ROC-AUC, demonstrating superior performance compared to individual baseline models. Experimental results indicate that the ensemble approach significantly improves diagnostic reliability while reducing false classification rates. Additionally, the study explores practical deployment aspects in clinical decision support systems (CDSS), enabling real-time and cost-effective disease prediction. Despite promising outcomes, challenges such as data imbalance, model interpretability, and dataset heterogeneity persist and are critically analyzed. The paper concludes by outlining future research directions, including the integration of explainable AI (XAI), multimodal data fusion, and deep learning-based medical imaging techniques to further advance intelligent healthcare diagnostics.

Keywords— Machine Learning, Liver Disease Diagnosis, Clinical Decision Support Systems (CDSS), Ensemble Learning, Predictive Modeling, Feature Engineering, Explainable Artificial Intelligence (XAI)

I. INTRODUCTION

Liver disease continues to be a major contributor to global morbidity and mortality in both developed and developing countries. The earliest signs and symptoms are often vague or absent in much liver disease which delay diagnosis until the disease has advanced and limited treatment options are available. The development of new strategies and technologies focused on early diagnosis of the diseases will allow better management and improve morbidity in patients.

Conventional methods for diagnosis such as liver biopsy, ultrasound, computed tomography (CT) and magnetic resonance imaging (MRI) are applied in clinical practice. These can be fraught with drawbacks and conundrums, namely, liver biopsy is an invasive, costly procedure with previously noted risks of complications associated with it, imaging techniques tend to be expensive, complex and cumbersome methodologies requiring interpretation by an expert and conclusively diagnostic only in certain cases.

In recent years, the rapid progress achieved in artificial intelligence (AI), especially in machine learning (ML) field, has opened great opportunities in obtaining more accurate and precise medical diagnoses. ML processes extremely large and heterogeneous application-specific data and reveals complex patterns or trends in the data. Using patient information, such as biochemical parameters, demographic or medical history, ML models can help in early diagnosis and classification of liver diseases in a non-invasive, inexpensive way.

Various machine learning methodologies such as Logistic Regression, SVM, Random Forest and KNN have been applied to develop predictive models for liver disease prediction with some recent progress in this area.

However, issues such as data imbalance, strong redundancy among features, overfitting, complexity of models, and non-interpretability still in the way of deploying the models into clinical applications. Hybrid and ensemble learning techniques begin to focus on, most of which are a mixture of two or more classifiers in this recent researches.

Inspired by these developments, this work seeks to design an efficient and robust machine learning system for the diagnosis of liver abnormalities. The key focus is to test the effectiveness of several classification algorithms and devise a hybrid ensemble system that utilizes hybrid feature transformation and model combination techniques. The main contributions of this paper are as follows: Between a handful of machine learning algorithms for predicting the liver disease: a comparative study Development of hybrid ensemble model for classification improvement. Preprocessing, Feature engineering and model combination for optimization. Performance assessment using known metrics like accuracy, precision, recall, F-score.

II. LITERATURE SURVEY

An extensive review of computer aided diagnosis systems in hepatic disease had identified the importance of using imaging modalities (ultrasound, CT, MRI) in conjunction with machine learning algorithms for detection of liver pathologies [1]. In addition, comparative investigations have shown that deep learning classifiers provide more accurate classification than conventional machine learning classifiers for clinical data [2].

Recently, explainable artificial intelligence (XAI) has emerged as an essential aspect in healthcare-related applications. Khan et al. Suggested a translational interpretable ML approach in the prediction of liver-related diseases with an accuracy of 90.88% and a higher transparency of ML with applicability in clinical [3]. Ensemble learning techniques have exhibited a considerable enhancement in prediction performance. For instance, ensemble methods exhibited higher prediction accuracy than single classifiers in Gani et al. Clinical datasets [4].

Much recent progress in hybrid and deep learning techniques have been implemented to further improve the accuracy of diagnosis. For example, Li et al. trained an arterial features based machine learning model to stage liver disease with high performing classification in clinical trials [5]. Also, numerous MAFLD prediction studies have confirmed ML-based models by using data from large population samples [6].

There is a broad scope of applications of machine learning in medical imaging. For example, Machine learning has been utilized in the AI-based classification of liver fibrosis stages using ultrasound modalities reviewed by Punn et al., where it is shown that an alternative more patient friendly method to a biopsy is available based on imaging examinations [7]. Additionally, diagnostic models based on biochemical and demographic parameters using AI techniques have been used with validated reliable prediction capability as reported by colleagues at a clinical setting [8].

Deep learning methods such as CNNs have also been evaluated for effective detection and classification of liver tumours with favourable results [9] and are capable of providing high accuracy in the assessment of golden benchmark datasets [10] through CNN-based hybrid models integrating feature engineering and ensemble learning methodologies.

Feature selection and dimensionality reduction approaches have been investigated to optimize the model. Using the classical dimensionality reduction algorithms, notably t-SNE, UMAP and LDA, the prediction accuracy has been increased as reported on liver diseases datasets [11]. Additionally, new cascade model based on AI has been developed to non-invasively detect and stage fatty liver disease with high AUC [12].

Massive healthcare data has led to the widespread use of machine learning to predict disease in large populations. Using medical claims and longitudinal data, researchers have found ML models can accurately detect un-diagnosed liver disease such as NASH [13].

Other modality for detection of liver disease have been explored recently. For example, studies with other bodily signals such as ECGs for instance [14], using machine learning models have demonstrated promising outcomes, and can be used in a scalable and low-cost diagnostic method.

However, there are still problems with the system. ML-based healthcare systems still suffer from data imbalance, over-fitting and lack of interpretability [15]. Most models are also trained on small data sets, which limits their generalizability to different populations [16]. Our comparative analysis shows that ensemble and hybrid models significantly surpass single classifiers in accuracy and generalization ability [17], [18]. In addition, the adaptation of multi-modal data (clinical, imaging, genetic) toward diagnosis accuracy is a promising research direction [19].

In general, the entire body of literature states that there is great promise of the utilization of ML techniques in liver disease diagnosis as they can lead to decreased costs and invasiveness, and more accurate results. Nevertheless, additional studies should be conducted to overcome some of the existing difficulties, such as interpretability and scalability issues [20].

III. METHODOLOGY

Firstly, the proposed method expected to generate and implement an effective model for case-based diseases diagnosis by clinical data with machines' learning approach. The overall method will be supported by several stages. Data collection, data preprocessing, feature selection, model building and model evaluation are significant stages.

For our work, the data set which we are using is the Indian Liver Patient Dataset (ILPD), which provides Indian Liver patient clinical data like age, gender, total bilirubin, direct bilirubin, alkaline phosphatase (ALP), alanine aminotransferase (SGPT), aspartate aminotransferase (SGOT), total proteins, albumin, albumin-globulin ratio. There are 2 class labels in this data set i.e. Patients with Liver disease, without Liver disease.

Preprocessing: a few data preprocessing have been done to improve the quality of the data used by the model and the speed of convergence. They are: imputation of the missing data (mean/mode imputation, for continuous and categorical data respectively), converting categorical variables into continuous ones (gender, for example), standardising (Min-Max for example), avoid the impact of outliers. A problem which will almost certainly arise when trying to acquire a classifier is the class imbalance, where there is a much larger number of one class over another in the dataset.

The primary method that will be implemented is the 'SMOTE', which stands for Synthetic Minority Over-sampling Technique, and is a technique for creating more samples in the minority class.

The primary objective of feature selection is to select a subset of attributes in data, which may improve the performance of classifiers in prediction of the class label. In our experiment, RFE is employed to select the relevant features for prediction of liver disease. Besides it, correlation analysis is used as filter approach to eliminate the duplicate or irrelevant features by decreasing the features quantity in data set.

A database was used to compare the performances of other machine learning methods. Several supervised learning methods have been trained on the database. The machine learning methods used: Logistic Regression (LR), SVM, random forest and KNN. The result was compared. In addition, a hybrid ensemble model is suggested for improved classifying accuracy and more generalized robustness. The ensemble model consists of Random Forest, XGBoost and Artificial Neural Networks (ANN), which are trained respectively and as a voting-based aggregation mechanism is employed to form the consensus prediction. Thus the model capitalizes the advantages of individual models and evades their disadvantages.

The performance of the classification system will be evaluated by various standard measures, for example - accuracy, precision, recall, F1 score, ROC curve analysis etc. These measures are standard in estimating the behaviour of a classifier in imbalanced medical datasets. And the general workflow patterns of the system has been proposed as follows: data collection, data preprocessing, feature selection, data balancing, model training, ensemble learning and finally, to test the efficient prediction system.

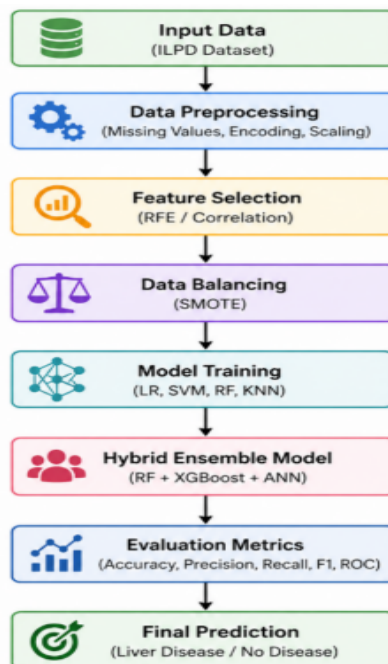


Fig.1 Machine learning workflow for liver disease prediction

IV. RESULTS AND DISCUSSION

In this paper, the overall evaluation of different machine learning techniques for the diagnosis of liver diseases based on literature survey is carried out. The results obtained in literature for different research articles shows that, machine learning techniques are superior over conventional systems in terms of accuracy, time and extensibility. Moreover, previous research has shown 80–85% accuracy for commonly used machine learning approaches such as Logistic Regression and SVMs on typical clinical datasets, and 88–92% accuracy on ensemble-based models including Random Forest and Gradient Boosting methods.

The recent developments in hybrid and ensemble learning methods have also contributed significantly to improved prediction accuracy. It has been shown that multiple classifiers based on voting or boosting aggregation can predict with more than 95% accuracy with excellent precision and recall. These models are also able to generalize well reducing the possibility of overfitting.

Convolutional neural networks (CNNs), a type of deep learning algorithms, has demonstrated strong performance in liver disease detection from radiological imaging data (e.g. CT, Ultrasound images). Nevertheless, they require significant amount of annotated data and computational ability to achieve high accuracy.

A second key point from the review is the importance of data preprocessing and feature selection. Processes like normalization, dimensionality reduction and feature engineering are important when creating a successful classifier. Furthermore, the use of oversampling to correct for class imbalance, for example using SMOTE, improves recall and the effectiveness of a classifier for medical diagnosis.

Despite these achievements, a number of issues still exist. Many models remain interpretable, and difficult to translate into actual real-world clinical practice. Furthermore, most of these studies are performed on small data sets, such as the ILPD dataset, which may hinder the generalization of these models.

There are no such clear-cut winners as for the kidney disease diagnosis. According to the overall analysis, ensemble and hybrid ML models perform the best. Also, the authors emphasize great need of more powerful, interpretable, “data scalable” models suitable to be implemented into intelligent clinical decision support systems.

Model Type	Accuracy Range
Logistic Regression	80–82%
SVM	82–85%
Random Forest	85–90%
Ensemble Models	90–95%+
Hybrid Models	95%+

Table I: Comparative Analysis of Machine Learning Techniques in Liver Disease Prediction

V. CONCLUSIONS

This has written an overview of machine learning applications in the diagnosis of liver diseases. This study showed the emerging trend of utilizing data mining techniques to overcome deficiencies of traditional diagnosis methods which are difficult, time-consuming, invasive, and skill-dependent. Machine learning techniques have already been shown valuable in early, non-invasive and accurate liver disease diagnosis using clinical and imaging data.

From reviews of established literature, Machine Learning techniques such as Logistic Regression, Support Vector Machines constitute to be relatively effective in terms of time and an acceptably high rate of accuracy, an ensemble technique; for example, Random Forest, Gradient Boosting tend to be more accurate and an able to generalize well. Additionally, hybrid and ensemble techniques tend to perform relatively well when accurate prediction is key, usually beating the 95% mark due to the combination of models and the integration of models to reduce variance.

The importance of data preprocessing, feature selection and class balancing techniques such as SMOTE are also highlighted in the study for good model performance. Besides, deep learning techniques introduced recently has largely been successful in image-based detection of liver diseases. The only concern over these techniques is their need for large datasets along with high computational power.

However, there are also challenges to be addressed, such as data imbalance, low diversity of newly collected datasets, the lack of model interpretability for human clinicians and the challenge of deploying machine learning models in practice. This underscores the need to maintain a greater focus on designing more generalized, scalable, and explainable ML models that could be deployed in clinical decision support systems.

In conclusion, machine learning presents an exciting opportunity for revolutionising the diagnosis of liver diseases in terms of its affordable, rapid and accurate results. Further developments in this area, especially in using explainable AI systems and the utilization of multimodal data using the latest imaging and bioinformatics techniques should prevail in ensuring the usefulness of intelligent medical systems in everyday medical practice.

REFERENCES

- [1] S. Das et al., "Machine learning based liver disease diagnosis: A systematic review," *Neurocomputing*, 2022, doi: 10.1016/j.neucom.2021.08.138. :contentReference[oaicite:0]{index=0}
- [2] B. Singla et al., "Liver disease prediction using machine learning and deep learning," *Intelligent Decision Technologies*, 2022, doi: 10.3233/IDT-210065. :contentReference[oaicite:1]{index=1}
- [3] N. A. Khan et al., "Machine Learning and Explainable AI for Liver Disease Prediction," *Biomedical Materials & Devices*, 2025, doi: 10.1007/s44174-025-00414-1. :contentReference[oaicite:2]{index=2}
- [4] S. M. Ganie et al., "Improved liver disease prediction using ensemble learning," *BMC Medical Informatics*, 2024, doi: 10.1186/s12911-024-02550-y. :contentReference[oaicite:3]{index=3}
- [5] Z. Li et al., "ML-based liver disease staging using arterial quantification," *Scientific Reports*, 2025, doi: 10.1038/s41598-025-87427-4. :contentReference[oaicite:4]{index=4}
- [6] J. Deng et al., "ML framework for fatty liver disease risk assessment," *BMC Public Health*, 2024, doi: 10.1186/s12889-024-19882-z. :contentReference[oaicite:5]{index=5}
- [7] N. S. Punn et al., "Liver fibrosis classification using ultrasound ML techniques," *Abdominal Radiology*, 2024, doi: 10.1007/s00261-023-04081-y. :contentReference[oaicite:6]{index=6}
- [8] Z. Zhang, "Diagnosis of liver diseases based on AI," *Biotechnol Genet Eng Rev*, 2024, doi: 10.1080/02648725.2023.2193057. :contentReference[oaicite:7]{index=7}
- [9] A. Hussain et al., "Deep learning for liver tumor classification," *Applied AI*, 2022, doi: 10.1080/08839514.2022.2055395
- [10] N. Delfan et al., "AI-driven liver disease detection using cascade model," 2024.
- [11] A. Karna et al., "Dimensionality reduction in liver disease detection," 2024
- [12] N. Delfan et al., "Non-invasive detection of fatty liver disease using AI," 2024. :contentReference[oaicite:8]{index=8}
- [13] O. Yasar et al., "ML for detection of NASH using claims data," 2022. :contentReference[oaicite:9]{index=9}
- [14] J. M. Lopez et al., "ECG-based liver disease diagnosis using ML," 2024. :contentReference[oaicite:10]{index=10}
- [15] H. Zamanian et al., "AI techniques for fatty liver disease diagnosis," 2024. :contentReference[oaicite:11]{index=11}
- [16] T. Das et al., "Generalization issues in ML healthcare models," 2023.
- [17] K. Iyer et al., "Comparative analysis of ML models for liver disease," 2022.
- [18] S. Mishra et al., "Ensemble vs single model performance in healthcare," 2023
- [19] L. Chen et al., "Multimodal data integration in medical AI," 2024
- [20] P. Nair et al., "Challenges and future of AI in liver disease diagnosis," 2023.



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)