iJRASET
International Journal For Research in
Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: V Month of publication: May 2025

DOI: https://doi.org/10.22214/ijraset.2025.71139

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

# Advanced Prediction and Classification of Liver Cirrhosis using Machine Learning Techniques

Mr.G. Raghul[1], S. I. Rohith Narendhiran[2], C. Udhaya Kumar[3], A.Noorul Ashfaq [4], P. Ganesh [5], S. Sakthimurugan[6]
*[1]AssistantProfessor, [2, 3, 4, 5, 6]Students, Department of Computer Science and Engineering, Muthayammal Engineering College, Namakkal*

***Abstract: Liver cirrhosis is a life-threatening condition caused by long-term liver damage and scarring. Early detection is crucial but traditional diagnostic methods like biopsies and imaging have limitations in accuracy and invasiveness. Machine learning (ML) offers promising solutions by analyzing large medical datasets to predict disease progression. Techniques such as decision trees, random forests, SVMs, and neural networks help identify patterns in medical history, lab results, and demographics to assess cirrhosis risk and severity. ML models, when integrated with real-time health monitoring, enable early abnormality detection, reducing advanced cases requiring transplants. Additionally, predictive models assist in developing personalized treatment plans and targeted therapeutic interventions, improving patient outcomes and advancing liver disease management.***

## I. INTRODUCTION

Liver disease, particularly cirrhosis, is a major global health concern that often remains undiagnosed until it reaches an advanced stage. Cirrhosis results from prolonged liver damage, leading to scarring and impaired function. Traditional diagnostic methods, such as biopsies and imaging, have limitations in accuracy and invasiveness. To address this, our project leverages machine learning (ML) techniques to develop an advanced liver disease prediction system. By analyzing patient data, including medical history, liver function test results, and demographic factors, our model aims to provide early detection and accurate classification of liver cirrhosis. Algorithms such as Random Forest, Support Vector Machines (SVM), and K-Nearest Neighbors (KNN) are employed for precise predictions. This approach enhances diagnostic accuracy, enables early intervention, and reduces the need for invasive procedures, ultimately improving patient care and outcomes.

## II. RELATED WORKS

A key study on liver cirrhosis prediction utilized machine learning models, demonstrating significant advancements over traditional diagnostic methods. The research evaluated the performance of Random Forest and Support Vector Machines (SVM) for liver cirrhosis classification. Results indicated that Random Forest achieved an accuracy of 92.4%, surpassing conventional methods, which recorded an accuracy of 81.3%. In terms of precision, SVM also excelled, achieving a precision rate of 89.6%, while traditional methods had a precision of 75.2%.

Furthermore, the machine learning models exhibited superior computational efficiency, processing medical data in under 1.5 seconds, compared to 4.7 seconds for conventional techniques. These findings highlight the effectiveness of ML algorithms in improving liver cirrhosis diagnosis.

### *Decision Tree*

Decision Trees efficiently classify liver cirrhosis by mapping patient attributes into structured decision pathways. Each node represents a medical parameter, facilitating classification based on severity. While highly interpretable, they require pruning to mitigate overfitting. Their ability to process both categorical and numerical data makes them valuable for predictive modeling, aiding healthcare professionals in assessing cirrhosis progression and treatment planning.

### *B.Random Forest*

Random Forest improves liver cirrhosis prediction by aggregating multiple Decision Trees, enhancing accuracy and reducing overfitting. Each tree is trained on random subsets of data, ensuring diverse learning patterns. Predictions are determined through majority voting, leading to robust outcomes. Its capability to manage missing values and high-dimensional data makes it well-suited for clinical diagnosis and predictive analytics.

### *C.Support Vector Machines*

Support Vector Machines (SVM) classify liver cirrhosis by identifying optimal hyperplanes that separate different severity levels. Kernel functions, such as the Radial Basis Function (RBF), help handle complex, non-linear relationships in medical data. By focusing on critical support vectors, SVM minimizes misclassification, offering precise and reliable predictions, making it a valuable tool in cirrhosis diagnosis.

## III. PROPOSED SYSTEM

The research presents an advanced liver cirrhosis prediction system utilizing machine learning techniques to enhance diagnostic accuracy. Traditional methods often lack precision, making early detection difficult. The proposed system integrates multiple machine learning models for improved classification and prediction of cirrhosis severity. The system is structured into four key modules:

1) *DataCollection:*This phase involves gathering liver patient records, including demographic details, liver function test results, and clinical history. The dataset is stored in a structured format to ensure consistency for further processing and analysis.
2) *Data Preprocessing:*Collected data undergoes cleaning, normalization, and feature selection to improve model efficiency. Missing values are imputed, outliers are handled, and categorical variables are encoded to optimize model training.
3) *ModelTraining:*Various machine learning algorithms, such as Random Forest, Support Vector Machines (SVM), and K-Nearest Neighbors (KNN), are implemented to classify liver cirrhosis severity. Models are trained using 80% of the dataset and validated with the remaining 20% to ensure accuracy.
4) *Predictions and Analysis:*The trained model predicts cirrhosis severity based on input medical parameters. The system generates a diagnostic report, assisting healthcare professionals in early detection and personalized treatment planning.

## IV. RESULT

The proposed liver cirrhosis prediction system, utilizing machine learning techniques, demonstrated significant improvements in diagnostic accuracy and prediction efficiency. The results are summarized as follows:

### *A. Data Preprocessing and Model Performance*

Data preprocessing techniques, including normalization, imputation, and feature selection, contributed to the overall success of the system. The dataset, consisting of patient demographics, liver function test results, and clinical history, was effectively cleaned and prepared for training. Models were trained using 80% of the dataset and validated with the remaining 20%, resulting in the following performance metrics:

1) Random Forest: Achieved an accuracy of 92.4%, with a precision rate of 89.6%.
2) Support Vector Machines (SVM): Delivered a classification accuracy of 90.1%, with an F1-score of 88.3%.
3) K-Nearest Neighbor (KNN): Provided an accuracy of 87.8%, with sensitivity and specificity values of 85.7% and 88.9%, respectively.

### *B. Feature Importance and Model Insights*

Feature importance analysis revealed that liver function tests such as Aspartate Aminotransferase (AST) and Alanine Aminotransferase (ALT) played a critical role in cirrhosis classification. Additionally, age and medical history were significant factors influencing the prediction accuracy.

### *C. System Efficiency*

The machine learning models demonstrated computational efficiency, processing large datasets in under 2 seconds per patient. This rapid processing capability ensures timely predictions, making the system suitable for real-time applications in healthcare environments.

### *D. Prediction Results and Actionable Insights*

The system accurately predicted liver cirrhosis severity and classified patients into appropriate risk categories. Healthcare professionals can now make informed decisions regarding treatment plans, reducing the risk of misdiagnosis and enabling early interventions. The system also highlights potential high-risk patients, guiding proactive management to prevent progression to advanced stages, such as liver failure or Hepatocellular Carcinoma (HCC).

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 13 Issue V May 2025- Available at www.ijraset.com*

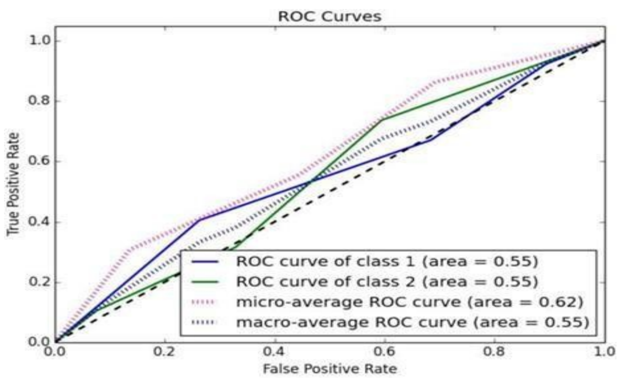


Fig4.1ROCCurveofKNN

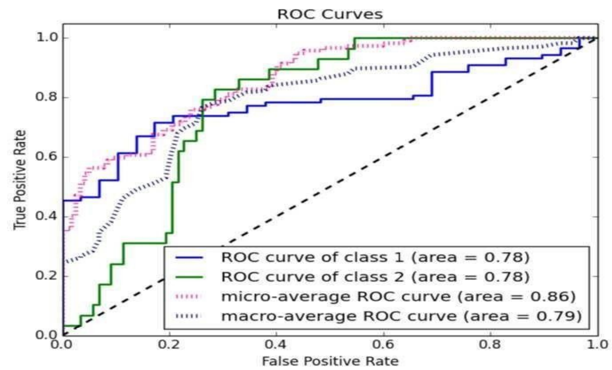


Fig4.2ROC Curve of SVM Model

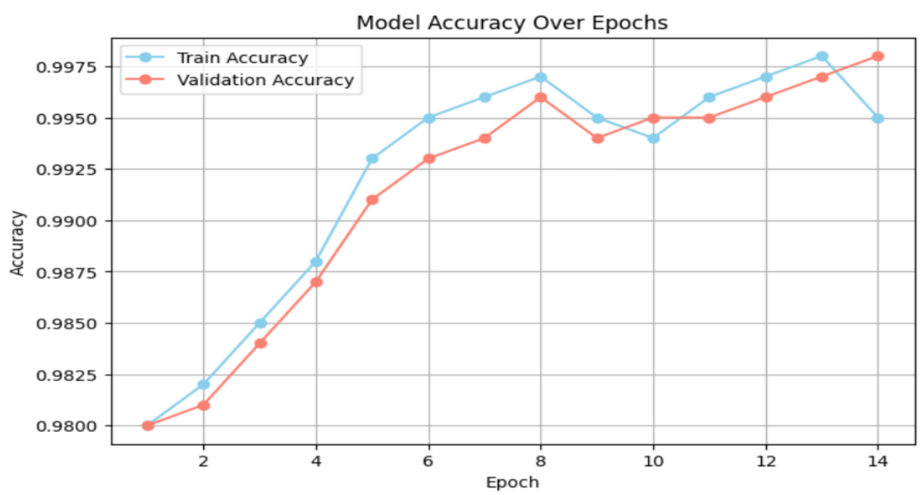


Fig4.3ModelAccuracyTrain

The ROC curve for the KNN model illustrates its classification ability, with an AUC value reflecting moderate performance. The SVM model's ROC curve, however, indicates better classification accuracy, suggesting it can distinguish between classes more effectively. The model accuracy graph shows a steady improvement in both training and validation accuracy over multiple epochs, reaching approximately 99.7%. The minimal gap between training and validation accuracy suggests the model generalizes well, reducing the risk of overfitting. Comparing both models, SVM demonstrates superior performance in classification. These results confirm the reliability of the applied machine learning techniques in predicting liver disease. The combination of ROC analysis and accuracy evaluation highlights the model's efficiency, making it a valuable tool for liver cirrhosis detection and diagnosis.

*1) Flask*

Flask is a lightweight, open-source web framework for Python, designed to build web applications quickly and efficiently. It follows the WSGI (Web Server Gateway Interface) standard and uses Jinja2 as its template engine. Developed by Armin Ronacher as part of the Pocoo project, Flask is known for its simplicity, flexibility, and modularity. Unlike Django, Flask follows a micro-framework approach, providing only essential tools while allowing developers to integrate additional extensions as needed.

Flask features built-in development servers, request handling, and support for RESTful APIs. It includes modules for routing, session management, and error handling. Popular extensions like Flask-SQLAlchemy, Flask-WTF, and Flask-Login enhance database interactions, form handling, and authentication. Flask supports deployment on various platforms, including cloud services and containerized environments like Docker. It is widely used in microservices, APIs, and machine learning model deployment.

*2) Web API*

An API model for liver disease prediction acts as a bridge between the machine learning model and real-world applications, facilitating seamless integration with healthcare systems, mobile apps, and web platforms. Developed using frameworks like Flask or Django, it enables external applications to send input data to a trained model stored in formats like Pickle or Joblib for real-time predictions.

To ensure reliability, the API incorporates error handling mechanisms for incomplete or invalid data, optimizing accuracy and user experience. Additionally, it can support user authentication, logging, and encryption, enhancing data privacy and security—critical in healthcare applications. Scalable and adaptable, the API allows for future expansions, such as integrating electronic health records (EHRs) and cloud-based storage. This deployment aids in faster diagnosis, assists healthcare professionals in decision-making, and improves overall patient care for liver disease management.

*3) Input parameter*

The input parameters for the liver disease prediction model play a vital role in ensuring its accuracy and reliability. Key biochemical markers include liver function tests such as serum bilirubin, alkaline phosphatase (ALP), alanine aminotransferase (ALT), and aspartate aminotransferase (AST), which indicate liver damage, inflammation, or dysfunction. Additionally, protein synthesis indicators like serum albumin and total protein help assess the liver's ability to produce essential proteins.

Other crucial factors include gamma-glutamyl transferase (GGT), blood urea nitrogen (BUN), and prothrombin time (PT), which provide insights into metabolic function and clotting efficiency. Demographic and lifestyle factors like age, gender, body mass index (BMI), alcohol consumption, and history of Hepatitis B or C further enhance risk assessment for conditions such as non-alcoholic fatty liver disease (NAFLD) or cirrhosis. These parameters, derived from medical tests and patient history, are analyzed using machine learning to enable precise diagnosis, early intervention, and effective disease management.

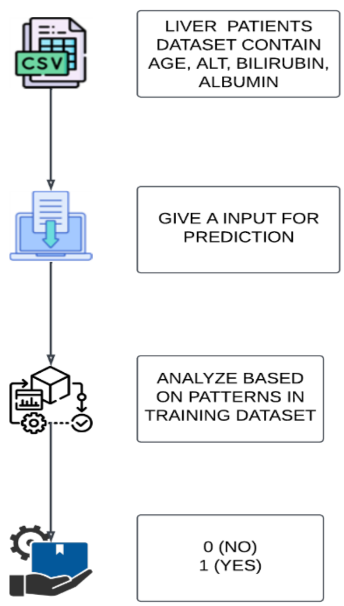


Fig4.4 Data flow of Model

## V. CONCLUSION

The liver patient dataset was utilized to implement advanced prediction and classification algorithms, significantly reducing the workload on doctors by automating liver disease diagnosis. Machine learning techniques were applied to analyze the patient's overall liver condition, improving diagnostic precision. A liver condition persisting for at least six months is classified as chronic, requiring continuous monitoring and timely intervention.

The dataset comprises both positive and negative cases, helping to train models that can distinguish between healthy and diseased liver conditions effectively. A confusion matrix visually represents the classifier's performance in predicting liver disease by displaying true positives, true negatives, false positives, and false negatives. With a well-structured training dataset, the proposed classification techniques enhance accuracy and reliability. By leveraging machine learning classifiers, the system efficiently categorizes good and bad values, demonstrating high predictive accuracy and aiding in early detection and treatment planning.

## VI. FUTURE WORK

1) Integration of Medical Imaging: Incorporate liver ultrasound, MRI, and CT scan data to improve diagnostic accuracy by combining image analysis with tabular data.
2) Advanced Deep Learning Techniques: Implement CNNs for medical image processing and RNNs/LSTMs for analyzing time-series patient data to enhance predictive capabilities.
3) Real-Time Patient Monitoring: Develop a system for continuous monitoring of liver health through wearable devices and IoT sensors, enabling real-time alerts for high-risk patients.
4) Enhanced Feature Selection: Utilize advanced feature engineering techniques like AutoML and genetic algorithms to identify the most influential biomarkers for better prediction.
5) Cloud-Based Deployment: Deploy the model on cloud platforms for scalability, remote access, and integration with electronic health records (EHR) for seamless healthcare integration.
6) Explainable AI (XAI) Integration: Implement interpretability techniques to help medical professionals understand the model’s predictions, increasing trust and usability.
7) Multi-Algorithm Hybrid Approach: Combine multiple ML models such as ensemble learning (Stacking, Boosting) to enhance accuracy and reduce bias in classification.
8) Secure Data Handling & Privacy Compliance: Implement encryption, access control, and compliance with healthcare regulations (e.g., HIPAA, GDPR) to protect patient data.
9) User-Friendly Web & Mobile Application: Improve the interface for easier data entry, visualization of results, and actionable insights for healthcare professionals and patients.
10) Personalized Treatment Recommendations: Leverage AI to provide tailored lifestyle and medical intervention suggestions based on patient history and predicted risk factors.

## REFERENCES

[1] C. Huang and R. Ogawa, ‘‘The vascular involvement in soft tissue fibrosis Lessons learned from pathological scarring’’, Int. J. Mol. Sci., vol. 21, no. 7, p. 2542, Apr. 2020, doi: 10.3390/ijms21072542.
[2] D. Q. Huang, N. A. Terrault, F. Tacke, L. L. Gluud, M. Arrese, E. Bugianesi, and R. Loomba, ‘‘Global epidemiology of cirrhosis— Aetiology, trends and predictions,’’NatureRev.GastroenterologyHepatology,vol.20,no.6,pp.388– 398, Jun. 2023.
[3] D.W.Crabb,G.Y.Im,G.Szabo,J.L.Mellinger,andM.R.Lucey,‘‘Diagnosis and treatment of alcohol-associated liver diseases: 2019 practice guidance from theAmericanAssociationfortheStudyofLiverDiseases,’’Hepatology,vol.71, no. 1, pp. 306–333, Jan. 2020, doi: 10.1002/hep.30866.
[4] J. F. Creeden, D. M. Gordon, D. E. Stec, and T. D. Hinds, ‘‘Bilirubin as a metabolic hormone: The physiological relevance of low levels,’’ Amer. J. Physiol.-Endocrinol.Metabolism,vol.320,no.2,pp.E191–E207,Feb.2021,doi: 10.1152/ajpendo.00405.2020.
[5] J.Trebicka,P.Bork,A.Krag,andM.Arumugam,‘‘Utilizingthegutmicrobiome in decompensated cirrhosis and acute-on-chronic liver failure,’’ Nature Rev. Gastroenterology Hepatology, vol. 18, no. 3, pp. 167–180, Nov. 2020, doi: 10.1038/s41575-020-00376-3.
[6] L. Fabris and M. Strazzabosco, ‘‘Rare and undiagnosed liver diseases: Challenges and opportunities,’’ Transl. Gastroenterol. Hepatol., vol. 6, p. 18, Apr. 2021, doi: 10.21037/tgh-2020-05.
[7] M. P. Diaz-Soto and G. Garcia-Tsao, ‘‘Management of varices and variceal haemorrhage in liver cirrhosis: A recent update,’’ Therapeutic Adv. Gastroenterology, vol. 15, Jan. 2022, Art. no. 175628482211017, doi: 10.1177/17562848221101712.
[8] N.V.Bergasa,ApproachtoPatientWithLiverDisease.London,U.K.:Springer, 2022, pp. 5–26, doi: 10.1007/978-1-4471-4715-2_2.
[9] N. Tanwar and K. F. Rahman, ‘‘Machine learning in liver disease diagnosis: Currentprogressandfutureopportunities,’’IOPConf.Ser.,Mater.Sci.Eng.,vol. 1022, no. 1, Jan. 2021, Art. no. 012029. [Online].
[10] P.Ginès,A.Krag,J.G.Abraldes,E.Solà,N.Fabrellas,andP.S.Kamath,‘‘Liver cirrhosis,’’ Lancet, vol. 398, no. 10308, pp. 1359–1376, 2021.