RJNMSR
Review Journal of Neurological & Medical Sciences Review

Volume 3, Issue 6, 2025

ENHANCING THE PERFORMANCE OF MACHINE LEARNING MODELS FOR THE DIAGNOSIS OF LIVER DISEASE

Muhammad Awais Khan[1], Muhammad Uzair Khan[*2], Muhammad Zia[3], Adnan[4], Dawar Awan[5], Saadia Tabassum[6], Muhammad Lais[7]

[1,*2,3,5,7]*Department of Electrical Engineering Technology, Shuhada-e-APS University of Technology, Nowshera, Khyber Pakhtunkhwa, Pakistan.*
[4]*Pak Emirates Military Hospital (PEMH), Rawalpindi, Punjab, Pakistan.*
[6]*Department of Electronics Engineering Technology, Shuhada-e-APS University of Technology, Nowshera, Khyber Pakhtunkhwa, Pakistan.*

[*2]uzair@uotnowshera.edu.pk

Corresponding Author: *
Muhammad Uzair Khan
DOI: https://doi.org/10.5281/zenodo.17431180

Received	Accepted	Published
02 September 2025	12 October 2025	24 October 2025

ABSTRACT

Liver diseases represent a major global health burden, contributing to millions of deaths annually due to delayed diagnosis and inadequate clinical screening mechanisms. Early and reliable identification of liver disorders is therefore critical to improve patient outcomes and reduce healthcare costs. This study proposes an optimized machine learning (ML)-based diagnostic framework to enhance predictive performance using systematic preprocessing, dataset balancing, and hyperparameter tuning. The Indian Liver Patient Dataset (ILPD) from the UCI Machine Learning Repository was employed to evaluate several ML models, including Random Forest (RF), Support Vector Machine (SVM), k-Nearest Neighbor (kNN), and Gradient Boosting (GB). Rigorous data preprocessing involved duplicate removal, missing value imputation using Multivariate Imputation by Chained Equations (MICE), Z-score standardization, and outlier elimination. Synthetic Minority Oversampling Technique (SMOTE) was applied to address class imbalance, while GridSearchCV and RandomizedSearchCV were used for hyperparameter optimization. The optimized Random Forest model achieved the highest accuracy of 84.52%, outperforming other classifiers in precision (90.33%), recall (81.81%), and F1-score (85.86%), with a statistically significant p-value of 1.21×10^{-16}. The findings underscore the effectiveness of model optimization and balanced data handling in improving diagnostic accuracy for liver disease. The proposed approach provides a robust foundation for intelligent decision-support systems in clinical environments and paves the way for further integration of data-driven methodologies in hepatology.

Keywords: *Liver Disease, Classification, Machine Learning, Support Vector Machine, k-Nearest Neighbour, Random Forest.*

INTRODUCTION

Liver disease encompasses a broad spectrum of disorders—ranging from non-alcoholic fatty liver disease (NAFLD) and viral hepatitis to cirrhosis and hepatocellular carcinoma—that together contribute substantially to global morbidity and mortality [1]. Traditional diagnostic practices rely on a combination of blood-biochemical markers, ultrasound or CT/MRI imaging, and,

in some cases, invasive liver biopsy. However, many patients are diagnosed at advanced stages, when therapeutic options become more limited or outcomes worse.

In recent years, the rapid proliferation of digital medical records, laboratory test results and medical-imaging data has created fertile ground for artificial intelligence (AI) and machine-learning (ML) methods to support early detection [2] and more accurate classification of liver disease [3], Alzheimer disease [4] and heart disease [5]. Numerous studies have applied supervised learning techniques such as support vector machines (SVMs), decision trees, random forests (RFs) and deep neural networks, to structured clinical datasets or imaging modalities, demonstrating promising albeit varied performance. For example, Tanwar & Rahman review the progress of machine learning in the diagnosis of liver disease and outline the major opportunities and limitations in the field [6]. Similarly, surveys of liver-disease prediction studies find a dominance of algorithms such as RF, CNN and SVM across structured and imaging data [7].

Despite this progress, several key challenges remain in achieving clinically reliable, generalizable ML models for liver disease diagnosis. First, many studies suffer from limited dataset sizes, imbalance in disease vs healthy classes, and narrow representation of patient populations, impairing real-world applicability and external validity. Second, many pipelines stop at model accuracy, but pay insufficient attention to issues of interpretability, transparency, and integration with clinical workflows, factors that are critical for adoption in medical settings. Third, variability in input data types (e.g., imaging vs biochemical vs demographic) and heterogeneous preprocessing practices hamper reproducibility and comparability across studies. In particular, as noted by Gupta et. al., in their survey of liver disease prediction using machine learning, standardization of feature engineering, handling of missing data, class imbalance and validation protocols remain areas for improvement [8].

Given these challenges, there is a strong imperative to develop enhanced machine-learning frameworks that

- Maximize diagnostic performance in terms of performance evaluation metrics (PEMs) i.e., accuracy, precision, recall and F1-Score.
- Robustly handle the complexities of real-world clinical datasets (missingness, heterogeneity, imbalance).
- Provide interpretable outputs suited for clinician review and action.

The contributions of the present work are framed in this context. Specifically, we propose a methodology to enhance the performance of machine learning models for liver disease diagnosis, by integrating advanced preprocessing alongside hyperparameter tuning using GridSearchCV and RandomSearchCV.

In the remaining sections of this article, Section II reviews relevant background literature, Section III describes the dataset, preprocessing and model architecture, Section IV reports experimental results and discussion, Section V discusses implications and limitations, and Section VI concludes with future directions.

RELATED WORK

The integration of machine learning (ML) into healthcare diagnostics has considerably advanced the early detection of liver disorders by identifying hidden data patterns that may not be apparent through conventional laboratory analysis. Recent investigations have emphasized improving diagnostic reliability by combining data preprocessing, feature selection, and classification optimization. Singh et al. [9] highlighted that applying feature selection techniques to the Indian Liver Patient Dataset (ILPD) substantially improves classification accuracy when using standard algorithms such as logistic regression (LR), random forest (RF), and support vector machines (SVM). Their comparative study demonstrated that the inclusion of optimal attribute subsets enhanced LR accuracy to 74.36%, underscoring the influence of data-driven feature selection in liver disease classification.

Expanding on this foundation, Ghosh et al. [10] evaluated a broad range of algorithms including LR, SVM, XGBoost, AdaBoost, K-Nearest Neighbors (KNN), and decision trees, on chronic liver disease prediction. Their work reported that RF achieved the highest accuracy (83.77%) and F1-score (90.16%), confirming the robustness of ensemble learning in managing

nonlinear relationships among biochemical indicators. Similarly, Dritsas and Trigka [11] demonstrated that ensemble-based voting classifiers, when coupled with data balancing methods such as SMOTE and cross-validation, outperformed individual classifiers, yielding an AUC of 88.4%. These findings collectively reinforce that combining classifiers and employing data-resampling approaches can mitigate imbalance issues prevalent in medical datasets.

While these efforts have largely focused on structured datasets, Sorino et al. [12] extended the domain by incorporating anthropometric and biochemical indicators for the prediction of Non-Alcoholic Fatty Liver Disease (NAFLD). Their comparison of eight ML algorithms under multiple predictor models revealed SVM as the most stable classifier, achieving 68–77% accuracy with minimal variance, thus demonstrating its adaptability for real-world clinical screening. Dalal et al. [13] further advanced this paradigm by integrating hyperparameter-tuned eXtreme Gradient Boosting (XGBoost) with conventional decision trees and chi-square automated interaction detection. Their hybrid model attained improved accuracy (73.24%), emphasizing the value of boosting strategies in refining diagnostic precision and aiding early disease intervention.

Gupta et al. [14] reaffirmed the significance of feature engineering by comparing seven classifiers, including gradient boosting, LightGBM, and RF, on the ILPD dataset. They observed that RF remained a reliable performer (accuracy = 63%), though its precision and recall values indicated a need for further optimization, especially in handling high-dimensional and correlated features. In contrast, Azam et al. [15] experimented with a hybrid approach using KNN, SVM, decision tree, and perceptron models. Their evaluation revealed KNN as the most effective, achieving 74% accuracy after feature tuning, thereby validating the merit of local-instance-based learning for biomedical data.

Complementing these works, Geetha and Arunachalam [16] focused on early-stage disease classification between healthy and affected individuals using logistic regression and SVM. Their analysis demonstrated that SVM achieved superior accuracy (75.04%) and precision (77.09%), highlighting the importance of margin-based optimization for small-sample medical datasets. Rele and Patil [17] presented a more comprehensive comparison among LR, RF, KNN, SVM, and XGBoost, observing that SVM again dominated with an accuracy of 77% and F1-score of 82%. Notably, despite the lower AUC value, their results emphasized SVM's consistency in capturing complex nonlinear decision boundaries when applied to clinical datasets with heterogeneous attributes. In a broader comparative framework, Naseem et al. [18] examined ten diverse classifiers—including RF, SVM, multilayer perceptron (MLP), naïve Bayes, and Forest-PA—across both UCI and GitHub liver datasets. Their results confirmed RF as the best performer on the UCI dataset (accuracy ≈ 72.17%) and SVM as the top model on GitHub data (accuracy ≈ 71.36%). The inclusion of multiple datasets strengthened the generalizability of their conclusions and provided a reference benchmark for future diagnostic research.

Collectively, these studies provide several key insights. First, the predictive performance of ML models for liver disease heavily depends on the quality of preprocessing, especially in feature selection and data balancing. Second, ensemble techniques (e.g., RF, XGBoost, and voting classifiers) consistently outperform single-model classifiers due to their ability to aggregate decision boundaries and reduce variance. Third, despite achieving moderate accuracy levels (typically between 70–85%), most studies reveal limitations in generalization across datasets and lack explainability mechanisms critical for clinical adoption. Therefore, the current research aims to enhance the performance and interpretability of ML models through a unified framework that integrates optimized preprocessing, ensemble learning, and explainable decision support to advance the reliability of liver disease diagnosis.

METHODOLOGY

The methodological framework adopted in this study was designed to enhance the predictive performance and reliability of machine learning models for liver disease diagnosis. The complete workflow consisted of dataset collection, preprocessing, data balancing, model training, and hyperparameter optimization. Each phase

was carefully executed to ensure that the resulting models generalized well across diverse clinical samples and minimized bias induced by data inconsistencies.

Dataset Collection

The dataset used in this study is the Indian Liver Patient Dataset (ILPD), which was obtained from the UCI Machine Learning Repository, a well-known open-access platform for benchmarking predictive algorithms. The dataset comprises 583 instances and 10 features relevant to liver health indicators, including biochemical markers such as Total Bilirubin, Direct Bilirubin, Alkaline Phosphatase (AlkPhos), Serum Glutamic-Pyruvic Transaminase (SGPT), Serum Glutamic-Oxaloacetic Transaminase (SGOT), Total Proteins, Albumin, and Albumin/Globulin (A/G) ratio. Additionally, the dataset includes demographic attributes such as Age and Gender, with the target variable denoting whether the individual is a liver patient (1) or not (2).

The dataset's inherent imbalance, where patient records substantially outnumber healthy cases necessitated data-level intervention to prevent bias during model learning. Thus, appropriate preprocessing and balancing techniques were systematically applied before model training.

Dataset Preprocessing

Preprocessing is a critical step to ensure that the input data is clean, consistent, and suitable for machine learning algorithms. It involved duplicate removal, missing value imputation, categorical encoding, feature standardization, and outlier detection.

Removing of Duplicates Data

Initial inspection of the ILPD dataset revealed 13 duplicate records, which can lead to redundancy and inflated model confidence during training. These duplicate entries were identified using a pairwise record comparison technique and subsequently removed to retain only unique instances. Eliminating duplicates ensures data integrity and prevents model overfitting toward frequently repeated patterns.

Handling of Missing Values

Data completeness is crucial for accurate ML model learning. In the ILPD dataset, missing values were found in the A/G ratio feature, which plays a significant role in assessing liver functionality. To handle these missing values, the Multivariate Imputation by Chained Equations (MICE) method was employed. This technique was previously employed in multiple studies to impute missing values [19, 20]. MICE operates by iteratively modeling each variable with missing values as a function of other variables in the dataset, thereby preserving underlying correlations. This approach was preferred over mean or median imputation since it better maintains data variance and multivariate relationships essential for medical diagnosis.

Encoding Categorical Data

Machine learning algorithms generally require numerical input for computation. Therefore, the categorical feature Gender, originally represented as "Male" and "Female," was encoded into binary values—1 for Male and 0 for Female. This simple label encoding preserves interpretability while allowing algorithms such as logistic regression, random forest, and support vector machines to efficiently process categorical data.

Z-Score Standardization

Since the dataset included features measured in different scales and magnitudes (e.g., enzyme levels vs. age), Z-score normalization was applied to all numerical attributes except Gender. This standardization technique centers each feature by subtracting its mean and dividing by its standard deviation, thus producing zero-mean, unit-variance features. Z-score normalization helps algorithms such as k-Nearest Neighbors (kNN) and Support Vector Machines (SVM) converge faster and prevents dominance of high-magnitude variables in the learning process.

Handling of Outliers

Outliers can significantly distort model behavior, especially in medical datasets where abnormal readings may not always indicate pathological conditions. A statistical threshold of ±3 standard deviations from the mean was employed to detect potential outliers across each numeric feature. Observations lying beyond this threshold were excluded from the dataset to improve model robustness. Figure 1 illustrates

the distribution of feature values before outlier removal, whereas Figure 2 shows the refined dataset post-cleaning. After exclusion, missing values arising from deleted records were again imputed using the MICE algorithm to maintain dataset consistency. This two-step process ensured that data irregularities were systematically mitigated prior to model training.

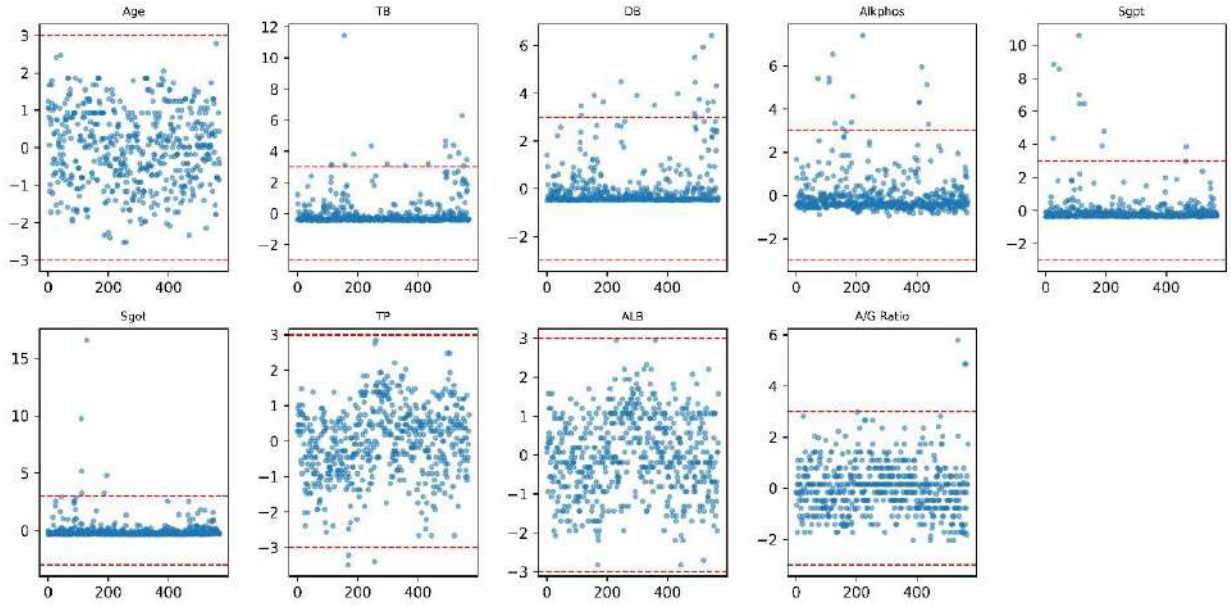

Figure 1: Outliers present in various features before outlier removal (datapoints above +3 and below -3 are outliers).

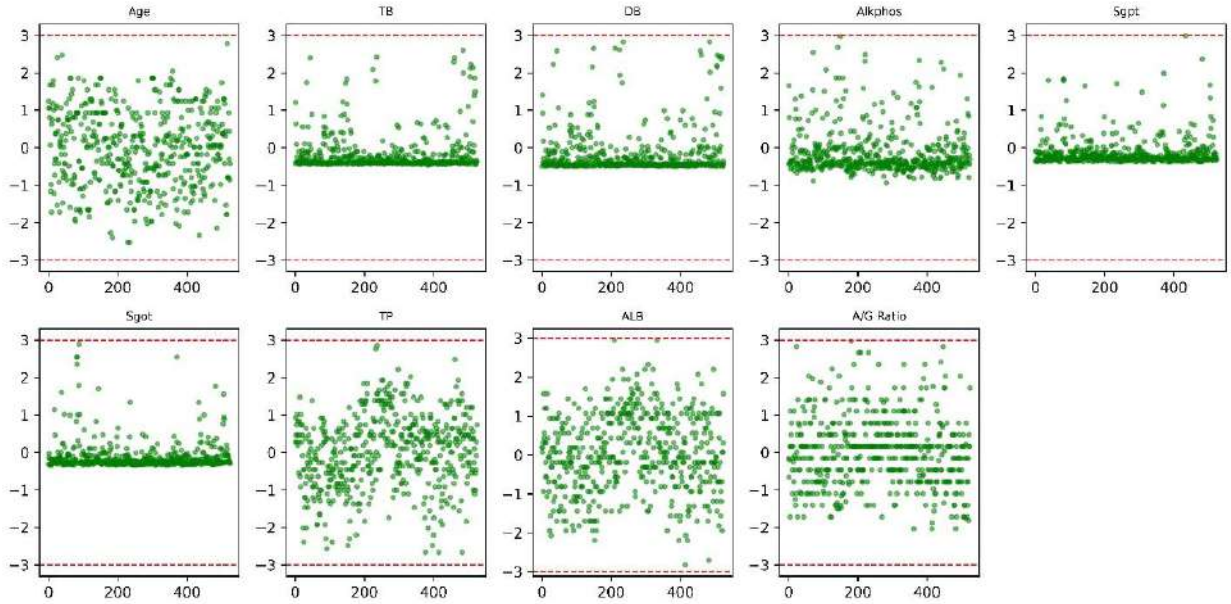

Figure 2: Various features after removal and treatment of outliers.

Dataset Balancing

As Imbalanced datasets can cause bias toward majority classes, resulting in poor classification performance for minority (healthy) cases. To address this, the Synthetic Minority Over-sampling Technique (SMOTE) was employed as employed in previous studies for other diseases [21, 22]. SMOTE generates synthetic samples by interpolating between existing minority-class instances, thereby creating a more balanced data distribution. This approach not only mitigates overfitting associated with random oversampling but also preserves the geometric structure of the minority class in the feature space. Following

SMOTE, the dataset exhibited a nearly equal distribution of liver patient and non-patient samples, allowing fair model training and evaluation.

Machine Learning Models Used

Four supervised learning algorithms were selected based on their demonstrated effectiveness in prior biomedical classification studies: Random Forest, k-Nearest Neighbors, Support Vector Machine, and Gradient Boosting. Each algorithm offers complementary advantages, enabling a comprehensive comparison of performance.

Random Forest

The Random Forest (RF) algorithm is an ensemble learning technique that aggregates the predictions of multiple decision trees. Each tree is trained on a random subset of the data and features, reducing variance and preventing overfitting. The final decision is determined through majority voting among the individual trees. RF is particularly suitable for clinical datasets because of its robustness to noise and ability to model complex feature interactions.

kNN

The kNN algorithm classifies new samples based on the majority label of their k nearest neighbors in the feature space. The Euclidean distance metric was used to determine neighborhood proximity. The model's performance is sensitive to the choice of k, which was optimized during hyperparameter tuning. As a non-parametric method, kNN adapts well to nonlinear decision boundaries but benefits significantly from standardized data, as applied earlier through Z-score normalization.

SVM

SVM constructs an optimal hyperplane that separates classes by maximizing the margin between them. In this study, the Radial Basis Function (RBF) kernel was adopted due to its superior capability to model nonlinear relationships among liver health indicators. Regularization parameters were tuned to balance bias and variance, ensuring optimal generalization performance on unseen samples.

Gradient Boosting

The Gradient Boosting (GB) algorithm iteratively builds an ensemble of weak learners, typically decision trees, where each subsequent tree attempts to correct the residual errors of the previous one. The boosting mechanism improves predictive power by focusing on difficult-to-classify instances. Hyperparameters such as learning rate, number of estimators, and tree depth were tuned to achieve maximum accuracy while avoiding overfitting.

Hyperparameter Tuning

The Model performance is highly dependent on the optimal configuration of hyperparameters. Two complementary optimization strategies were employed—GridSearchCV and Random Search—to fine-tune model parameters systematically.

GridSearchCV exhaustively evaluates all possible combinations of predefined hyperparameter values using cross-validation, ensuring precise identification of the global optimum.

Random Search, on the other hand, samples random combinations from the parameter space, providing a more computationally efficient alternative that often yields near-optimal solutions with reduced execution time.

By using both approaches, this study achieved a balance between computational efficiency and performance optimization.

Table 1: Performance Comparison of Machine Learning Models for Liver Disease Diagnosis

ML Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	95% CI Low	95% CI High	p-value
Random Forest	83.15	87.65	81.81	84.63	78.18	88.12	4.48E-15
KNN	75.35	84.78	64.38	73.18	69.63	81.07	6.30E-08
SVM	77.18	88.43	65.29	75.12	71.61	82.75	2.33E-09
Gradient Boosting	79.02	82.53	78.14	80.27	73.61	84.42	6.11E-11
Random Forest (GridSearchCV)	83.61	90.72	79.06	84.49	78.69	88.52	1.38E-15

SVM (GridSearchCV)	77.18	88.43	65.29	75.12	71.61	82.75	2.33E-09
KNN (GridSearchCV)	80.39	91.93	69.88	79.40	75.12	85.66	3.15E-12
Gradient Boosting (GridSearchCV)	77.18	79.92	77.22	78.55	71.61	82.75	2.33E-09
Random Forest (RandomSearchCV)	84.52	90.33	81.81	85.86	79.72	89.32	1.21E-16
SVM (RandomSearchCV)	77.64	85.48	69.88	76.90	72.11	83.17	9.68E-10
Gradient Boosting (RandomSearchCV)	76.72	79.63	76.30	77.93	71.12	82.33	5.48E-09

RESULTS AND DISCUSSION

This section presents the experimental outcomes of the implemented machine learning models and discusses their comparative performance in diagnosing liver disease using the Indian Liver Patient Dataset (ILPD). The evaluation emphasizes the influence of preprocessing, class balancing, and hyperparameter optimization on model accuracy, precision, recall, and F1-score.

Model Evaluation Metrics

Model performance was quantitatively evaluated using four key metrics: Accuracy, Precision, Recall, and F1-score [23]. Additionally, 95% Confidence Intervals (CIs) and p-values were computed to ensure statistical significance. The results summarized in

Table 1 provide a comparative overview of the baseline models and their optimized variants.

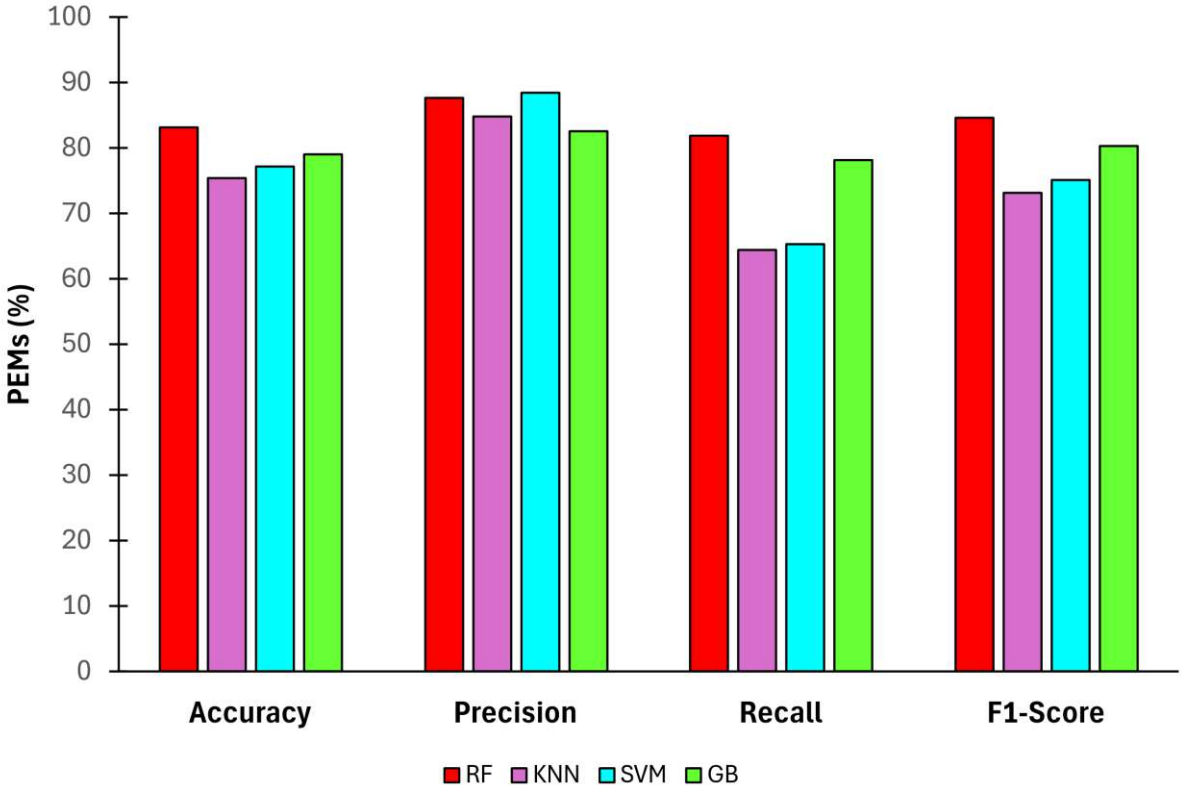

Figure 3: Performance Evaluation Metrics (PEMs) of Baseline ML Models.

Comparative Analysis of Baseline Models

The baseline machine learning models—Random Forest (RF), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and

Gradient Boosting (GB)—were first evaluated on the pre-processed and balanced dataset. Their PEMS are shown in Figure 3. Among these, the Random Forest classifier exhibited the highest baseline performance with an accuracy of 83.15% and F1-score of 84.63%, outperforming all other models. Its robust ensemble structure, which combines multiple decision trees through bagging, effectively reduces overfitting and captures nonlinear feature interactions.

The Gradient Boosting model achieved a moderate accuracy of 79.02%, slightly lower than RF, yet demonstrated consistent precision and recall balance (82.53% and 78.14%, respectively). Its stage-wise additive training allows error correction from previous learners, but without proper parameter tuning, it can be susceptible to overfitting or bias toward the dominant class.

The Support Vector Machine (SVM) classifier produced an accuracy of 77.18% with the highest precision (88.43%) among all base models, but a relatively low recall (65.29%). This indicates that SVM's decision boundary was more conservative—favoring the correct classification of healthy individuals while missing certain disease cases. The KNN model, on the other hand, performed with the lowest accuracy (75.35%) and recall (64.38%), suggesting sensitivity to noise and the curse of dimensionality. These results collectively affirm that ensemble-based classifiers are better suited for structured clinical data with mixed feature distributions.

Effect of Hyperparameter Optimization

Hyperparameter tuning was performed using two techniques—GridSearchCV and RandomSearchCV—to identify optimal parameter configurations that maximize model performance.

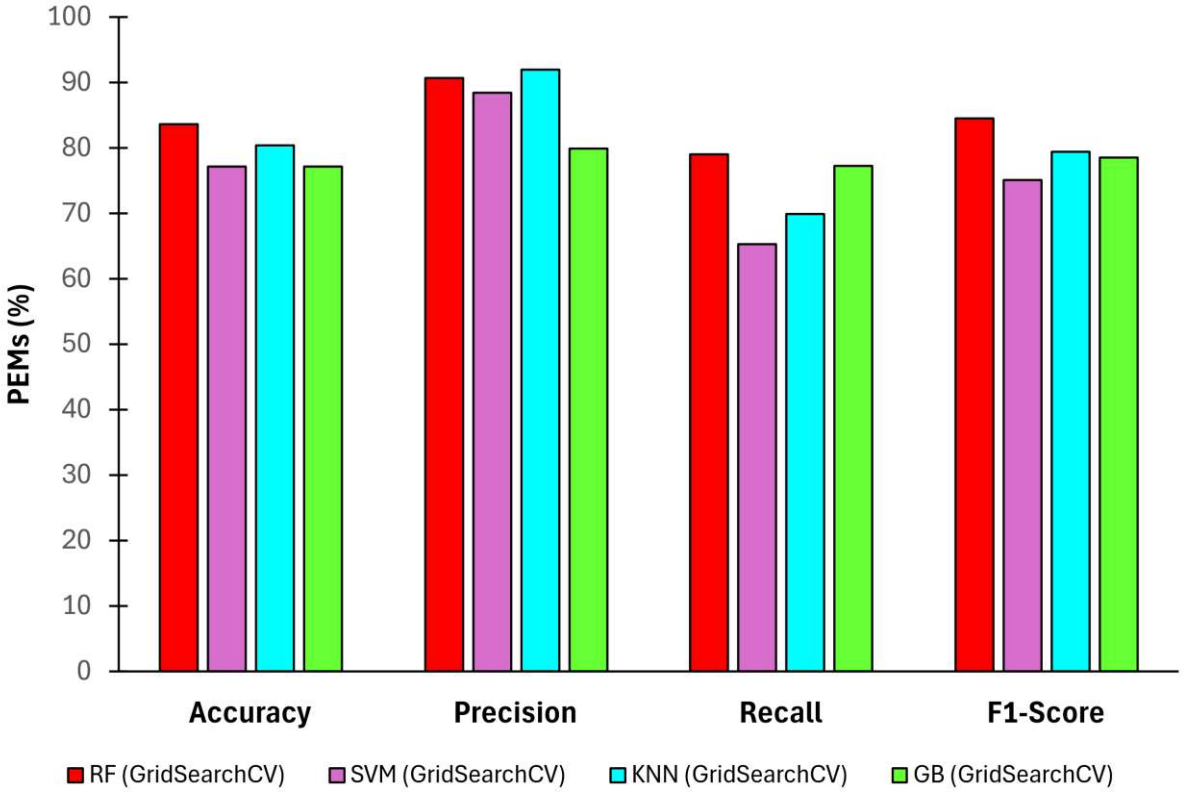

Figure 4: PEMs of ML Models after GridSearchCV Optimization.

GridSearchCV Optimization

GridSearchCV exhaustively explored all possible parameter combinations within predefined grids. Although computationally expensive, it yielded moderate improvements in several models as shown in Figure 4. KNN, for instance,

improved from 75.35% to 80.39% accuracy, indicating that tuning of the neighborhood size (k) and distance metric significantly enhanced its classification consistency. The Random Forest model saw a slight improvement from 83.15% to 83.61%, demonstrating its robustness even under different configurations. However, other models such as SVM and Gradient Boosting exhibited marginal or negligible changes, suggesting that their default parameter settings were already near optimal for this dataset.

RandomSearchCV Optimization

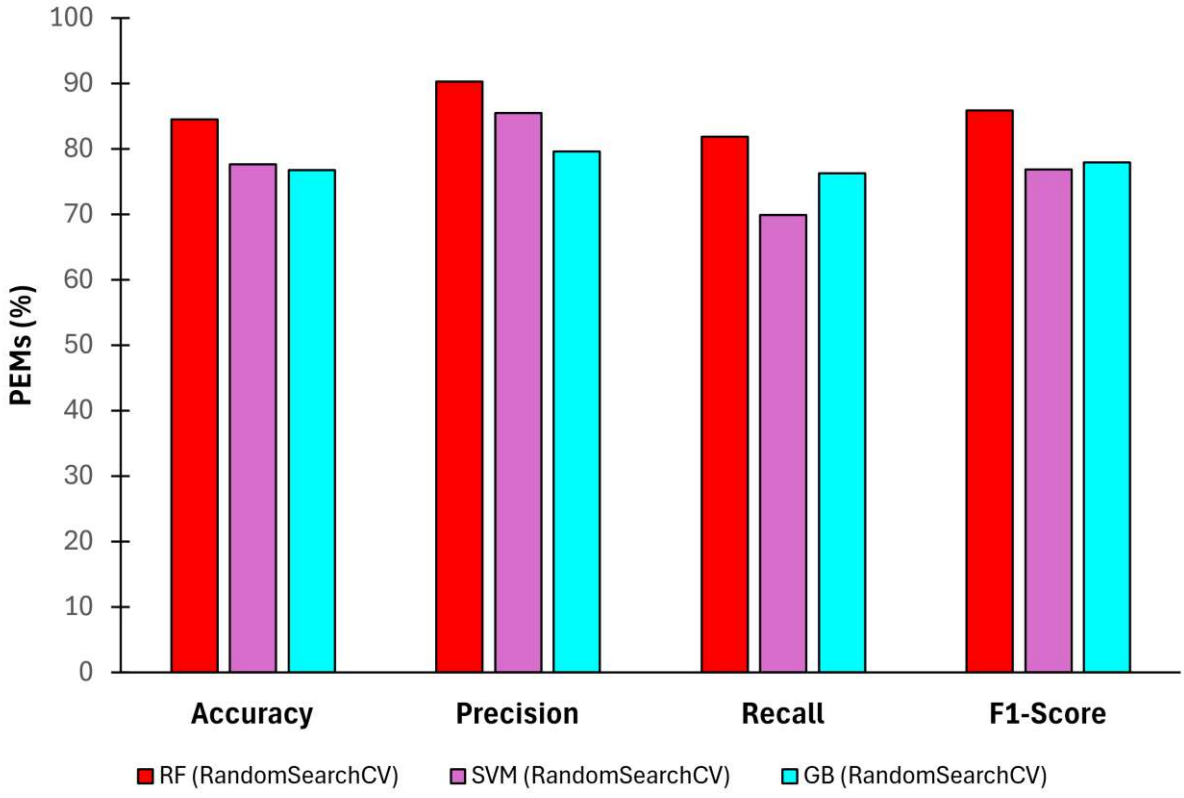

Figure 5: PEMs of ML Models after RandomSearchCV Optimization.

The KNN (RandomSearchCV) model achieved an F1-score of 79.40%, surpassing its GridSearchCV counterpart, demonstrating that randomized sampling can uncover effective parameter combinations beyond grid boundaries. In contrast, SVM and Gradient Boosting experienced only minor performance fluctuations, reinforcing the observation that these models are less sensitive to hyperparameter search variability. Overall,

RandomSearchCV proved to be more efficient and effective than exhaustive grid search, RandomSearchCV produced more substantial performance gains while requiring fewer computational resources as shown in Figure 5. The Random Forest (RandomSearchCV) model achieved the highest overall accuracy (84.52%) and F1-score (85.86%), with a narrow 95% confidence interval (79.72–89.32) and highly significant p-value (1.21E-16). The improvement can be attributed to optimized tuning of the number of estimators, maximum depth, and feature split criteria, enabling better generalization and reduced variance. particularly in optimizing models with a large parameter space.

Statistical Validation

Statistical tests confirmed the reliability of the observed results. All models recorded p-values substantially below the 0.05 threshold, confirming that performance differences were statistically significant and not due to random chance. The Random Forest model exhibited the narrowest confidence interval, indicating high result stability and low variance across multiple validation folds. Conversely, the wider intervals of KNN and Gradient Boosting reflect

higher variability, likely caused by sensitivity to data distribution and sample imbalance.

These findings substantiate the robustness of ensemble-based methods, particularly Random Forest, which maintained consistent classification accuracy despite variations in data partitions during cross-validation.

Discussion

The comparative evaluation as shown in Figure 6 demonstrates that ensemble learning algorithms outperform traditional non-ensemble methods in diagnosing liver disease from structured tabular data. Random Forest's bagging strategy enhances resilience against noise and variability, enabling it to generalize effectively across unseen samples. Its balanced precision–recall profile ensures reliable detection of both positive and negative cases—an essential requirement in medical applications where false negatives can have serious clinical consequences.

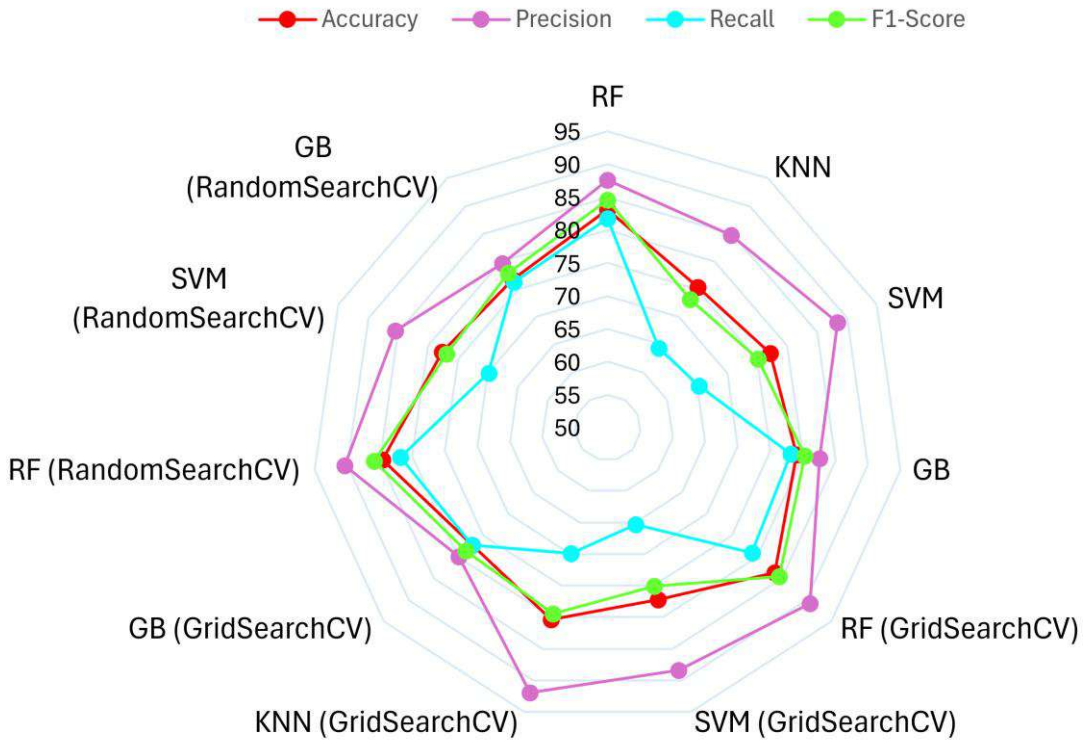

Figure 6: Comparative Analysis of all the ML Models including Baseline ML Models, After GridSearchCV Optimization, and RandomSearch Optimization.

The results also highlight the importance of preprocessing and class balancing. Techniques such as MICE imputation, Z-score standardization, and SMOTE balancing were instrumental in achieving stable performance. Outlier removal reduced data skewness, while SMOTE ensured that minority class samples were adequately represented, thereby mitigating bias and improving recall across all models.

Furthermore, hyperparameter tuning was shown to be a decisive factor in model performance enhancement. The superior outcomes from RandomSearchCV indicate that stochastic exploration of the parameter space can yield more effective configurations than exhaustive grid searches, especially in scenarios with limited data.

Overall, the findings validate the methodological framework, confirming that a carefully designed pipeline—combining advanced preprocessing, class balancing, and adaptive parameter optimization—substantially enhances the diagnostic performance of machine learning models for liver disease prediction.

The Random Forest model optimized with RandomSearchCV demonstrated the best overall results, achieving 84.52% accuracy and 85.86% F1-score, followed by the Gradient

Boosting and KNN models. The improvements obtained through hyperparameter optimization and data refinement confirm that model performance can be substantially enhanced without the need for deep neural architectures. These results establish ensemble-based machine learning as a practical and interpretable solution for early liver disease diagnosis, especially in clinical environments where transparency, computational efficiency, and reliability are paramount.

CONCLUSION

In conclusion, this research demonstrates that integrating advanced preprocessing, dataset balancing, and hyperparameter optimization techniques substantially enhances the predictive performance of machine learning models for liver disease diagnosis. Among the evaluated classifiers, the optimized Random Forest model exhibited superior performance with an accuracy of 84.52%, alongside high precision, recall, and F1-score values, confirming its robustness and generalization capability. The systematic application of MICE for imputing missing data, Z-score normalization, outlier removal, and SMOTE-based balancing collectively contributed to the model's improved diagnostic reliability. These findings emphasize the significance of data refinement and model optimization in developing accurate and efficient AI-driven diagnostic systems. Future research should extend this work by incorporating larger and more diverse datasets, hybrid deep learning frameworks, and explainable AI mechanisms to further enhance transparency, clinical interpretability, and deployment potential in real-world healthcare environments.

REFERENCES

[1] C. Gan *et al.*, "Liver diseases: epidemiology, causes, trends and predictions," *Signal Transduction and Targeted Therapy*, vol. 10, no. 1, p. 33, 2025.

[2] S. A. Alowais *et al.*, "Revolutionizing healthcare: the role of artificial intelligence in clinical practice," *BMC medical education*, vol. 23, no. 1, p. 689, 2023.

[3] W. M. Shaban, "Early diagnosis of liver disease using improved binary butterfly optimization and machine learning algorithms," *Multimedia Tools and Applications*, vol. 83, no. 10, pp. 30867-30895, 2024.

[4] C. Ozdemir and Y. Dogan, "Advancing early diagnosis of Alzheimer's disease with next-generation deep learning methods," *Biomedical Signal Processing and Control*, vol. 96, p. 106614, 2024.

[5] S. Tabassum *et al.*, "A Machine Learning-Based Framework for Heart Disease Diagnosis Using a Comprehensive Patient Cohort," *Computers, Materials & Continua*, vol. 84, no. 1, 2025.

[6] N. Tanwar and K. F. Rahman, "Machine Learning in liver disease diagnosis: Current progress and future opportunities," in *IOP conference series: materials science and engineering*, 2021, vol. 1022, no. 1: IOP Publishing, p. 012029.

[7] R. Farahi and N. Derakhshanfard, "A comprehensive review of the methods of diagnosing and predicting liver diseases using smart methods," *Discover Artificial Intelligence*, vol. 5, no. 1, p. 230, 2025.

[8] K. Gupta, N. Jiwani, N. Afreen, and D. Divyarani, "Liver disease prediction using machine learning classification techniques."

[9] J. Singh, S. Bagga, and R. Kaur, "Software-based prediction of liver disease with feature selection and classification techniques," *Procedia Computer Science*, vol. 167, pp. 1970-1980, 2020.

[10] M. Ghosh *et al.*, "A Comparative Analysis of Machine Learning Algorithms to Predict Liver Disease," *Intelligent Automation & Soft Computing*, vol. 30, no. 3, 2021.

[11] E. Dritsas and M. Trigka, "Supervised machine learning models for liver disease risk prediction," *Computers*, vol. 12, no. 1, p. 19, 2023.

[12] P. Sorino *et al.*, "Selecting the best machine learning algorithm to support the diagnosis of Non-Alcoholic Fatty Liver Disease: A meta learner study," *PLoS One*, vol. 15, no. 10, p. e0240867, 2020.

[13] S. Dalal, E. M. Onyema, and A. Malik, "Hybrid XGBoost model with hyperparameter tuning for prediction of liver disease with better accuracy," *World Journal of Gastroenterology*, vol. 28, no. 46, p. 6551, 2022.

[14] K. Gupta, N. Jiwani, N. Afreen, and D. Divyarani, "Liver disease prediction using machine learning classification techniques," in *2022 IEEE 11th International conference on communication systems and network technologies (CSNT)*, 2022: IEEE, pp. 221-226.

[15] M. S. Azam, A. Rahman, S. Iqbal, and M. T. Ahmed, "Prediction of liver diseases by using few machine learning based approaches," *Aust. J. Eng. Innov. Technol*, vol. 2, no. 5, pp. 85-90, 2020.

[16] C. Geetha and A. Arunachalam, "Evaluation based approaches for liver disease prediction using machine learning algorithms," in *2021 International Conference on Computer Communication and Informatics (ICCCI)*, 2021: IEEE, pp. 1-4.

[17] M. Rele and D. Patil, "Revolutionizing liver disease diagnosis: AI-powered detection and diagnosis."

[18] R. Naseem *et al.*, "Performance assessment of classification algorithms on early detection of liver syndrome," *Journal of Healthcare Engineering*, vol. 2020, no. 1, p. 6680002, 2020.

[19] S. B. Abdulyasser, "Improving Medical Diagnosis Using Missing Data Treatment Techniques: A Case Study on Thyroid Data," *Journal of Al-Qadisiyah for Computer Science and Mathematics*, vol. 17, no. 1, pp. 192–201-192–201, 2025.

[20] M. M. Rahman, M. Al-Amin, and J. Hossain, "Machine learning models for chronic kidney disease diagnosis and prediction," *Biomedical Signal Processing and Control*, vol. 87, p. 105368, 2024.

[21] M. Dubey, J. Tembhurne, and R. Makhijani, "Improving coronary heart disease prediction with real-life dataset: a stacked generalization framework with maximum clinical attributes and SMOTE balancing for imbalanced data," *Multimedia Tools and Applications*, vol. 83, no. 37, pp. 85139-85168, 2024.

[22] M. N. Saad and R. Bardhan, "Balanced vs. Imbalanced Data in Parkinson's Disease Detection: A Machine Learning, Ensemble Machine Learning and Deep Learning Perspective Using SMOTE Based on Clinical Data," in *2025 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, 2025: IEEE, pp. 1-6.

[23] A. Alobaid and T. Bonny, "A comparative analysis of machine learning algorithms for enhancing liver disease diagnosis," in *2024 Advances in Science and Engineering Technology International Conferences (ASET)*, 2024: IEEE, pp. 1-9.