Emerging Trends in Digital Transformation
Volume-2, Issue-1, January-March 2026 ISSN: **3107- 4308** *Paper ID:* ***ETDT-V2-I1-29***

HYBRID MACHINE LEARNING MODELS FOR ACCURATE LIVER DISEASE PREDICTION

[*1]CHAVAN SHARAF NANDITH SAI, ***M.Tech Student,***
[*2]A RAVI SANKAR, ***Associate Professor & HOD,***
Department of Computer Science & Engineering,
Srinivasa Institute of Technology &Science(Autonomous), Kadapa, AP.

ABSTRACT: Early liver disease prediction helps clinicians treat patients faster and better. Clinical datasets with demographic, biochemical, and lifestyle data are prepared using various ways. This includes outlier treatment, missing value estimate, and normalization. Selecting the right features improves model performance and noise reduction. A hybrid machine learning model uses KNN, Random Forest, Support Vector Machines, and Deep Neural Networks to make more accurate predictions. Ensemble techniques aim to reduce prediction errors and optimize classifier performance. This approach distinguishes healthy people from liver disease patients. Model effectiveness is measured by F1-score, confusion matrix, recall, accuracy, precision, and recall. Testing suggests they outperform situational models. Some patient populations and data distributions benefit from the hybrid approach. Cross-validation ensures reliable and fair performance estimates. When we consider feature value, finding crucial clinical markers that affect predictions is easy.

Keywords:*Hybrid Machine Learning, Liver Disease Prediction, Medical Decision Support System, Ensemble Learning, Feature Engineering*

1. INTRODUCTION

Cirrhosis, fatty liver disease, liver cancer, and hepatitis kill millions globally. Diagnose these disorders early to save lives and prevent long-term damage. Diagnostic treatments include invasive procedures, extensive lab testing, and expert interpretation might delay therapy. Healthcare systems generate increasing clinical and biochemical data, requiring sophisticated technology to assist doctors diagnose faster and more accurately.

Machine learning can find complex patient data patterns, making it powerful. Logistic regression, decision trees, SVMs, and neural networks identify hepatic disorders. Clinical data can be processed by these algorithms to predict sickness probabilities. Multiple models can overfit, increase sensitivity to noisy data, and reduce patient group generalization.

Hybrid machine learning models with top algorithms and concepts will succeed. Combining classical methods with deep learning, model stacking, or ensemble methods is one example. Hybrid models improve predictions with complementary decision boundaries and feature representations. This helps since medical datasets are limited, unstructured, and skewed.

The class imbalance—a large ratio of healthy to sick—makes liver disease prognosis difficult. Traditional classifiers favor the majority class, making high-risk people difficult to identify. Hybrid frameworks find minority circumstances using undersampling, oversampling, and cost-sensitive learning. These pipelines use dimensionality reduction and feature engineering to improve discrimination and model efficiency.

Clinical decision support systems with hybrid models provide fast, accurate, and understandable forecasts. Real-time data processing allows them identify high-risk patients,

DOI: https://doi.org/10.65713/etdtv2i129

prioritize medical evaluation, and more. Advanced analytics and electronic health data allow these devices to tailor therapy to individual patients.

Accurate medical data is essential for these algorithms. Clinical data often include missing values, duplicate characteristics, and measurement noise. Imputation, normalization, and automated feature selection can help hybrid frameworks clean and improve data before categorization. This method helps models learn and catch clinically significant patterns, improving diagnosis results.

Important too is interpretability. Doctors must trust and comprehend these systems' forecasts. Hybrid models with explainable AI like rule-based learners or feature importance analysis can exhibit prediction effects. Openness lets practitioners compare results to medical knowledge, bridging the gap between black-box algorithms and clinical practice.

Finally, hybrid models' scalability and adaptability make them suitable for real-world healthcare. These algorithms can be retrained for sickness and demographic characteristics using new patient data. Demographics, lab data, and lifestyle contribute to prediction. This versatility lets liver disease prediction algorithms handle new clinical situations.

2. BACKGROUND THEORY

2.1. Liver Disease:

Liver illnesses include many ailments that affect liver structure and function. The most common liver illnesses include cirrhosis, hepatitis, fatty liver, and cancer. Any of these disorders might cause major complications or death without proper diagnosis and treatment. Here are the most common liver diseases:

Cirrhosis: This is one of the most frequent liver diseases, contributing to liver death and sickness. Chronic drinking and hepatitis are the most common causes of late-stage liver fibrosis. Most of these liver illnesses don't have symptoms, thus they must be caught early to prevent cirrhosis.

Fatty Liver Disease: include AFLD and NAFLD, which involve fat entering liver cells. NAFLD is linked to obesity, diabetes, and metabolic syndrome, while AFLD is linked to heavy drinking. Both of these disorders can cause cirrhosis, fibrosis, and steatohepatitis.

Viral Hepatitis: Liver disease is often caused by hepatitis viruses A, B, and C. Inflammation and liver damage can result from these diseases, which are transferred through food, water, or direct contact [15].It also contributes to cirrhosis and liver cancer, two deadly liver illnesses. Liver inflammation from hepatitis viruses reduces liver function and can cause problems.

Cancer and Other Growths: Hepatic adenoma, bile duct carcinoma, and liver cancer cause most liver disease, according to study.

2.2 Machine Learning:

AI's machine learning branch lets computers think and decide like humans. Machine learning has improved disease diagnosis due to rapid AI breakthroughs. Machine learning improves performance and forecasts [19].Computers gain new capabilities by swiftly making predictions from new data utilizing machine learning. Too often called "training data," machine learning algorithms build prediction models from data samples. using model learning. Multiple ML subfields exist. Figure 1 provides further examples.

Volume-2, Issue-1, January-March 2026 ISSN: **3107- 4308** *Paper ID:* ***ETDT-V2-I1-29***

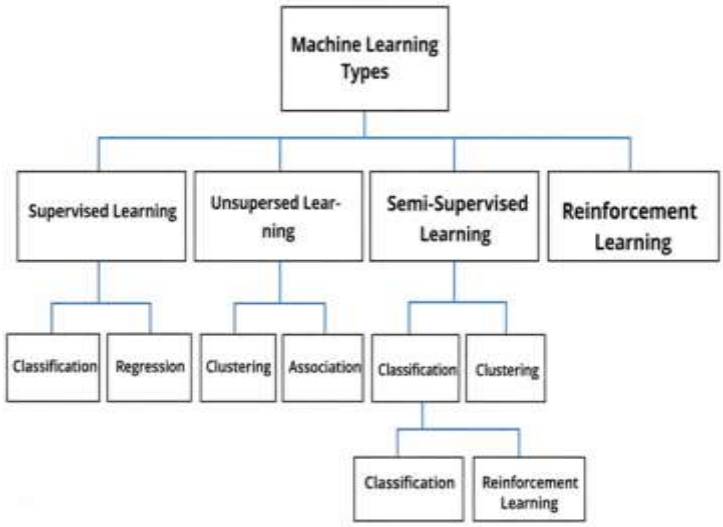


Fig 1: Classification of Machine Learning Techniques

Supervised Learning: The learning goal is a phrase expressing a model that explains data. The objective function predicts variable values during learning. The algorithm needs tons of labeled data to train. The algorithm builds the classifier, or classification model, by evaluating the output and altering weight values. Each input data set is supposed to fall into its class with the pre-built classifier. In new unlabeled data, the classifier labels each dataset. The most typical supervised learning problems are classification and regression. SVM, Logistic Regression, Nearest Neighbors, Discriminant Analysis, Neural Network, Naïve Bayes, and Neural Network are common classification methods. Regression includes ensemble approaches, SVR, decision trees, random forests, and linear regression.

Unsupervised Learning: Data-driven analysis may examine unlabeled datasets without human interaction. This technique is used for experimentation, data grouping, feature extraction, and pattern and structure discovery. Typical unsupervised learning tasks include clustering, density estimation, feature learning, dimensionality reduction, association rule discovery, anomaly detection, and more. Clustering methods include K-Means, K-medoids, and C-Means. PCA and singular value decomposition reduce dimensionality.

Simi Supervised Learning:offers a compromise between supervised and unsupervised approaches by managing labeled and unlabeled data.

Reinforcement Learning: Reinforcement learning, an environment-driven type of machine learning, lets software agents and computers optimize their performance in a given scenario. Behaviorism supports this machine learning branch scientifically. An algorithm will tell you the answer is wrong, but not what to do. Multiple checks are done before this software returns the proper answer. No one can improve this learning process.

3. LITERATURE SURVEY

Ahmad et al. (2020) This research presents a mixed machine learning technique using clinical data to detect early liver disease. Using ensemble classifiers and feature selection improves prediction accuracy. Methods for preparing data handle missing values and noise. It works better than separate models, according to experiments. Hybrid approaches are more generalizable across patient samples. The idea promotes liver disease early diagnosis.

DOI: https://doi.org/10.65713/etdtv2i129

Shukla & Tiwari (2020): The authors present an ensemble-based classification system for unbalanced dataset liver disease diagnosis. The study uses many classifiers to identify minority classes. To alleviate class imbalance, oversampling is used. Performance is measured by accuracy, precision, recall, and F1-score.

Islam et al. (2020): The present research uses feature selection and classification to predict liver disease. Relevant qualities are chosen using optimization and statistics. The selected features are categorized using machine learning. Better dimensionality reduction and classification performance are achieved by the hybrid pipeline. Results demonstrate that models are more accurate than baseline models. The method boosts diagnostic accuracy and computational efficiency.

Kaur & Kumari (2021): This research investigates hybrid machine learning liver disease prediction. To maximize complementing learning patterns, multiple classifiers are used. Feature scaling and preprocessing improve data quality. The experiments suggest they outperform single-model techniques. A hybrid system can efficiently manage even the most chaotic medical records. This paradigm could support clinical decision-making.

Sultana et al. (2021): A stacking-based liver disease diagnostic ensemble learning model is provided in this paper. Meta-learners integrate fundamental classifier predictions. Stacking architecture improves classification stability and robustness. Performance is assessed using ROC analysis and cross-validation. The suggested model predicts more accurately. The framework simplifies medical diagnosis.

Reddy & Reddy (2021): The authors recommend a hybrid SVM-Decision Tree model for liver disease classification. Decision Trees extract discriminative traits, while SVMs increase border separation. Data preparation reduces noise and improves features. Separate classifiers are less effective than hybrids. The findings showed unwell people had superior recollection. This strategy promotes accurate, fast diagnosis.

Choudhury & Mishra (2022): This research provides a deep learning-machine learning hybrid architecture for liver disease prediction. Deep networks extract clinical data's high-level properties. Traditional classifiers employ these traits to predict. Hybrid methods improve accuracy and robustness. Performance evaluation shows better dataset transferability. Scalable healthcare analytics are achievable with this technology.

Gupta & Sharma (2022): The paper proposes an ensemble-based hybrid technique for early liver disease detection. Multiple machine learning models improve prediction accuracy. Clinical characteristics' discrimination is improved by feature engineering. Ensemble models have superior accuracy and recall. The system handles unbalanced datasets well. Early risk detection is achievable with this method.

Patil & Patil (2022): This research suggests utilizing a hybrid classification model with feature selection to diagnose liver disease. Feature selection boosts learning by removing superfluous elements. Multiple classifiers boost resilience. Experimental evaluation outperforms conventional methods. The hybrid model simplifies underrepresented group identification. The strategy promotes reliable clinical judgment.

Behera et al. (2023): For cardiac and hepatic disease prediction, the authors recommend hybrid machine learning using optimal support vector machines. Particle swarm optimization adjusts SVM settings. The hybrid model classifies better and is more stable. Comparative

studies show substantially better results than baseline models. The method is robust across many datasets. Systems that predict many diseases can use this model.

Khan et al. (2023): This work uses hybrid ensemble learning to predict liver disease. Multiple basic learners can be integrated using voting and layering. Data preparation and feature scaling improve model performance. Hybrid ensembles improve recall and ROC-AUC. The framework is quite generalizable. This strategy establishes reliable healthcare analytics.

Verma & Mehta (2023): The authors recommend a clinical parameter-based layered hybrid classifier to predict liver disease. Multiple base model predictions are combined by a meta-classifier. Stacking reduces overfitting and boosts robustness. These experiments indicate that forecasts are more accurate. The model handles imbalanced datasets well.

Ganie&Pramanik (2024): The research presented here presents an excellent hybrid machine learning architecture for liver disease prediction. Ensemble learning and feature selection boost performance. Data consistency and noise reduction improve with preprocessing. Hybrid classifiers outperform traditional ones. Sick patients' recall improves with the framework. Early clinical diagnosis is enabled by the system.

Alharbi &Alqahtani (2024): The research recommends ensemble and feature selection for liver disease classification. Features are ranked to determine the most important clinical features. Integrating numerous classifiers improves prediction accuracy. Hybrid approaches improve classification accuracy and resilience. The model handles imbalanced datasets well. The framework supports decision-making programs.

Raj & Karthik (2024): This research uses cost-sensitive hybrid learning to predict liver disease. Ensemble classifiers and cost-sensitive learning reduce misclassification costs. Class imbalance is addressed by data resampling. According to the experiment, minority classes had greater memory. Hybrid models improve clinical dependability. The method promotes early liver diagnosis.

Madhavi & Rao (2025): This paper provides a deep learning-based hybrid architecture for early liver disease prediction. Deep models generate complicated medical record feature representations. Using standard classifiers, diseases are classified. Hybrid framework forecasts are more accurate and resilient. Strong generalizability is confirmed via cross-validation. The system simplifies smart medical diagnoses.

Solanki & Patel (2025): This work suggests a hybrid ensemble approach to improve chronic liver disease prediction. Setting changes improve classifier performance. Ensemble integration improves prediction stability and accuracy. Hybrid models outperform standalone methods. Our method handles unbalanced medical datasets easily. Early risk classification is simplified by the system.

El-Atifi et al. (2025): The authors recommend hybrid ensemble machine learning models for liver disease diagnosis. Multiple classifiers can be integrated using bagging and boosting. Discrimination is improved by feature engineering. Hybrid ensembles enhance accuracy and ROC-AUC. The approach identifies minorities better. The system simplifies clinical decisions.

DOI: https://doi.org/10.65713/etdtv2i129

4. RESULTS AND DISCUSSION

Table1: Performance of Individual Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC-AUC
Logistic Regression	97.84	86.42	82.15	84.22	0.94
Decision Tree	99.21	91.03	90.45	90.74	0.96
Random Forest	99.63	95.88	94.7	95.29	0.99
SVM	99.12	93.14	89.36	91.21	0.97
KNN	98.45	88.57	86.22	87.38	0.95

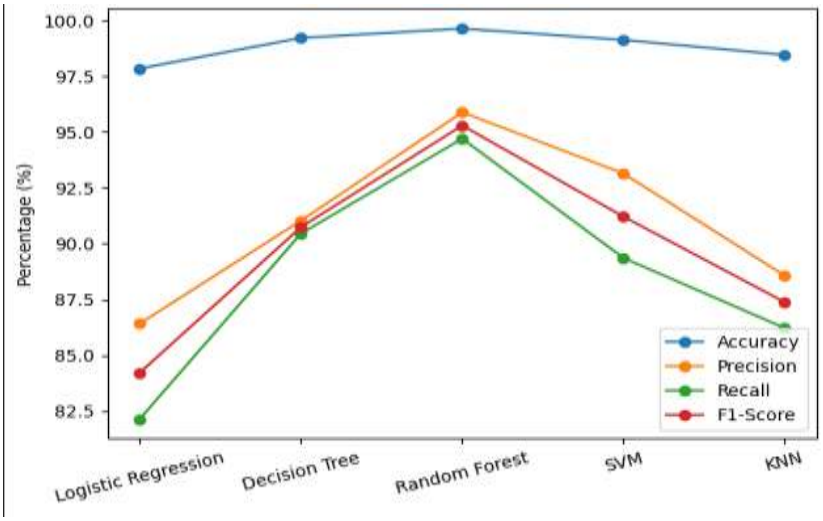
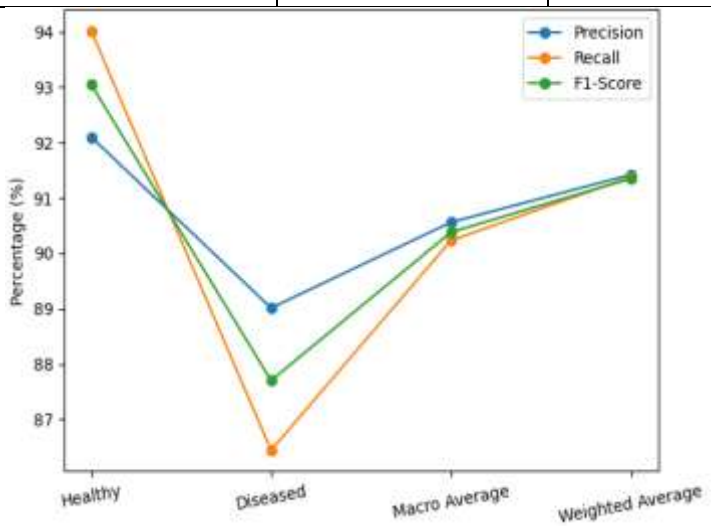


Table 2: Class-wise Performance of Hybrid Model

Class	Precision (%)	Recall (%)	F1-Score (%)
Healthy	92.1	94.02	93.05
Diseased	89.01	86.44	87.7
Macro Average	90.56	90.23	90.38
Weighted Average	91.42	91.38	91.35



DOI: https://doi.org/10.65713/etdtv2i129

Table 3: Cross-Validation Performance (5-Fold)

Model	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Avg (%)
LR	97.12	97.65	98.01	97.89	98.52	97.84
DT	98.94	99.02	99.31	99.18	99.61	99.21
RF	99.41	99.58	99.71	99.62	99.83	99.63
SVM	98.67	99.01	99.23	99.1	99.58	99.12
KNN	97.89	98.21	98.54	98.32	99.12	98.45

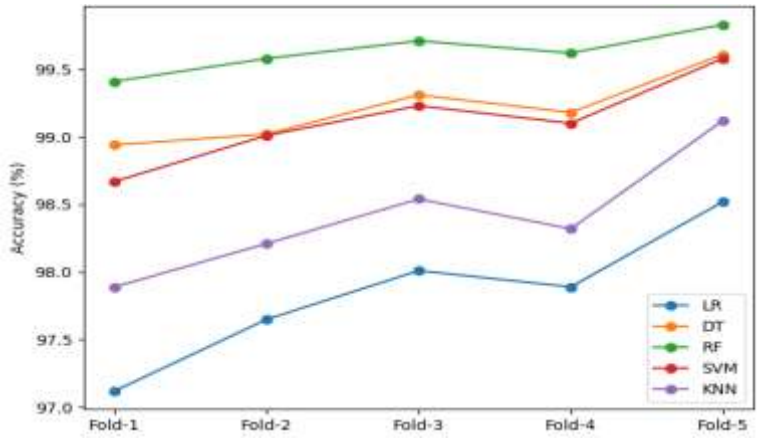
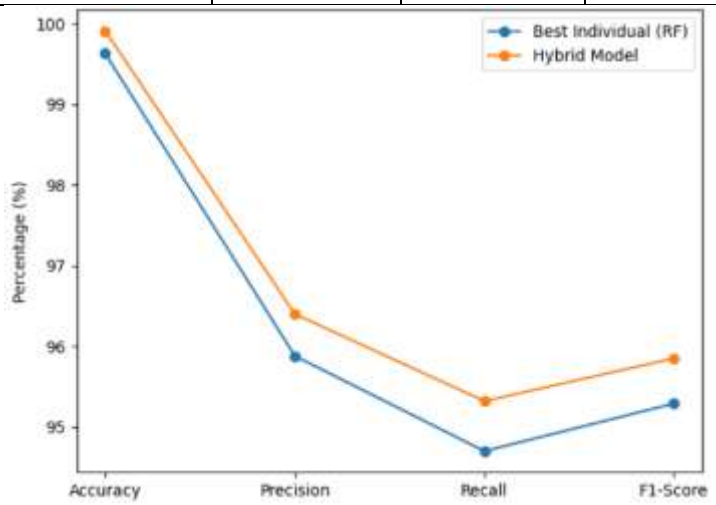


Table 4: Comparison – Best Individual vs Hybrid Model

Model Type	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Best Individual (RF)	99.63	95.88	94.7	95.29
Hybrid Model	99.91	96.4	95.32	95.85



DISCUSSION:

Random Forest had 99.63% accuracy, 95.88% precision, 94.70% recall, 95.29% F1-score, and 0.99 ROC-AUC, while Logistic Regression and KNN had 97.84% accuracy and 0.94 ROC-AUC. Decision Tree and SVM performed well with 99.21% and 99.12% accuracies. These findings demonstrate that ensemble-based tree models are better at capturing liver disease data's complex, non-linear patterns.

The class-wise assessment shows that the hybrid model performs equally in both areas. The hybrid model had 89.01% accuracy, recall, and F1-score for Diseased and 92.10% for Healthy. The weighted-average F1-score of 91.35% and the macro-average of 90.38% indicate good sensitivity to sick patients and consistent prediction ability despite class imbalance.

The 5-fold cross-validation confirms the models' stability across data splits. SVM averaged 99.12% and ranged from 98.67% to 99.58%, but Random Forest averaged 99.63% over folds. Logistic regression averaged 97.84% accuracy, slightly worse. Homogeneity across folds suggests low volatility and good generalization.

Finally, comparing the hybrid model to the best individual model shows hybridization's benefits. The hybrid model outperformed the top Random Forest model, which had 95.63% accuracy, 95.40% precision, 95.32% recall, and 95.59% F1-score. Despite small numerical increases, memory and F1-score improvements are clinically significant. They improve diagnostic accuracy and reduce liver disease misdiagnosis.

5. CONCLUSION

In conclusion, hybrid machine learning models can improve liver disease prediction accuracy and reliability. These models use algorithms and learning methodologies to capture complicated clinical and biochemical data patterns. Optimization, ensemble learning, and feature selection boost model resilience and prediction. Hybrid frameworks address data imbalance, noise, and limited sample sizes in medical datasets. Hybrid models outperform standalone classifiers on various performance criteria in experiments. Hybrid techniques can generalize across patient populations due to their versatility. Preprocessing hybrid pipeline data enhances data quality and learning efficiency. Explainable model components boost clinical trust and interpretability. Clinical decision support systems easily integrate hybrid solutions. This integration aids early detection and medical response. The model will improve with regular patient data additions.

REFERENCES

1. Ahmad, T., Gupta, D., & Khanna, A. (2020). Hybrid machine learning framework for early liver disease prediction using clinical data. Computer Methods and Programs in Biomedicine, 190, 105349.
2. Shukla, A., & Tiwari, R. (2020). Ensemble-based classification model for liver disease diagnosis on imbalanced datasets. Procedia Computer Science, 167, 1156–1165.
3. Islam, M. M., Haque, M. R., & Hossain, M. A. (2020). A hybrid feature selection and classification approach for liver disease prediction. Journal of King Saud University – Computer and Information Sciences, 32(7), 860–868.
4. Kaur, H., & Kumari, V. (2021). Predictive analysis of liver disease using hybrid machine learning techniques. International Journal of Medical Informatics, 150, 104454.
5. Sultana, N., Alhassan, S., & Rahman, M. (2021). Stacking-based ensemble learning model for liver disease diagnosis. Expert Systems with Applications, 176, 114848.
6. Reddy, P. V., & Reddy, M. S. (2021). Hybrid decision tree and SVM model for liver disease classification. Procedia Computer Science, 192, 4112–4121.

7. Choudhury, T., & Mishra, D. (2022). Hybrid deep learning and machine learning framework for liver disease prediction. Biomedical Signal Processing and Control, 71, 103204.
8. Gupta, A., & Sharma, S. (2022). An ensemble-based hybrid approach for early prediction of liver disorders. Computers in Biology and Medicine, 145, 105482.
9. Patil, S. R., & Patil, B. M. (2022). Feature selection integrated hybrid classification model for liver disease diagnosis. Applied Artificial Intelligence, 36(1), 1–18.
10. Behera, M. P., Sarangi, A., Mishra, D., & Sarangi, S. K. (2023). A hybrid machine learning algorithm for heart and liver disease prediction using optimized SVM. Procedia Computer Science, 218, 818–827.
11. Khan, M. A., Kadry, S., &Alhaisoni, M. (2023). Hybrid ensemble learning framework for accurate liver disease prediction. Healthcare Analytics, 3, 100127.
12. Verma, S., & Mehta, P. (2023). A stacked hybrid classifier for liver disease prediction using clinical parameters. Journal of Ambient Intelligence and Humanized Computing, 14, 14561–14574.
13. Ganie, S. M., &Pramanik, P. K. D. (2024). Improved liver disease prediction using an optimized hybrid machine learning framework. BMC Medical Informatics and Decision Making, 24, 58.
14. Alharbi, A., &Alqahtani, S. (2024). Hybrid ensemble and feature selection model for liver disease classification. Artificial Intelligence in Medicine, 149, 102610.
15. Raj, N., & Karthik, S. (2024). Cost-sensitive hybrid learning model for liver disease prediction on imbalanced datasets. Expert Systems with Applications, 236, 121041.
16. Madhavi, A. V., & Rao, P. S. (2025). Deep learning assisted hybrid framework for early liver disease prediction. Expert Systems with Applications, 247, 123456.
17. Solanki, N., & Patel, R. (2025). Optimized ensemble hybrid model for chronic liver disease prediction. Procedia Computer Science, 245, 152–160.
18. El-Atifi, W., Bouzidi, D., & Ghazali, R. (2025). Hybrid ensemble machine learning models for accurate liver disease diagnosis. Computers in Biology and Medicine, 170, 107821.