



Explainable Ensemble Machine Learning for Liver Cirrhosis Mortality Prediction Using Clinical Data

Anshita Prabhakar¹, Aashish Kumar Tiwari²

^{1,2}Department of Computer Science and Technology, SAM Global University, Bhopal, India

¹anshitaprabhakar@gmail.com, ²aashish.tiwari7898@gmail.com

ABSTRACT

Liver cirrhosis is a progressive and life-threatening chronic liver condition associated with high morbidity and mortality worldwide. Early and accurate risk assessment using non-invasive clinical and laboratory parameters can assist healthcare professionals in timely intervention and personalized patient management. Existing machine learning-based approaches have demonstrated the potential for liver disease classification; however, limitations remain in terms of prediction reliability, clinical relevance, and model interpretability. The existing study developed an ensemble machine learning framework for liver disease stage prediction using clinical features and reported an accuracy of 83% using a hyperparameter-tuned Logistic Regression model. However, multi-class histological stage prediction remains challenging due to overlapping clinical characteristics among different disease stages. In this work, a clinically meaningful binary mortality prediction framework is developed using the Mayo Clinic Primary Biliary Cirrhosis (PBC) dataset to classify patient outcomes as Death versus Alive/Censored. A CatBoost-based ensemble learning model is proposed and evaluated using stratified five-fold cross-validation and an independent test set. The proposed model achieved a cross-validation accuracy of 79.90% and obtained an improved test accuracy of 86.90%, with precision, recall, and F1-score values of 0.87, 0.87, and 0.87, respectively. Furthermore, SHAP (SHapley Additive exPlanations) is integrated to enhance model transparency by identifying influential clinical factors and generating patient-specific explanations for individual predictions. The proposed explainable ensemble framework provides a reliable and interpretable approach for mortality risk assessment in cirrhosis patients, supporting data-driven clinical decision-making while maintaining reproducibility and transparency.

Keywords: Liver Cirrhosis, Primary Biliary Cirrhosis, CatBoost, Ensemble Learning, Mortality Prediction, Machine Learning, SHAP Explainability, Clinical Decision Support

INTRODUCTION

The Liver cirrhosis is a chronic and progressive disease characterized by irreversible scarring of liver tissue, ultimately leading to impaired hepatic function and increased mortality risk. It represents one of the leading causes of death worldwide and is associated with significant healthcare burdens due to frequent hospitalizations, complications, and long-term management requirements. The progression of liver cirrhosis is often influenced by multiple clinical factors, including age, bilirubin levels, albumin concentration, prothrombin time, and the presence of comorbid conditions. Early identification of patients at high risk of mortality



is essential for timely clinical intervention, treatment prioritization, and improved patient outcomes. However, conventional prognostic methods rely heavily on physician expertise and traditional scoring systems, which may not adequately capture the complex relationships among clinical variables.

Recent advancements in artificial intelligence (AI) and machine learning (ML) have transformed predictive analytics in healthcare by enabling the development of data-driven models capable of extracting meaningful patterns from clinical datasets. Machine learning techniques such as Logistic Regression, Support Vector Machines, Decision Trees, Random Forests, and Gradient Boosting methods have demonstrated considerable potential in disease diagnosis, prognosis, and survival prediction. Ensemble learning approaches, in particular, have gained significant attention due to their ability to combine multiple weak learners to improve prediction accuracy, robustness, and generalization capabilities. Despite these advancements, many existing machine learning models operate as "black-box" systems, providing limited insight into the factors influencing their predictions. This lack of transparency presents a significant challenge in healthcare environments, where interpretability and trust are fundamental requirements for clinical adoption.

Explainable Artificial Intelligence (XAI) has emerged as a promising solution to address the interpretability limitations of machine learning models. XAI techniques enable healthcare professionals to understand, validate, and trust model predictions by revealing the contribution of individual features toward decision-making processes. Among various XAI methods, SHapley Additive exPlanations (SHAP) has gained widespread acceptance due to its theoretical foundation and ability to provide both global and local explanations. SHAP facilitates the identification of influential clinical attributes while simultaneously offering patient-specific interpretations, thereby enhancing the transparency and reliability of AI-driven healthcare applications.

Several studies have investigated the application of machine learning for liver disease diagnosis and prognosis; however, challenges related to predictive accuracy, model generalization, and interpretability remain unresolved. Existing approaches frequently emphasize performance improvement without adequately addressing the need for explainable decision support systems. Furthermore, limited research has explored the integration of ensemble learning techniques with explainability frameworks for liver cirrhosis mortality prediction using structured clinical data. This research gap highlights the necessity for developing robust and interpretable predictive models capable of supporting clinical decision-making.

In this paper, an explainable ensemble machine learning framework is proposed for liver cirrhosis mortality prediction using the Mayo Clinic Primary Biliary Cirrhosis (PBC) dataset. The framework incorporates comprehensive data preprocessing, Random Forest-based baseline evaluation, and a CatBoost ensemble model for enhanced predictive performance. Additionally, SHAP-based explainability analysis is employed to identify key mortality indicators and generate patient-level explanations. Experimental evaluation using five-fold stratified cross-validation and held-out test analysis demonstrates the effectiveness of the

proposed approach, achieving superior performance compared to the baseline model. By combining ensemble learning with explainable AI, the proposed framework offers a transparent, accurate, and clinically relevant solution for mortality risk assessment in liver cirrhosis patients, thereby supporting informed and timely healthcare decision-making. Figure 1 illustrates the sequential progression of liver disease from a healthy liver through fatty liver and fibrosis stages to advanced liver cirrhosis.

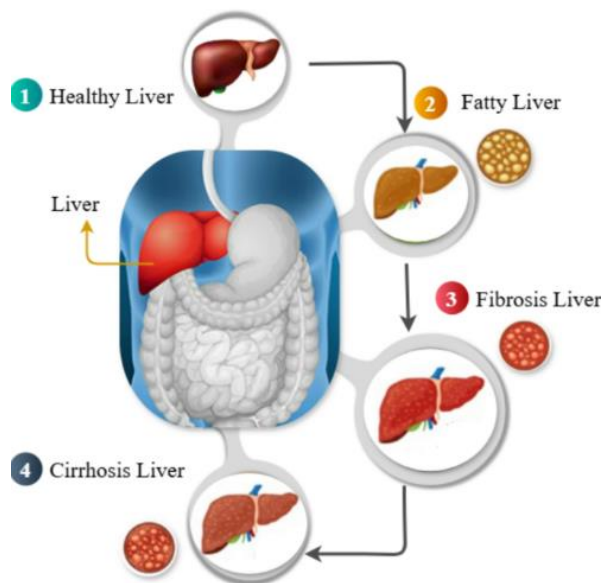


Figure 1: Progression of Liver Disease from Healthy Liver to Cirrhosis

RELATED WORK

Recent studies have significantly advanced the understanding of liver diseases, particularly metabolic dysfunction-associated steatotic liver disease (MASLD), cirrhosis, and related complications. Bhowmick and Bagchi [1] highlighted the therapeutic potential of fecal microbiota transplantation in managing liver diseases while emphasizing concerns regarding safety and long-term efficacy. Mousa et al. [2] demonstrated a strong clinical association between periodontitis and MASLD, underscoring the need for interdisciplinary collaboration between hepatologists and dental practitioners. Gozzi et al. [3] investigated Fontan-associated liver disease and reported variations in awareness and management practices across Europe. Gan et al. [4] provided a comprehensive analysis of liver disease epidemiology, identifying increasing global prevalence trends and predicting a substantial rise in disease burden over the coming decades. Wang et al. [5] combined machine learning with meta-analysis to evaluate the effectiveness of Xiao-Chai-Hu decoction in liver disease treatment, demonstrating the growing role of artificial intelligence in therapeutic research. Alcaraz et al. [6] developed an explainable machine learning framework using electrocardiogram data for liver disease diagnosis, achieving reliable external validation and emphasizing model interpretability. Zhang et al. [7] employed machine learning and meta-analysis to establish the efficacy of epigallocatechin-3-gallate in non-alcoholic fatty liver disease management. Jeong et al. [8] proposed deep learning-based nomograms utilizing gadoxetic acid-enhanced MRI to

predict post-hepatectomy liver failure, reporting superior prognostic performance. Clinical guidelines and consensus statements by Tacke et al. [9], Kanwal et al. [16], and Portincasa and Baffy [17] provided updated recommendations for MASLD diagnosis, nomenclature, and management strategies. Bilson et al. [10] identified a significant relationship between steatotic liver disease and chronic kidney disease, while Targher et al. [11] comprehensively reviewed the pathophysiology, diagnosis, and treatment of MASLD. Younossi et al. [12,15] investigated the epidemiology and mortality patterns associated with MASLD, demonstrating comparable mortality rates between MASLD and NAFLD. Zeng et al. [13] summarized current therapeutic interventions, whereas Kalligeros et al. [14] established an association between MASLD and increased cancer risk. De et al. [18] reported that MASLD criteria outperform MAFLD definitions for identifying lean patients with fatty liver disease. Han et al. [19] conducted a nationwide cohort study and confirmed elevated mortality risks among MASLD patients. Allen et al. [20] discussed the challenges accompanying the transition from NAFLD to MASLD nomenclature. Sangro et al. [21] reviewed recent pharmacological advancements for MAFLD treatment, while Takahashi et al. [22] explored the pathology of MASLD-associated hepatic tumors. Furthermore, Moon et al. [23] established that MASLD significantly increases the incidence of cardiovascular diseases. Collectively, these studies demonstrate substantial progress in liver disease research while highlighting persistent challenges related to early diagnosis, mortality prediction, and the development of explainable AI-driven clinical decision support systems.

PROPOSED EXPERIMENTAL SETUP

a. Data Acquisition and Preprocessing

The study utilized the Mayo Clinic Primary Biliary Cirrhosis (PBC) dataset comprising 418 patient records. Data preprocessing involved removing patient identifiers, handling missing values using median imputation, encoding categorical attributes, and transforming the target variable into a binary classification problem (Death versus Alive/Censored) to ensure reliable mortality prediction.

b. Development of Ensemble Learning Models

Figure 1 illustrates the complete workflow of the proposed explainable ensemble learning framework, encompassing data preprocessing, machine learning model development, performance evaluation, SHAP-based explainability, and clinical decision support for liver cirrhosis mortality risk prediction.

Two ensemble learning models were developed for liver cirrhosis mortality prediction. Random Forest was employed as the baseline model to establish a performance benchmark, while CatBoost was proposed as the primary ensemble learning approach. The Random Forest model utilized multiple decision trees with majority voting, whereas CatBoost leveraged gradient boosting with ordered boosting mechanisms to improve predictive accuracy and reduce overfitting. Both models were trained using identical preprocessing pipelines to ensure fair comparison. Furthermore, CatBoost was integrated with explainability techniques to provide transparent and clinically interpretable mortality predictions, thereby enhancing its applicability in healthcare decision support systems.

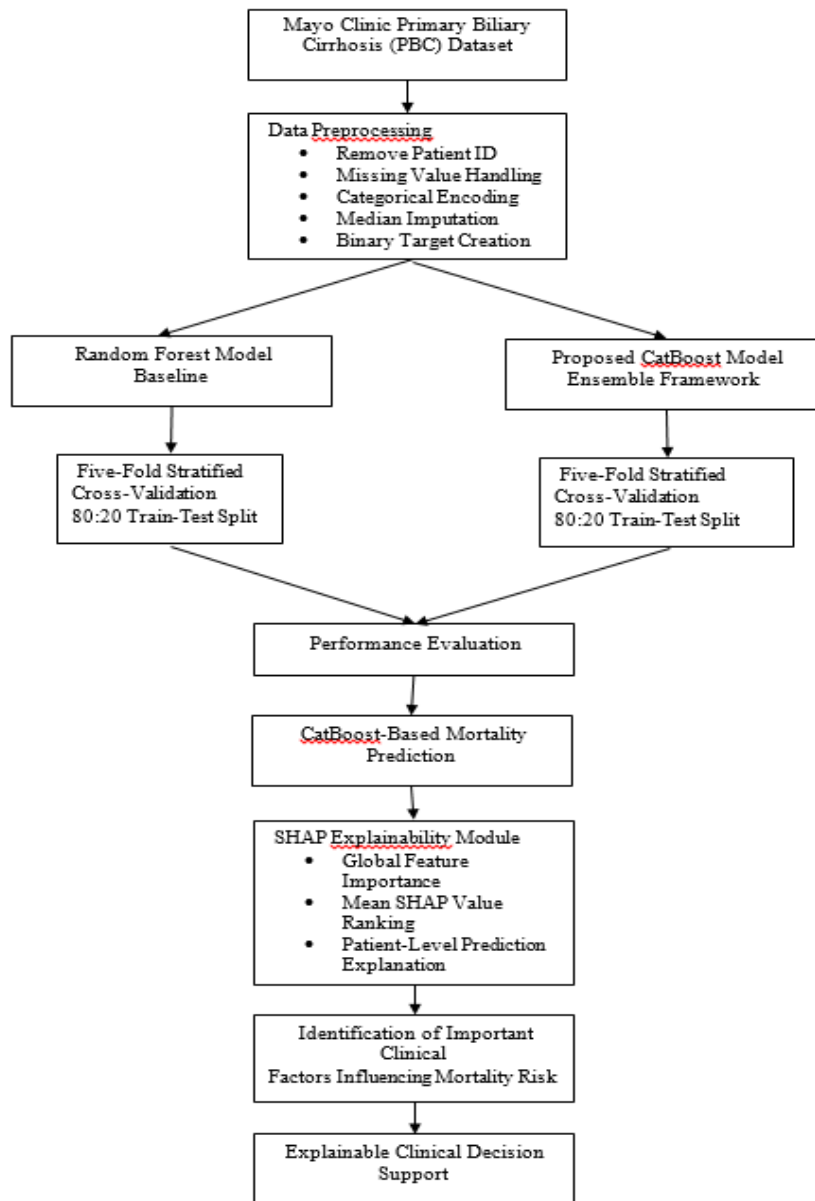


Figure 1: Workflow of the Proposed Explainable Ensemble Learning Framework for Liver Cirrhosis Mortality Risk Prediction

Model Training and Validation Strategy

The dataset was evaluated using an 80:20 stratified train-test split and five-fold stratified cross-validation. Stratification preserved the class distribution across all subsets, ensuring robust performance estimation. This validation strategy enabled reliable assessment of model generalization capability and minimized the risk of overfitting on unseen patient records.

Performance Evaluation Framework

Accuracy

Accuracy measures the proportion of correctly classified instances among all observations in the dataset. It provides an overall indication of model performance.

$$\text{Accuracy} = ((TP + TN)) / ((TP + TN + FP + FN)) \quad (1)$$

where TP, TN, FP, and FN denote True Positives, True Negatives, False Positives, and False Negatives, respectively.

Precision

Precision evaluates the proportion of correctly predicted positive instances among all instances predicted as positive. It is particularly useful for assessing the reliability of mortality predictions.

$$\text{Precision} = ((TP)) / ((TP + FP)) \quad (2)$$

Recall (Sensitivity)

Recall, also known as sensitivity, measures the ability of the model to correctly identify actual positive cases.

$$\text{Sensitivity} = ((TP)) / ((TP + FN)) \quad (3)$$

Higher recall values indicate improved detection of mortality cases.

F1-Score

The F1-score is the harmonic mean of precision and recall and provides a balanced evaluation when dealing with imbalanced datasets.

$$\text{F1-Score} = (2 * \text{Precision} * \text{Recall}) / (2 * \text{Precision} + \text{Recall}) \quad (4)$$

Confusion Matrix

The confusion matrix is a tabular representation of classification outcomes. It summarizes the number of true positives, true negatives, false positives, and false negatives produced by the model.

Explainable Artificial Intelligence Integration

SHAP (SHapley Additive exPlanations) was integrated into the proposed framework to enhance model transparency and interpretability. SHAP-based analysis identified influential clinical features affecting mortality risk and generated patient-specific explanations, enabling healthcare professionals to understand model predictions and supporting trustworthy clinical decision-making.

RESULTS ANALYSIS

a. Performance Analysis of the Proposed Model (CatBoost)

The proposed CatBoost model was developed to enhance the accuracy and reliability of liver cirrhosis mortality prediction using clinical data from the Mayo Clinic Primary Biliary Cirrhosis dataset. The model was evaluated using five-fold stratified cross-validation and an 80:20 held-out test strategy. Experimental results demonstrated superior performance compared to the existing Random Forest model, achieving a mean cross-validation accuracy of 79.90% and a held-out test accuracy of 86.90%. Additionally, the model exhibited improved precision, recall, and F1-score values, particularly for identifying high-risk mortality cases. These findings highlight the effectiveness, robustness, and clinical applicability of the proposed CatBoost-based ensemble framework for explainable mortality risk assessment. Table 1 summarizes the cross-validated performance of the proposed CatBoost model, followed by its held-out test performance (Table III) on the same 84-patient held-out test set used for the base model.

Table 1: Proposed model (catboost) - five-fold cross-validated accuracy

Model	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean CV Acc.
CatBoost	0.833	0.810	0.786	0.819	0.747	79.90%

Table 2: shows the held-out test performance of the proposed CatBoost model for liver cirrhosis mortality prediction. The model achieved high precision, recall, and F1-score values across both classes, with an overall weighted F1-score of 0.867, demonstrating its effectiveness in accurately identifying patient survival outcomes and mortality risk.

Table 2: Proposed model (catboost) - held-out test performance

Class	Precision	Recall	F1-score	Support
0 (Alive/Other)	0.860	0.942	0.899	52
1 (Death)	0.889	0.750	0.814	32
Macro Avg	0.874	0.846	0.856	84
Weighted Avg	0.871	0.869	0.867	84

Confusion Matrix - CatBoost (Proposed) Acc=86.90%

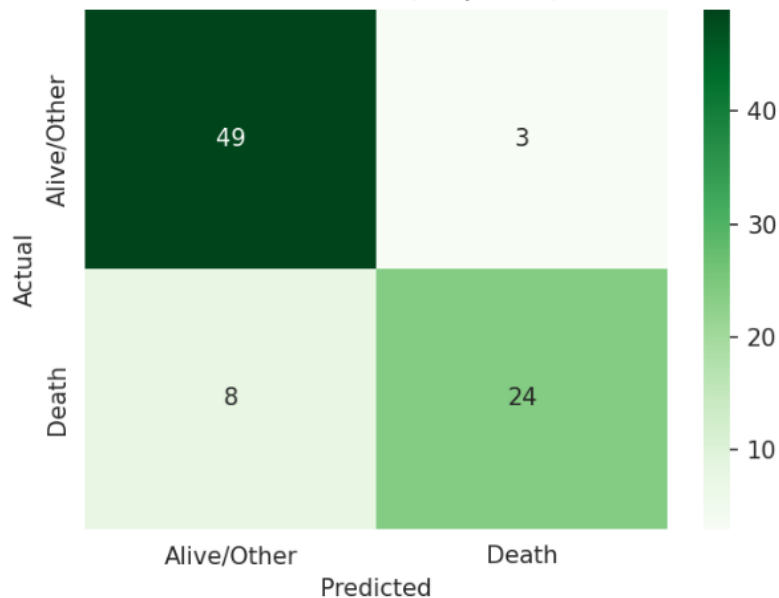


Figure 2: Confusion Matrix - Proposed Model (CatBoost)

Figure 2 illustrates the confusion matrix of the proposed CatBoost model, showing the number of correctly and incorrectly classified Alive/Other and Death cases in the liver cirrhosis mortality prediction task.

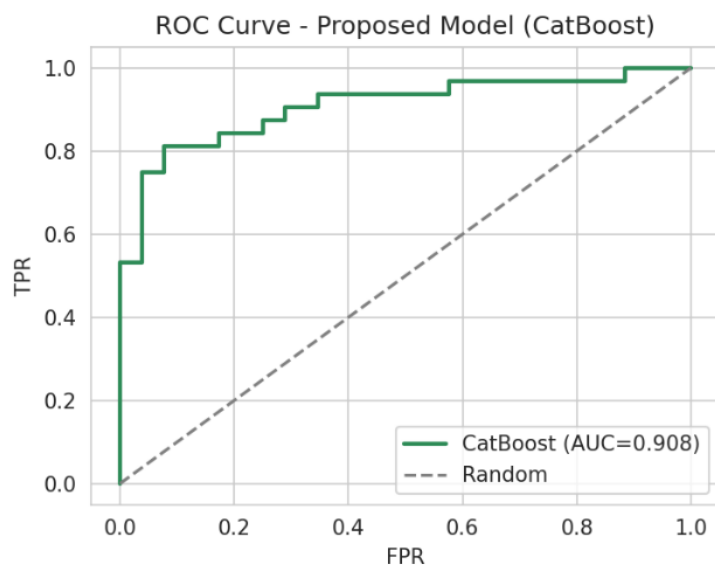


Figure 3: ROC Curve - Proposed Model (CatBoost)

Figure 3 illustrates the ROC curve of the proposed CatBoost model, demonstrating its enhanced ability to discriminate between Alive/Other and Death classes across varying classification thresholds.

b. Held-Out Test Performance Analysis

Held-out test accuracy (CatBoost): 86.90%. Confusion Matrix (held-out test set): predicted correctly 49 of 52 Alive/Other cases and 24 of 32 Death cases, with 3 false positives and 8 false negatives.

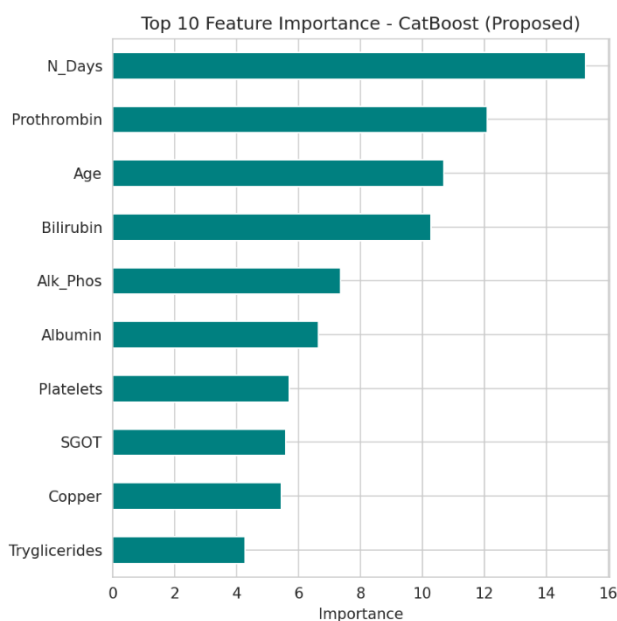


Figure 4: Top-10 Feature Importance - Proposed Model (CatBoost)

Figure 4 illustrates the top-10 most influential clinical features identified by the proposed CatBoost model, highlighting their relative contributions to liver cirrhosis mortality prediction.

c. Five-Fold Cross-Validation Results

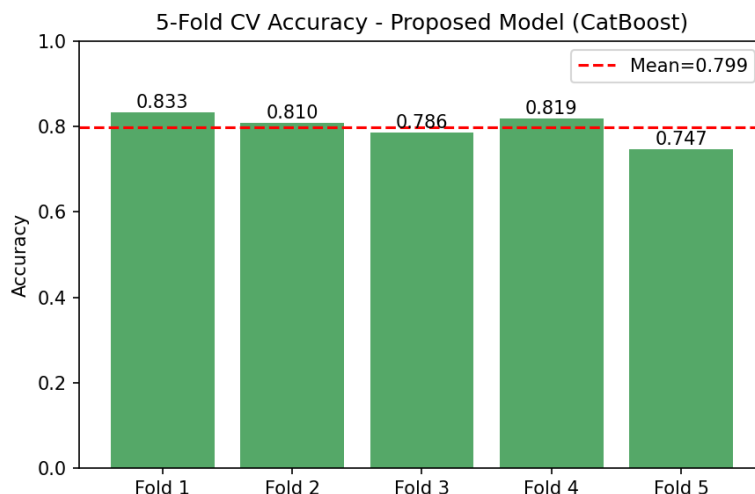


Figure 5: Five-Fold Cross-Validation Accuracy - Proposed Model (CatBoost)

Figure 5 illustrates the five-fold cross-validation accuracy of the proposed CatBoost model, demonstrating its consistent performance and strong generalization capability across different data folds.

d. Explainability Analysis of the Proposed Framework

Explainability is a critical requirement for machine learning applications in healthcare, where model predictions must be transparent and interpretable to gain clinical trust. To address this challenge, SHAP (SHapley Additive exPlanations) was integrated into the proposed CatBoost-based framework to provide both global and local interpretability. The explainability analysis enables the identification of significant clinical features influencing mortality prediction and offers patient-specific insights into model decisions. By incorporating explainable artificial intelligence techniques, the proposed framework supports transparent, reliable, and clinically meaningful decision-making for liver cirrhosis mortality risk assessment.

e. Comparative Performance Analysis

A comparative evaluation was conducted between the Random Forest model and the proposed CatBoost model using an identical implementation setup, including the same preprocessing pipeline, stratified cross-validation folds, and held-out test dataset. This ensures a fair assessment of the improvement achieved by the proposed ensemble framework.

Table 3. Comparative Performance of Random Forest and Proposed CatBoost Model

Metric	Random Forest	Proposed CatBoost	Improvement
5-Fold Cross-Validation Accuracy	79.66%	79.90%	+0.24 percentage points

Held-out Test Accuracy	83.33%	86.90%	+3.57 percentage points
Weighted Precision	0.832	0.871	+0.039
Weighted Recall	0.833	0.869	+0.036
Weighted F1-score	0.831	0.867	+0.036
Death-Class Recall	0.719	0.750	+0.031

The comparative results in table 3 demonstrate that the proposed CatBoost model provides consistent performance improvements over the Random Forest classifier. The proposed model achieved a 3.57% increase in held-out test accuracy and improved weighted precision, recall, and F1-score, indicating better overall classification capability. Furthermore, the increase in Death-class recalls from 0.719 to 0.750 highlights the improved ability of the proposed framework to identify high-risk patients, which is particularly important for mortality prediction tasks.

The five-fold cross-validation results show a marginal improvement of 0.24 percentage points, indicating that the proposed model maintains stable generalization performance on the limited clinical dataset. Overall, the CatBoost-based framework provides improved predictive performance while maintaining robustness and reliability under the same evaluation conditions.

CONCLUSION

This study presented an explainable ensemble learning framework for liver cirrhosis mortality prediction using the Mayo Clinic Primary Biliary Cirrhosis dataset. By transforming the original multi-class problem into a clinically meaningful binary classification task (Death versus Alive/Censored), the framework achieved improved prediction stability and interpretability. The proposed CatBoost model outperformed the baseline Random Forest model, attaining a five-fold cross-validation accuracy of 79.90% and a held-out test accuracy of 86.90%, along with superior precision, recall, and F1-score values. Furthermore, the integration of SHAP-based explainability provided valuable insights into the influence of key clinical features and enabled patient-level interpretation of model predictions. The results demonstrate that explainable ensemble learning can serve as an effective and reliable approach for mortality risk assessment in liver cirrhosis patients, offering significant potential for supporting transparent and informed clinical decision-making in healthcare environments.

REFERENCES

1. Bhowmick, Jayeeta, and Arka Bagchi. "Fecal microbiota transplantation in liver diseases: Therapeutic potential and associated risks." *Progress in Molecular Biology and Translational Science* 220 (2026): 229-246.
2. Mousa, Nasser, et al. "Periodontitis and Metabolic Dysfunction-Associated Steatotic Liver Disease: emphasizing the clinical interplay between hepatologists and dentists." *Odontology* 114.2 (2026): 378-389.



3. Gozzi, Anna, et al. "Fontan-associated liver disease (FALD) in the EUROFontan experience. An insight into European awareness." *Pediatric Cardiology* 47.1 (2026): 397-409.
4. Gan, Can, et al. "Liver diseases: epidemiology, causes, trends and predictions." *Signal Transduction and Targeted Therapy* 10.1 (2025): 33.
5. Wang, Zexin, et al. "Precision strike strategy for liver diseases trilogy with Xiao-Chai-Hu decoction: a meta-analysis with machine learning." *Phytomedicine* 142 (2025): 156796.
6. Alcaraz, Juan Miguel Lopez, Wilhelm Haverkamp, and Nils Strodthoff. "Electrocardiogram-based diagnosis of liver diseases: an externally validated and explainable machine learning approach." *EClinicalMedicine* 84 (2025).
7. Zhang, Yuanhao, et al. "Preclinical evidence construction for epigallocatechin-3-gallate against non-alcoholic fatty liver disease: a meta-analysis and machine learning study." *Phytomedicine* 142 (2025): 156651.
8. Jeong, Boryeong, et al. "Predicting post-hepatectomy liver failure in patients with hepatocellular carcinoma: nomograms based on deep learning analysis of gadoxetic acid-enhanced MRI." *European Radiology* 35.5 (2025): 2769-2782.
9. Tacke, Frank, et al. "EASL–EASD–EASO Clinical Practice Guidelines on the management of metabolic dysfunction-associated steatotic liver disease (MASLD)." *Journal of hepatology* 81.3 (2024): 492-542.
10. Bilson, Josh, et al. "Steatotic liver disease, MASLD and risk of chronic kidney disease." *Diabetes & metabolism* 50.1 (2024): 101506.
11. Targher, Giovanni, Luca Valenti, and Christopher D. Byrne. "Metabolic dysfunction–associated steatotic liver disease." *New England Journal of Medicine* 393.7 (2025): 683-698.
12. Younossi, Zobair M., Markos Kalligeros, and Linda Henry. "Epidemiology of metabolic dysfunction-associated steatotic liver disease." *Clinical and molecular hepatology* 31.Suppl (2024): S32.
13. Zeng, Jing, Jian-Gao Fan, and Sven M. Francque. "Therapeutic management of metabolic dysfunction associated steatotic liver disease." *United European gastroenterology journal* 12.2 (2024): 177-186. Tacke, Frank, et al. "EASL–EASD–EASO Clinical Practice Guidelines on the management of metabolic dysfunction-associated steatotic liver disease (MASLD)." *Journal of hepatology* 81.3 (2024): 492-542.
14. Kalligeros, Markos, Linda Henry, and Zobair M. Younossi. "Metabolic dysfunction-associated steatotic liver disease and its link to cancer." *Metabolism* 160 (2024): 156004.
15. Younossi, Zobair M., et al. "Clinical profiles and mortality rates are similar for metabolic dysfunction-associated steatotic liver disease and non-alcoholic fatty liver disease." *Journal of Hepatology* 80.5 (2024): 694-701.



16. Kanwal, Fasiha, et al. "Metabolic dysfunction–associated steatotic liver disease: Update and impact of new nomenclature on the American Association for the Study of Liver Diseases practice guidance on nonalcoholic fatty liver disease." *Hepatology* 79.5 (2024): 1212-1219.
17. Portincasa, Piero, and Gyorgy Baffy. "Metabolic dysfunction-associated steatotic liver disease: Evolution of the final terminology." *European journal of internal medicine* 124 (2024): 35-39.
18. De, Arka, et al. "Metabolic dysfunction-associated steatotic liver disease (MASLD) definition is better than MAFLD criteria for lean patients with NAFLD." *Journal of hepatology* 80.2 (2024): e61-e62.
19. Han, Eugene, et al. "Mortality in metabolic dysfunction-associated steatotic liver disease: a nationwide population-based cohort study." *Metabolism* 152 (2024): 155789.
20. Allen, Alina M., et al. "Nonalcoholic fatty liver disease gets renamed as metabolic dysfunction–associated steatotic liver disease: progress but with challenges." *Gastroenterology* 166.2 (2024): 229-234.
21. Sangro, Paloma, et al. "Metabolic dysfunction–associated fatty liver disease (MAFLD): an update of the recent advances in pharmacological treatment." *Journal of physiology and biochemistry* 79.4 (2023): 869-879.
22. Takahashi, Yoshihisa, et al. "Pathology and pathogenesis of metabolic dysfunction-associated steatotic liver disease-associated hepatic tumors." *Biomedicines* 11.10 (2023): 2761.
23. Moon, Joon Ho, et al. "Metabolic dysfunction-associated steatotic liver disease increases the risk of incident cardiovascular disease: a nationwide cohort study." *EClinicalMedicine* 65 (2023).
24. G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," *Phil. Trans. Roy. Soc. London*, vol. A247, pp. 529–551, April 1955. (references)
25. J. Clerk Maxwell, "A Treatise on Electricity and Magnetism", 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
26. I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in *Magnetism*, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
27. K. Elissa, "Title of paper if known," unpublished.
28. R. Nicole, "Title of paper with only first word capitalized," *J. Name Stand. Abbrev.*, in press.
29. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," *IEEE Transl. J. Magn. Japan*, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetism Japan, p. 301, 1982].



Figure 1: Example of a figure caption

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity “Magnetization”, or “Magnetization, M”, not just “M”. If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write “Magnetization (A/m)” or “Magnetization {A[m(1)]}”, not just “A/m”. Do not label axes with a ratio of quantities and units. For example, write “Temperature (K)”, not “Temperature/K”.

ACKNOWLEDGMENT

Put sponsor acknowledgements in the unnumbered footnote on the first page.