FMDB ***Transactions on Sustainable Biomedical Engineering***

Leveraging Machine Learning for the Prognosis of Chronic Liver Disease

Abeer A. Amer[1,*]

[1]Department of Computer and Information System, Faculty of Management Science, Sadat Academy for Management Science, Alexandria, Alexandria Governorate, Egypt.
abamer.2000@gmail.com[1]

Abstract: This research applies machine learning (ML) techniques to predict liver disease using readily available blood test data, aiming to support early diagnosis and timely medical intervention. The study evaluates four widely used ML algorithms Random Forest, Support Vector Machine (SVM), Logistic Regression, and XG-Boost—using a dataset from the UCI Machine Learning Repository. To evaluate each model's effectiveness, several performance metrics were used, including the confusion matrix, precision, recall, F1 score, accuracy, and ROC_AUC, enabling a comprehensive comparison of classification performance. In the first phase of experimentation, the dataset was divided into 90% for training and 10% for testing. Under this configuration, the Random Forest model delivered the highest accuracy of 93%, followed by XG-Boost with 90%, while both Logistic Regression and SVM achieved 74% accuracy. Recognising the influence of data volume on model learning, the dataset was further split into 95% for training and 5% for testing. This modification resulted in notable performance improvements across all models. Random Forest achieved an impressive 98% accuracy, XG-Boost improved to 95%, Logistic Regression increased to 81%, and SVM rose to 79%, demonstrating the positive impact of additional training data on predictive performance.

Keywords: Chronic Liver Disease; Random Forest; Logistic Regression; Support Vector Machine (SVM); Machine Learning; Blood Test Data; Early Diagnosis; Confusion Matrix.

Received on: 18/06/2025, **Revised on:** 01/09/2025, **Accepted on:** 02/10/2025, **Published on:** 09/03/2026

Journal Homepage: https://www.fmdbpub.com/user/journals/details/FTSBE

DOI: https://doi.org/10.69888/FTSBE.2026.000672

Cite as: A. A. Amer, "Leveraging Machine Learning for the Prognosis of Chronic Liver Disease," *FMDB Transactions on Sustainable Biomedical Engineering*, vol. 1, no. 1, pp. 46–55, 2026.

1. Introduction

A Health Information System (HIS) is a comprehensive, organised system for managing all types of healthcare data, from patient records to operational and administrative records. It entails the methodical gathering, processing, reporting, and effective application of health-related data, enabling organisations to improve the efficiency and overall efficacy of health services at multiple administrative and clinical tiers [1]. As populations grow, healthcare needs grow, technology improves, and expectations for care quality rise, healthcare settings are becoming more complicated. This makes the function of strong information systems even more important. The Hospital Information System is at the heart of this larger framework. It is a very important part of medical informatics. The main purpose of a Hospital Information System is to use electronic data processing to make patient care better and make administrative tasks easier [2]. In today's healthcare world, it's very important to have

*Corresponding author.

rapid access to patient information and ensure it is gathered correctly. With a well-connected computerised system, healthcare personnel can quickly access patient data, review medical histories, track treatment progress, and make decisions that directly affect patient outcomes. A Hospital Information System must meet several important requirements to perform these tasks well. These include secure and accurate data storage, reliability in daily use, quick access to information, strong security measures to prevent unauthorised access to data, and cost-effectiveness in routine operation [4].

There are many benefits to using a whole hospital information system that improve both patient care and hospital management. One of the main benefits is that it makes it easier to access doctors' data, allowing you to maintain thorough records that include categories based on demographics, gender, age, medical issues, and other important factors. This is especially useful in outpatient settings, where the accuracy and timeliness of medical information are critical to continuity of care [6]. Also, internet-based technologies let people access this information from anywhere, so clinicians and administrators can get the information they need no matter where they are. This adaptability improves how things work, speeds up decision-making, and enhances care coordination. Another major benefit of a Hospital Information System is that it improves patient care. It helps hospital officials make decisions by serving as a decision-support system. This helps create comprehensive healthcare policies based on accurate and up-to-date data [7]. Hospitals may analyse patient trends, assess how well different treatments work, and make decisions based on hard data rather than guesswork. HIS tools also support key hospital functions, including finance, dietary management, engineering services, and medical aid. This centralised, organised approach enables managers to monitor resource use, reduce waste, and improve workflow efficiency. A facility Information System also provides healthcare leaders with a comprehensive, detailed view of how the facility has grown and developed over time [8].

Hospitals may assess how well they are fulfilling their goals, identify areas for improvement, and prepare for the future by systematically collecting and analysing data. HIS tools also make it much easier to keep track of how people use drugs and to study how well they work. These qualities are very important for reducing harmful drug interactions, keeping patients safe, and encouraging the use of safer, cheaper medications. Another major benefit of implementing HIS is improved documentation quality [9]. When electronic systems replace paper records, transcription errors occur far less often, and the integrity of the information is much stronger. Less redundant information and more reliable patient records make clinical workflows run more smoothly and patient histories more accurate. Hospital software is made to be easy to use and to reduce mistakes that happen when handwriting is hard to read or when data is entered by hand. Modern computerised systems work quite well, swiftly retrieving patient information and reducing the time required for administrative duties [10]. Healthcare organisations collect large volumes of data every day as part of routine clinical procedures. Most of the time, this data is stored in electronic health record (EHR) systems [12]. Even though huge amounts of information are available, they are often underused. But as data science applications become more common in healthcare, these data may be analysed to yield insights that improve the quality of care [14].

In the modern world, technology has come a long way, allowing people to gather data, clean it, identify patterns, and uncover hidden insights that would otherwise go undiscovered. This includes the ability to handle missing data, a major challenge in medical research that can significantly affect the accuracy of diagnostic models. Medical diagnoses are very significant for improving patient care, advancing research, and helping policymakers in the healthcare field [11]. Health practitioners use many different pathology methods and diagnostic tools to determine what's wrong with a patient and the best way to treat them [15]. These technologies create large data sets that can be used to make predictive tools that help doctors make decisions more quickly and accurately. Machine learning has had a significant impact on the biomedical sector over the last few years, especially in predicting and diagnosing diseases early and making therapeutic decisions. Researchers and doctors are very interested in improving the accuracy and speed of disease detection, and machine learning models have significant potential to achieve this [16]. The use of machine learning techniques makes the diagnosis procedure more objective and less dependent on human judgement alone [3]. These methods also make it easier to find medical concerns and lower the cost of diagnosis. One of the best uses of data mining and machine learning is to get useful information from large datasets more quickly, cheaply, and reliably than ever before [7]. Health systems worldwide are struggling as healthcare expenditures are rising, especially for older people. But new developments in information technology hold promises for improving the efficiency of healthcare delivery.

Information technology is still making big changes to the quality and availability of healthcare services [6]. As more and more doctors work to improve medical diagnoses, artificial intelligence and machine learning algorithms are being integrated into clinical processes to help anticipate how diseases will evolve. Machine learning models can help identify those most likely to develop major medical conditions by combining patient data and diagnostic information from multiple sources [4]. These high-risk individuals can then be targeted with early interventions to slow the onset of disease, improve health outcomes, and lower total healthcare costs. An examination of the current literature indicates that many machine learning models have been utilised to predict liver disease and other health issues. Earlier research has employed algorithms including C5.0, CHAID, Naïve Bayes, Decision Tree, Multi-layer Perceptron, k-NN, Random Forest, Logistic Regression, NBTree, SVM, AdaBoost, LogitBoost, Bagging, Grading, J48, Random Tree, REPTree, Decision Stump, and Hoeffding Tree. Many of these studies used datasets from the UCI machine learning repository, such as the liver patient dataset, which is often used and includes 416 patients with liver disease and 167 without [13]. This dataset contains 441 men and 142 women and was collected in the Andhra Pradesh region of northeastern India. Each entry has unambiguous class values that show the status of liver disease [8].

The majority of research assessing classification performance used metrics such as the confusion matrix, precision, recall, F1 score, accuracy, and ROC_AUC to evaluate predictive performance. This study seeks to determine the most effective machine learning method for early diagnosis of liver disease, building on previous research. Early diagnosis is very important for preventing disease progression, improving patient outcomes, and reducing healthcare expenditures. This study aims to create a robust predictive tool to aid healthcare decision-making by comparing algorithms from prior studies and consolidating the two most accurate models into a cohesive framework. The goals of this study are to identify the best algorithm for early disease detection, reduce the risk of worsening, improve health outcomes and costs, help healthcare organisations run their businesses more smoothly, support decision-making, and make patients more loyal and happier [11]. Healthcare organisations can turn vast amounts of clinical data into useful information by leveraging modern machine learning technology within a Health Information System. This gives doctors the information they need to make smart, timely decisions, which makes the organisation more efficient and, ultimately, leads to better health outcomes for patients. Healthcare systems can work towards a future where early detection, effective resource management, and personalised care are the norm by continuing to study and improve predictive models. This will be possible because of the potent combination of HIS infrastructure and machine learning innovation [5].

2. Materials and Methods

The study was split into three phases. Before applying machine learning techniques, the dataset is described and preprocessed, and the importance of features is evaluated as a first step. The next step is to build and deploy different machine learning models. The last step is to analyse the performance of each model with the help of confusion matrix, precision, recall, F1 score, accuracy and ROC_AUC. Finally, the evaluation phase is done and the two best models are selected to predict liver disease by the comparative analysis in the Figure 1. The study utilised Python for model implementation.

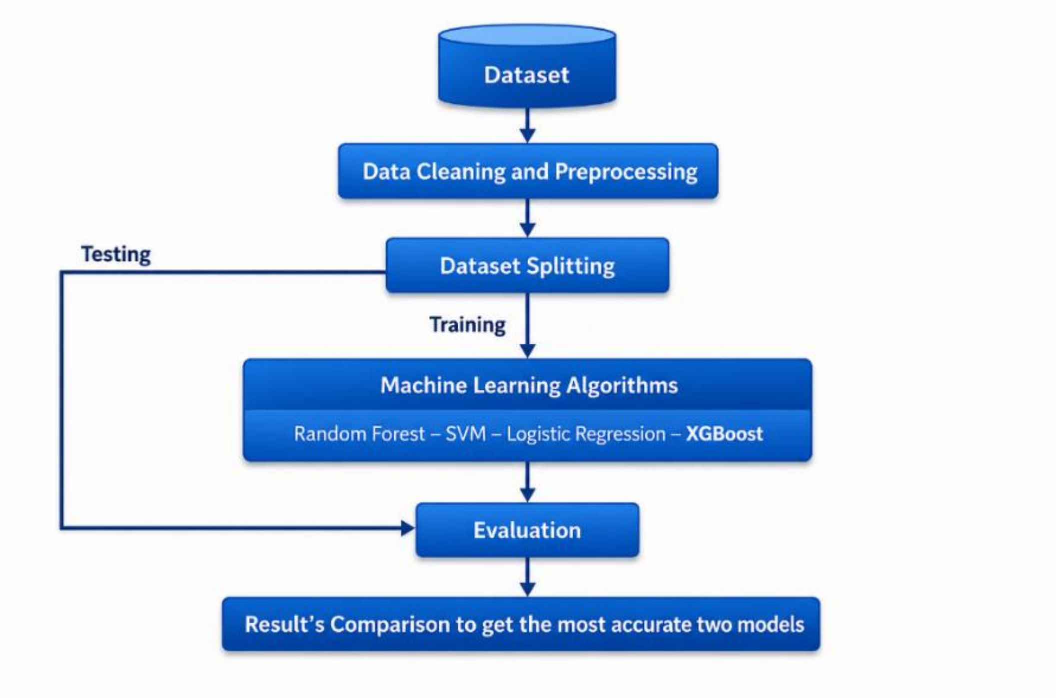

Figure 1: Methodology of the proposed model

2.1. Dataset Description

The Indian liver patient dataset, downloaded from the UCI machine learning repository, was used in this study. It includes 416 liver patients and 167 suspicious livers who are not considered patients. Also, it is composed of 441 males and 142 females. This data was collected from the Andhra Pradesh area in north-east India and labelled with class values indicating whether each patient has liver disease. Table 1 below presents the attributes, their data types, and annotations.

Table 1: Attributes of the dataset

Data	Column	Non-Null Count	Data Type
0	Age	583 non-null	int64
1	Gender	583 non-null	int64
2	TB.	583 non-null	float64
3	DB.	583 non-null	float64
4	Alkphos.	583 non-null	int64
5	SGPT.	583 non-null	int64
6	SGOT.	583 non-null	int64
7	TP.	583 non-null	float64
8	AL.	583 non-null	float64
9	A/G. Ratio	579 non-null	float64
10	Class	583 non-null	int64

A sample of the liver disease dataset used in this study is presented in Table 2. Table 2 presents demographic data, such as Age and gender, and a handful of biochemical parameters, such as Total Bilirubin (TB), Direct Bilirubin (DB), Alkaline Phosphotase (Alkphos), SGPT, SGOT, Total Protein (TP), Albumin (Alb) and Albumin/Globulin (A/G) Ratio. The attribute Class is the classification target variable. The sample records show varying levels of liver function indicators across patients, demonstrating the variety of clinical characteristics in the dataset. These features are important predictors for the detection and characterisation of liver disease patterns in machine learning models.

Table 2: Liver patient records dataset

No.	Age	Gender	TB.	DB.	Alkphos.	SGPT.	Sgot.	TP.	Alb.	A/G. Ratio	Class
0	65	1	0.7	0.1	187	16	18	6.8	3.3	0.90	1
1	62	2	10.9	5.5	699	64	100	7.5	3.2	0.74	1
2	62	2	7.3	4.1	490	60	68	7.0	3.3	0.89	1
3	58	2	1.0	0.4	182	14	20	6.8	3.4	1.00	1
4	72	2	3.9	2.0	195	27	59	7.3	2.4	0.40	1

2.2. Data Cleaning and Preprocessing

Data quality is indeed a crucial factor for the data mining process, particularly for disease prediction and diagnosis. To obtain a high quality of data, there are several steps to follow such as data cleaning which is very important to improve the accuracy of the prediction. Hence, the researcher employed common procedures and techniques of data cleaning that can be used to improve data quality:

- **Handling Missing Values:** Some attributes in this dataset had some missing values. In this study, the mean was used to fill in missing values in the datasets. This method was chosen because it is a simple and widely used technique for handling missing data.

Researchers observe a strong correlation among some variables in Table 3, so researchers need to drop some of these variables as features should be independent.

Table 3: Heatmap of the correlation between the dataset attributes based

Feature	Age	Gender	TB.	DB.	Alkphos.	SGPT.	SGOT.	TP.	ALB.	A/G. Ratio	Class
Age	1.000000	0.056560	0.011763	0.007529	0.080425	-0.086883	-0.019910	-0.187461	-0.265924	-0.216089	0.137351

Class	A/G. Ratio	ALB.	TP.	SGOT.	SGPT.	Alkphos.	DB.	TB.	Gender
0.137351	-0.216089	-0.265924	-0.187461	-0.019910	-0.086883	0.080425	0.007529	0.011763	0.056560
0.082416	-0.003404	-0.093799	-0.089121	0.080336	0.082332	-0.027496	0.100436	0.089291	1.000000
0.220208	-0.206159	-0.222250	-0.008099	0.237831	0.214065	0.206669	0.874618	1.000000	0.089291
0.246046	-0.200004	-0.228531	-0.000139	0.257544	0.233894	0.234939	1.000000	0.874618	0.100436
0.184866	-0.233960	-0.165453	-0.028514	0.167196	0.125680	1.000000	0.234939	0.206669	-0.027496
0.163416	-0.002374	-0.029742	-0.042518	0.791966	1.000000	0.125680	0.233894	0.214065	0.082332
0.151934	-0.070024	-0.085290	-0.025645	1.000000	0.791966	0.167196	0.257544	0.237831	0.080336
-0.035008	0.233904	0.784053	1.000000	-0.025645	-0.042518	-0.028514	-0.000139	-0.008099	-0.089121
-0.161388	0.686322	1.000000	0.784053	-0.085290	-0.029742	-0.165453	-0.228531	-0.222250	-0.093799
-0.162319	1.000000	0.686322	0.233904	-0.070024	-0.002374	-0.233960	-0.200004	-0.206159	-0.003404
1.000000	-0.162319	-0.161388	-0.035008	0.151934	0.163416	0.184866	0.246046	0.220208	0.082416

- **Outlier Detection and Treatment:** Researchers have skewed features, as shown below. According to Data Transformation, they need to fix this by using the log1p transformation to scale these features.

Then, researchers split the data into a majority and a minority using 2 datasets, and merged the majority with the up-sampled minority. And researchers are working on two options for dataset splitting: 90% training and 10% testing, and 95% training and 5% testing.

2.3. Model Selection

In this study, researchers have employed various machine learning techniques, including Random Forest, SVM, Logistic Regression, and XG-Boost, as shown in Figure 2.

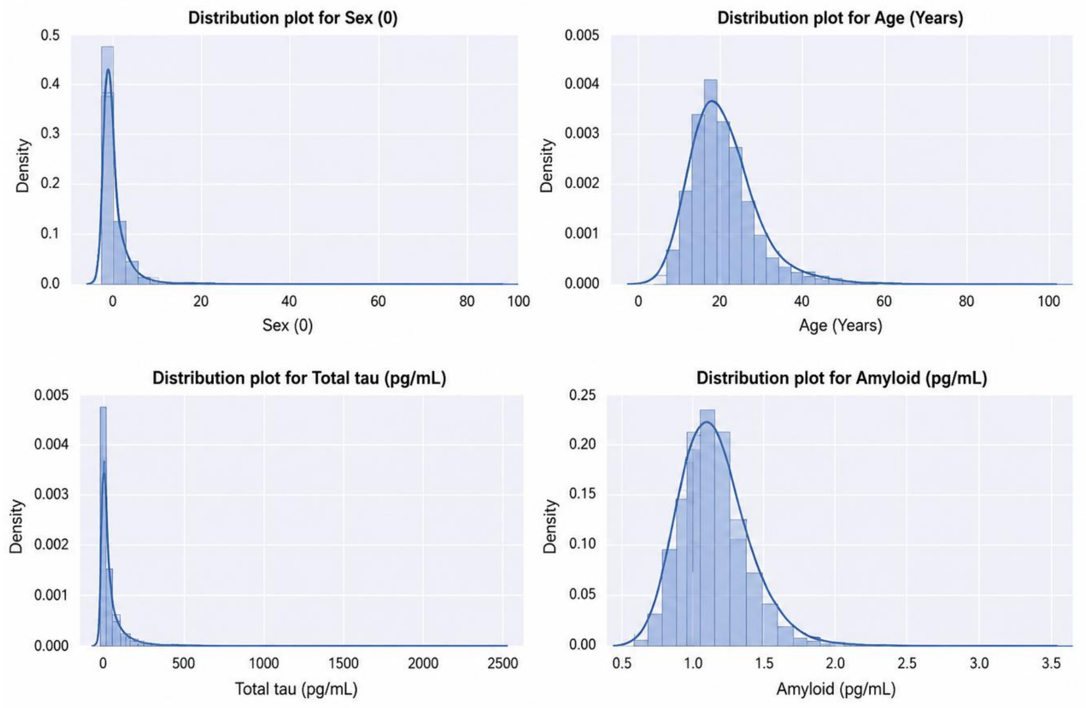

Figure 2: Distribution plot of techniques

2.4. Evaluation

Researchers used performance metrics such as the confusion matrix, precision, recall, F1 score, accuracy, and ROC_AUC.

Table 4: Confusion matrix for classification model evaluation

Actual / Predicted	Positive	Negative
Positive Actual	True Positives	False Positives
Negative Actual	False Positives	True Positive

- **A Confusion Matrix** is a Table 4 often used in machine learning classification tasks to visualise an algorithm's performance. It is especially useful for assessing the performance of binary classification models.
- **Layout of the Matrix:** The confusion matrix is typically arranged as follows.
- **Derived Metrics:** From the confusion matrix, several performance metrics can be calculated.
- **Accuracy:** Overall, how often is the model correct? (TP + TN) / (TP + TN + FP + FN).
- **Precision:** When it predicts the positive class, how often is it correct? TP / (TP + FP).
- **Recalling (Sensitivity):** How often does it correctly identify actual positives? TP / (TP + FN).
- **F1 Score:** A balance between precision and recall. 2 * (Precision * Recall) / (Precision + Recall).
- **Specificity:** How often does it correctly identify actual negatives? TN / (TN + FP).
- **ROC_AUC score:** is a performance measurement for classification problems at various threshold settings. ROC stands for "Receiver Operating Characteristic," and AUC stands for "Area Under the Curve." This score is used to evaluate a classification model's predictive power and is particularly useful for binary classification problems.

2.4.1. Understanding ROC Curve

- **ROC Curve:** It's a plot with the True Positive Rate (TPR, or Recall) on the y-axis and the False Positive Rate (FPR) on the x-axis at various threshold settings.

- **True Positive Rate (TPR):** It's calculated as TPR = TP/(TP + FN), where TP is the number of true positives, and FN is the number of false negatives.
- **False Positive Rate (FPR):** It's calculated as FPR = FP/(FP + TN), where FP is the number of false positives, and TN is the number of true negatives.
- **Thresholds**: The ROC curve illustrates the performance of a classification model at all classification thresholds. Lowering the classification threshold classifies more items as positive, thus increasing both False Positives and True Positives.

2.4.2. AUC - Area Under the ROC Curve

- **AUC Score:** It measures the entire two-dimensional area underneath the entire ROC curve from (0,0) to (1,1).

2.4.3. Interpretation

- **AUC = 1:** The model makes perfect predictions.
- **0.5 < AUC < 1:** The model is better than random guessing. The higher the AUC, the better the model is at distinguishing between positive and negative classes.
- **AUC = 0.5:** The model's predictions are no better than random guessing.
- **AUC < 0.5:** The model is worse than random guessing. This usually suggests that something is wrong with the model or the data.

3. Results

After applying the previous phases, from data cleaning and preprocessing to model deployment, researchers found the following results, as shown in Tables 5 and 6. By using the dataset, splitting 90% training and 10% testing.

Table 5: Results of the techniques

ML. Algorithms	Accuracy	Precision		Recall		F1-Score		ROC_AUC Score
		0	1	0	1	0	1	
RF	0.93	0.94	0.92	0.94	0.92	0.94	0.92	0.93
SVM	0.74	0.70	0.86	0.94	0.49	0.80	0.62	0.71
LR	0.74	0.74	0.74	0.83	0.62	0.78	0.68	0.73
XG-Boost	0.90	0.91	0.89	0.91	0.89	0.91	0.89	0.90

Figure 3 shows that the RF and SVM models had the highest True Positive (TP) values (44), while XGBoost had a TP of 43. SVM had the highest number of False Positives (FPs) at 19, whereas RF had the fewest at 3, suggesting fewer false-positive predictions.

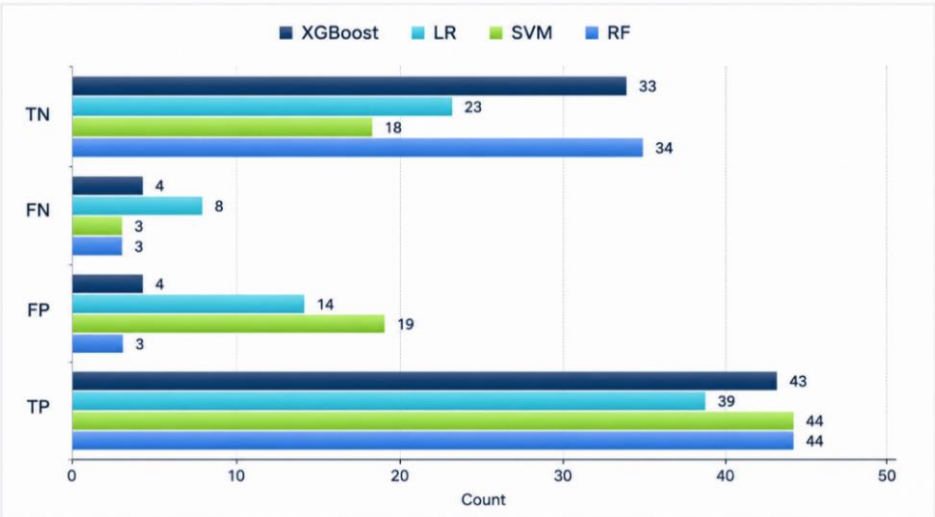

Figure 3: Results of the techniques

Moreover, RF had the highest True Negative (TN) value (34) and the lowest False Negative (FN) count (3), indicating the most balanced classification performance among the investigated models. The confusion matrix results are shown in Figures 3 and 4 using the dataset, with 95% training and 5% testing.

Table 6: Results of the techniques

ML. Algorithms	Accuracy	Precision		Recall		F1-Score		ROC-AUC Score
		0	1	0	1	0	1	
RF	0.98	0.96	1	1	0.93	0.98	0.97	0.97
SVM	0.79	0.82	0.71	0.85	0.67	0.84	0.69	0.76
LR	0.81	0.88	0.71	0.81	0.8	0.85	0.75	0.81
XG-Boost	0.95	0.93	1	1	0.87	0.96	0.93	0.93

In Figure 4, XGBoost and RF achieved the highest True Positive (TP) score of 27, indicating strong detection capability. RF achieved the lowest FP count (1) and no FN cases, indicating better classification accuracy and the confusion matrix results as follows in Figure 4.

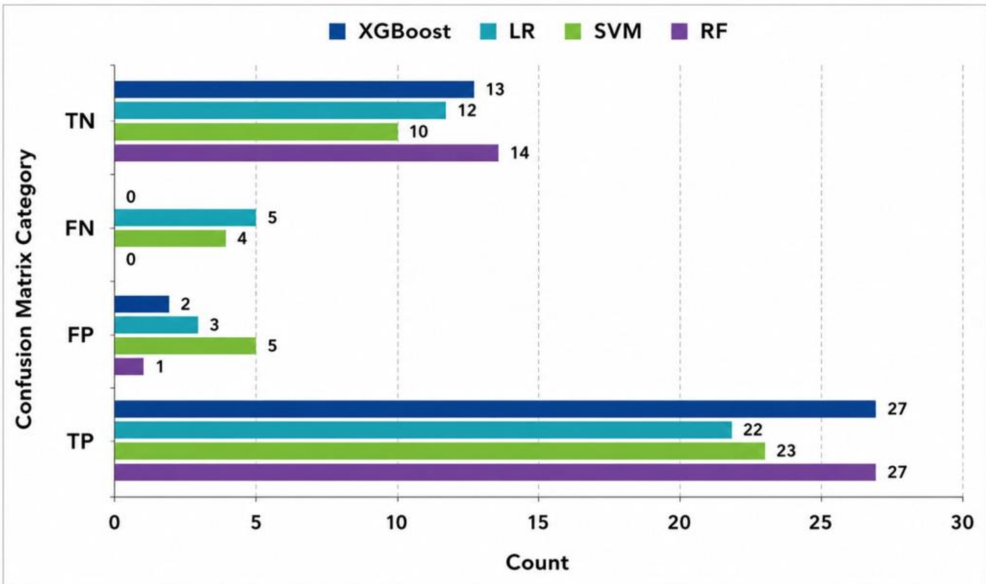

Figure 4: Results of the techniques

In addition, RF achieved the highest True Negative (TN) rate (14), indicating it correctly identifies negative instances compared to other models.

4. Conclusion

This study seeks to elucidate and evaluate several prediction methodologies that facilitate early diagnosis and prognosis of liver disease, underscoring the importance of prompt detection to improve health outcomes and mitigate long-term medical complications. The study assesses the efficacy of several Machine Learning algorithms for classifying the presence of illness by analysing hepatitis-related indicators across multiple independent variables using conventional performance criteria. The objective is not only to discern the most precise models but also to understand how dataset construction and algorithmic architecture affect the reliability of predictions in practical applications. Researchers used several categorisation criteria, such as accuracy, precision, and recall, to assess the model's performance. These numbers show how well each algorithm identifies liver disease while keeping false positives and false negatives in check. The Random Forest method performed best, demonstrating its ability to handle complex, nonlinear interactions in the dataset. The Random Forest algorithm achieved 93% accuracy when the data were split into 90% for training and 10% for testing. This was better than the other models. XGBoost was next with 90% accuracy, and Logistic Regression and Support Vector Machine were each at 74%. These results show that traditional and ensemble-based models differ in their ability to detect patterns in medical data. When the dataset was split into 95% for training and 5% for testing, all models performed better. This shows that having more data to learn from is helpful. With this setup, Random Forest achieved an impressive 98% accuracy, indicating it can handle larger training sets effectively.

XG-Boost also improved, achieving 95% accuracy, indicating it is a good gradient boosting classifier. The accuracy of Logistic Regression increased to 81%, and that of SVM increased to 79%. This shows that simpler linear models perform better with larger training sets, but they still don't capture complex decision boundaries as well as more complex ensemble methods. These results show that model choice significantly affects diagnostic accuracy and that algorithms such as Random Forest and XG-Boost may be better suited for medical prediction tasks. The findings of this study underscore the importance of selecting appropriate Machine Learning methodologies for clinical prediction and illustrate the potential of AI-driven strategies to assist healthcare practitioners in early diagnosis of liver disease. The study also shows that the dataset's size and quality significantly affect the model's performance. This means that more diversified and complete data could lead to even better results. In the future, researchers will add more data to the dataset to improve rule-generation accuracy and provide a more representative sample of patients with liver disease. Also, exploring other ways to weight the data might improve classification, allowing the models to handle datasets or factors that don't contribute equally to disease prediction. Using innovative classification methods and hybrid models could also help the research by improving the accuracy and reliability of early diagnosis. The paper aims to advance more effective, data-driven healthcare technologies that help doctors improve patient outcomes in the diagnosis and treatment of liver disease through these ongoing efforts.

Acknowledgement: The author expresses sincere gratitude to Sadat Academy for Management Sciences for providing the academic support, facilities, and resources necessary for the successful completion of this research.

Data Availability Statement: The data that support the results and conclusions of this study are available from the corresponding author upon reasonable request.

Funding Statement: This research was conducted without receiving any dedicated financial assistance, grant, or funding from public, commercial, or non-profit organizations.

Conflicts of Interest Statement: The author declares that there are no financial, personal, institutional, or professional conflicts of interest that could have influenced the design, execution, interpretation, or reporting of this research.

Ethics and Consent Statement: This study was conducted in accordance with established ethical standards. Participants were informed that their responses would remain confidential and anonymous.

References

1. M. Abdar, M. Z. Moghadam, R. Das, and I.-H. Ting, "Performance analysis of classification algorithms on early detection of liver disease," *Expert Systems with Applications*, vol. 67, no. 1, pp. 239–251, 2017.
2. H. Jin, S. Kim, and J. Kim, "Decision Factors on Effective Liver Patient Data Prediction," In*ternational Journal of Bio-Science and Bio-Technology*, vol. 6, no. 4, pp. 167–178, 2014.
3. S. N. N. Alfisahrin and T. Mantoro, "Data Mining Techniques for Optimization of Liver Disease Classification," *International Conference on Advanced Computer Science Applications and Technologies*, Kuching, Malaysia, 2013.
4. M. Ghosh, M. M. S. Raihan, M. Raihan, L. Akter, A. K. Bairagi, S. S. Alshamrani, and M. Masud, "A Comparative Analysis of Machine Learning Algorithms to Predict Liver Disease," *Intelligent Automation and Soft Computing*, vol. 30, no. 3, pp. 917–928, 2021.
5. F. J. John Joseph, K. Chinnusamy, J. Jeganathan, A. J. Obaid, and S. S. Rajest, Eds., Optimizing Patient Outcomes Through Multi-Source Data Analysis in Healthcare, *IGI Global*, United States of America, 2025.
6. C. C. Wu, W. C. Yeh, W. D. Hsu, M. M. Islam, P. A. Nguyen, T. N. Poly, Y. C. Wang, H. C. Yang, and Y. C. Li, "Prediction of fatty liver disease using machine learning algorithms," *Computer Methods and Programs in Biomedicine*, vol. 170, no. 3, pp. 23–29, 2019.
7. A. J. Obaid, M. Muthmainnah, S. S. Rajest, and M. Baron, Eds., AI in Health and Human-Centric Systems. Advances in Computational Intelligence and Robotics, *IGI Global*, United States of America, 2025.
8. A. K. M. S. Rahman, F. M. J. M. Shamrat, Z. Tasnim, J. Roy, and S. A. Hossain, "A Comparative Study on Liver Disease Prediction Using Supervised Machine Learning Algorithms," *International Journal of Scientific & Technology Research*, vol. 8, no. 11, pp. 419–422, 2019.
9. M. Pasha and M. Fatima, "Comparative Analysis of Meta Learning Algorithms for Liver Disease Detection," *Journal of Software*, vol. 12, no. 12, pp. 923–932, 2017.
10. S. K. Sehrawat, "Leveraging AI for early detection of chronic diseases through patient data integration," *AVE Trends in Intelligent Health Letters*, vol. 1, no. 3, pp. 125–136, 2024.
11. N. Nahar and F. Ara, "Liver Disease Prediction By Using Different Decision Tree Techniques," *International Journal of Data Mining & Knowledge Management Process (IJDKP)*, vol. 8, no. 2, pp. 1-9, 2018.

12. M. B. Priya, P. L. Juliet, and P. R. Tamilselvi, "Performance Analysis of Liver Disease Prediction Using Machine Learning Algorithms," *International Research Journal of Engineering and Technology (IRJET)*, vol. 5, no. 1, pp. 206–211, 2018.
13. A. N. Arbain and B. Y. P. Balakrishnan, "A Comparison of Data Mining Algorithms for Liver Disease Prediction on Imbalanced Data," *International Journal of Data Science and Advanced Analytics*, vol. 1, no. 1, pp. 1-11, 2019.
14. J. Soni, U. Ansari, D. Sharma, and S. Soni, "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction," *International Journal of Computer Applications*, vol. 17, no. 8, pp. 43–48, 2011.
15. P. Thangaraju and R. Mehala, "Novel Classification based approaches over Cancer Diseases," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 4, no. 3, pp. 294–297, 2015.
16. H. K. Sindhu, M. S. Koti, S. Priyadarshini, R. Regin, A. M. Suhail, and S. Jaber, "A machine learning approach to analyze and predict liver inflammation using Python," *AVE Trends in Intelligent Health Letters*, vol. 2, no. 2, pp. 85–93, 2025.