

Evaluating Tree-Based and Probability-Based Learning in Liver Disease Prediction: A Comparison of Decision Tree and Naïve Bayes Algorithms

I.I. Akuji^{1*}, T.A. Ahmed², and A.O. Giwa-Raheem³

¹Department of Computer Science, Faculty of Science and Computing, Abdurasaq Abubakar Toyin University, Oke-Ogba, Kwara State, Nigeria

²Department of Computer Science, School of ICT and Management Science, Kwara State College of Arabic and Islamic Legal Studies, Ilorin, Nigeria

³Department of Computer Science, Faculty of Computing, Engineering and Technology, Al-Hikmah University, Ilorin, Nigeria

Abstract

INTRODUCTION: Liver disease imposes a heavy global burden, particularly in developing countries such as Nigeria, where early detection through traditional methods remains challenging due to resource constraints and late symptom onset. Machine learning algorithms, including tree-based and probability-based algorithms have shown promise in enhancing the early prediction of liver disease. However, existing studies lack a comparative analysis between tree-based and probability-based techniques, which is crucial for determining the most effective approach. Therefore, this study addresses this gap by evaluating and comparing tree-based versus probability-based algorithms for liver disease prediction.

OBJECTIVES: The specific objectives are to: explore and preprocess liver disease patient dataset from Kaggle using label encoding, and mean imputation techniques; implement and evaluate standalone Decision Tree and Naïve Bayes classifiers via key metrics including accuracy, precision, recall, F1-score; perform comparative analysis between the two models; and validate the model's generalizability using 10-fold cross validation technique.

METHODS: Techniques such as label encoder, simple imputer were employed for categorical variable conversion, and handling missing values, respectively. Tree-based (Decision Tree) algorithm, and Probability-based (Naïve Bayes) algorithm were employed to develop predictive models for liver disease prediction, evaluated using accuracy, precision, recall, and F1-score metrics.

RESULTS: Results reveal that Decision Tree achieved 98% across metrics, outperforming Naïve Bayes (66% accuracy, 74% precision, 66% recall, and 63% f1-score). The Decision Tree model after 10-fold cross-validation maintained 98% across all metrics, affirming its outstanding generalization capability. On the other hand, the Naïve Bayes classifier showed performance improvements following 10-fold cross-validation, with a 1% increase in accuracy and precision, and a 2% increase in recall and f1-score.

CONCLUSION: These results affirm tree-based superiority yet highlight hybrids' potential for efficient, and effective diagnostics. Therefore, this study suggests future work to explore advanced hybrids such as stacking Decision Tree-Naïve Bayes, deep feature engineering, and validation on diverse datasets for clinical deployment in high-burden regions.

Keywords: Liver disease, Machine Learning, Tree-based Machine Learning, Probability-based Machine Learning, Decision Tree, Naïve Bayes

Received on 24 April 2026, accepted on 06 October 2026, published on 09 October 2026

Copyright © 2026 I. I. Akuji *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [CC BY-NC-SA 4.0](#), which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi: 10.4108/eetismmla.12753

*Corresponding author. Email: akujiismaheel@gmail.com

1. Introduction

The liver stands as a cornerstone of human physiology, executing multifaceted roles essential for homeostasis: producing bile for fat digestion, synthesizing proteins like albumin and clotting factors, storing and releasing glucose to regulate blood sugar, detoxifying harmful substances such as ammonia and drugs, and metabolizing nutrients and hormones. Despite these critical functions, liver diseases such as viral hepatitis, alcoholic liver disease, non-alcoholic fatty liver disease (NAFLD), autoimmune disorders, genetic conditions like hemochromatosis, and hepatocellular carcinoma pose a mounting global health crisis [1].

Early detection is paramount for effective management, yet traditional diagnostics which rely on invasive biopsies, costly imaging (e.g., ultrasound, CT), and blood tests prone to false negatives, hindering timely interventions in resource-limited settings. The advent of digital health technologies and data mining has revolutionized this landscape, particularly through machine learning (ML), a subset of artificial intelligence enabling systems to learn patterns from historical data for predictive tasks with minimal human input [2][3]. Supervised ML algorithms have excelled in liver disease prediction using datasets like the Indian Liver Patient Dataset (ILPD) and BUPA, analyzing biomarkers such as age, gender, bilirubin, alkaline phosphatase, alanine aminotransferase (SGPT), aspartate aminotransferase (SGOT), total proteins, albumin, albumin-globulin ratio, and selector variables [4][5][6]. Studies report accuracies from 94-100% accuracy with Decision Tree, and 31-67% accuracy with Naïve Bayes, yet challenges persist: tree-based methods risk overfitting noisy data, while probability-based approaches falter on feature correlations, underscoring the need for comparative analysis between the two approaches on the same dataset.

This study addresses these gaps by evaluating tree-based (Decision Tree) versus probability-based (Naïve Bayes) algorithms head-to-head, then developing machine learning models to compare Decision Tree's hierarchical accuracy and interpretability with Naïve Bayes' probabilistic efficiency and speed. Such comparative analysis facilitates proper justification to determine the model with superior performance for clinical deployment. The specific objectives are follows: exploration and preprocess of liver disease patient dataset from Kaggle using label encoding, and missing values handling techniques; implement and evaluate standalone Decision Tree and Naïve Bayes classifiers via key metrics including accuracy, precision, recall, F1-score; perform comparative analysis between the two models; and validate the model's generalizability and propose deployment guidelines for healthcare systems.

2. Literature Review

Liver disease prediction represents a critical application of machine learning in healthcare, targeting early detection of conditions like cirrhosis and failure that cause high morbidity in developing nations, including Nigeria [1]. This domain demands models that balance accuracy, and speed for clinician adoption, making tree-based (e.g., Decision Tree) and probability-based (e.g., Naïve Bayes) comparisons essential to optimize real-world deployment.

2.1. Related Works

Studies are categorized into two forms, namely probability-based and tree-based studies. The former comprises studies that employed Naïve Bayes algorithm while the later covers studies that utilized decision tree, and random forest algorithms. Additionally, the last part of this sub-section compares the performance of probability-based and tree-based literatures.

2.1.1 Literatures on Probability-Based Algorithms

Probability-based algorithms like Naïve Bayes leverage probabilistic independence assumptions to classify liver disease efficiently, prioritizing speed and interpretability for clinical use. [7] developed predictive models using Logistic Regression, Naïve Bayes, and KNN on the Liver Cirrhosis dataset, where Naïve Bayes achieved only ~31% accuracy, underscoring its weakness in handling correlated physiological features like bilirubin levels. Similarly, [8] applied Bernoulli Naïve Bayes to the Indian Liver Patient Dataset (ILPD) with 580 patients and 11 attributes, attaining 67% accuracy, a recall of 0.82 for at-risk patients, but poor recall (0.28) for healthy cases after data cleaning and normalization. Strengths here include fast training and transparent probabilistic outputs, ideal for medical decision-support prototypes, though limitations in specificity highlight sensitivity to feature dependencies. [9] further tested Naïve Bayes on the BUPA dataset (345 instances, 6 variables) via 5-, 10-, and 20-fold cross-validation in Orange, yielding accuracies of 64.6%, 66.7%, and 64.3%, respectively, with 10-fold CV performing better than some peers, yet still moderate due to oversimplified independence assumptions.

2.1.2 Literatures on Tree-Based Algorithms

Tree-based algorithms, such as Decision Tree and its variants like J48, build hierarchical structures for intuitive decision paths in liver disease prediction, often delivering high accuracy through recursive partitioning. [10] achieved 100% accuracy with Decision Tree and Random Forest on liver datasets, praising their robustness while suggesting hybrids for gains. [11] paired Decision Tree with LASSO feature selection, reaching 94.29% accuracy and demonstrating its strength in feature-driven predictions. [12] reported 95.04% accuracy with J48, emphasizing tree models' interpretability via visual splits. [13] found Random Forest at 98.14% on 615-patient data, outperforming ANN and SVM. Strengths of tree-based methods include high accuracy, inherent feature

importance ranking, and human-readable rules, making them suitable for healthcare diagnostics. However, weaknesses like overfitting on noisy data and sensitivity to small dataset shifts necessitate pruning or ensemble, as noted across studies.

2.1.3 The Need for Tree-Based and Probability-Based Comparison

Comparing tree-based (e.g., Decision Tree) and probability-based (e.g., Naïve Bayes) algorithms head-to-head, as proposed here, is essential: trees deliver superior accuracy (up to 100%; [10]) and interpretability via rule-based splits but risk overfitting on noisy data, while Naïve Bayes offers rapid training and transparent probabilistic outputs (e.g., 67% accuracy; [8]) yet underperforms (~31-67%; [7]) due to independence assumptions faltering on correlated biomarkers. Dataset heterogeneity and varying preprocessing (e.g., normalization, feature selection) amplify these disparities, making standardized evaluation on identical datasets and conditions imperative to identify true relative strengths, and guide reliable clinical deployment.

3. Methodology

The framework of the proposed models is depicted in Figure 1.

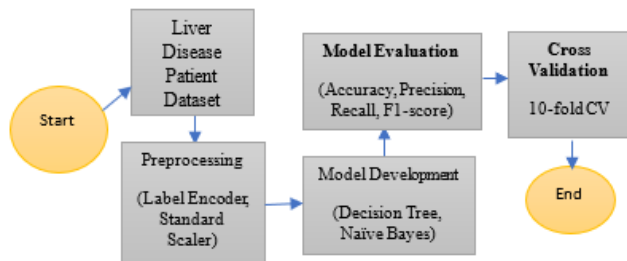


Figure 1. Proposed Model Framework

Figure 1 shows that the liver disease patient dataset is pre-processed using label encoder for categorical variable conversion and simple imputer for handling missing values, followed by model development using decision tree and naïve bayes, and model evaluation using accuracy, precision, recall and f1-score.

3.1 Data Exploration

This study used a public dataset from Kaggle Repository with 11 features including target feature and 30,691 instances of liver status instances spanning 21,917 class 1 and 8,774 class 2. The features with their descriptions are presented in Table 1.

Table 1. List of Features and Descriptions

S/N	Feature name	Description
1	Age	Patient's age in years
2	Gender	Patient's sex (male/female);
3	Total_Bilirubin	Total serum bilirubin level (mg/dL)
4	Direct_Bilirubin	Conjugated (direct) bilirubin level
5	Alkaline Phosphatase	Alkaline phosphatase (ALP) activity (IU/L)
6	Serum glutamic-pyruvic transaminase (ALT or SGPT)	Sgpt_Alanine_Aminotransferase
7	Serum glutamic-oxaloacetic transaminase (AST or SGOT)	Sgot_Aspartate_Aminotransferase
8	Total serum protein concentration	Total_Proteins
9	Serum albumin level	ALB_Albumin
10	Ratio of albumin to globulin in serum	Albumin_and_Globulin_Ratio
11	Target label indicating liver disease status (e.g., "Normal" vs "Liver Patient" or graded severity level).	Result

3.2. Data Preprocessing

Data preprocessing was conducted to improve the quality, consistency, and suitability of the dataset for classification. The preprocessing stage involved data cleaning, categorical variable conversion, and class balancing to ensure that the machine learning models could learn effectively from the available data.

3.2.1 Data Cleaning

Data cleaning was performed to address missing values within the dataset. Missing or null values can negatively affect the performance of machine learning algorithms by introducing bias and reducing prediction accuracy. This affects several features including gender, Total_Bilirubin, Direct_Bilirubin, Total_Proteins, ALB_Albumin among others. To handle this issue, the Simple Imputer technique was employed using the mean imputation strategy. In this approach, missing numerical values were replaced with the mean value of the corresponding attribute. Mean imputation helps preserve the overall distribution of the data while ensuring that no records are discarded due to incomplete information. As a result, the dataset became complete and suitable for subsequent analysis and model training.

3.2.2 Categorical Variable Conversion

Since machine learning algorithms require numerical input, categorical variable which is Gender was transformed into numerical representations using the Label Encoder technique. Label encoding assigns a unique numerical value to each category within a feature. For example, categorical values such as "Male," and "Female" were encoded as 0, and 1, respectively. This conversion enabled the classification algorithms to process categorical information effectively while maintaining the distinctions among category levels.

3.2.3 Class Balancing

Class imbalance leads to biased model predictions toward the majority class. To address this issue, the Synthetic Minority Oversampling Technique (SMOTE) was applied. SMOTE balances the dataset by generating synthetic instances of the minority class (class 2) rather than simply duplicating existing samples. This process improves the classifier's ability to learn patterns associated with underrepresented classes, thereby enhancing prediction performance, reducing bias, and improving model generalization.

The data preprocessing techniques applied include data cleaning, categorical variable conversion, and class balancing.

- Data clearing: replacement of null with mean using mean imputation simple imputer.
- Categorical variable conversion: label encoder.
- Class balancing: Synthetic Minority Oversampling Technique.

3.3 Model Development

Model development followed a structured pipeline on the preprocessed Liver Disease Patient Dataset. The dataset after preprocessing reduced to 11010 instances vs 11 features. Data was split into training (80%) and testing (20%) sets. Default hyperparameters were used initially, as shown in Table 2.

Table 2. Models Parameters

Model	Key Perparameters (Default)	Fit Pseudocode
Decision Tree	criterion='gini', max_depth=None, random_state=1	decision_tree.FIT(x_train, y_train)
Naïve Bayes	var_smoothing=1e-09	naive_bayes.FIT(x_train, y_train)

3.4 Model Evaluation

Accuracy measures the proportion of all correctly predicted cases (both liver-disease and non-liver cases) out of the total number of samples analyzed, shown in (1):

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \dots \dots \dots (1)$$

Precision is the positive patterns that every Precision reflects how many of the patients predicted as having liver disease truly have the disease, thereby minimizing false alarms, shown in (2):

$$Pre = \frac{TP}{TP + FP} \dots \dots \dots (2)$$

Recall, also known as sensitivity, indicates the proportion of actual liver-disease cases that the model successfully identifies and is critical for early detection, because missing true positive cases (false negatives) can delay treatment, shown in (3):

$$Recall = \frac{TP}{TP + FN} \dots \dots \dots (3)$$

The F1-score combines both precision and recall into a single value using their harmonic mean, which is particularly useful when there is a trade-off between these two metrics, shown in (4):

$$F1 - score = 2 \times \frac{Pre \times Recall}{Pre + Recall} \dots \dots \dots (4)$$

where:

- TP denotes True Positive (correctly predicted liver-disease cases)
- TN denotes True Negative (correctly predicted healthy cases)
- FP denotes False Positive (healthy cases wrongly classified as liver-disease)
- FN denotes False Negatives (actual liver-disease cases wrongly classified as healthy)

3.5 Cross Validation

A 10-fold stratified cross-validation procedure was used to evaluate the generalization performance and reliability of the developed classification models. The training dataset was divided into ten approximately equal subsets (folds), while preserving the proportion of samples in each class within every fold. During each iteration, nine folds were used to train the model and the remaining fold was used for validation. This process was repeated ten times so that each fold served once as the validation set. The resulting performance measures, including accuracy, precision, recall, and F1-score, were averaged across the ten folds to obtain a more robust estimate of model performance than that provided by a single train-test split. Stratified K-fold validation is appropriate for binary and multiclass

classification because it preserves the class distribution in each fold.

The validation procedure was implemented using StratifiedKFold with `n_splits=10`, `shuffle=True`, and `random_state=42`. Setting `shuffle=True` randomized the class-specific samples before fold construction, reducing the possibility that the original ordering of observations affected the validation outcome. The fixed random seed (`random_state=42`) ensured that the same fold partitions could be reproduced in subsequent experiments. This procedure helped assess whether the observed model performance was consistent across different training and validation subsets and provided evidence for identifying potential overfitting.

4. Results and Discussion

This section covers results of dataset preprocessing including missing values handling and class balancing, dataset exploration, feature selection, model development, model evaluation and cross validation.

4.1. Results of Data Preprocessing

4.1.1 Missing Values Handling

The list of features with individual frequency of missing values is presented in Figure 2.

Age	0
Gender	0
Total_Bilirubin	325
Direct_Bilirubin	298
Alkphos_Alkaline_Phosphatase	390
Sgpt_Alamine_Aminotransferase	264
Sgot_Aspartate_Aminotransferase	223
Total_Protiens	195
ALB_Albumin	227
Albumin_and_Globulin_Ratio	269
Result	0
dtype: int64	

Figure 2. List of Features and Frequency of Missing Values before Handling Missing Values

Figure 2 shows that age, gender and result features contain no missing values unlike others.

Age	0
Gender	0
Total_Bilirubin	0
Direct_Bilirubin	0
Alkphos_Alkaline_Phosphatase	0
Sgpt_Alamine_Aminotransferase	0
Sgot_Aspartate_Aminotransferase	0
Total_Protiens	0
ALB_Albumin	0
Albumin_and_Globulin_Ratio	0
Result	0
dtype: int64	

Figure 3. Visualization of Features after Missing Values Handling

Figure 3 shows that none of the features contains missing value represented by zero '0' across features in the dataset.

4.1.2 Class Distribution

After data preprocessing, the dataset reduced to 11,010. The class distribution for the target variable Result is nearly balanced (Figure 4), with 5520 instances labeled as class 0 and 5490 instances labeled as class 1, giving a ratio of approximately 1:1 between the two classes; this indicates that the dataset does not exhibit strong class skewness, which is favorable for standard classification algorithms.

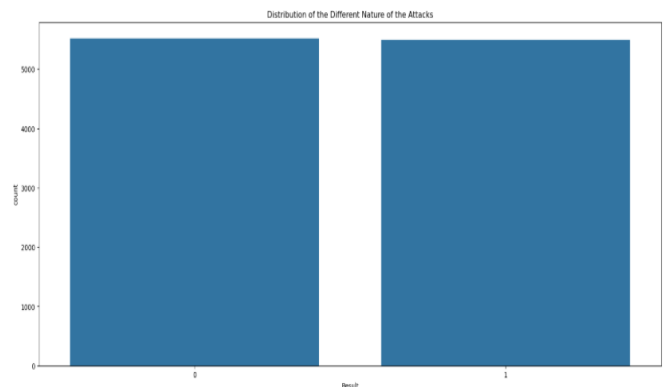


Figure 4. Class Distribution of the Liver Disease Dataset

4.2 Correlation Analysis using Heatmap

Figure 5 reveals that the strongest linear relationships are among the liver-biochemistry markers rather than with demographic variables. Age and Gender show very weak correlations with all other features (most absolute values below 0.03), indicating these demographic variables are almost independent of the laboratory measurements and the outcome Result. Total bilirubin and direct bilirubin are

highly positively correlated (≈ 0.89), as expected since direct bilirubin is a component of total bilirubin, and both are moderately correlated with alkaline phosphatase and the transaminases (SGPT & SGOT), suggesting a coherent cholestatic-hepatocellular pattern. The two transaminases are strongly correlated (≈ 0.79), reflecting their shared role as markers of hepatocellular injury, while proteins and albumin are highly positively correlated (≈ 0.80) and albumin is moderately correlated with the albumin–

globulin ratio (≈ 0.68), consistent with known protein-metabolism physiology.

Finally, the outcome Result shows only modest negative correlations with bilirubin-related markers (≈ -0.21 to -0.31) and slightly positive correlations with albumin and albumin–globulin ratio (≈ 0.18), indicating that higher bilirubin and lower albumin tend to be associated with liver disease but the relationships are not extremely strong.

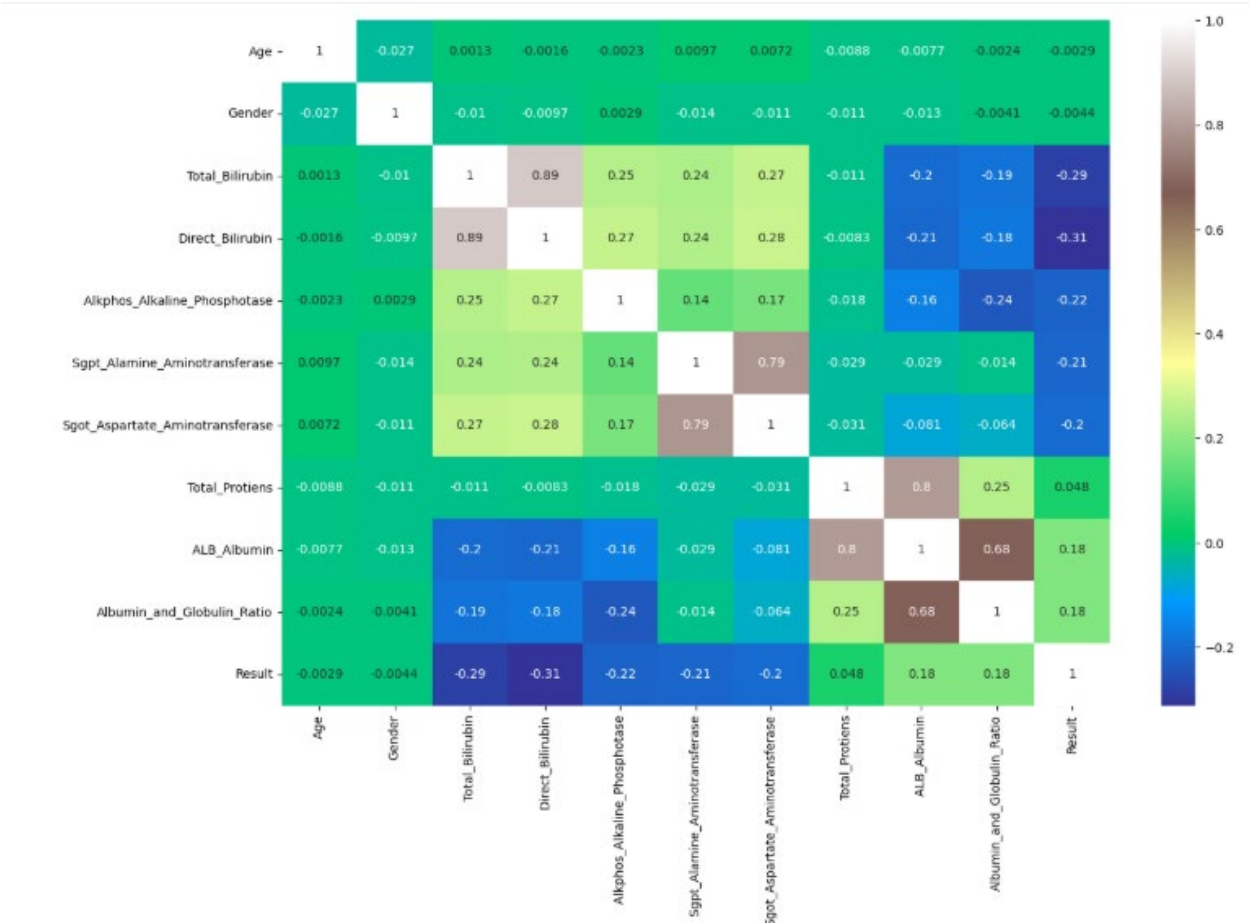


Figure 5. Correlation Heatmap

4.3. Model Results

4.3.1 Decision Tree

Figure 6 indicates excellent performance on the Decision Tree algorithm.

	precision	recall	f1-score	
0	0.97	0.98	0.98	
1	0.98	0.97	0.98	
accuracy				0.98
macro avg	0.98	0.98		0.98
weighted avg	0.98	0.98		0.98

Figure 6. Decision Tree Classification Report

As shown in Figure 6, the model achieves an overall accuracy of 98%, with both class-0 and class-1 prediction showing very high precision (0.97–0.98), recall (0.97–0.98), and F1-score (0.98), which means it correctly identifies positive and negative cases with low rates of false positives and false negatives. The macro and weighted averages are also 0.98 across precision, recall, and F1-score, confirming that the model is well-balanced and performs consistently for both classes rather than favoring one over the other.

4.3.2 Naïve Bayes

The Naïve Bayes classifier in Figure 7 performs noticeably worse than the decision tree on this liver-disease task.

	precision	recall	f1-score
0	0.88	0.37	0.52
1	0.60	0.95	0.74
accuracy			0.66
macro avg	0.74	0.66	0.63
weighted avg	0.74	0.66	0.63

Figure 7. Naïve Bayes Classification Report

Figure 7 reveals that the model achieves an overall accuracy of only 66%. The macro and weighted averages (precision ≈ 0.74 , recall ≈ 0.66 , F1 ≈ 0.63) show that the model is highly imbalanced in behavior between the two classes and overall, less reliable than the decision tree, despite still being able to capture some signal for the positive class.

4.3.3 Cross Validation

To evaluate the reliability of the developed models' generalization performance across multiple subsets of the available data and to determine whether the 98% baseline Decision Tree accuracy reflects genuine predictive capability rather than overfitting, this study applies 10-fold cross-validation (CV). This approach reduces dependence on a single train-test split and provides a more robust assessment of potential overfitting.

As shown in Figure 8, the 10-fold accuracies ranged from 97.39% to 98.98%, which indicates only small performance variation across the ten validation subsets. This suggests that the model is robust and is not overly dependent on a particular training-data split. Specifically, the Decision Tree model after 10-fold CV achieved 98% across accuracy, precision, recall, and F1-score. This implies that the model is reliable for the liver disease classification task and can be considered a strong candidate model because it combines high predictive accuracy with interpretability.

Accuracy for each fold:
[0.983 0.9898 0.9796 0.9751 0.9785 0.983 0.9864 0.9739 0.9751 0.9819]

Final Test-Set Accuracy:
0.9805

Final Test-Set Classification Report for Decision Tree + 10-fold CV:

	precision	recall	f1-score
0	0.98	0.99	0.98
1	0.99	0.98	0.98
accuracy			0.98
macro avg	0.98	0.98	0.98
weighted avg	0.98	0.98	0.98

Figure 8. Decision Tree 10-fold Cross Validation Classification Report

On the other hand, Figure 9 reveals that the accuracies of the Naïve Bayes after 10-fold CV ranged from 65.87% to 71.20%, indicating relatively stable but substantially lower performance than the Decision Tree model. Moreover, the final classification report of the model showed 67% accuracy, 75% precision, 68% recall, and 65% f1-score. Overall, the Decision Tree is therefore the more appropriate model, as it provides much stronger and more balanced classification performance.

Accuracy scores for the 10 folds:
[0.6803 0.6599 0.6599 0.6587 0.6667 0.6746 0.6757 0.712 0.6712 0.6927]

Final Test Accuracy:
0.6748

Final Test Classification Report for Naive Bayes + 10-fold CV:

	precision	recall	f1-score
0	0.89	0.40	0.55
1	0.61	0.95	0.74
accuracy			0.67
macro avg	0.75	0.68	0.65
weighted avg	0.75	0.67	0.65

Figure 9. Naïve Bayes 10-fold Cross Validation Classification Report

4.3.4 Comparison of Results

This section covers comparative analysis of the performance of the Decision Tree and Naïve Bayes classifiers before and after the application of cross-validation (CV) using four evaluation metrics: accuracy, precision, recall, and F1-score.

Table 3. Comparison of the Decision Tree and Naïve Bayes Results before and after CV

Classifiers	Decision Tree		Naïve Bayes	
	Before CV	After CV	Before CV	After CV
Accuracy	98%	98%	66%	67%
Precision	98%	98%	74%	75%
Recall	98%	98%	66%	68%
F1-score	98%	98%	63%	65%

Table 3 demonstrates the impact of cross-validation especially on the performance of the Naïve Bayes classifier. The Decision Tree classifier maintained a consistent performance of 98% across accuracy, precision, recall, and F1-score before and after cross-validation, indicating excellent stability and generalization capability. Conversely, the Naïve Bayes classifier experienced modest performance improvements following cross-validation, with accuracy increasing from 66% to 67%, precision from 74% to 75%, recall from 66% to 68%, and F1-score from 63% to 65%. Despite these improvements, the performance gap between the two classifiers remained substantial. The results suggest that the Decision Tree algorithm is more effective in modelling the liver dataset and capturing its underlying relationships, whereas the Naïve Bayes classifier is limited by its conditional independence assumption.

Therefore, the Decision Tree classifier is identified as the optimal classification model for this study due to its superior predictive performance, robustness, and consistency across validation scenarios.

Table 4. Comparison of Results with Existing Studies

S/N	Studies	Methodology (Algorithms & Dataset)	Results
1	[7]	Naïve Bayes; Liver Cirrhosis dataset	Naïve Bayes: ~31% accuracy
2	[8]	Bernoulli Naïve Bayes; ILPD (580 patients, 11 attributes)	67% accuracy; recall 0.82 (at-risk), 0.28 (healthy)
3	[9]	Naïve Bayes; BUPA (345 instances, 6 variables; 10-fold CV)	Accuracies: 64.6%, 66.7%, 64.3%
4	[10]	Decision Tree	Decision Tree: 100% accuracy

5	[11]	Decision Tree + LASSO feature selection	94.29% accuracy
6	[12]	J48 (Decision Tree variant)	95.04% accuracy
7	This study's best result	Decision Tree	Accuracy: 98%, Precision: 98%, Recall: 98%, F1-score: 98%.

Table 4 juxtaposes this study's Decision Tree results (98% across accuracy, precision, recall, F1-score) against prior literature, revealing tree-based models' consistent superiority over probability-based ones, though dataset variations temper direct equivalency. Probability-based studies [8][9] yield modest outcomes: Rahman et al. (2024) report ~31% Naïve Bayes accuracy on the Liver Cirrhosis dataset; [8] achieve 67% on ILPD (580 patients, 11 biomarkers like bilirubin/albumin, post-cleaning/normalization), with strong at-risk recall (0.82) but weak healthy detection (0.28); [9] achieves average ~65% on smaller BUPA (345 instances, 6 variables via cross-validation). These results lag due to Naïve Bayes' independence flaws on correlated liver features. Conversely, tree-based benchmarks [10][11][12] shine: [10] hit 100% Decision Tree; [11] reach 94.29% with LASSO-enhanced Decision Tree; [12] score 95.04% via J48. This study's 98% aligns closely, competitive yet realistic without extreme tuning. Overall, Tree-based dominance across studies affirms their fit for liver prediction, with this work's 98% validating generalizability without prone to overfitting (unlike 100% peaks).

5. Conclusion and Recommendations

In conclusion, this study demonstrates Decision Tree's superior performance over Naïve Bayes for liver disease prediction, aligning with literature trends where tree-based algorithms excel in handling correlated biomarkers on datasets like ILPD, though probability-based methods offer speed for prototypes, highlighting the need for standardized evaluations amid dataset variances. These findings advance ML for early detection, potentially reducing global morbidity in resource-limited settings.

For future studies, this study recommends developing hybrid Decision Tree-Naïve Bayes models via stacking or boosting to fuse trees' accuracy with Bayes' efficiency, tested on larger, diverse datasets, and incorporating explainability tools (SHAP/LIME) for transparency.

Data Availability

The dataset used for this paper are accessible through the link below:

<https://www.kaggle.com/datasets/abhi8923shriv/liver-disease-patient-dataset>

References

- [1] Neha, T., & Khandakar, F. Machine Learning in liver disease diagnosis: Current progress and future opportunities. *IOP Conf. Series: Materials Science and Engineering*. 2021; 1–19.
- [2] Ismail, I. A., Chukwuka, E. E., & Abiola, M. K. Improving business insights using machine learning algorithms. *Global Science Journal (GSJ)*. 2024; 12(2).
- [3] Akuji, I., Eletu, I. I., & Olasunkanmi, Y. Artificial intelligence and machine learning in modern technology: A comprehensive review. *Journal of Modern Technology*, 2024; 4, 156-162.
- [4] Samiksha, H., & Kamini, S. Prediction of Liver Disease using Machine Learning Algorithms. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*. 2020; 9(9), 57–59.
- [5] Jayakumar, S., Senthil, J., Ganesh, R. M., & Chellapan, N. Liver Disease Prediction Using Machine Learning Classification. *Webology*. 2021; 18, 441–452.
- [6] Mohammed, A., Hikmat, U., & Ahsan, M. An Empirical Comparative Analysis Using Machine Learning Techniques for Liver Disease Prediction. *International Journal of Healthcare Information Systems and Informatics*. 2021; 16(4), 1–12.
- [7] Rahman, F., Das, D., Sami, A., Podder, P., & Michael, D. L. Liver cirrhosis prediction using logistic regression, naïve bayes and KNN. *International Journal of Science and Research Archive*. 2024; 12(1), 2411-2420.
- [8] Nurhalisa, V., & Fajri, I. N. Liver Disease Classification using the Naive Bayes. *Sistemasi: Jurnal Sistem Informasi*, 2025; 14(4), 1517-1523.
- [9] Güre, Ö. B. Classification of Liver Disorders Diagnosis using Naïve Bayes Method. *Bitlis Eren Üniversitesi Fen Bilimleri Dergisi*. 2024; 13(1), 153-160.
- [10] Boggarapu, S., Nitesh, S., & Sasi, R. Implementation of Liver Disease Prediction Using Machine Learning. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*. 2021; 7(2), 606–616.
- [11] Saima, A., Javed, S., Tafsirul-Islam, N., Fahmida, M., Shakil, M., Imran, M., & Abdulla, M. Supervised machine learning based liver disease prediction approach with LASSO feature selection. *Bulletin of Electrical Engineering and Informatics*. 2021; 10(6), 3369–3376.
- [12] Vasan, D., Suyan, R., & Dinesh, K. Liver disease prediction using machine learning. *International Journal of Advance Research, Ideas and Innovations in Technology*. 2019; 5(2), 1584–1588.
- [13] Fahad, M., Easin, H., Morgan, W., & Hafiz, K. Statistical Machine Learning Approaches to Liver Disease Prediction. *Livers*. 2021; 1, 294–312.