Journal of Medical, Clinical and Surgical Case Reports

Research Article

WRJMCSCR-26-025

Implementation and Study on Liver Cirrhosis Disease Diagnosis Prediction using a Method for Machine Learning Algorithms: A Comparative Approach Analysis

Rajani Kumari* **and Daya Shankar Singh**

Madan Mohan Malaviya University of Technology, Gorakhpur, India

***Correspondence:** Rajani Kumari, Madan Mohan Malaviya University of Technology, Gorakhpur, India; E-mail: kumarirajani035-@gmail.com; DOI: https://doi.org/10.56147/jmcscr.2.2.25

Citation: Kumari R, Singh DS (2026) Implementation and Study on Liver Cirrhosis Disease Diagnosis Prediction using a Method for Machine Learning Algorithms: A Comparative Approach Analysis. J Med Clin Surg Case Reports 2: 25.

Abstract

Background: We explored anonymized patient records from the 'Aadarshvelu Liver-Cirrhosis-Stage-Classification' Kaggle dataset, which is rich in varied laboratory and clinical details. By visualizing model performance through confusion matrices, we uncovered how the quality of the training data shapes classifier accuracy.

Method: Our findings reveal that machine learning not only boosts diagnostic accuracy but also makes liver cirrhosis detection more cost-effective. By testing models on real clinical data, we identified the most powerful predictive techniques and crafted actionable recommendations for healthcare. These innovations pave the way for earlier, more reliable diagnoses, sharpen clinical decision-making and inspire the development of automated tools to manage liver disease.

Findings: The experimental dataset is constructed from real clinical medical records and the model's empirical and comparative experiments are carried out, achieving an accuracy of 94.05% and a macro-averaged F1 of 95.05%.

Interpretation: The Liver Cirrhosis Disease Diagnosis Prediction coding model, based on labeled attention and proposed in this paper, can effectively improve the performance of coding medical records using Machine Learning and compared with the best-performing algorithm, achieves better results.

Keywords: Liver Disease (liver patient data); Prediction; Classification; Computational intelligence; Machine learning; Supervised learning; Logistic regression; Nearest neighbors; Support vector machine; Random Forest; Naïve bayes; Adaboost; Gradient boosting; MLP classifier

Received date: July 10, 2026; **Accepted date:** August 05, 2026; **Published date:** August 14, 2026

Introduction

Chronic liver disease and cirrhosis represent the twelfth leading cause of mortality in the United States, as reported by the National Center for Health Statistics [6]. Mortality rates may be underreported. This study seeks to provide additional insight into the true impact of liver-related mortality across the United States. Recent research highlights that liver-related deaths in the United States are frequently underestimated. Worldwide, liver cirrhosis stands as a major yet preventable threat, claiming over a million lives across 187 countries from 1980-2010 [1]. While better data collection and advanced modeling have sharpened mortality estimates, inconsistent approaches still cloud our understanding. (development, validation and evaluation of a simple machine learning model to predict cirrhosis mortality) machine learning is a promising tool for predicting and diagnosing disease risk, but its accuracy depends on sound model selection, high-quality data and clinical context. Many studies rely on small or skewed datasets, limiting their reach. Despite machine learning's success in other fields, cost-

 Copyright © 2026 | https://medical-clinical-surgery-reports.wren-research-journals.com/1

effectiveness and seamless clinical integration often slip through the cracks. (Science Intelligence Diagnosis and Prognostic Stratification of Hepatocellular disease: Evidence for Clinical features Applications and Future Perspectives) Current efforts aim to reduce diagnostic costs and improve accuracy, yet hurdles remain to making models interpretable and practical.

Comparative studies typically pit algorithms like Logistic Regression, k-Nearest Neighbors, Support Vector Machines, Random Forests, Naïve Bayes, AdaBoost, Gradient Boosting and MLP Classifiers against each other using metrics such as accuracy, precision, recall, F1 score and ROC curves [5]. More research is needed to align algorithms with specific clinical needs and to test their generalizability. Ultimately, while machine learning holds great promise for liver disease detection and management, rigorous evaluation is vital to ensure lasting benefits in healthcare.

Background

Data mining

Data mining methods have become valuable tools for clinicians, with many studies applying these techniques to disease risk prediction, prevention and management [11]. According to a study by Brockmann and Hunt, data mining methods were used to identify significant patterns in heart rate variability across different time points and exercise intensities during treadmill running. Create and use a data mining method to study which medical features are present or missing in breast cancer datasets. This can help find important patterns and make diagnoses more accurate [18].

The authors employed multiple machine learning and data mining techniques to analyze Clinical information and Artery Disease (CAD) in patients [21]. The study evaluated patient characteristics, including age, gender, blood pressure, cholesterol levels, Electrocardiogram (ECG) results, chest pain type and additional clinical features. Among all classification methods tested, the proposed model achieved the highest diagnostic accuracy.

Knowledge Discovery in Databases (KDD) includes data mining, as shown in **Figure 1** data mining is the main extraction step in the KDD23 process, which also involves transforming and interpreting data. Data mining methods have become valuable tools for clinicians, with many studies applying these techniques to disease risk prediction, prevention and management 11. Data mining involves hypothesis testing and model interpretation. In this study, data mining identified significant patterns in heart rate variability 12, 19. The liver, a powerhouse among human organs, sits in the upper right abdomen just below the diaphragm. Harnessing clinical and laboratory data to classify and predict liver disease sharpens diagnostic precision and guides better decisions.

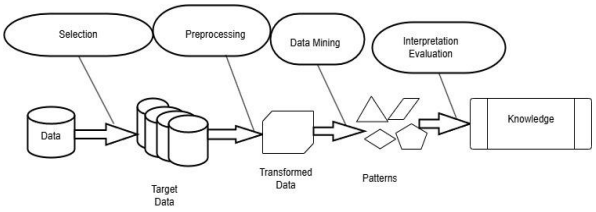


Figure 1: The data mining or Knowledge Discovery in Databases (KDD) process 23.

As shown in **Figure 2**, liver disease does not always2 lead to death. Many patients do better when they receive proper screening, diagnosis and treatment. Without coordinated plans, strong government support and enough funding for care and research, liver diseases will likely continue to cause preventable deaths and put pressure on national health care systems [2].

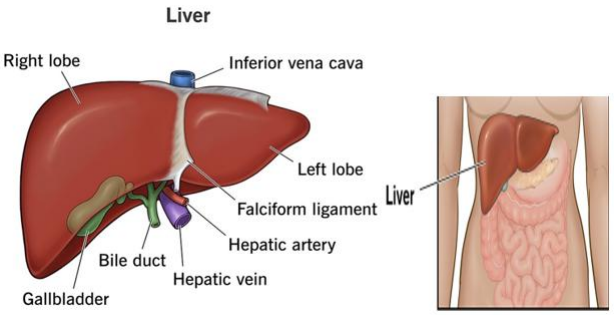


Figure 2: The location of the liver in the human body.

This vital organ orchestrates hormone and protein production, maintains blood glucose levels and regulates blood clotting.

According to Mayo Clinic, liver disease can result from many different causes. The paper in **Figure 3** titled "Analysis of liver disorder using data mining algorithm" was proposed by 3.

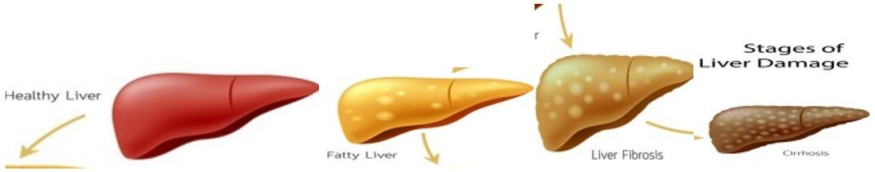


Figure 3: Stage of liver damage.

Copyright © 2026 | https://medical-clinical-surgery-reports.wren-research-journals.com/1

This likely includes a figure illustrating the typical data analysis process or comparison of algorithmic accuracies

Liver cirrhosis presents unique challenges

Machine learning in medical diagnosis. We tackled imbalanced datasets with random under-sampling and SMOTE, boosting classifier performance. Blood test data proved invaluable for segmenting cirrhosis and supporting precise diagnoses, underscoring their critical role in evaluating liver health. Causes of Liver Disease 40 are show in **Table 2**.

Table 2: Clotting profile.

		Test	Finding
Prothrombin Time (PT)	Prolonged	NA	NA
INR	Increased (key marker of severity)	NA	NA
Test	-	-	-
Finding	-	-	-

Infection

The liver can become infected by parasites. Blood tests are very important in **Table 3** because they reveal hidden details about the liver, especially when it is damaged. Contagion is contracted through contaminated food or water or by contact with an infected person. Hepatitis A, C and B are liver infections that can affect people.

Table 3: Complete Blood Count (CBC).

Parameter	Finding
Platelet count	↓ (early sign of cirrhosis)
Hemoglobin, WBC	↓ (anemia)

Immune system abnormality

Certain diseases affect the immune system, causing it to attack the body's tissues, including the liver. Such conditions may include autoimmune hepatitis.

Inheritance

Rare genetic factors inherited from one or both parents can lead to the accumulation of various compounds in the liver, resulting in hepatic damage.

Cancer and related conditions

Malignancies contributing to liver disease include liver adenoma, bile duct cancer and primary liver cancer.

Other causes

Additional etiologies include prolonged alcohol consumption, hepatic steatosis (Nonalcoholic Fatty Liver Disease (NAFLD)), certain medications and specific herbal formulations.

Risk aspects

Factors that might raise the risk of liver diseases are excessive usage of liquor being overweight, type 2 diabetes, tattoos, body piercings, drug injection with used needles, transfusion of blood, exposure to foreign blood, unprotected intercourse, exposure to chemicals and inheritance. Compounds and substances produced by the liver include bilirubin, albumin, alkaline phosphatase, aspartate aminotransferase and globulin. These are essential to normal hepatic function.

Bilirubin

Bilirubin is a yellow compound produced when heme breaks down in vertebrates. It is present in bile and urine. If bilirubin levels get too high, it can lead to disease. Bilirubin causes the yellow color in bruises and the yellowing of skin seen in jaundice. When bilirubin breaks down, it forms stercobilin, which gives feces their brown color and urobilin, which gives urine its straw-yellow color.

Alkaline phosphatase

In humans, alkaline phosphatase is found in all tissues, but it is most common in the liver, intestinal lining, bile ducts, bone, kidney and placenta. In the blood, the main types of alkaline phosphatase are from the skeleton and the liver. A hybrid intelligent system integrates machine learning and artificial intelligence to diagnose microalbuminuria in individuals type 2 diabetes. Diabetes, obesity, smoking and age have differential impacts on men and women [14]. This approach eliminates the need for clinicians to directly measure urinary albumin [13]. The integration of various artificial intelligence and machine learning techniques enhances diagnostic accuracy. In children, most alkaline phosphatase comes from the skeleton. Most mammals, including humans, have these types of alkaline phosphatases:

ALPI

This is an intestinal protein with a molecular mass of 150 kDa.

ALPL

This type is not specific to one tissue and is mainly found in the liver, kidney and bone.

ALPP

This type is found in the placenta and is also called the Regan isozyme.

GCAP

This type is found in germ cells and is a form of alkaline phosphatase.

Aspartate Aminotransferase (AST)

This enzyme is found mostly in the heart and liver, with small amounts in the kidneys and muscles. Normally, AST levels are low in the blood. When muscle or liver cells are

Copyright © 2026 | https://medical-clinical-surgery-reports.wren-research-journals.com/1

damaged, AST enters the blood. The AST testing is used to detect liver damage or problems. Serum is the most abundant albumin protein in the blood. They bind to Thyroxine (T4), water, cations such as Ca2+ and Na+, hormones, fatty acids, bilirubin and drugs. Their main job is to regulate and maintain oncotic pressure in the blood show in **Table 7**.

Table 7: Example correlation matrix (Pearson's r).

Feature 1	Feature 2	Correlation	Interpretation
	Direct bilirubin	0.88	Strong positive correlation direct bilirubin is a component of total bilirubin.
Albumin	A/G ratio	0.72	Strong positive albumin directly influences the A/G ratio.
ALT	AST	0.81	Strong positive both are liver enzymes elevated in liver damage.
Total protein	Albumin	0.69	Moderate to strong positive albumin contributes to total protein.
Age	Other attributes	0.25	Weak positive age has a limited direct linear association. Attributes.

Globulin

Globulins are protein particles that are heavier than albumin and although they do not dissolve in pure water, they do in weak salt solutions. Some globulins are produced by the liver. According to the Cleveland Clinic, normal serum globulin levels are between 2.0 and 3.5 grams per deciliter. There are various types, such as alpha, beta and gamma-globulins. Imbalances, including excess production, can affect liver health and certain liver diseases are characterized by changes in the proportions of these proteins. Additionally, advances in diagnostic technology have enabled the use of machine learning algorithms in equipment such as CT and ECG machines. These algorithms diagnose various diseases in clinical research [36]. Cirrhosis is the end-stage of liver disease. It has an asymptomatic stage, called compensated cirrhosis, followed by a rapidly progressive phase with liver dysfunction, called decompensated cirrhosis. The most severe condition is Hepatocellular Carcinoma (HCC) is show in **Table 1**, also called primary liver cancer [3,7].

AST/SGOT.8 to 48 U/L, ALP.40 to 129 U/L, Albumin. 3.5 to 5.0 (G/DL), Bilirubin. 0.1 to 1.2 (Mg/DL)

Total protein

GGT.8 to 61 U/L, LD.122 to 222 U/L, PT.9.4 to 12.5 seconds

Blood tests used to diagnose/assess liver cirrhosis

Liver Function Tests (LFT)

Table 1: These are the most important blood tests.

	What it indicates in cirrhosis	Test
ALT (SGPT)	Liver cell damage (may be normal in advanced cirrhosis)	-
AST (SGOT)	Often higher than ALT in cirrhosis	-
ALP	Bile duct damage	-
Total & direct bilirubin	↑ causes jaundice	-
Albumin	↓ indicates poor liver function	-
Total protein	↓ in chronic liver disease	-
Test	-	-

Related Work

Yan et al.

proposed to investigate the correlation between Child–Pugh degree and the four examinations of Traditional Chinese Medicine (TCM) inspection, listening/smelling, inquiry and palpation in patients with liver cirrhosis and to evaluate whether TCM examination indicators can reflect the severity and progression of liver cirrhosis as classified by the Child-Pugh scoring system [24,37,38]. The authors developed a fuzzy logic system to classify patients with liver cirrhosis into Child-Pugh Classes A, B or C. This method helps clinicians assess disease severity by accounting for uncertainty in clinical measurements [24].

Anu Sebastian and Surekha Mariam Varghese

Said that Survival analysis is a common procedure in the medical field. Predicting life expectancy is a highly significant factor in the decision-making of both patients and doctors. In addition to the correct diagnosis of the ailment, accurate classification and assessment are necessary to commence appropriate and effective treatment. The Child-Pugh classification has been used for many years for the assessment of the severity and prognosis of chronic liver disease and cirrhosis of the liver. It is considered a standard for classifying patients with end-stage liver disease. Fuzzy logic is a powerful technique for modeling complex, imprecise systems efficiently. Intelligent procedure based on Fuzzy logic could be used for the Child-Pugh classification of patients suffering from liver cirrhosis. The life expectancy of patients with different classes of cirrhosis also differs. This paper

Copyright © 2026 | https://medical-clinical-surgery-reports.wren-research-journals.com/1

presents a fuzzy-intelligent technique for classifying and predicting the life expectancy of patients with liver cirrhosis [25].

Insha Arshad and Chiranjit Dutta

Said that alcohol is consumed in excess by millions of people across the world. Alcohol consumption is directly linked to life-threatening liver diseases such as cirrhosis, which may ultimately lead to death. Early detection of liver disease caused by overconsumption of alcohol would help save the lives of many people 26. Detecting liver disease early can lead to timely diagnosis and may result in full recovery in some patients. This paper proposes Detecting and predicting the presence of liver disease using data mining algorithms. We will build a decision tree for the dataset and then generate the rules. After determining the rules, we will use different data mining algorithms to train and test the dataset to detect liver disease. The data were collected from the UCI repository and our training dataset was developed. It consists of 7 different attributes having 345 instances. In the dataset, different categories of blood tests are taken into consideration, which are directly linked to liver diseases that may arise due to excessive alcohol consumption, along with the frequency of alcohol consumption. Based on the type of liver disease detected, a prognosis may be suggested.

Ramkumar N, et al.

Cancer is one of the most dangerous diseases worldwide. It can be organs like the liver, breast and bones. Liver problems are especially serious and often lifelong. Symptoms of liver disease include jaundice, loss of appetite, yellow urine, vomiting, pain in the upper right abdomen, fever and an enlarged liver 27. When cancer starts in the liver, it is called primary liver cancer. If cancer spreads from another part of the body to the liver, it is called secondary liver cancer. The liver is a vital organ in the human body. The American Cancer Society reports that liver cancer rates are highest in Micronesia, Eastern Asia, Southeast Asia and Northern Africa. The most common type of liver is hepatocellular, which is more common in men than in women. Excessive alcohol use is a main cause of liver cancer. Today, data mining and artificial intelligence are used to help predict liver cancer. For example, Bayes' theorem in WEKA can be used to calculate the probability of liver cancer.

Hassoon M, et al.

Current medical research emphasizes diagnosing illnesses by identifying the most influential features for recognition. Medical Data Mining (MDM) is a recent advancement that employs data mining techniques such as classification and clustering to group diseases and symptoms, thereby facilitating diagnosis [28]. This study presents a novel approach to di liver disease, designed to assist clinicians and patients in identifying symptoms, reducing diagnostic time and lowering mortality risk. The proposed method enhances the rules generated by the Boosted C5.0 classification algorithm by applying a Genetic Algorithm (GA), thereby increasing both diagnostic speed and accuracy. Rather than generating rules with an evolutionary algorithm, the genetic algorithm refines and simplifies the rule set produced by another algorithm. A 2021 Science Direct article reviewed various machine learning methods for liver disease prediction, but did not provide detailed comparisons or demonstrate accuracy improvements over previous studies. The experimental results indicate superior performance and faster operation compared to existing methods. Specifically, diagnostic accuracy improves from 81%-93% [29].

Liver function tests and their interpretation

BR Thapa and Anuj Walia

Saying that Liver Function Tests (LFTs) are a helpful screening tool and an effective modality for detecting hepatic dysfunction. Since the liver performs a number of functions, no single test is sufficient to estimate liver function. Clinicians often encounter reports that do not align with the patient's clinical condition and they have difficulty interpreting LFTs [30]. An attempt is being made to study and understand LFTs and to simplify their interpretation using algorithms.

Model selection

Ch Shravya, Pravalika K and Subhani S

Said, the most exciting phase in building any machine learning model is the selection of an algorithm 31. We can use multiple data mining techniques on large datasets. But at a high level, it can be classified into two groups: Supervised and unsupervised learning. Supervised learning is a method in a machine is trained on data with well-labeled input and output. The model can be trained on this training data and can process future data to predict outcomes. They are grouped into Regression and Classification techniques. A regression problem is one in which the result is a real or continuous value, such as “salary” or “weight”. A classification problem is one where the result is a category, like filtering emails as “spam” or “not spam”.

Description of the classification algorithms

Breiman L, Random forests

In random forests, each decision tree uses a random vector from a shared distribution. As more trees are added, the generalization error of the forest levels off at a certain value [32]. This error depends on how strong each tree’s decisions are and how similar the trees results are. When features are randomly chosen to split nodes, the error rates are close to those from AdaBoost (Y. Freund & R. Schapire, Machine Learning: Proceedings of the Thirteenth International Conference), but random forests are better at handling noise. Internal estimates keep track of error strength and correlation, showing how results change as

Copyright © 2026 | https://medical-clinical-surgery-reports.wren-research-journals.com/1

more features are used for splitting. These estimates also help measure the importance of each variable. The same method can be used for regression tasks.

XNB: Explainable class-specific Naive-Bayes classifier Aguilar-Ruiz JS, Romero C andrea C, said

High-dimensional datasets are increasingly prevalent, making feature reduction essential to mitigate overfitting and enhance model accuracy. Many dimensionality reduction techniques use the same feature set across all classes, potentially overlooking class-specific characteristics. The Explainable Class-Specific Naïve Bayes (XNB) classifier addresses this limitation through two primary strategies: employing Kernel Density Estimation to improve the accuracy of posterior probability calculations and selecting distinct feature subsets for each class to emphasize the most relevant variables. Empirical evaluations of high-dimensional genomic data demonstrate that XNB achieves performance comparable to that of traditional Naïve Bayes while enhancing interpretability. By identifying the most significant features for each class, XNB generates concise, class-specific feature sets and clarifies the model's decision-making process. This methodology is particularly valuable in domains where both predictive accuracy and interpretability are critical 35.

$$P\left(\frac{A}{B}\right) = \frac{P\left(\frac{B}{A}\right) * P(A)}{P(B)}$$

Christopher R. Stephens, Hugo Flores Huerta and Ana Ruíz Linares state that the Naive Bayes Approximation (NBA) and its classifier are widely utilized and demonstrate strong performance across various domains [1-4,35]. The algorithm assumes that all input features are independent of one another, given the class label. In reality, this is not always true since many features in real-world data are related [34].

Although the NBA assumes feature independence, which may appear counterintuitive, researchers have offered multiple explanations for its effectiveness and have developed numerous variants. In their study, the authors introduce local error measures associated with likelihood functions for attribute subsets and individual classes. They demonstrate how these local errors aggregate to form a global error across all attributes. This framework enables the identification of conditions under which errors offset or accumulate, allowing researchers to anticipate their impact on classifier performance. These diagnostic tools provide a more nuanced understanding of the circumstances in which the NBA is effective or limited. Additionally, the authors employ these diagnostics to inform feature selection and illustrate their application in a straightforward extension of NBA on several real-world datasets.

K-Nearest Neighbors (KNN)

The k-Nearest Neighbors (kNN) model is a popular and effective tool for classification 33. Still, it has two main drawbacks: it can be slow with large datasets, like those in dynamic web mining and it is very sensitive to the choice of k. In this study, we introduce a new kNN-based classification method to tackle these issues. Our approach builds a kNN model and uses it for classification, with k automatically chosen to achieve the best accuracy for each dataset. This means manual parameter tuning is the finish line, making the process more efficient. We tested our method on public datasets from the UCI Machine Learning Repository. The results show that our kNN-based model matches the classification accuracy of C5.0 and standard kNN, but is more efficient than standard kNN.

Decision Tree (DT)

One of the most popular techniques used in the data mining process is classification and trees. After completing a classification process, a decision is naturally made and both labels are correctly included in the name, though they are usually used separately (*e.g.,* classification trees or decision trees). From now on, it will just say decision trees, since they represent the model's final goal. The greatest benefit of using decision trees lies in their flexibility and ease of understanding. These are mainly an overview of the conceptual steps for building and applying a decision tree to real-world problems. Predictions with higher confidence scores are selected as new labeled examples15.These newly labeled samples are subsequently incorporated into the training set.

Hervonenkis AY, early history of support vector machines

Many of the ideas now being developed in the framework of support vector machines were first proposed by Vapnik VN and Chervonenkis AY (Institute of Control Sciences of the Russian Academy of Sciences, Moscow, Russia), in the framework of the generalized portrait method for computer learning and pattern recognition. The development of these ideas started in 1962 and they were first published in 1964. SVM makes the optimization problem8 computationally feasible- We introduce additional features of the loss function, such as continuity and relate them to analogous features associated with the loss function [36]. Since the results of the section will be used throughout this book in history, we recommend that even experienced readers become familiar with the material.

Linear data set (SVM) Kernel trick-linear kernel-k $(x_i,y_j)=x_i,x_j$

Poly kernel with degree d: k=Gaussian (radial basis function) kernel with width k $(x_i,x_j)=(1+x_i+x_j)^d$

Sigmoid with parameter k & c =) =k(xi,xj)

Copyright © 2026 | https://medical-clinical-surgery-reports.wren-research-journals.com/1

$$=\tanh(k.xi.xj+c)$$

$$wT.x+b=0$$

$$r=wT.xi+b$$

$$x = \frac{\mathrm{wT.\,xi} + \mathrm{b}}{||\mathrm{w}||}$$

Hyperplane: Classifier Linear Support Vector Machine (LSVM)

$$f(x) = sign(wT.x + b)$$

Logistic regression

According to a recent study, machine learning techniques that use serum bile acids can help predict cirrhosis in patients with chronic hepatitis B [1]. Existing models predict disease categories but do not clearly show the chance of developing the disease. This paper main a model to predict the probability of liver disease. According to a report in Frontiers in Medicine, the model uses logistic regression and was tested on the Indian Liver Patient Dataset (ILPD), which contains 583 records with 11 different attributes, including demographic data like age and gender, as well as nine progressions and makes more timely decisions. In the proposed model, we have leveraged logistic regression. As the general formula of logistic regression in terms of the sigmoid function are show in Eq. 1 and Eq. 2

$$P=\frac{1}{1+\exp^{-z}} \ldots (1)$$

Where represents the probability of an event's occurrence (dependent variable), whereas is the linear combination of the independent variables4. Can also be expressed as follows:

$$z = \beta 0 + \beta 1v1 + \beta 2v2 + \cdots + \beta nvn \ldots (2)$$

By using the linear combination function in Eq. (1) with the test results as input to predict liver disease progression, the model calculates the chance of disease progression based on these tests. The proposed method uses a machine learning system that cleans the data, applies eight algorithms and evaluates their performance using several metrics to identify the best model for predicting liver disease. The process includes data cleaning, feature selection and classification.

In **Figure 4** Key to the advancement has been the development of a deeper understanding and theoretical analysis of critical issues in algorithmic construction and learning theory [9].

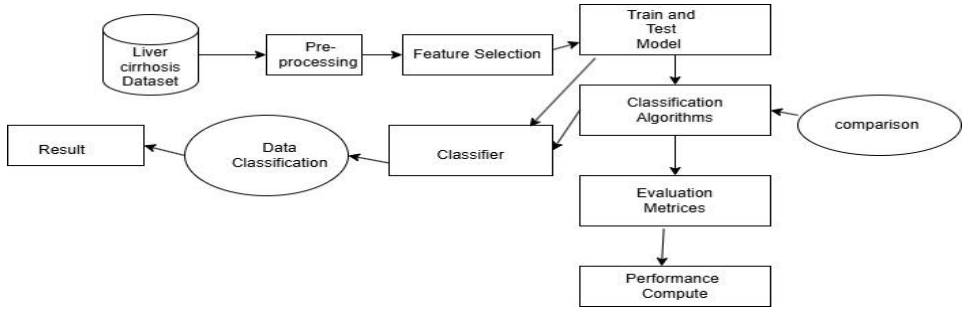


Figure 4: Step out of our proposed work.

Materials and Methodology

This section describes the data and classification methods used in the prediction tests. This study introduces an automated system that uses machine learning to diagnose Coronary Artery Disease (CAD) by analyzing heart rate signals [17]. The primary objective is to predict clinical outcomes for patients with liver disease using data mining techniques. An additional aim is to reduce the time between liver tests and the receipt of results. A confusion matrix is employed to quantify correct and incorrect predictions. Previous studies have examined various classification and prediction methods, including algorithmic combinations to assess accuracy.

Measurement of classification techniques

In this study, we used real-world data to assess how well different classification methods perform. We evaluated those using metrics such as accuracy, sensitivity, specificity, precision and F1 score. The study demonstrates that collecting behavioral data during the clock-drawing process provides valuable diagnostic indicators [16]. Machine learning models that utilize these nuanced behaviors identify cognitive impairment more accurately than models that rely solely on the completed drawing.

These measures come from the confusion matrix. True Positive (TP) means the prediction correctly says a patient has liver disease. A False Positive (FP) occurs when a prediction incorrectly indicates that a patient has liver disease. True Negative (TN) means the prediction correctly says a patient does not have liver disease. A False Negative (FN) is a prediction that incorrectly indicates a patient does not have liver disease.

This study uses records in **Figure 5** from patients with Indian liver7 ailments to support clinical decision-making. The answer is that the method's accuracy grows as models

Copyright © 2026 | https://medical-clinical-surgery-reports.wren-research-journals.com/1

in illness prediction improve. Precision measures the proportion of the model's positive predictions that are correct. To calculate precision, we look at true positives, true negatives, false positives and false negatives.

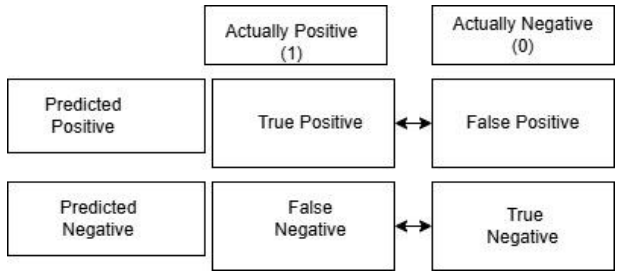


Figure 5: Prediction form for liver cirrhosis.

Methodology

The paper concisely outlines its methodological approach.

Data collection and description

In this experiment, we collected the dataset by downloading the Aadarshvelu liver cirrhosis stage classification dataset for patients from 5. We used 10-fold cross-validation for each algorithm. They were split into 10 parts in the dataset that was generated, with 20,000 for training and 5,000 for testing. (Namineni & Sreerama, 2023) This involved 19 features contributing to the final accuracy metric using nineteen cross-validated models.

x_train: (20000 rows × 2 columns)

x_test: (5000 rows × 2 columns)

The instances in this dataset correspond to individual patient records, with each record providing a unique observation related to liver cirrhosis. These records include 19 distinct clinical and demographic features relevant for classification tasks, such as laboratory measurements and patient characteristics, enabling comprehensive analysis for predictive modeling. This machine learning algorithm classifies a new patient's condition by finding and comparing it to the most similar patients in the dataset22.The dataset was obtained from Kaggle, specifically the 'Aadarshvelu Liver-Cirrhosis-Stage-Classification' dataset and comprises 25,000 medical patient records. Each record represents a patient entry or observation related to liver cirrhosis. The dataset includes 19 features (columns) relevant to classification tasks as show in Fig. 4 and Fig. 5, with sample rows provided for illustration, 5. These include both clinical and demographic variables relevant to liver disease and cirrhosis staging. The features typically represent data collected for cirrhosis and are derived from the Mayo Clinic Primary Biliary Cirrhosis (PBC) dataset and similar clinical data sources (as seen in related Kaggle uses) as show in **Table 4**.

Table 4: Features like clinical status.

Numerical	N_Days, age, bilirubin, cholesterol, albumin, copper, Alk_Phos, SGOT, Tryglicerides, platelets, prothrombin, stage
Categorical	Status, drug, sex, ascites, hepatomegaly, spiders and edema
Note: This is generally numeric (int or float) and most classification models try to predict it.	

Status: Patient status (*e.g.,* death, censored, transplant); Drug: Drug or treatment indicator; Sex: Patient sex; Ascites: Presence of ascites; Hepatomegaly: Liver enlargement; Spiders: Spider angiomas present and; Edema: Edema presence in **Table 5**.

Table 5: Categorical (non-numeric) features: These are typically encoded later for ML (*e.g.,* label encoding or one-hot encoding).

Feature	Type	Description
Status	Categorical	Patient status category (*e.g.,* compensated *vs.* decompensated)
Drug	Categorical	Type of medication or drug treatment
Sex	Categorical	Patient's gender
Ascites	Categorical	Whether ascites (fluid in abdomen) is present
Hepatomegaly	Categorical	Whether liver enlargement is present
Spiders	Categorical	Presence of spider angiomas (skin vascular signs)
Edema	Categorical	Presence of edema (fluid/ swelling)

Context

Cirrhosis results from prolonged liver damage, which can be caused by hepatitis or alcohol abuse. The data are sourced from a Mayo Clinic study on Primary Biliary Cirrhosis (PBC) of the liver conducted from 1974 to 1984. The dataset was manually cleaned and synthetic data were incorporated to increase the sample size 5.

Features like clinical or numeric features. Information: these attributes are as follows

- **N_Days:** Number of days between registration and the earlier of death, transplantation or study analysis time in 1986.
- **Status:** Status of the patient C (censored), CL (censored due to liver tx) or D (death).
- **Drug:** Type of drug, D-penicillamine or placebo.
- **Age:** Age in days.
- **Sex:** M (male) or F (female).
- **Ascites:** Presence of ascites N (No) or Y (Yes).
- **Hepatomegaly:** Presence of hepatomegaly N (No) or Y (Yes).

Copyright © 2026 | https://medical-clinical-surgery-reports.wren-research-journals.com/1

Journal of Medical, Clinical and Surgical Case Reports

- **Spiders:** Presence of spiders N (No) or Y (Yes).
- **Edema:** Presence of edema N (no edema and no diuretic therapy for edema), S (edema present without diuretics or edema resolved by diuretics) or Y (edema despite diuretic therapy).
- **Bilirubin:** Serum bilirubin in [mg/dl].
- **Cholesterol:** Serum cholesterol in [mg/dl].
- **Albumin:** Albumin in [gm/dl].
- **Copper:** Urine copper in [ug/day].
- **Alk_Phos:** Alkaline phosphatase in [U/liter].
- **SGOT:** SGOT in [U/ml].
- **Tryglicerides:** Triglycerides in [mg/dl].
- **Platelets:** Platelets per cubic [ml/1000]
- **Prothrombin:** Prothrombin time in seconds [s]
- **Stage:** Histologic stage of disease (1-3].

Tasks performed: Classification, Feature Type Integer and Real. This dataset contains 11 parameters are show in **Table 6**, whereas we chose 10 for further analysis and 1 as the target class.

Table 6: Numerical features, measured clinically many are blood test results.

Feature	Type	Description
N_Days	Integer	Number of days since reference (*e.g.,* registration)
Age	Integer	Age of patient (in days or years depending on dataset version)
Bilirubin	Float	Bilirubin level in blood (indicator of liver function)
Cholesterol	Float	Cholesterol measure
Albumin	Float	Albumin level (protein produced by liver)
Copper	Float	Copper level
Alk_Phos	Float	Alkaline phosphatase (liver enzyme)
SGOT	Float	Serum glutamic-oxaloacetic transaminase (liver enzyme)
Tryglicerides	Float	Triglycerides level in blood
Platelets	Float	Platelet count in blood
Prothrombin	Float	Prothrombin time (blood clotting measure)
Stage	Integer/float	Target label stage of liver cirrhosis
Note: These numerical measurements are important clinical indicators of liver health/disease severity.		

Target variable stage

The class label indicates the cirrhosis stage (severity).

Data preprocessing and feature selection

Preprocessing

The feature was performed using a combination of statistical criteria and clinical relevance. Initially, a correlation matrix was generated and visualized as a heatmap to assess relationships among all features. Features with a Pearson correlation coefficient below 0.2 with the target variable (cirrhosis stage) were considered to have a weak association and were candidates for omission. In addition, pairs of features with a high intercorrelation (above 0.8) were examined to avoid multicollinearity, retaining only the variable with stronger clinical justification or predictive value. (Multiparametric ultrasonics of significant liver fibrosis: A machine learning-based analysis [10]. The rationale for omitting features was guided by both data-driven thresholds and domain knowledge, including the established relevance of liver function enzymes and key demographic characteristics for cirrhosis staging. The entire preprocessing and feature selection process was implemented in Python using Google Colab (online).

- Performed data cleaning and preprocessing (handling missing values, removing irrelevant data).
- Used correlation heatmap to select the most relevant features and remove weakly correlated ones.
- This reduces dimensionality and avoids noise.

Cirrhosis occurs due to in **Figure 6** show in above chronic liver disease, leading to extensive scarring [5]. Common causes include hepatitis and prolonged alcohol consumption. The dataset analyzed in this study originates from a Mayo Clinic5 investigation of Primary Biliary Cirrhosis (PBC) conducted between 1974 and 1984. I manually cleaned the dataset and supplemented it with synthetic data to increase the sample size.

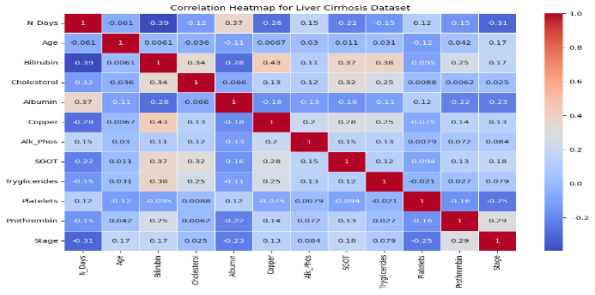


Figure 6: Heat map.

Copyright © 2026 | https://medical-clinical-surgery-reports.wren-research-journals.com/1

- Correlation matrix.
- Pearson correlation-heat map: Importance of attribute.
- Data acquisition-Liver disease dataset. Data cleaning-remove value (null value).
- Feature extraction-identify important attributes (Pearson).

Sex

F22137 M2863

Name, count, dtype: int64

In **Figure 7 and 8** From 1974 to 1984, 424 PBC patients referred to the Mayo Clinic5 qualified for the randomized placebo-controlled trial testing the drug D-penicillamine.

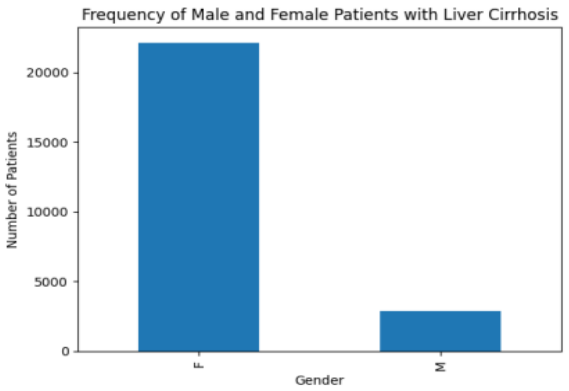


Figure 7: Comparison with the frequency of males and females.

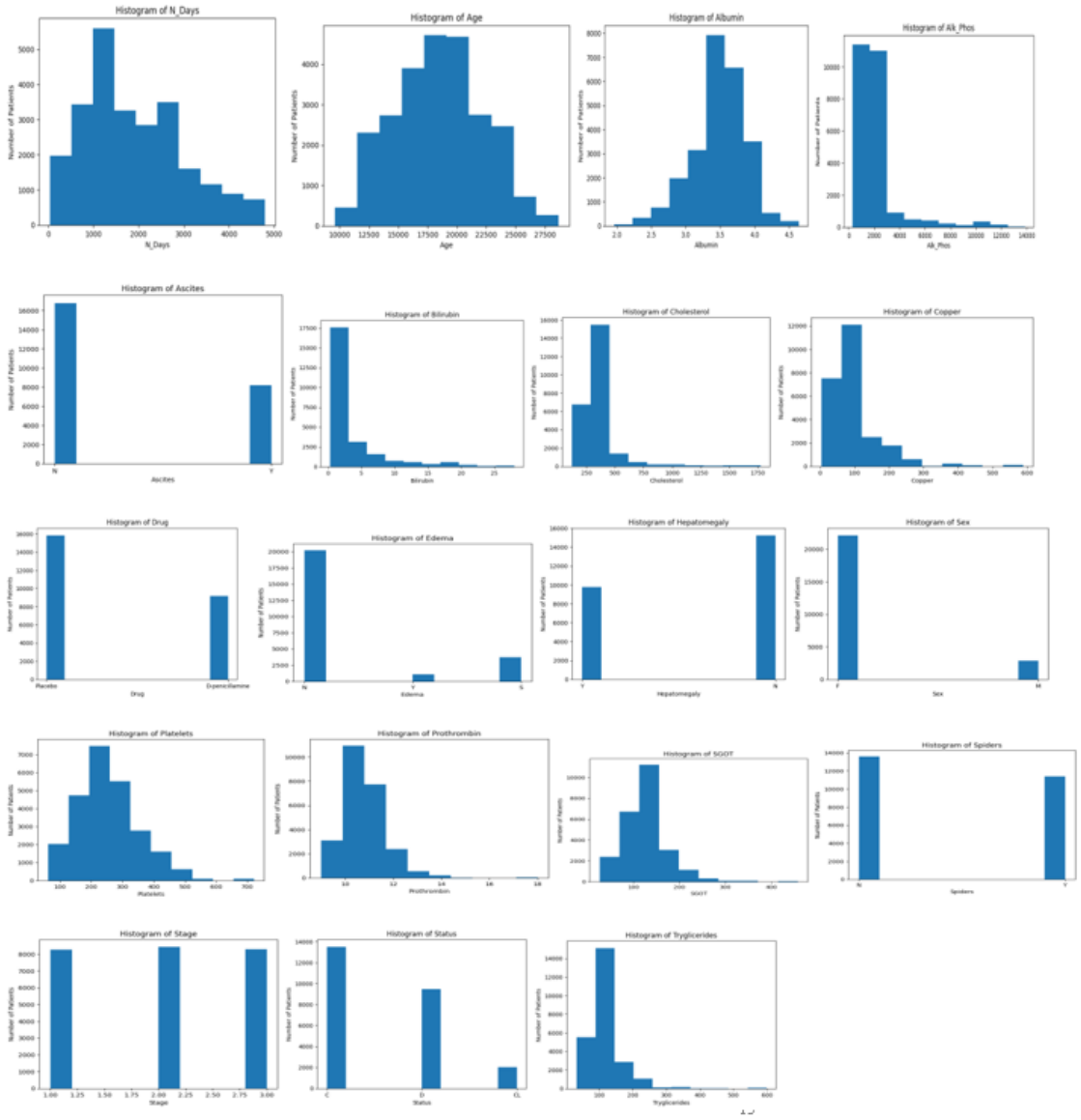


Figure 8: Histogram of various patients' frequency distributions for the attribute.

Copyright © 2026 | https://medical-clinical-surgery-reports.wren-research-journals.com/1

Extraction

Identify important attributes, Output of Feature Extraction.

- X_scaled → Extracted & normalized features
- y → Liver cirrhosis stage/disease class.

The Implementation of Classification

Algorithms

Classification algorithms used

Implemented eight supervised learning algorithms for prediction.

Algorithms tested

The proposed work used eight algorithms: Logistic Regression, K-Nearest Neighbors, Support Vector Machine, Random Forest, Naïve Bayes, AdaBoost, Gradient Boosting and MLP Classifier.

Performance evaluation

Evaluation criteria

We analyze the eight algorithms described above through experimental comparison. Models were compared based on a confusion matrix, which can be used to calculate the following performance evaluation metrics for a model: accuracy, precision, recall and F1 score.

Accuracy

The affectability test indicates the rate of effective identification of the patient's liver disease. It mainly demonstrates the test's positive instances. It is additionally known as the Recall and True Positive Rate (TPR).

It is the fraction of predictions our model got right. The higher the value of accuracy, the better the model. Accuracy = number of correctly predicted records / Total number of records.Accuracy=(TP+TN) / (TP+TN+FP+FN).

Precision

It is the number of positive cases that were correctly predicted. Higher precision represents a better model. Precision = TP / (TP+FP).

Recall (sensitivity) precision/recall

It is the number of actual positive cases that were correctly predicted. It is called the True Positive Rate (TPR) or Sensitivity. Higher values imply a better model. Recall = TP / (TP+FN). Precision is defined as the ratio of positives to all positives, *i.e.,* the number of true positives divided by the total number of positives. Recall is defined as the number of relevant instances in the search space that the system finds in the retrieved data.

Specificity sensitivity/specificity

It is the proportion of actual negatives that were correctly predicted. A higher specificity value implies a better model. Specificity = TN / (TN+FP).

Sensitivity

Is the true positive rate the ratio of correctly identified positive elements? is the true negative rate, the proportion of negative elements correctly identified.

F1 score

It is the harmonic mean of a classification model's precision and recall. The F1 score is a measure of the model's performance. F1score = $((recall{-}1{+}\ precison{-}1)$ / 2) −1 (11) = 2. $((precision.recall)$ / $(precision + recall))$

ROC Curve / AUC

Area under the ROC curve: The area enclosed by the ROC graph is used to evaluate the accuracy of a detector classifier or model, including the ratio of true positives to false positives. As its shape is closer to the upper-left vertex of the coordinate system, it performs well and its size. Evaluation Metrics: Model performance was systematically assessed using a classification report that included accuracy, precision, sensitivity (also called recall), specificity, F1 score and support. In addition, the area under the characteristic curve (ROC/AUC) was computed to evaluate the discriminative ability of the classifiers. These metrics enabled a rigorous comparison of predictive models based on their ability to better identify and distinguish between patient classes. Specificity, F1 score, support and ROC/AUC.

Performance and evaluation confusion matrix

According to a report by logistic regressions is the most effective model for predicting liver cirrhosis, achieving an accuracy of about 55% **(Table 8)** [5].

Table 8: Model: Logistic regression classification report.

	Precision	Recall	F1-score	Support
1	0.5	0.52	0.51	2480
2	0.5	0.47	0.49	2532
3	0.63	0.65	0.64	2488
Accuracy			0.55	7500
Macro avg	0.55	0.55	0.55	7500
Weighted avg	0.55	0.55	0.55	7500

According to an article by Rahman, logistic regression was found to be the most effective model for predicting liver cirrhosis, achieving an accuracy of approximately 88 percent **(Table 9)**.

Copyright © 2026 | https://medical-clinical-surgery-reports.wren-research-journals.com/1

Table 9: Model: KNN classification report.

	Precision	Recall	F1-score	Support
1	0.86	0.87	0.87	2480
2	0.85	0.87	0.86	2532
3	0.92	0.9	0.91	2488
Accuracy	-	-	0.88	7500
Macro avg	0.88	0.88	0.88	7500
Weighted avg	0.88	0.88	0.88	7500

According to a study by 5, Support Vector Machine (SVM), Logistic Regression and Decision Tree algorithms were compared for liver disease prediction using a Kaggle dataset with 20,000 training records and about 1,000 test records **(Table 10)** [5].

Table 10: Model: SVM.

	Precision	Recall	F1-score	Support
1	0.75	0.72	0.74	2480
2	0.69	0.75	0.72	2532
3	0.84	0.81	0.82	2488
Accuracy	-	-	0.76	7500
Macro avg	0.76	0.76	0.76	7500
Weighted avg	0.76	0.76	0.76	7500

According to research by the random forest model achieved accuracies ranging from 95% in classifying liver disease **(Table 11)** [31].

Table 11: Model: Random forest classification report.

	Precision	Recall	F1-score	Support
1	0.95	0.94	0.94	2480
2	0.93	0.94	0.94	2532
3	0.96	0.96	0.96	2488
Accuracy			0.95	7500
Macro avg	0.95	0.95	0.95	7500
Weighted avg	0.95	0.95	0.95	7500

According to a study by Dwi Yuni Utami, the Naive Bayes algorithm achieved percent accuracy when predicting liver disease using a dataset from the UCI machine learning repository **(Table 12)** [47].

Table 12: Model: Naive Bayes classification report.

	Precision	Recall	F1-score	Support
1	0.41	0.78	0.54	2480
2	0.45	0.17	0.25	2532
3	0.63	0.45	0.53	2488
Accuracy	-	-	0.47	7500
Macro avg	0.5	0.47	0.44	7500
weighted avg	0.5	0.47	0.44	7500

Classification report for AdaBoost (Anjan). According to a recent article, gradient boosting achieved the highest accuracy among the boosting algorithms tested for predicting chronic liver disease **(Table 13)**.

Table 13: Model: AdaBoost classification report.

	Precision	Recall	F1-score	Support
1	0.59	0.56	0.57	2480
2	0.56	0.57	0.56	2532
3	0.66	0.67	0.67	2488
Accuracy			0.6	7500
Macro avg	0.6	0.6	0.6	7500
Weighted avg	0.6	0.6	0.6	7500

Classification report for the gradient boost. According to a study by gradient boosting achieved the highest accuracy among all evaluated algorithms for predicting chronic liver disease **(Table 14)** [5].

Table 14: Model: Gradient boosting classification report.

	Precision	Recall	F1-score	Support
1	0.85	0.8	0.83	2480
2	0.79	0.83	0.81	2532
3	0.87	0.88	0.87	2488
Accuracy			0.84	7500
Macro avg	0.84	0.84	0.84	7500
Weighted avg	0.84	0.84	0.84	7500

Classification report for MLP Classifier. The accuracy of the MLP Classifier 82.5% **(Table 15)** [6,39].

Table 15: Model: MLP classifier.

	Precision	Recall	F1-score	Support
1	0.83	0.79	0.81	2480
2	0.8	0.82	0.81	2532
3	0.87	0.89	0.88	2488
Accuracy			0.83	7500
Macro avg	0.83	0.83	0.83	7500
Weighted avg	0.83	0.83	0.83	7500

Above show in **Figure 9** cirrhosis results from prolonged liver failure, leading to extensive scarring, often due to conditions like hepatitis or chronic alcohol consumption [5]. The data are from a Mayo Clinic study on Primary Biliary Cirrhosis (PBC) of the liver conducted from 1974-1984.The heatmap in **Figure 6** reveals clusters of features with strong connections, while others show weaker ties. By removing less-correlated columns in showing **Figure 9**, we sharpened the model's focus and

Copyright © 2026 | https://medical-clinical-surgery-reports.wren-research-journals.com/1

boosted its predictive accuracy for liver disease [5]. This study explores diverse strategies for diagnosing liver cirrhosis, ranging from medical image analysis to pattern discovery in patient records. We blended gender, symptoms, habits and treatments using artificial intelligence, image analysis and data science. Eight machine learning models were tested on real clinical data to predict cirrhosis recurrence and patient survival. Performance was measured using precision, recall, F1-score, support, confusion matrices and ROC analysis, with the ROC curve providing a clear view of how effectively blood tests and models detect disease show in **Figure 10** [5].

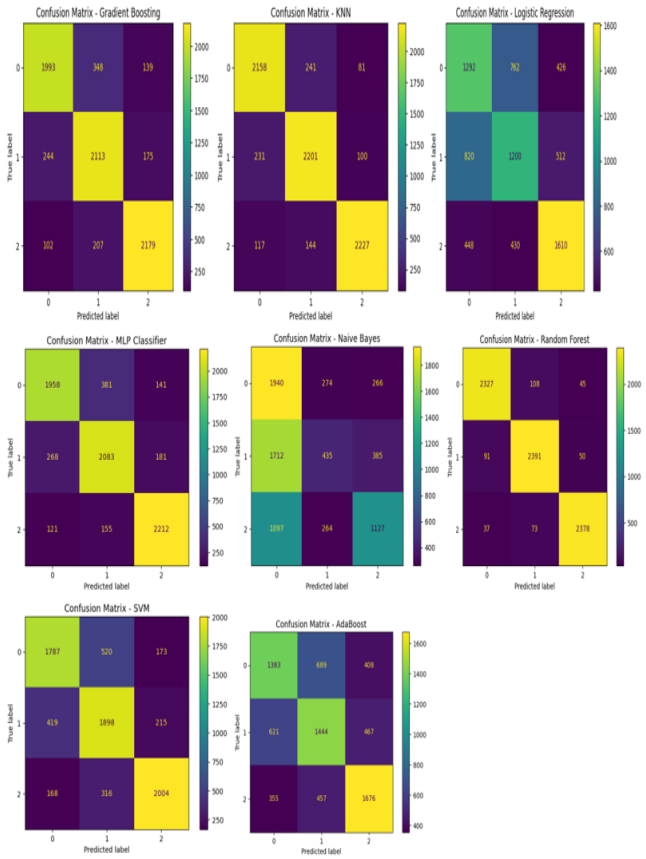


Figure 9: A comparative study of learning algorithms through a confusion matrix over the liver cirrhosis dataset.

Training and testing phase

- These consist of training and testing sets, which are splits.
- Each classifier was trained with the selected features.
- Predictions were made on the test data.

Proposed outcome

- This comparison identified the most accurate rate algorithm for predicting liver disease.
- The proposed system is intended to provide clinicians with a reliable tool for a more confident diagnosis of liver disease.

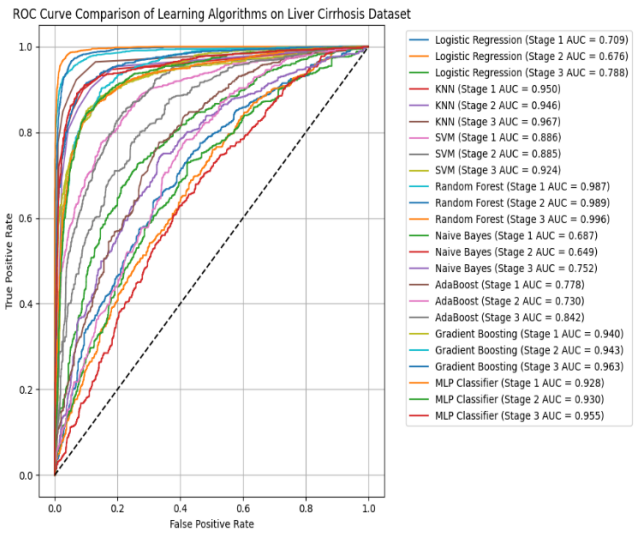


Figure 10: A comparative study of learning algorithms through ROC over the liver cirrhosis dataset.

Limitations and Result

This study has several limitations that must be considered. First, although the dataset used is sizable and diverse, it may not fully represent all populations affected by liver cirrhosis worldwide. The inclusion of synthetic data to increase sample size could introduce biases or distributional differences not observed in real patient populations, potentially affecting model accuracy in clinical practice. Additionally, the retrospective nature of the data, drawn primarily from the Aadarshvelu Liver-cirrhosis-stage-classification Kaggle dataset and similar sources, may limit its generalizability across geographic regions, clinical settings or patient subgroups are show in **Figures 11 and 12**. There is also potential for class imbalance, as some cirrhosis stages may be underrepresented. Finally, while the study benchmarks multiple algorithms, the results are based on internal cross-validation rather than external validation with independent clinical datasets. Future research should focus on validating these findings across broader, more varied clinical environments and on assessing their long-term clinical impact.

Counts:

Sex: F22137 M2863

Name: Count, dtype: int64

Percentages: Sex

F88.55 M11.45.

Copyright © 2026 | https://medical-clinical-surgery-reports.wren-research-journals.com/1

Journal of Medical, Clinical and Surgical Case Reports

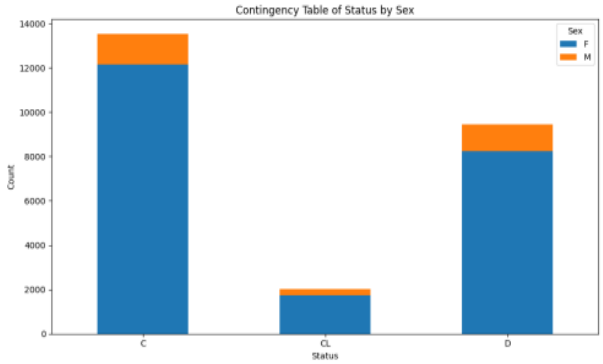


Figure 11: Contingency table of status by sex for count in male and female (C, D & CL).

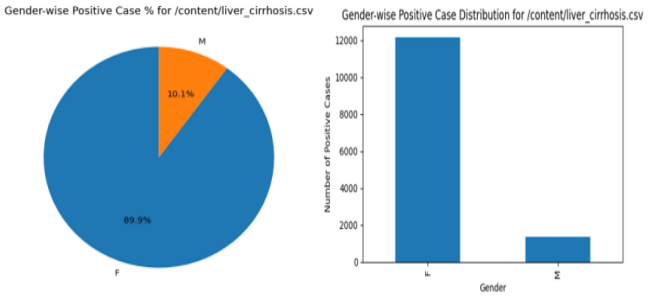


Figure 12: Positive case distribution gender-wise.

Total records (rows)

The dataset contains approximately 25,000 records (rows) of patient data (ILPD (Indian Liver Patient Dataset with clinical and Demographic information.

Total attributes (columns)

There are 19 columns (features) describing each patient's (Alvarez, n.d.; Huang et al. 2026) data for classifying liver cirrhosis stage. Key Notes (for thesis/paper use).

- r>0.7 → Strong correlation
- 0.3 ≤ r ≤ 0.7 → Moderate correlation
- r<0.3 → Weak correlation.

High correlations (*e.g.,* bilirubin pairs, AST-ALT) may indicate multicollinearity, which is important for ML model selection. Each machine learning classification technique is evaluated against alternatives for accuracy in a table. **Tables 8 to 15** compare predicted and actual results. Key advantages of our approach include enhanced classification, earlier risk prediction and high accuracy. Notably, our random forest model achieved 95% accuracy, surpassing the performance reported in prior studies on liver disease prediction using machine learning [7,28]. Previous work typically reported accuracies of 81%-90% across various algorithms and data sources. By benchmarking our results against these studies as showing in **Figure 15**, we demonstrate that our methodology offers substantial improvements in predictive performance [25]. We evaluated leading classification algorithms for accuracy, precision, sensitivity and specificity using a comprehensive dataset of liver patients to ensure a robust comparative analysis. Replace 'Age' and 'Bilirubin' with your actual feature column names from your DataFrame in show **Figure 14** [5]. Replace 'Status' with your actual target column name are show in **Figure 13**.

X = df[['Age', 'Bilirubin']]

Y= df['Status']

Split into and 20% testing data, 80% training

Habit_columns = ['Drug']

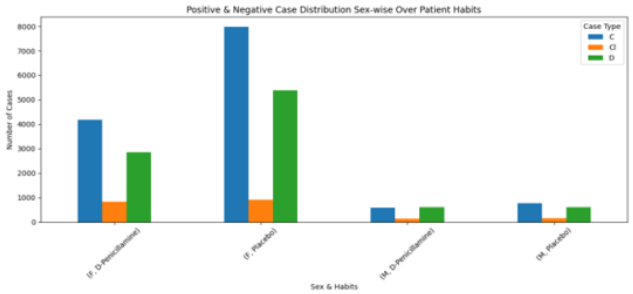


Figure 13: Positive and negative case distribution, gender or sex wise, over the patients' habits.

Symptom_columns = ['Ascites', 'Hepatomegaly', 'Spiders', 'Edema']

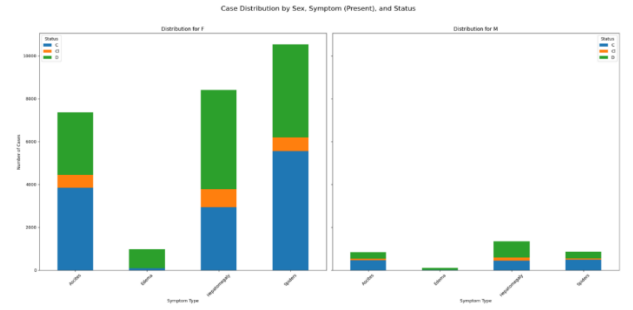


Figure 14: Positive and negative case distribution, gender-wise, over the patient's symptom status

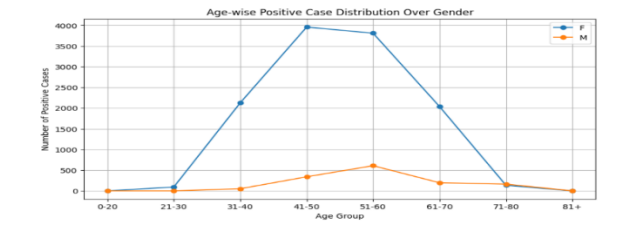


Figure 15: Age-wise distribution of positive cases by gender.

Copyright © 2026 | https://medical-clinical-surgery-reports.wren-research-journals.com/1

Table 16 shows a head-to-head comparison of machine learning methods, measuring accuracy, precision, sensitivity and specificity across gender, symptoms and habits in the liver dataset. Eight models, including Logistic Regression, K-Nearest Neighbors, Support Vector Machine, Random Forest, Naive Bayes, AdaBoost, Gradient Boosting and MLP Classifier, were evaluated and delivered impressive results. These methods are more accurate than Random Forest and the KNN Model. This study offers a sweeping analysis of classification algorithms for early cirrhosis detection, spotlighting key blood tests such as ALT/SGPT, AST/SGOT, ALP, Albumin, Bilirubin, Total Protein, GGT, LD and PT. Looking ahead, research should delve into deep learning, advanced ensemble methods, creative feature engineering and validation on independent clinical datasets.

Table 16: Comparing accuracy for ML methods.

ML Models	Accuracy
LR	54.5
KNN	87.5
SVM	75.5
RF	94.5
NB	46.5
Adaboost	59.5
Gradiant boosting	83.5
MLP classifier	82.5

Conclusion and Future Scope

For these machine learning solutions to achieve real impact, clinical integration and interpretability are essential. Predictive models should be designed to fit naturally into existing clinical workflows, ideally as decision-support tools embedded in electronic health record systems. For example, clinicians could receive risk scores or stage predictions as part of routine test-result reviews, thereby streamlining patient triage and management decisions. Interpretability is equally important; models like decision trees and logistic regression can provide transparent logic and feature contributions, help clinicians trust and validate predictions. Even with more complex models such as Random Forests or Gradient Boosting, built-in feature importance rankings and visual tools like SHAP (Shapley Additive explanations) plots can help illuminate the factors influencing each prediction. Close collaboration between data scientists and clinicians is vital for selecting clinically meaningful features, tailoring model outputs to provider needs and ensuring they are presented in a clear, actionable manner. Making models more interpretable and working closely with clinical experts will be key for real-world adoption. The data reveal a higher prevalence of cirrhosis, suggesting an increased risk. Future efforts should aim to build even more accurate predictive models, expand the range of algorithms and boost predictive power by focusing on advanced techniques. Careful study of clinical features, symptoms and biomarkers remains vital. Early detection and severity classification tools can empower clinicians to make smarter decisions and improve patient care; though detecting early cirrhosis remains a challenge 5. By harnessing information about symptoms, habits and patient records, diagnostic systems can be strengthened to speed up treatment. In this study, machine learning algorithms were compared for predicting cirrhosis using 19 features from about 25,000.

Data Availability

The data supporting this study are not publicly available due to hospital restrictions. The model code that supports this study is available from the first author upon reasonable request.

References

1. Mokdad AA, Lopez AD, Shahraz S, Lozano R, Mokdad AH, et al. (2014) Liver cirrhosis mortality in 187 countries between 1980 and 2010: A systematic analysis. BMC Medicine 12: 14.
2. LDCR (2015) Liver disease in Canada report 2015.
3. Rajeswari P, Reena GS (2010) Analysis of liver disorder using data mining algorithm. Global Journal of Computer Science and Technology 10: 48-52.
4. Abdalrada AS, Yahya OH, Alaidi AHM, Hussein NA, Alrikabi HT, et al. (2019) A predictive model for liver disease progression based on logistic regression algorithm. Periodicals of Engineering and Natural Sciences 7: 1255-1264.
5. Dickson,E., Grambsch,P., Fleming,T., Fisher,L. and Langworthy,A, Cirrhosis Patient Survival Prediction 2023, UCI Machine Learning Repository.
6. "Aadarshvelu/liver-cirrhosis-stage-classification" liver_cirrhosis.csv (2.35 MB).
7. http://www.kaggle.com/dataset/aadarshvelu/Liver-Cirrhosis-Stage-Classification/data.
8. K. Sumeet, J.J. Larson, B. Yawn, T.M. Therneau. W.R. Kim, Underestimation of liver-related mortality in the United States. 2013, Gastroenterology;145:375–382, e371–372.
9. L. A. Auxilia. Accuracy Prediction Using Machine Learning Techniques for Indian Patient Liver Disease. 2nd International Conference on Trends in Electronics and Informatics (ICOEI), IEEE, 2018.
10. Hashem, M. Esraa, S. Mai. A study of support vector machine algorithm for liver disease diagnosis. American Journal of Intelligent Systems. 2014, 4.1; 9-14.
11. P. Sajda. Machine learning for detection and diagnosis of disease. Annu. Rev. Biomed, 2006, Eng. 8; 537-565.
12. S. M. Mahmud, et al. Machine Learning-Based Unified Framework for Diabetes Prediction. Proceedings of the International Conference on Big Data Engineering and

Copyright © 2026 | https://medical-clinical-surgery-reports.wren-research-journals.com/1

Technology, ACM 2018.

13. Das. R. A comparison of multiple classification methods for diagnosis of Parkinson's disease. Expert Systems with Applications. 2010, 37, 1568-1572.

14. Kasabov, n, & Capecci, e. Spiking neural network methodology for modeling. classification and understanding of EEG spatio-temporal data measuring cognitive process. Information Science. 2015, 294, 565-575.

15. Marateb, H.R. mansoorian, m. Faghihimani, E., Amini. M., & farina, d.(2014). A hybrid intelligent system for diagnosing microalbuminuria in type 2 diabetes patients without having to measure urinary albumin,Computers in Biology and Medicine. 1 February 2014, Volume 45,Pages 34-42.

16. Nahar, j., imam, t., tickle, k. s., & Chen, Y. p.p. Association rule mining to detect factors which contribute to heart disease in males and females, Expert Systems with Applications. 2013, 40, 1086-1093.

17. Tanha, J., van Someren, M., & Afsarmanesh, h. Semi-supervised self-training for decision tree classifiers, International Journal of Machine Learning and Cybernetics. 2015, 1-16.

18. Souillard- Mandar, W., Davis, R., Rudin, C., Au, R., Libon, D. j., Swenson, R., et al. Learning classification models of cognitive condition from subtle behaviors in the digital clock drawing test, Machine learning. 2015,1-49.

19. Patidar, S., & Pachori, R. b., & acharya, u. r. Automated diagnosis of coronary artery disease using tunable-Q wavelet transform applied on heart rate signals. Knowledge-based systems. 2015, 82, 82,1-10.

20. Rodríguez-Jiménez, J. M., Cordero, P. Enciso, M., Mora, A. Data mining algorithms to compute mixed concepts with negative attributes: an application to breast cancer data analysis. Mathematical Methods in the Applied Science 2016.

21. Han, J., Kamber, M. Data Mining: Concepts and Techniques. Elsevier. 2001.

22. Marinakis, Y., Marinaki, M., & Matsatsinnis, N. A stochastic nature-inspired metaheuristic for clustering analysis. International journal of business intelligence and data mining. 2008. 3. 30-44.

23. Alizadehsani, R, Habibi, J., Hosseini, M. j., Mashyekhi, H., Boghrati, R., Ghandeharioun, A., et al. A data mining approach for diagnosis of coronary artery disease, Computer methods and programs in biomedicine. 2013, 111, 52-61.

24. Deekshatulu, B. L., & Chandra, P. Classification of heart disease using k-nearest neighbor and genetic algorithm. Procedia Technology. 2013. 10, 85-94.

25. Gullo, F. from patterns in data to knowledge discovery: what data mining can do.Physics Procedia. 2015. 63, 18-22.

26. Sebastian A, Varghese SM (2016) Fuzzy logic for Child-Pugh classification of patients with cirrhosis of the liver. International Conference on Information Science (ICIS). IEEE.

27. Arshad, Insha, et al. Liver disease detection due to excessive alcoholism using data mining techniques. International Conference on Advances in Computing and Communication Engineering (ICACCE). IEEE. 2018.

28. Ramkumar, N. et al. Prediction of liver cancer using Conditional probability Bayes theorem, International Conference on Computer Communication and Informatics (ICCCI). IEEE. 2017.

29. Hassoon, Mafazalyaqeen, et al. Rule optimization of boosted C5.0 classification using a genetic algorithm for liver disease prediction. International Conference on Computer and Applications (ICCA). IEEE. 2017.

30. M. Abdar, M. Zomorodi-Moghadam, R. Das and I.-H. Ting. Performance analysis of classification algorithms on early detection of liver disease. Expert Systems with Applications. 2017, vol. 67, pp. 239–251.

31. B.R. Thapa and Anuj Walia,Liver Function Tests and their Interpretation. Indian J Pediatr. 2007,74 (7): 663-671.

32. Ch. Shravya, K. Pravalika, Shaik Subhani. Prediction of Breast Cancer Using Supervised Machine Learning Techniques. International Journal of Innovative Technology and Exploring Engineering (IJITEE). April 2019, Volume-8 Issue-6, ISSN: 2278-3075.

33. L. Breiman. Random Forests. Machine Learning. 2001, 45(1), 5–32.

34. G. Guo, H. Wang, D. Bell, Y. Bi, K. Greer. KNN Model-Based Approach in Classification. Springer. Berlin, Heidelberg. 2003, pp. 986–996.

35. Decision trees.

36. Classification and Decision Trees.

37. Christopher R. Stephens1,2 · Hugo Flores Huerta1,3 · Ana Ruíz Linares3,4. When is the Naive Bayes approximation not so naive? Received: 30 September 2015 / Accepted: 30 June 2017 / Published online: 21 July 2017.

38. Chervonenkis AY. Early history of support vector machines.Springer. empirical inference. pp 13–20:.

39. Logistic Regression, Retrieve from: https://www.saedsayad.com/logistic _regression.htm, last Accessed: 5 Octobor 2019.

40. Yan et al. Correlation between Child-Pugh degree and the four examinations of traditional Chinese medicine with liver cirrhosis. 2008.

41. Clinical and Laboratory Data Ricardo Ribeiro1,2, Rui Marinho3, José Velosa3, Fernando Ramalho3 and J. Miguel Sanches1, Institute for Systems and Robotics / Instituto Superior Técnico 2 Escola Superior de Tecnologia da Saúde de Lisboa 3 Liver Unit, Diffuse Liver Disease Classification from Ultrasound Surface Characterization. Department of Gastroenterology and Hepatology. Hospital de Santa Maria, Medical School of Lisbon, Lisbon, Portugal, IbPRIA 2011, LNCS 6669, pp. 167–175.

Copyright © 2026 | https://medical-clinical-surgery-reports.wren-research-journals.com/1

42. Yugal Kumar and G. Sahoo. Prediction of different types of liver diseases using rule-based classification model. Department of Information Technology. Birla Institute of Technology. Mesra, Ranchi. Jharkhand India, Technology and Health Care. 2013, 21, 417–432.

43. Tokala S, Hajarathaiah K, Gunda SRP, Botla S, Nalluri L, et al. (2023) Liver disease prediction and classification using machine learning techniques. International Journal of Advanced Computer Science and Applications 14.

Copyright © 2026 | https://medical-clinical-surgery-reports.wren-research-journals.com/1