iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** I **Month of publication:** January 2026

DOI: https://doi.org/10.22214/ijraset.2026.76976

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue I Jan 2026- Available at www.ijraset.com

A Multi-Metric Evaluation Framework for Enhancing Supervised Learning Models in Liver Disease Prediction

Mr. Ashish Vishvakarma[1], Dr. Sandhya Dwivedi[2]
[1]*Research Scholar, Department of Computer Science and Engineering, Asian International University, Manipur, India*
[2]*Associate Professor, Department of Computer Science and Engineering, Asian International University, Manipur*

Abstract: ***Early detection of liver disease is very important because it can help reduce the risk of liver problems. Supervised Machine Learning (ML) models help in predicting liver disease, but their performance depends on how they are evaluated and improved. In this research study, we examine different supervised learning models to understand how accurately they can predict liver disease. in this study, we use multiple evaluation methods, such as precision, recall curves, and K-Fold cross-validation to test the models more effectively, especially when the dataset is unbalanced. Several ML algorithms like Logistic Regression, Random Forest, Support Vector Machine, and Decision Tree are trained using real medical data. Many existing studies rely heavily on a single metric, such as accuracy. Our results show that using different evaluation metrics gives a better understanding of each model's performance and helps improve prediction accuracy. Overall, the study shows that using a combination of evaluation techniques can lead to more reliable Machine learning models for liver disease diagnosis.***
Keywords: Liver Disease Prediction, Machine Learning, Cross Validation, Logistic Regression, Random Forest, Supervised Machine Learning

I. INTRODUCTION

Liver disease remains a critical global health challenge for substantial mortality and healthcare. Early detection plays a vital role in preventing the progression of liver disorders; traditional diagnostic procedures are often costly and time-consuming. In the growing structured healthcare data, machine learning (ML) has emerged as a powerful tool for supporting clinical decision-making through automated disease prediction systems.

Existing studies on liver disease prediction, including the KNN-Based work, are limited by the use of small, outdated datasets and by the evaluation of only a single machine learning model. Lack of validation approach: the author used only 50% for training and 50% for testing, which is weak for medical data; no K-fold cross-validation or statistical testing was performed to ensure model reliability [6].

This study proposes an enhanced evaluation methodology incorporating ROC-AUC, precision, Recall curves, and cross-validation to optimize supervised learning models for liver disease prediction. These metrics collectively provide deeper insight into model capability, sensitivity to positive cases, and generalization performance[7]. Machine learning algorithms on real-world liver disease datasets, the work aims to identify the most robust model configuration and highlight the significance of multi-metric evaluation in healthcare applications.

A. Primary Liver Diseases

1) Fatty Liver A reversible disease known as fatty liver occurs when large cholesterol fat vacuoles form in liver cells as a result of limited setting. It might happen to those who have a high amount of alcohol use as well as to those who have never consumed alcohol[14].

2) Cirrhosis is one of the most dangerous liver conditions. It is a procedure used to identify all types of liver disorders distinguished by notable cell loss. The liver gets hard and leathery as it steadily shrinks in size. Under liver cirrhosis, the regeneration process persists, but the progressive loss of liver cells outweighs cell replacement.

3) Hepatitis is typically brought on by a virus that is transferred by excessive contamination or close contact with infected bodily fluids that are infected[2].

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue I Jan 2026- Available at www.ijraset.com

II. LITERATURE REVIEW

This research paper uses Artificial Immune System (AIS) algorithms to detect liver disorders from blood test data. Earlier studies that used only the standard AIS method, two improved real-valued negative selection algorithms are proposed and compared with artificial neural network (ANN) and common classifiers. Both AIS and ANN perform better than traditional methods such as Naïve Bayes, decision trees, and SVM. ANN provides the highest detection rate, 83.73% , it fails when trained only on normal data. In contrast, the modified variable AIS algorithms can be trained using only a normal sample and still perform competitively, achieving an 81118% detection rate [1].

In this paper, Unsupervised machine learning techniques are used to predict liver disease. This performance is evaluated using metrics such as V-measure, Completeness, Homogeneity, Adjusted Rand Index, Adjusted Mutual Information, and Silhouette Coefficient. Among these, the Silhouette coefficient is the key factor as it reflects accuracy and optimal cluster count. Based on the K-means is chosen as the best method. Future work includes applying these techniques to diagnose other organs (Heart, Brain, Lungs) and integrating the approach with methods used to predict liver transplantation success[2].

In this work, several Classifiers were applied to a liver disease dataset to predict patient outcomes using a developed software tool. The dataset was processed in 10-fold cross-validation. Results with and without feature selection were compared based on accuracy and execution time, along with metrics like the kappa statistic, correctly classified instances, and mean absolute error. Logistic regression with feature selection produced the best accuracy, and feature selection also reduces execution time for all classifiers. finally , and intelligent liver Disease Prediction system(IPDPS) was developed using software engineering principles[3].

This proposed work generates synthetic samples for the minority class and adjusts the model's cost function; our system reduces the impact of class imbalance. This creates a more balanced dataset and improves model performance. Testing on independent shows clear gains in accuracy, precision, recall, and F1–score [4].

This paper has looked at and analysed the prediction of liver disease in patients. This paper examines and analyses the prediction of liver illness in patients after the data has been cleaned using a variety of techniques, including imputation of missing values with and dummy encoding. The data was cleaned using a variety of methods, including dummy encoding after the median was used to remove any missing values.[5].

III. PROPOSED METHODOLOGY

There are several stages in the research methodology. Each component of the methodology is as follows.

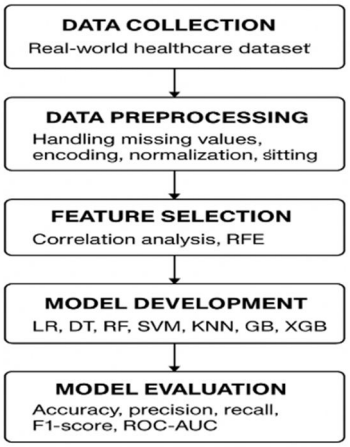


Figure: 1

A. Data Collection

Data selection is crucial for choosing important records for analysis and for using different data mining techniques to gain useful or constructive knowledge.

B. Data Preprocessing

This method is employed to extract the missing values from the data and use the median to impute the null values. Four missing values for the albumin and globulin ratios in the dataset of Indian patients with liver disease have been filled in using the median value [5].

C. Future Selection

Feature selection is the process of reducing the number of input variables so that the machine learning algorithm can train the model more quickly. It simplifies the computation and facilitates interpretation[12].

D. Model Development

Logistic regression, decision trees, support vector machines, and other supervised algorithms were used in the model development phase to verify that the model worked effectively and predicted liver disease.

E. Model Evaluation

One of the most important aspects of any research project is model evaluation. You might get results that are in line with your needs when you assess your model using a few common assessment measures. In this study, Accuracy, precision, recall, F1-Score, and ROC-AUC the suggested model using the following metrics in comparison to other models[13].

IV. DESCRIPTION OF THE DATASET

The Indian Liver Patient Dataset (ILPD) dataset is used to extract databases with 900 records or entries in order to address the problem addressed in this paper. In 900 Patients, there were 400 Female Patients and 500 Male Patients, aged 20 years to 80 years old. The source of this dataset is the UCI Machine Learning Repository. The entire ILPD dataset contains information on 850 active liver patients. A selector is a class label that divides individuals into groups according to whether or not they have liver disease. Age, gender, total bilirubin, alkaline phosphatase, alanine, aminotransferase, and albumin are among the attributes listed in the dataset.

```
      Result
0          1
1          1
2          1
3          1
4          1
..       ...
895        1
896        1
897        2
898        2
899        2

[900 rows x 11 columns]
```

Figure 2

```
Index(['Age ', 'Gender of the patient', 'Total Bilirubin', 'Direct Bilirubin',
       ' Alkphos Alkaline Phosphotase', ' Sgpt Alamine Aminotransferase',
       'Sgot Aspartate Aminotransferase', 'Total Protiens', ' ALB Albumin',
       'A/G Ratio Albumin and Globulin Ratio'],
      dtype='object')
     Age  Gender of the patient  Total Bilirubin  Direct Bilirubin  \
0     65                 Female              0.7               0.1
1     26                 Female              0.9               0.2
2     29                 Female              0.9               0.3
3     74                 Female              1.1               0.4
4     40                 Female              0.9               0.3
..   ...                    ...              ...               ...
895   46                   Male             15.8               7.2
896   52                   Male              1.8               0.8
897   42                   Male              0.8               0.2
898   42                   Male              0.8               0.2
899   62                   Male              0.7               0.2
```

Figure 3

Figure 3 , Age, Gender, Total Bilirubin, Direct Bilirubin, Alkaline Phosphatase, SGPT, SGOT, Protein ALB, and Ratio Albumin and Globulin Ratio are among the eleven columns in this dataset that contain liver attributes.

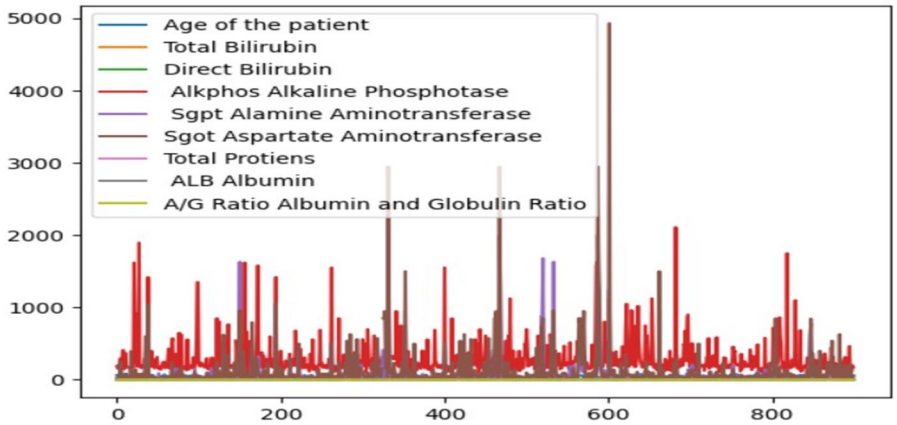


Figure:4

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue I Jan 2026- Available at www.ijraset.com

A. Performance Evaluation Matrices

Model evaluation is a crucial part of any research endeavour. When you evaluate your model using a few standard evaluation metrics, you may find results that satisfy your needs. The following metrics are used in this study to estimate the suggested model in comparison to other models[8]. There are some parameters to evaluate Performance matrices.

Tr Po: - True Positive
Tr Ne: - True Negative
Fa Po: - False Positive
Fa Ne: - False Negative

Accuracy is the most commonly used evaluation metric in a classification problem. it measures the correctly predicted cases out of the total cases.

$$\text{Accuracy} = \frac{(\text{TrPo}) + (rTNe)}{\text{Total Number of Predictions}}$$

B. Recall (True Positive rate)

$$Recall = \frac{(\text{TrPo})}{(\text{TrPo}) + (FaNe)}$$

It measures how well the model identifies actual positives

Precision:

$$Precision = \frac{(\text{TrPo})}{(\text{TrPo}) + (FaPo)}$$

F1-Score:

$$F1\,Score = 2 * \frac{\text{Precision} \,.\, \text{Recall}}{\text{Precision} + \text{Recall}}$$

Provide a balance between precision and Recall

C. Confusion Matrix

The positive Rate (Recall/Sensitivity)

Fa Po R: - False Positive Rate :

$$\boldsymbol{FaPoR} = \frac{\text{FaPo}}{\text{FaPo} + \text{TrNe}}$$

FaNeR:- False Negative Rate :

$$\boldsymbol{FaNeR} = \frac{\text{FaNe}}{\text{FaNe} + \text{TrPo}}$$

D. ROC-AUC

Receiver Operating Characteristic (ROC) Area Under the Curve(AUC) it is one the most important performance evaluation metrics for classification models, especially medical prediction problems like liver disease.

The ROC curve is a graph that shows how a model performs at different classification thresholds it plots: X-axis—False positive (FaPoR) , Y-axis—True Positive (TrPoR)

It is a single number between 0 and 1 that summarizes performances:

AUC Score	Meaning
0.90-1.00	Outstanding model
0.80-0.89	Decent Model
0.70-0.79	Fair Model
0.60-0.69	poor model
0.50	No better than random guessing

V. MACHINE LEARNING ALGORITHMS

Supervised Learning Algorithms used labelled training data to determine the mapping function that converts input variables (X) into the output variable (Y) in order to make the best prediction. Like Logistic Regression, Support Vector Machine, decision tree, and Random Forest. To put it another way, it solves for f in the following equation:

$$Y = f(X).$$

This enables us to produce outputs that are accurate when new inputs are provided[9].

A. Analysis Of Machine Learning Model Performance Using The Provided Matrices

Model performance summary based on the chart given below:

Table: 1

Metric	LR	SVM	DT	RF
Accuracy	0.6778	0.6889	0.7944	0.8389
precision	0.7215	0.7118	0.837	0.8613
Recall	0.8906	0.9453	0.8828	0.9219
F1-Score	0.7972	0.8121	0.8593	0.8906

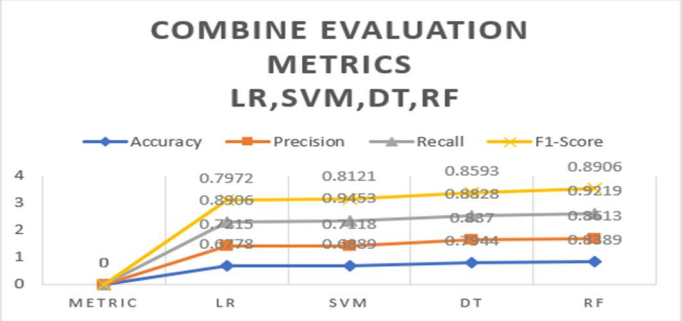


Figure: 5

Result: Among all evaluated models (LR, SVM.DT, RF), the Random Forest classifier demonstrated the best overall performance with the highest accuracy, Precision, and F1-Score. Although SVM achieved the highest recall, Random Forest provided a more balanced and reliable classification performance, making it the most effective model for predicting liver disease in this dataset.
Interpretation of Combined Confusion Metrics visualizes the confusion matrices of the four Machine Learning models evaluated for liver disease prediction.

Table: 2

Confusion Metrics (LR)

Actual\Predicted	Positive (1)	Negative (0)
Positive (1)	8	44
Negative (0)	114	14

Confusion Metrics (SVM)

Actual\Predicted	Positive (1)	Negative (0)
Positive (1)	121	7
Negative (0)	49	3

Confusion Metrics (DT)

Actual\Predicted	Positive (1)	Negative (0)
Positive (1)	113	15
Negative (0)	22	30

Confusion Metrics (RF)

Actual\Predicted	Positive (1)	Negative (0)
Positive (1)	118	10
Negative (0)	19	33

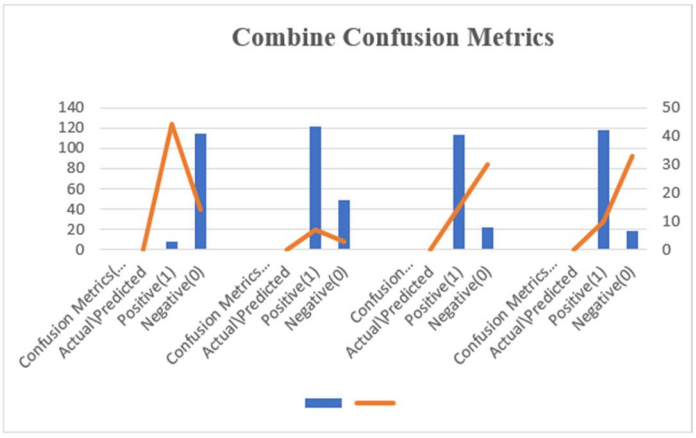


Figure: 6

Result: The Combined confusion metrics demonstrate that the Random Forest and Decision Tree models provide a balanced performance with high true positives and true negatives. SVM achieves the lowest false negatives, making it the most effective for capturing actual disease cases. Logistic Regression shows the weakest performance, particularly due to higher false negatives. Overall, Random Forest provides the best balance between sensitivity and specificity.

VI. DISCUSSION

This study's results align with contemporary literature, which emphasizes the necessity of combining multiple evaluation metrics and validation, including K-Fold cross-validation, to enhance reliability[10]. This multi-metric framework ensures that predictive models are not only accurate but also trustworthy, generalizable, and clinically relevant.

Sustainability Goal:

The long-term sustainability goal of this research is aligned with United Nations Sustainable Development Goal(SDG) 3: Good Health and Well-being (SDG) 9: Research and innovation in technology by developing machine learning models that improve early detection and diagnosis of liver disease. This study supports the creation of accurate and scalable healthcare technologies. This multi-metric evaluation framework ensures that these predictive tools are not only technically effective but also suitable for real-world medical environments.

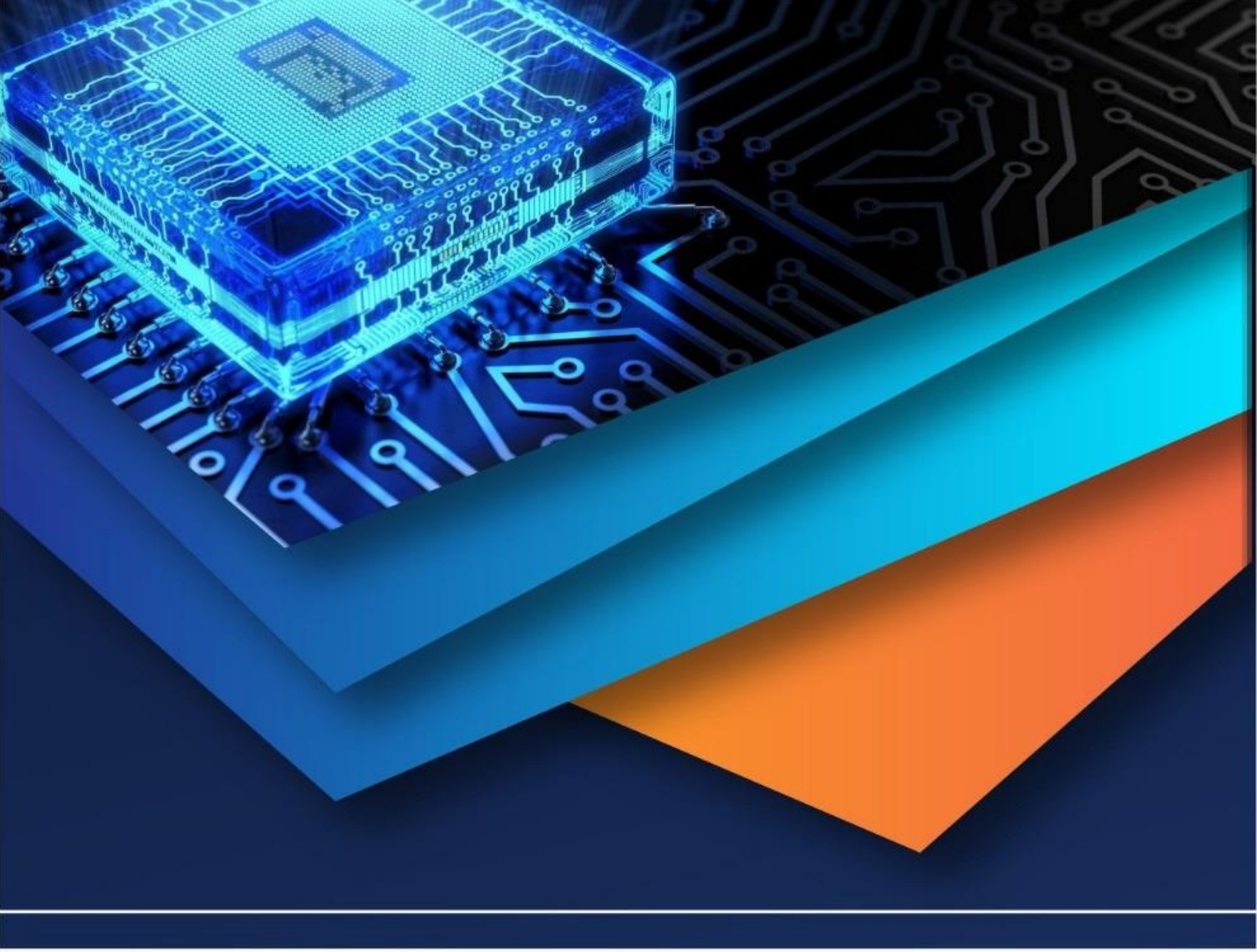
VII. CONCLUSION

In order to achieve successful outcomes in the healthcare industry, this research paper's main conclusion takes into account dataset quality, interpretability, and clinical integration when utilizing machine learning models, primarily Random Forest. When it comes to diagnosing liver diseases, Random Forest has demonstrated high accuracy. Due to poor performance with this dataset, SVM and Decision Tree may be a backup option.

Future research can explore the application of Artificial Neural Network (ANN) and Convolutional Neural Network(CNN). ANN can capture deeper nonlinear relationships within structured medical data, while CNN, used for image-based Analysis, can be CT scans. Advanced deep learning models may improve predictive performance and support decision-making in the healthcare domain.

REFERENCES

[1] Dixon, S., & Yu, X. H. (2015, August). Liver disorder detection based on artificial immune systems. In 2015 11th International Conference on Natural Computation (ICNC) (pp. 743-748). IEEE.
[2] Vats, V., Zhang, L., Chatterjee, S., Ahmed, S., Enziama, E., & Tepe, K. (2018, December). A comparative analysis of unsupervised machine techniques for liver disease prediction. In 2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT) (pp. 486-489). IEEE.
[3] Rani, R., Jaiswal, G., Nancy, Lipika, Bhushan, S., Ullah, F., ... & Diwakar, M. (2025). Enhancing liver disease diagnosis with hybrid SMOTE-ENN balanced machine learning models—an empirical analysis of Indian patient liver disease datasets. Frontiers in Medicine, 12, 1502749.
[4] Singh, J., Bagga, S., & Kaur, R. (2020). Software-based prediction of liver disease with feature selection and classification techniques. Procedia Computer Science, 167, 1970-1980.
[5] Gupta, K., Jiwani, N., Afreen, N., & Divyarani, D. (2022, April). Liver disease prediction using machine learning classification techniques. In 2022 IEEE 11th International conference on communication systems and network technologies (CSNT) (pp. 221-226). IEEE.
[6] Sharshar, E. T., Amin, H., & Badr, N. (2023). Survey of Liver Fibrosis Prediction Using Machine Learning Techniques. International Journal of Intelligent Computing and Information Sciences, 23(2), 1-12.
[7] Kalaiselvi, R., Meena, K., & Vanitha, V. (2021, October). Liver disease prediction using machine learning algorithms. In 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA) (pp. 1-6). IEEE.
[8] Durai, V., Ramesh, S., & Kalthireddy, D. (2019). Liver disease prediction using machine learning. Int. J. Adv. Res. Ideas Innov. Technol, 5(2), 1584-1588.
[9] Tokala, S., Hajarathaiah, K., Gunda, S. R. P., Botla, S., Nalluri, L., Nagamanohar, P., ... & Enduri, M. K. (2023). Liver disease prediction and classification using machine learning techniques. International Journal of Advanced Computer Science and Applications, 14(2).
[10] Kodinariya, T. M., & Gondaliya, N. (2024). Machine learning application in liver disease prediction. Journal of Electrical Systems, 20(10s), 1835-1844.
[11] Ganie, S. M., Dutta Pramanik, P. K., & Zhao, Z. (2024). Improved liver disease prediction from clinical data through an evaluation of ensemble learning approaches. BMC Medical Informatics and Decision Making, 24(1), 160.
[12] Tripathi, A., Ragiri, P. R., Jain, D., & Yadav, T. (2025). Machine Learning-based Predictive Models for Early Diagnosis of Liver Disease: MACHINE LEARNING-BASED MODELS FOR EARLY DIAGNOSIS OF LIVER DISEASE. Journal of Scientific & Industrial Research (JSIR), 84(5), 575-583.
[13] Md, A. Q., Kulkarni, S., Joshua, C. J., Vaichole, T., Mohan, S., & Iwendi, C. (2023). Enhanced preprocessing approach using ensemble machine learning algorithms for detecting liver disease. Biomedicines, 11(2), 581.
[14] Weng, S., Hu, D., Chen, J., Yang, Y., & Peng, D. (2023). Prediction of fatty liver disease in a Chinese population using machine-learning algorithms. Diagnostics, 13(6), 1168.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY