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

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

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

Vaibhav Kalyanrao Tawale[1], Madhubala Chaudhari[2], Priyanka Tawale[3]
[1, 2, 3]Deogiri Institute of Engineering and Management Studies, Chhatrapati Sambhajinagar , Maharashtra, 431005

Abstract:** Liver related illnesses mean a large portion of annual deaths. They can soonbecomegreatly damaging, however they tend to be joined with other symptoms and gounnoticedfor the early period of the illness. They are often left untreated, and continueuntil theuncontrollable damage is done. Traditional liver illnesses can be diagnosed throughnon-invasive blood tests. More recently machine learning has become the mainchoiceofalgorithm which can be used to help in the way of diagnosis, prediction, prevention, andprognosis of many diseases. This paper presents a comprehensive investigation of machine learning-basedtechniques for the diagnosis of liver diseases using clinical datasets. Multiplesupervisedlearning algorithms, including Logistic Regression, Support Vector Machines(SVM), Random Forest, and K-Nearest Neighbors (KNN), are systematically evaluatedintermsofclassification performance. Furthermore, a hybrid ensemble framework is proposed, integrating feature engineering, class balancing techniques, and voting-basedmodel aggregation to enhance predictive accuracy and generalization capability. The proposed model is validated using standard performance metrics such asaccuracy, precision, recall, F1-score, and ROC-AUC, demonstrating superior performancecomparedto individual baseline models. Experimental results indicate that the ensembleapproachsignificantly improves diagnostic reliability while reducing false classificationrates.Additionally, the study explores practical deployment aspects in clinical decisionsupportsystems (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 concludesbyoutlining future research directions, including the integration of explainableAI (XAI), multimodal data fusion, and deep learning-based medical imaging techniquestofurtheradvance intelligent healthcare diagnostics.

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

I. INTRODUCTION

Liver disease continue to be a major contributor to global morbidity and mortalityinbothdeveloped and developing countries. The earliest signs and symptoms are oftenvagueor absent in much liver disease which delay diagnosis until the disease has advancedandlimited treatment options are available. The development of newstrategiesandtechnologies focused on early diagnosis of the diseases will allowbetter managementand improve morbidity in patients.

Conventional methods for diagnosis such as liver biopsy, ultrasound, computedtomography (CT) and magnetic resonance imaging (MRI) are applied in clinical practice.These can be fraught with drawbacks and conundrums, namely, liver biopsyisaninvasive, costly procedure with previously noted risks of complications associatedwithit, imaging techniques tend to be expensive, complex and cumbersome methodologiesrequiring interpretation by an expert and conclusively diagnostic only in certaincases.

In recent years, the rapid progress achieved in artificial intelligence (AI), especiallyinmachine learning (ML) field, has opened great opportunities in obtaining moreaccurateand precise medical diagnoses. ML processes extremely large and heterogeneousapplication-specific data and reveals complex patterns or trends in the data. Usingpatient information, such as biochemical parameters, demographic or medical history, ML models can help in early diagnosis and classification of liver diseasesinanoninvasive, inexpensive way.

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

However, issues suchasdataimbalance, strong redundancy among features, overfitting, complexity of models, andnon-interpretability still in the way of deploying the models into clinical applicationsHybrid and ensemble learning techniques begin to focus on, most of which aremixtureoftwo or more classifiers in this recent researches.

Inspired by these developments, this work seeks to design an efficient androbustmachine learning system for the diagnosis of liver abnormalities. The key focusistotestthe effectiveness of several classification algorithms and devise a hybridensemblesystem 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: acomparative 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 diseasehadidentified the importance of using imaging modalities (ultrasound, CT, MRI)inconjunction with machine learning algorithms for detection of liver pathologies[1]. Inaddition, comparative investigations have shown that deep learning classifiersprovidemore accurate classification than conventional machine learning classifiers for clinical data [2].

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

Much recent progress in hybrid and deep learning techniques have been implementedtofurther improve the accuracy of diagnosis. For example, Li et al trainedanarterial features based machine learning model to stage liver disease with highperformingclassification in clinical trials [5]. Also, numerous MAFLD prediction studieshaveconfirmed ML-based models by using data from large population samples [6].

There is a broad scope of applications of machine learning in medical imaging. Forexample, Machine learning has been utilized in the AI-based classification of liver fibrosisstages using ultrasound modalities reviewed by Punn et al., where it is shownthat analternative more patient friendly method to a biopsy is available based onimagingexaminations [7]. Additionally, diagnostic models based on biochemical anddemographic parameters using AI techniques have been used with validatedreliableprediction capability as reported by colleagues at a clinical setting [8].

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

Feature selection and dimensionality reduction approaches have been investigatedtooptimize the model. Using the classical dimensionality reduction algorithms, notablyt-SNE, UMAP and LDA, the prediction accuracy has been increased as reportedonliverdiseases datasets[11]. Additionally, new cascade model based on AI hasbeendeveloped to non-invasively detection and stage fatty liver disease with high AUC[12].

Massive healthcare data has led to the widespread use of machine learningtopredictdisease in large populations. Using medical claims and longitudinal data, researchershave found ML models can accurately detect un-diagnosed liver disease suchasNASH[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 machinelearningmodels have demonstrated promising outcomes, and can be used in a scalableandlow-cost diagnostic method.

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

In general, the entire body of literature states that there is great promise of theutilizationof ML techniques in liver disease diagnosis as they can lead to decreasedcostsandinvasiveness, and more accurate results. Nevertheless, additional studies shouldbeconducted to overcome some of the existing difficulties, such as interpretabilityandscalability issues [20].

III. METHODOLOGY

Firstly, the proposed method expected to generate and implement an effectivemodel forcase-based diseases diagnosis by clinical data with machines' learning approach. Theoverall method will be supported by several stages. Data collection, data prepossessing 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, directbilirubin, alkaline phosphatase (ALP), alanine aminotransferase (SGPT), aspartateaminotransferase (SGOT), total proteins, albumin, albumin-globulin ratio. Thereare2class 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 qualityof thedata used by the model and the speed of convergence. They are: imputationthemissingdata (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 istheclassimbalance, where there is a much larger number of one class over another in thedataset.

The primary method that will be implemented is the 'SMOTE', which stands for SyntheticMinority Over-sampling Technique, and is a technique for creating more samplesintheminority class.

The primary objective of feature selection is to select a subset of attributes indata, whichmay improve the performance of classifiers in prediction of the class label. Inourexperiment, 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 duplicateorirrelevant features by decreasing the features quantity in data set.

A database was used to compare the performances of other machine learningmethods.Several supervised learning methods have been trained on the database. Themachinelearning methods used: Logistic Regression (LR), SVM, randomforest andKNN. Theresult was compare. In addition, a hybrid ensemble model is suggested for improved classifying accuracyandmore generalized robustness. The ensemble model consists of RandomForest, XGBoostand Artificial Neural Networks(ANN), which are trained respectively and as a voting-basedaggregation mechanism is employed to form the consensus prediction. Thusthemodel capitalizes the advantages of individual models and evades their disadvantages.

The performance of the classification system will be evaluated by variousstandardmeasures, for example - accuracy, precision, recall, F1 score, ROC curve analysisetc.These measures are standard in estimating the behaviour of a classifier inimbalancedmedical 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 efficint prediction system.

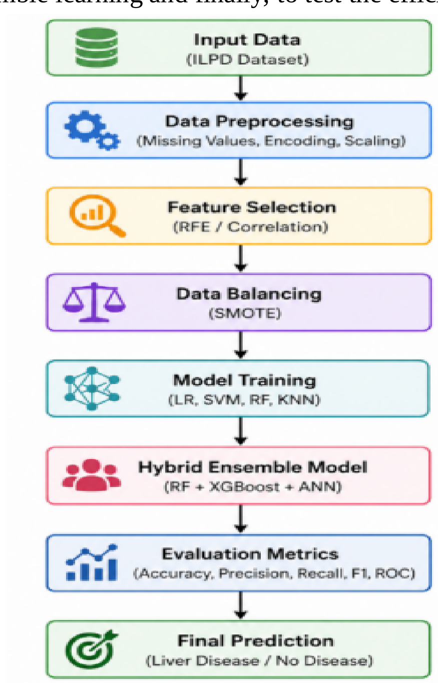


Fig.1 Machine learning workflow for liver disease prediction

IV. RESULTS AND DISCUSSION

In this paper, the overall evaluation of different machine learning techniquesfor thediagnosis of liver dis-eases based on literature survey is carried out. The resultsobtainedin literature for different research articles shows that, machine learning techniquesisSuperior over conventional systems in terms of accuracy, time and extensibility. Moreover, previous research has shown 80–85% accuracy for commonly usedmachinelearning approaches such as Logistic Regression and SVMs on typical clinical datasets,and 88–92% accuracy on ensemble-based models including Randomforest andGradient Boosting methods.

The recent developments in hybrid and ensemble learning methods have alsocontributedsignificantly to improved prediction accuracy. It has been shown that multipleclassifiersbased on voting or boosting aggregation can predict with more than 95%accuracywithexcellent precision and recall. These models are also able to generalize well reducingthepossibility of overfitting.

Convolutional neural networks (CNNs),a type of deep learning algorithms, hasdemonstrated strong performance in liver disease detection fromradiological imagingdata (e.g. CT, Ultrasound images). Nevertheless, they require significant amount ofannotated data and computational ability to achieve high accuracy.

A second key point from the review is the importance of data preprocessingandfeatureselection. Processes like normalization, dimensionality reduction and feature engineeringare important when creating a successful classifier. Furthermore, theuseofoversampling to correct for class imbalance, for example using SMOTE,improvesrecall and the effectiveness of a classifier for medical diagnosis

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

There are no such clear-cut winners as for the kidney disease diagnosis. Accordingtotheoverall analysis, ensemble and hybrid ML models perform the best. Also, theauthorsemphasize great need of more powerful, interpretable, “data scalable” models suitabletobe 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 DiseasePrediction

V. CONCLUSIONS

This has written an overview of machine learning applications in the diagnosisof liverdiseases. This study showed the emerging trend of utilizing data mining techniquestoovercome deficiencies of traditional diagnosis methods which are difficult, time-consuming, invasive, and skill-dependent. Machine learning techniques havealreadybeen shown valuable in early, non-invasive and accurate liver disease diagnosisusingclinical and imaging data.

From reviews of established literature, Machine Learning techniques suchasLogisticRegression, Support Vector Machines constitute to be relatively effective intermsoftime and an acceptably high rate of accuracy, an ensemble technique; for example, Random Forest, Gradient Boosting tend to be more accurate and an able togeneralizewell. Additionally, hybrid and ensemble techniques tend to performrelatively well whenaccurate prediction is key, usually beating the 95% mark due to the combinationofmodels and the integration of models to reduce variance.

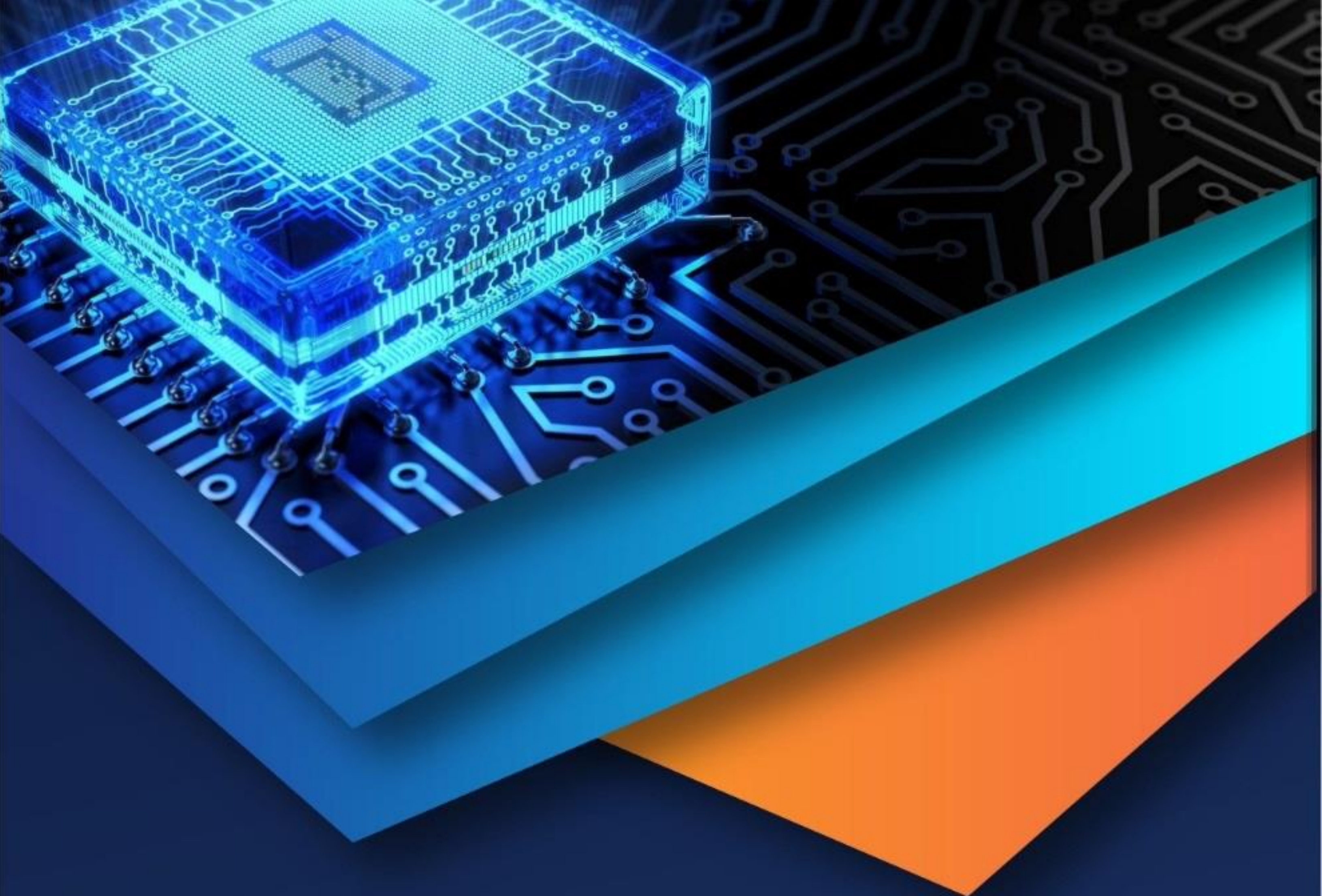
The importance of data preprocessing, feature selection and class balancingtechniquessuch as SMOTE are also highlighted in the study for good model performance. Besides, deep learning techniques introduced recently has largely been successful in imagebaseddetection of liver diseases. The only concern over these techniques is their needfor largedatasets along with high computational power.

However, there are also challenges to be addressed, such as data imbalance, lowdiversity of newly collected datasets, the lack of model interpretability for humanclinicians and the challenge of deploying machine learning models in practice. Thisunderscores the need to maintain a greater focus on designing more generalized, scalable, and explainable ML models that could be deployed in clinical decisionsupportsystems.

In conclusion, machine learning presents an exciting opportunity for revolutionisingthediagnosis of liver diseases in terms of its affordable, rapid and accurate results. Furtherdevelopments in this area, especially in using explainable AI systems and theutilizationof multimodal data using the latest imaging and bioinformatics techniques shouldprevail 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 DecisionTechnologies, 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