IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.15501

Data-Driven Analysis of Mortality and Survival Outcomes in Patients with Liver Cirrhosis: A Global Healthcare Analytics Study

Mehdi Mostofi[1], Prasanna Poreddy[2], Jayasri Katragadda[2], Nadeem Basha Syed[2], Sunandamala Bodaballa[2]

Assistant Professor, Master of Data Analytics, University of Niagara Falls Canada, Niagara Falls, Canada[1]

Graduate, Master of Data Analytics, University of Niagara Falls Canada, Niagara Falls, Canada [2]

Abstract: Liver cirrhosis is a progressive chronic disease associated with high morbidity and mortality, requiring accurate risk stratification to support clinical decision-making. This study presents a data-driven analytical framework integrating survival analysis, machine learning, and uncertainty quantification to improve mortality prediction in cirrhosis patients. A retrospective dataset of 418 patients was analyzed using Kaplan–Meier estimation and Cox proportional hazards modeling to evaluate survival patterns and identify significant predictors. For predictive modeling, multiple supervised learning algorithms—including Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, Gradient Boosting, and Extreme Gradient Boosting (XGBoost)—were implemented to classify patients into high- and low-risk groups. Model performance was assessed using accuracy, recall, F1-score, and ROC-AUC, with cross-validation employed to ensure robustness. To enhance reliability, conformal prediction was applied to quantify predictive uncertainty at a predefined confidence level. Results indicate that disease stage, age, bilirubin, and albumin are significant predictors of mortality. Ensemble models demonstrated superior predictive performance, with XGBoost achieving the highest recall and strong discrimination. Conformal prediction provided well-calibrated uncertainty estimates, improving the interpretability and trustworthiness of model outputs. The findings demonstrate that integrating statistical and machine learning approaches enhances mortality risk stratification and supports the development of reliable clinical decision-support systems.

Keywords: Liver Cirrhosis; Survival Analysis; Machine Learning; Mortality Prediction; Cox Proportional Hazards; XGBoost; Risk Stratification; Conformal Prediction

I. INTRODUCTION

Liver cirrhosis is a progressive chronic liver disease and a leading cause of global morbidity and mortality, with increasing prevalence driven by viral hepatitis, alcohol-related liver disease, and non-alcoholic fatty liver disease (NAFLD) [1], [2]. As cirrhosis progresses from compensated to decompensated stages, patients experience a substantial increase in complications such as ascites, hepatic encephalopathy, and variceal bleeding, leading to significantly reduced survival rates [3]. Accurate mortality risk stratification is therefore critical for improving patient outcomes, optimizing treatment prioritization, and supporting healthcare resource allocation.

Traditional prognostic models, including the Child–Pugh score and the Model for End-Stage Liver Disease (MELD), are widely used to estimate disease severity and short-term mortality risk [4]. These models rely on a limited set of clinical and biochemical variables and assume relatively simple relationships between predictors and outcomes. While clinically interpretable, they may not fully capture complex interactions and nonlinear patterns inherent in heterogeneous patient populations [5].

Recent advances in machine learning (ML) have enabled the development of predictive models that can capture high-dimensional and nonlinear relationships in clinical data. Ensemble learning methods such as Random Forest, Gradient Boosting, and Extreme Gradient Boosting (XGBoost) have demonstrated strong predictive performance in healthcare applications, including mortality prediction in cirrhosis patients [6], [7]. Studies have shown that ML models can outperform traditional statistical approaches in classification tasks by improving discrimination and sensitivity [8]. However, despite these advantages, ML models often suffer from reduced interpretability and limited transparency, which can hinder clinical adoption [9].

Survival analysis techniques remain fundamental for modeling time-to-event outcomes in medical research. Kaplan–Meier estimation provides a non-parametric approach for analyzing survival probabilities, while Cox proportional hazards modeling enables quantification of the effect of covariates on mortality risk [10]. These methods offer strong interpretability and statistical rigor but are primarily designed for inference rather than predictive classification, limiting their ability to fully leverage complex data structures [11].

In recent years, hybrid approaches that combine statistical modeling with machine learning have gained attention for improving predictive performance while maintaining interpretability [12]. Such approaches aim to integrate the strengths of both paradigms, enabling more robust risk stratification frameworks. Additionally, uncertainty quantification has emerged as a critical requirement in clinical decision-support systems. Conformal prediction methods provide distribution-free guarantees on prediction reliability, offering calibrated confidence measures that enhance trust in model outputs [13], [14].

Despite these advancements, existing studies often treat survival analysis and machine learning independently, with limited integration of uncertainty quantification techniques. Furthermore, many predictive models prioritize discrimination metrics such as ROC-AUC without adequately addressing reliability and calibration, which are essential for clinical applicability [15].

To address these gaps, this study proposes an integrated analytical framework that combines survival analysis, ensemble machine learning models, and conformal prediction to improve mortality risk stratification in patients with liver cirrhosis. The main contributions of this study are threefold: (1) evaluation of survival patterns using Kaplan–Meier estimation and Cox proportional hazards modeling; (2) development and comparison of multiple machine learning classifiers for mortality prediction; and (3) incorporation of conformal prediction to quantify uncertainty and enhance model reliability. The proposed framework aims to support more accurate, interpretable, and reliable clinical decision-making.

II. METHODOLOGY

This study follows a structured analytical framework aligned with the Cross-Industry Standard Process for Data Mining (CRISP-DM), encompassing data understanding, preparation, modeling, and evaluation. The objective is to develop an integrated approach for mortality risk stratification in patients with liver cirrhosis by combining survival analysis, machine learning, and uncertainty quantification.

A. Data Source and Study Design

The dataset used in this study was obtained from the publicly available UCI Machine Learning Repository and consists of anonymized clinical records of patients diagnosed with liver cirrhosis. The dataset includes 418 observations and 20 variables representing demographic characteristics, clinical indicators, biochemical laboratory measurements, and survival outcomes.

The study design is retrospective and observational. Each record corresponds to an individual patient with associated time-to-event information and outcome status. Death was defined as the primary event of interest, while censored observations include patients who remained alive or underwent liver transplantation during the study period.

B. Data Preprocessing and Feature Engineering

Data preprocessing was performed to ensure data quality and model compatibility. Missing values in numerical variables were imputed using median values to reduce sensitivity to skewed distributions, while categorical variables were imputed using the most frequent category.

To enhance interpretability, age was converted from days to years and further rescaled into decades for modeling purposes. Categorical variables, including sex and clinical indicators, were encoded numerically, while ordinal encoding was applied to variables representing severity levels such as edema.

The outcome variable was transformed into a binary classification target, where 1 indicates death and 0 represents censored observations. The dataset was partitioned into training (80%) and testing (20%) subsets using stratified sampling to preserve class distribution. Five-fold cross-validation was applied within the training data for robust model evaluation.

C. Survival Analysis

To analyze time-to-event outcomes, Kaplan–Meier estimation was used to evaluate survival probabilities across the study cohort and stratified by disease stages. The log-rank test was applied to assess statistically significant differences in survival distributions.

To quantify the effect of predictors on mortality risk, the Cox proportional hazards model was employed. This semi-parametric model estimates hazard ratios for demographic, clinical, and biochemical variables while allowing flexible baseline hazard specification. Model assumptions, including proportional hazards, were evaluated using residual-based diagnostics.

D. Machine Learning Models

To perform mortality risk classification, multiple supervised machine learning algorithms were implemented:

- Logistic Regression (baseline model)

ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.15501

- Decision Tree
- Random Forest
- Support Vector Machine (SVM)
- Gradient Boosting
- Extreme Gradient Boosting (XGBoost)
- Stacking Ensemble Model

These models represent diverse learning paradigms, including linear, tree-based, ensemble, and margin-based approaches. Hyperparameter tuning was conducted using grid search within the cross-validation framework to optimize model performance.

E. Model Evaluation

Model performance was assessed using multiple evaluation metrics, including accuracy, recall (sensitivity), precision, F1-score, and area under the receiver operating characteristic curve (ROC-AUC). Recall was prioritized as the primary metric due to its importance in minimizing false negatives in clinical risk prediction.

In addition to discrimination metrics, calibration performance was considered to ensure agreement between predicted probabilities and observed outcomes, supporting the reliability of predictions.

F. Uncertainty Quantification

To enhance model reliability, conformal prediction was applied to the Gradient Boosting model to quantify predictive uncertainty. This approach generates prediction sets with guaranteed coverage at a predefined confidence level, enabling more transparent and trustworthy decision-making in clinical contexts.

III. RESULTS

A. Descriptive Analysis and Data Distribution

The distribution of disease severity stages within the study cohort is illustrated in **Fig. 1**, which shows that Stage III and Stage IV patients constitute the majority of observations, indicating a higher representation of advanced disease cases. This distribution supports the suitability of the dataset for survival analysis and risk stratification.

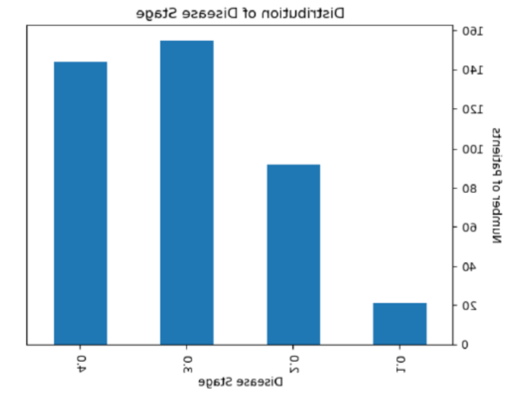

Fig. 1. Distribution of Liver Cirrhosis Disease Stages

The distribution of survival time across disease stages is presented in **Fig. 2**, where a progressive decline in median survival time is observed as disease severity increases. Patients in early stages exhibit longer survival durations, while advanced-stage patients demonstrate shorter survival times.

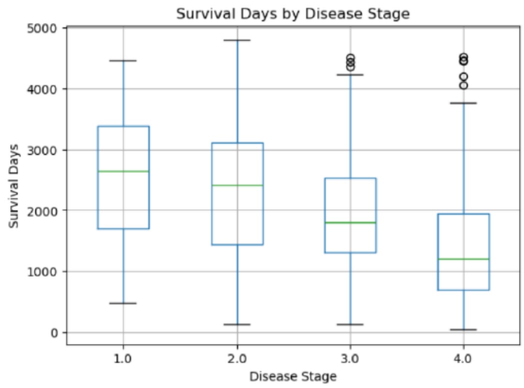

Fig. 2. Distribution of Survival Time Across Stages

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.15501

Correlation among independent variables is shown in **Fig. 3**, which indicates weak to moderate relationships among predictors, suggesting that multicollinearity is unlikely to significantly affect model performance.

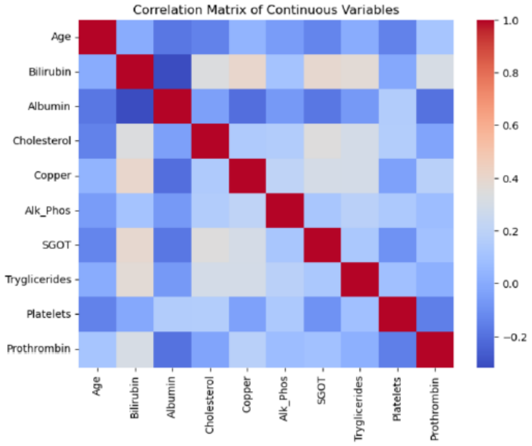

Fig. 3. Correlation Matrix

Variance Inflation Factor (VIF) analysis is presented in **Fig. 4**, confirming that all predictors remain within acceptable thresholds, indicating no significant multicollinearity.

	Variable	VIF
0	const	307.674300
1	Age	1.175043
2	Bilirubin	2.067321
3	Albumin	1.233176
4	Cholesterol	1.410626
5	Copper	1.371159
6	Alk_Phos	1.114205
7	SGOT	1.395189
8	Tryglicerides	1.348955
9	Platelets	1.173439
10	Prothrombin	1.270644

Fig. 4. VIF Plot

The distribution of key biochemical markers across outcome groups is illustrated in **Fig. 5** and **Fig. 6**. Patients who experienced mortality show higher bilirubin levels (Fig. 5) and lower albumin levels (Fig. 6), supporting their role as important mortality predictors.

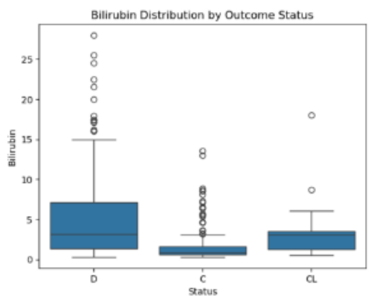

Fig. 5. Bilirubin Distribution by Outcome

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.15501

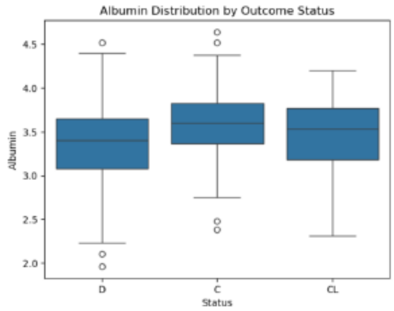

Fig. 6. Albumin Distribution by Outcome

B. Survival Analysis Results

The overall survival probability of the cohort is presented in **Fig. 7**, which demonstrates a steady decline in survival probability over time, reflecting increasing cumulative mortality risk.

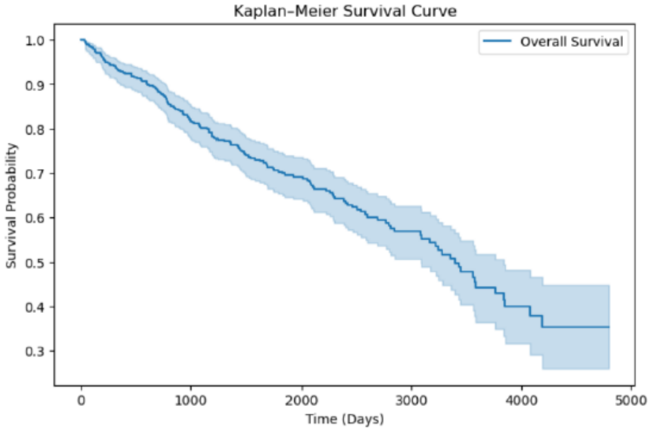

Fig. 7. Kaplan-Meier Overall Survival Curve

Survival curves stratified by disease stage are shown in **Fig. 8**, where a clear separation between stages is observed. Patients in Stage I exhibit the highest survival probabilities, whereas Stage IV patients show significantly reduced survival.

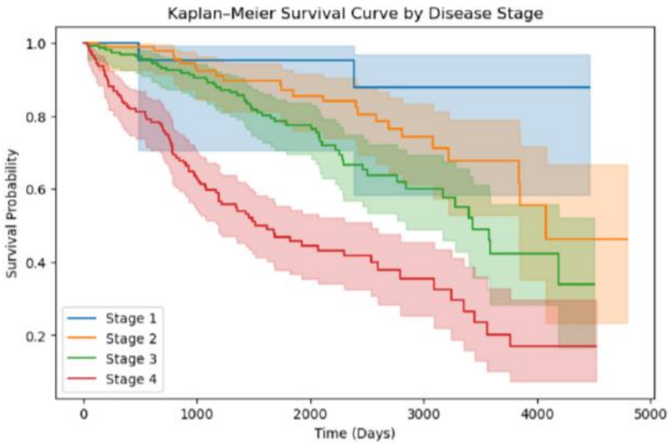

Fig. 8. Kaplan-Meier Survival by Stage

The statistical comparison of survival distributions using the log-rank test is presented in **Fig. 9**, indicating a statistically significant difference across disease stages ($p < 0.05$).

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 15, Issue 5, May 2026
DOI: 10.17148/IJARCCE.2026.15501

t_0	-1
null_distribution	chi squared
degrees_of_freedom	3
test_name	multivariate_logrank_test

	test_statistic	p	-log2(p)
0	68.30	<0.005	46.53

Fig. 9. Log-Rank Test Results

C. Cox Proportional Hazards Modeling

Multivariable Cox regression results are presented in **Fig. 10** and **Fig. 11**, showing the impact of predictors on mortality risk. Disease stage, bilirubin, albumin, and age are identified as significant factors.

	coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) lower 95%	exp(coef) upper 95%	cmp to	z	p	-log2(p)
Bilirubin	0.13	1.13	0.01	0.10	0.15	1.11	1.16	0.00	10.25	<0.005	79.48
Albumin	-1.29	0.27	0.20	-1.68	-0.91	0.19	0.40	0.00	-6.60	<0.005	34.49

Concordance	0.77
Partial AIC	1613.00
log-likelihood ratio test	137.94 on 2 df
-log2(p) of ll-ratio test	99.50

Fig. 10. Multivariate Cox Model (H2)

	coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) lower 95%	exp(coef) upper 95%	cmp to	z	p	-log2(p)
Age_10yr	0.00	1.00	0.00	0.00	0.00	1.00	1.00	0.00	4.87	<0.005	19.75
Sex	0.27	1.31	0.22	-0.17	0.71	0.85	2.03	0.00	1.22	0.22	2.16

Concordance	0.61
Partial AIC	1724.36
log-likelihood ratio test	26.58 on 2 df
-log2(p) of ll-ratio test	19.17

Fig. 11. Multivariate Cox Model (H3)

Univariate Cox model results for key biomarkers are shown in **Fig. 12** and **Fig. 13**, where bilirubin demonstrates stronger predictive performance compared to albumin.

	coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) lower 95%	exp(coef) upper 95%	cmp to	z	p	-log2(p)
Bilirubin	0.14	1.15	0.01	0.12	0.16	1.13	1.18	0.00	12.26	<0.005	112.41

Concordance	0.78
Partial AIC	1651.56
log-likelihood ratio test	97.38 on 1 df
-log2(p) of ll-ratio test	73.89

Fig. 12. Bilirubin Cox Model

IJARCCE

ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.15501

	coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) lower 95%	exp(coef) upper 95%	cmp to	z	p	-log2(p)
Albumin	-1.54	0.21	0.18	-1.90	-1.19	0.15	0.31	0.00	-8.46	<0.005	55.05

Concordance	0.70
Partial AIC	1684.45
log-likelihood ratio test	64.50 on 1 df
-log2(p) of ll-ratio test	49.88

Fig. 13. Albumin Cox Model

D. Predictive Modeling Performance

Comparative performance of machine learning models is summarized in **TABLE I**, which reports accuracy, recall, F1-score, ROC-AUC, and cross-validation results.

TABLE I. Model Performance Comparison

Rank	Model	Accuracy	Recall	F1 Score	ROC-AUC (Test)	Recall (CV Mean)
1	XGBoost	0.75	0.9375	0.7407	0.9087	0.8065
2	Logistic Regression	0.8333	0.8125	0.7879	0.8407	0.6895
3	Decision Tree	0.7262	0.7188	0.6667	0.7212	0.6597
4	Random Forest	0.8214	0.7188	0.7541	0.878	0.6742
5	SVM	0.7381	0.6875	0.6667	0.8377	0.7286
6	Gradient Boosting	0.7857	0.6875	0.7097	0.875	0.6898
7	Stacking	0.8214	0.6875	0.7458	0.8936	0.6214

Among the evaluated models, XGBoost achieved the highest recall, indicating superior performance in identifying high-risk patients. Logistic Regression demonstrated strong baseline performance, while ensemble methods provided balanced results across metrics.

The combined Receiver Operating Characteristic (ROC) curves for all models are illustrated in **Fig. 14**, showing that all models achieved strong discrimination.

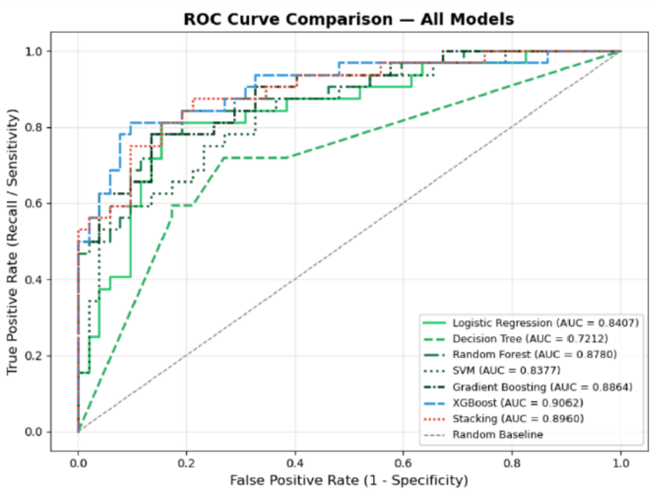

Fig. 14. Combined ROC Curve

Calibration performance was assessed to ensure alignment between predicted probabilities and observed outcomes, supporting the reliability of model predictions.

ISSN (O) 2278-1021, ISSN (P) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 15, Issue 5, May 2026
DOI: 10.17148/IJARCCE.2026.15501

IV. DISCUSSION

This study presents an integrated analytical framework combining survival analysis, machine learning, and uncertainty quantification to improve mortality risk stratification in patients with liver cirrhosis. The results demonstrate that both statistical and machine learning approaches provide complementary insights into patient outcomes, with important implications for clinical decision-making.

A. Interpretation of Key Findings

The survival analysis results confirm that disease severity is a primary determinant of patient outcomes. The Kaplan–Meier curves showed a clear gradient in survival probabilities across disease stages, with advanced-stage patients exhibiting significantly reduced survival. These findings are consistent with established clinical knowledge that disease progression is strongly associated with increased mortality risk.

The Cox proportional hazards model further quantified the impact of key predictors. Elevated serum bilirubin was associated with increased mortality risk, reflecting impaired hepatic excretory function, while higher albumin levels demonstrated a protective effect, indicating preserved liver synthetic capacity. Age was also identified as a significant predictor, highlighting increased vulnerability among older patients. These results align with prior studies emphasizing the importance of biochemical markers in cirrhosis prognosis.

B. Machine Learning Performance and Trade-Offs

The predictive modeling results indicate that machine learning approaches can effectively classify patients into high- and low-risk groups. Among the evaluated models, XGBoost achieved the highest recall, demonstrating superior sensitivity in identifying high-risk patients. This is particularly important in clinical settings, where failure to identify high-risk individuals (false negatives) can have serious consequences.

However, the results also highlight trade-offs between performance metrics. While ensemble models achieved higher recall and competitive ROC-AUC values, simpler models such as Logistic Regression provided strong baseline performance with greater interpretability. This suggests that model selection should be guided not only by predictive accuracy but also by clinical usability and transparency.

C. Integration of Survival Analysis and Machine Learning

A key contribution of this study is the integration of survival analysis with machine learning models. Survival analysis provides interpretable insights into time-to-event outcomes and quantifies the effects of predictors, while machine learning models enhance predictive discrimination and classification performance.

The combined approach enables a more comprehensive understanding of mortality risk, leveraging the strengths of both paradigms. This integration addresses a gap in existing research, where survival modeling and predictive classification are often treated independently.

D. Uncertainty Quantification and Model Reliability

The application of conformal prediction adds an important dimension to the analytical framework by providing calibrated uncertainty estimates. The observed empirical coverage of approximately 96% indicates that the prediction intervals are well-calibrated and reliable.

In clinical decision-support systems, uncertainty quantification is critical for ensuring trust in model predictions. By providing confidence measures alongside predictions, the proposed framework enhances transparency and supports more informed decision-making.

E. Clinical and Practical Implications

The findings of this study have several practical implications. First, the identification of key predictors such as bilirubin, albumin, disease stage, and age supports early identification of high-risk patients. Second, the integration of machine learning models enables more accurate risk classification, which can assist clinicians in prioritizing monitoring and interventions.

Furthermore, the incorporation of uncertainty quantification enhances the reliability of predictions, making the framework more suitable for real-world clinical applications. While the proposed approach is not intended to replace clinical judgment, it provides a valuable decision-support tool that can complement existing prognostic methods.

F. Limitations

Despite its contributions, this study has several limitations. The analysis is based on a single publicly available dataset, which may limit generalizability to broader patient populations. External validation using independent datasets is required to confirm the robustness of the proposed framework.

Additionally, missing values were imputed prior to model training, which may introduce minor information leakage. Future work should implement pipeline-based preprocessing within cross-validation to eliminate this risk. Finally, while machine learning models demonstrated strong predictive performance, further work is needed to enhance interpretability through techniques such as feature importance analysis.

G. Future Research Directions

Future research should focus on validating the proposed framework using larger and more diverse clinical datasets. The integration of explainable artificial intelligence (XAI) methods, such as SHAP, can further improve model transparency and clinical interpretability.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ※ Peer-reviewed & Refereed journal ※ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.15501

Additionally, incorporating longitudinal data and dynamic prediction models may enhance the ability to capture disease progression over time. Expanding the framework to include real-time clinical decision-support systems represents another important direction for future work. These findings reinforce the importance of combining statistical interpretability with predictive modeling in clinical risk assessment.

V. CONCLUSION

This study presented an integrated analytical framework for mortality risk stratification in patients with liver cirrhosis by combining survival analysis, machine learning, and uncertainty quantification. Kaplan–Meier estimation and Cox proportional hazards modeling identified disease stage, age, serum bilirubin, and serum albumin as significant predictors of mortality, providing interpretable insights into patient risk.

Comparative evaluation of supervised learning models demonstrated that ensemble methods, particularly XGBoost, achieve strong discriminative performance and high sensitivity in identifying high-risk patients, while simpler models such as Logistic Regression offer competitive baseline performance with greater interpretability. These findings highlight the importance of balancing predictive accuracy with clinical usability when selecting models for healthcare applications.

The incorporation of conformal prediction provided calibrated uncertainty estimates, enhancing the reliability and transparency of predictions. This contributes to the development of more trustworthy clinical decision-support systems by enabling risk predictions to be accompanied by confidence measures.

Overall, the results demonstrate that integrating statistical and machine learning approaches improves both interpretability and predictive performance in cirrhosis mortality analysis. The proposed framework has the potential to support early identification of high-risk patients and inform clinical decision-making. Future work should focus on external validation, pipeline-based preprocessing to eliminate potential data leakage, and the integration of explainable artificial intelligence techniques to further enhance model transparency and real-world applicability.

REFERENCES

[1]. S. K. Asrani, H. Devarbhavi, J. Eaton, and P. S. Kamath, "Burden of liver diseases in the world," *Journal of Hepatology*, vol. 70, no. 1, pp. 151–171, 2019.
[2]. P. Ginès, A. Krag, J. G. Abraldes, E. Solà, N. Fabrellas, and T. Kamath, "Liver cirrhosis," *The Lancet*, vol. 398, no. 10308, pp. 1359–1376, 2021.
[3]. F. Kanwal et al., "Risk of hospitalization and mortality in cirrhosis," *Hepatology*, vol. 71, no. 6, pp. 2082–2093, 2020.
[4]. P. S. Kamath and R. H. Kim, "The model for end-stage liver disease (MELD)," *Hepatology*, vol. 45, no. 3, pp. 797–805, 2007.
[5]. P. Jalan, R. Pavesi, and R. Moreau, "Acute-on-chronic liver failure: Pathophysiological basis of therapeutic options," *Blood*, vol. 138, no. 1, pp. 1–14, 2021.
[6]. X. Li et al., "Machine learning-based prediction of mortality in cirrhosis patients," *BMC Medical Informatics and Decision Making*, vol. 24, no. 1, pp. 1–12, 2024.
[7]. Y. Wang et al., "Prediction of short-term mortality in liver cirrhosis using machine learning models," *Scientific Reports*, vol. 14, no. 1, pp. 1–10, 2024.
[8]. H. Zhang et al., "Machine learning approaches for clinical risk prediction in liver disease," *IEEE Access*, vol. 13, pp. 1–12, 2025.
[9]. S. M. Lundberg et al., "From local explanations to global understanding with explainable AI for trees," *Nature Machine Intelligence*, vol. 2, no. 1, pp. 56–67, 2020.
[10]. T. M. Therneau and P. M. Grambsch, *Modeling Survival Data: Extending the Cox Model*, New York, NY, USA: Springer, 2000.
[11]. H. Kvamme, Ø. Borgan, and I. Scheel, "Time-to-event prediction with neural networks and Cox regression," *Journal of Machine Learning Research*, vol. 20, no. 129, pp. 1–30, 2019.
[12]. J. Katzman et al., "DeepSurv: Personalized treatment recommender system using Cox proportional hazards deep neural network," *BMC Medical Research Methodology*, vol. 18, no. 1, pp. 1–12, 2018.
[13]. A. N. Angelopoulos and S. Bates, "A gentle introduction to conformal prediction and distribution-free uncertainty quantification," *Annual Review of Statistics and Its Application*, vol. 9, pp. 379–405, 2022.
[14]. V. Vovk, A. Gammerman, and G. Shafer, *Algorithmic Learning in a Random World*, New York, NY, USA: Springer, 2005.
[15]. B. Van Calster et al., "Calibration: The Achilles heel of predictive analytics," *BMC Medicine*, vol. 17, no. 1, pp. 1–7, 2019.

DOI: 10.17148/IJARCCE.2026.15501

BIOGRAPHY

Mehdi Mostofi has academic and industrial experience in the areas of data analytics, machine learning, data visualization, and operations analytics. He earned his PhD degree in Mechanical Engineering and has extensive experience developing and teaching courses in predictive analytics, marketing analytics, operations analytics, data warehousing, advanced data visualization, statistics, and machine learning. Mehdi has taught across multiple Canadian colleges and universities and has supervised more than 400 students in applied analytics capstone projects. He also has several years of industry experience as a data scientist and analytics consultant, contributing to projects in the energy sector, automotive industry, and public-sector analytics. His work includes developing predictive models, conducting statistical analysis, and auditing large-scale Python and R codebases for enterprise clients. He holds professional certifications in Data Analytics and Supply Chain Management from the Six Sigma Global Institute, along with a Data Science Certificate from George Brown College and a Certificate in Adult Education from the Canadian College of Educators. He can be contacted at: mehdi.mostofi@unfc.ca.