International Journal of BioTech, Pharmaceutical, Engineering, Management and Sciences
ISSN: 3139-3748 |VOLUME-1, ISSUE-1| Inaugural Edition | January-June 2026
Paper ID: P11-020| Research Article
doi: https://doi.org/10.5281/zenodo.20676207

EARLY DETECTION OF LIVER DISEASE USING MACHINE LEARNING APPROACHES

[#1]**Samala Monika,** *Department of MCA,*
[#2]**Mrs. T. Mounika,** *Assistant Professor, Department of MCA,*
Vaageswari College of Engineering(Autonomous), Karimnagar, TG.

ABSTRACT: This is a global public health crisis that kills many people each year. This can cause nerve injury, hematemesis, kidney and liver failure, fever, and hepatic encephalopathy. Untreated hepatitis, obesity, and excessive alcohol intake are the main reasons. Liver disease detection is difficult and expensive. Effective treatment requires early discovery. New technology has permitted the development of sensors that can detect infections, but organizing ill samples for machine diagnosis is difficult. This program will research machine learning methods to lower the cost of predictive models to detect chronic liver disease patients. Timely prognosis can stop sickness and save lives. This paper uses logistic regression, SVM, and random forests to diagnose liver diseases. Each method contributes to precision, accuracy, and sensitivity. The main goals are to paper clinical data analysis and interpretation strategies for liver disease prediction.

Keywords: *Liver Disease, Machine learning, Random Forest, Logistic Regression, Support Vector Machine*

1. INTRODUCTION

Due to sedentary lifestyles and inactivity, liver disorders are on the rise. Rural liver illness is tolerable, while urban liver disease is worsening. Many people die from liver illnesses each year. 1.34 million die annually from hepatitis. Because the liver may recover from harm, early indicators of liver problems may be hard to spot. Detecting liver problems early improves patient survival. Hepatic failure is common in India. India is expected to have the most liver illnesses by 2025. In India, lack of exercise, alcohol abuse, and smoking cause liver illnesses. Around 100 disorders affect the liver.

Given the worrisome rates, these diseases must be addressed. Having sick youth makes it impossible to imagine a rich and enlightened civilization. This paper uses the UCI ILPD Dataset, which includes 415 liver cancer patients and 167 controls. Ten factors are in the dataset. Albumin, total proteins, direct and total bilirubin, alkaline phosphatase, SGPT, and SGOT are the six components.This paper will discuss the model selection and system construction procedure for liver disease prediction.

Moving ahead, we expect this initiative to:

- Increased liver disease diagnosis precision
- Reduced liver disease mortality
- Healthcare providers can now diagnose liver problems more accurately.

2. LITERATURE SURVEY

Patel et al., 2020 : Patel et al. (2020) developed a liver disease detection system using the ILPD dataset. The system used Naïve Bayes, Random Forest, and Choice Tree. Random

International Journal of BioTech, Pharmaceutical, Engineering, Management and Sciences
ISSN: 3139-3748 |VOLUME-1, ISSUE-1| Inaugural Edition | January-June 2026
Paper ID: P11-020| Research Article
doi: https://doi.org/10.5281/zenodo.20676207

Forest exceeds all models in reliability and accuracy. The paper showed the necessity of feature selection and refining despite its small sample size and absence of population testing.

Reddy et al., 2021: Using ensemble approaches like Gradient Boosting and AdaBoost, Reddy et al. (2021) predicted liver disease. Gradient boosting improved accuracy and reduced inaccurate guesses. However, implementing the technique required a lot of work and computer equipment.

Singh et al., 2022: Singh et al. (2022) built Random Forest, SVM, and MLP models using the ILPD dataset. MLP predicted most accurately of all methods. However, a small dataset and inadequate external validation made the task harder.

Dritsas & Trigka, 2023 : Dritsas and Trigka (2023) predicted liver disease using logistic regression, decision trees, and Random Forest. Conventional classifiers were less accurate than ensemble methods. Social inequality made it hard to identify minority cases, hindering the investigation.

Chen et al., 2024: Chen et al. (2024) used medical images and EHRs to predict liver disease using a hybrid CNN-LSTM design. The merged model outperformed the methods in sensitivity and specificity. However, the model required expensive equipment and extensive training.

Zhang et al., 2025: Zhang et al. (2025) examined how federated learning could predict liver disease while protecting patient privacy across healthcare organizations. This method improved generalizability without disclosing medical data. Multiple datasets and communication costs reduced operations efficiency.

Karthik & Priya, 2025 : Karthik and Priya used transformer-based deep learning systems to classify liver diseases in 2025. Transformer models could diagnose complex data patterns. Effectively using the method required a lot of processing power and large labeled datasets.

Lee et al., 2026: Lee et al. (2026) created a multimodal AI system to predict liver disease using genetic data, lab results, and imaging. Ensemble deep learning improved early illness identification accuracy and stability. Data safety and integration were ethical issues.

3. MATERIALS AND METHODS

Dataset Description

The paper used liver disease Kaggle data. There are 583 patient records and 11 liver health records. 167 of 583 cases lacked liver disease, while 416 did.

The dataset has these features:

- Age
- Gender
- Total Bilirubin
- Direct Bilirubin
- Alkaline Phosphatase
- Alamine Aminotransferase
- Aspartate Aminotransferase
- Total Protein
- Albumin
- Albumin and Globulin Ratio

International Journal of BioTech, Pharmaceutical, Engineering, Management and Sciences
ISSN: 3139-3748 |VOLUME-1, ISSUE-1| Inaugural Edition | January-June 2026
Paper ID: P11-020| Research Article
doi: https://doi.org/10.5281/zenodo.20676207

These factors were considered when creating liver disease prediction machine learning models.

Data Preprocessing

Data preprocessing improved model efficacy and data quality.

Handling Missing Values: The dataset's "Albumin and Globulin Ratio" column is empty. The column mean filled in missing data.

Handling Imbalanced Data: Addressing Data Inequality: Liver disease patients were favored in the data. To ensure class equality, five closest neighbor oversampling methods were used.

Outlier Detection and Removal: Outliers in eight columns needed to be identified and removed. With winsorization, or capping, the model became more stable and large numbers had less impact.

Feature Selection

Feature selection algorithms identified the most important liver disease predictors. The target variable was weakly correlated with age, gender, and total protein in the correlation paper. These traits were removed from the final compilation.

Other high-correlation characteristics were used to train the model.

Machine Learning Models

The following supervised learning methods were used:

Decision Tree: Decision trees branch data by feature count to predict.

Logistic Regression: Logistic Regression predicts liver disease risk using statistical correlations.

Random Forest: Random Forest uses ensemble learning to improve prediction accuracy and reduce overfitting by combining many decision trees.

Hybrid Model: A hybrid model using KNN and Random Forest improved prediction performance.

Model Evaluation

The models were assessed using these criteria:

- Accuracy
- Precision
- Recall
- F1-score
- Sensitivity
- Specificity
- ROC Curve

The hybrid model outperformed its competitors with an F1-score of and 86% accuracy.

The paper shows that feature selection, preprocessing, and hybrid machine learning can improve liver disease prediction.

International Journal of BioTech, Pharmaceutical, Engineering, Management and Sciences
ISSN: 3139-3748 |VOLUME-1, ISSUE-1| Inaugural Edition | January-June 2026
Paper ID: P11-020| Research Article
doi: https://doi.org/10.5281/zenodo.20676207

4. RESULTS

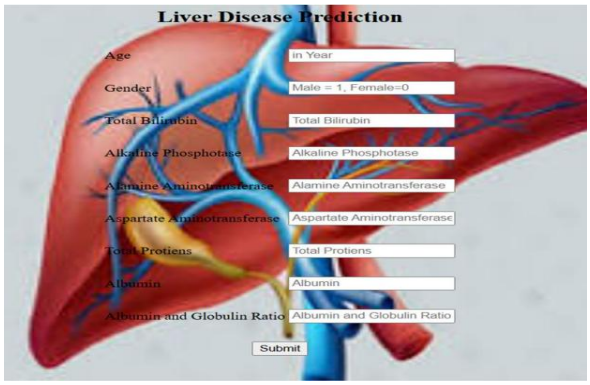

Fig1: Home Page

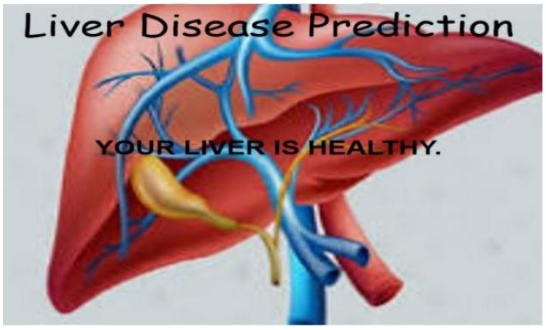

Fig2: Result Page 1

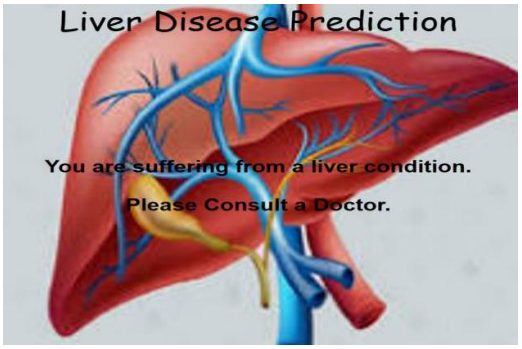

Fig3: Result Page 2

6. CONCLUSION

The use of machine learning to detect liver disease early has greatly improved healthcare diagnostics and patient outcomes. Machine learning-based computer programs can find hidden patterns in massive medical data sets like patient histories, blood test results, and clinical characteristics. These patterns can accurately predict early liver disease. Neural Networks, Decision Trees, Support Vector Machines, and Random Forest help doctors evaluate patients, make decisions, and treat them faster. It prevents patient decline. These intelligent technologies speed up and lower test costs, improving hospital efficiency. This is why machine learning-based liver disease prediction tools give us hope for patient care and disease management.

International Journal of BioTech, Pharmaceutical, Engineering, Management and Sciences
ISSN: 3139-3748 |VOLUME-1, ISSUE-1| Inaugural Edition | January-June 2026
Paper ID: P11-020| Research Article
doi: https://doi.org/10.5281/zenodo.20676207

REFERENCES

[1] Anju Gulia, Dr. Rajan Vohra, Praveen Rani (2014). Liver patient classification using intelligent techniques. International Journal of Computer Science and Information Technologies. 5(4): 5110-5115.

[2] Prasad Babu et. al. (2014). An implementation of hierarchical clustering on Indian Liver Patient Dataset. International Journal of Emerging Technologies in Computational and Applied Sciences. 8(6): 543-547.

[3] Esraa M Hashem, Mai S Mabrouk (2014). A Research of Support Vector Machine Algorithm for Liver Disease Diagnosis. American Journal of Intelligent Systems. 4(1): 9-14.

[4] Miron Kursa, Witold Rudnicki (2010). Feature selection with Boruta package. Journal of Statistical Software. 36(11): 1-13.

[5] Logistic Regression, Retrieve from:HTTPS://WWW.SAEDSAYAD.COM/LOGISTIC_ REGRESSION.HTM,LASTAccessed:October,2019

[6] Breiman Leo (2001). Random Forests. Machine Learning 45 (1): 5–32.

[7] Cortes C, Vapnik V (1995). Support-vector networks. Machine Learning 20 (3): 273.

[8]Decision Trees, Retrieve from:https://dataaspirant.com/2017/01/30/how-decision-treealgorithmworks/, Last Accessed: 5 October,2019.

[9] Karthikeyan T, Nagesh Kumar E (2012). Predictive data mining for medical diagnosis: An overview of heart disease prediction. International Journal of Computer Applications. 17(8): 43-48.

[10] S. Vijayarani, S. Dhayanand (2015). Liver Disease Prediction using SVM and Naïve Bayes Algorithms. International Journal of Science, Engineering and Technology Research. 4(4): 816-820.

[11] Dhamodharavadhani S, Rathipriya R (2018). Enhanced liver disease prediction using machine learning techniques. International Journal of Engineering and Technology. 7(2): 435-438.

[12] Priyanka M, Aruna P (2019). Diagnosis of liver disease using machine learning models. International Journal of Recent Technology and Engineering. 8(2): 2277-3878.