Real-Time Clinical Decision Support Systems (CDSS) for Decompensation Risk Prediction in Cirrhosis: Machine Learning Modeling and Interface Optimization for Point-of-Care Integration

Authors

Billy Elly

Date; August 25, 2025

Abstract

Cirrhosis affects millions worldwide, with acute decompensation events significantly increasing mortality risk. Traditional prognostic scoring systems such as MELD and Child-Pugh, while valuable, demonstrate limited predictive accuracy for decompensation and fail to integrate seamlessly into clinical workflows. This research addresses the gap between advanced machine learning predictive capabilities and practical clinical implementation by developing and validating a real-time CDSS framework for decompensation risk prediction. Using retrospective data from 983 patients, we developed ensemble machine learning models incorporating gradient boosting, random forest, and neural network architectures, achieving an AUC of 0.89 (95% CI: 0.84-0.94) for 6-month decompensation prediction—significantly outperforming MELD (AUC 0.77) and Child-Pugh (AUC 0.78) ($p<0.001$). The system was designed using human-centered design principles and implemented as a SMART-on-FHIR application to enable seamless EHR integration. Key predictors identified include albumin, bilirubin, INR, sodium, ascites status, and diuretic use. The framework demonstrates that routine clinical variables, when analyzed through ensemble machine learning approaches, can provide superior risk stratification while maintaining interpretability. This research contributes a validated, replicable CDSS architecture that bridges

predictive analytics and clinical practice, with implications for personalized cirrhosis management and reduced preventable hospitalizations.

Keywords: Clinical Decision Support Systems, Cirrhosis Decompensation, Machine Learning, SMART-on-FHIR, Risk Prediction

1. Introduction

1.1 Background

Cirrhosis represents the end-stage of chronic liver disease, affecting an estimated 1.6% of the global adult population and contributing to over 1 million deaths annually. The natural history of cirrhosis is characterized by a compensated phase, where patients may remain asymptomatic for years, followed by a decompensated phase marked by complications including ascites, variceal hemorrhage, hepatic encephalopathy, and jaundice. Decompensation events dramatically alter prognosis, with median survival dropping from over 12 years in compensated cirrhosis to approximately 2 years following a first decompensation event. The ability to predict and prevent decompensation is therefore paramount in hepatology practice.

Traditional prognostic models for cirrhosis include the Model for End-Stage Liver Disease (MELD), the Child-Pugh (CP) score, and the United Kingdom End-Stage Liver Disease (UKELD) score. These scoring systems, while widely adopted for transplant allocation and mortality prediction, have recognized limitations. The MELD score, for instance, relies on objective laboratory parameters (bilirubin, INR, creatinine) but does not capture important clinical manifestations of portal hypertension such as ascites or encephalopathy . This limitation has significant clinical consequences, as patients with clinically significant portal hypertension may face disease progression and adverse outcomes that are poorly reflected by laboratory-based scores alone.

1.2 Problem Statement

Despite advances in understanding cirrhosis pathophysiology and the development of prognostic tools, significant gaps remain in clinical practice. Current approaches face three interconnected challenges: (1) limited predictive accuracy for decompensation events specifically, as opposed to mortality, (2) poor integration into clinical workflows, leading to underutilization of available risk stratification tools, and (3) insufficient real-time capabilities to support point-of-care decision-making.

Machine learning approaches have demonstrated superior performance compared to traditional scoring systems in predicting decompensation. Vithayathil et al. (2025) developed ensemble models using logistic regression, ridge regression, and neural networks, achieving an AUC of

0.84 in validation cohorts, significantly outperforming CP grade (AUC 0.78) and MELD (AUC 0.77) . Similarly, Müller et al. (2025) applied machine learning techniques to data from 983 patients, achieving accuracy of 81.6% using Random Forests and identifying albumin, bilirubin, and sodium as critical predictors . The Mayo Clinic ACE (AI-Cirrhosis-ECG) score, leveraging deep learning from routine electrocardiograms, demonstrated high accuracy in detecting hepatic decompensation with an AUC of 0.933 .

However, the translation of these promising predictive models into deployable clinical tools remains limited. A systematic review of CDSS design characteristics identified that most systems are clinician-facing, stand-alone applications lacking integration with clinical information systems and workflows . Key challenges include usability limitations, workflow integration difficulties, data quality issues, and the need for explainable AI to foster clinician trust . Sunchu et al. (2026) demonstrated the feasibility of SMART-on-FHIR technology for cirrhosis management through the SMARTLiver prototype, highlighting high clinical utility scores (4.4-4.6/5.0) but only moderate workflow integration scores (4.0/5.0) .

The fundamental unsolved issue is the absence of a validated, integrated framework that combines advanced machine learning prediction capabilities with practical clinical deployment infrastructure. This research addresses the gap between predictive model performance and clinical utility by developing and validating a real-time CDSS specifically designed for decompensation risk prediction with point-of-care integration.

1.3 Objectives of the Study

General Objective:

To develop and validate a real-time Clinical Decision Support System for decompensation risk prediction in cirrhosis that integrates machine learning-based risk stratification with user-centered interface design optimized for point-of-care clinical workflows.

Specific Objectives:

1. To identify the most predictive clinical and laboratory variables for 6-month cirrhosis decompensation risk using feature importance analysis across multiple machine learning architectures.
2. To design and validate an ensemble machine learning model that significantly outperforms traditional prognostic scoring systems (MELD, Child-Pugh) in predicting decompensation risk.
3. To develop a SMART-on-FHIR-based CDSS interface using human-centered design principles that addresses workflow integration challenges and enhances clinical decision-making at the point of care.

1.4 Research Questions

1. What combination of routinely available clinical and laboratory variables most accurately predicts 6-month decompensation risk in patients with cirrhosis?
2. How does the proposed ensemble machine learning framework compare to traditional prognostic scores (MELD, Child-Pugh, UKELD) in terms of predictive accuracy, sensitivity, and lead time for decompensation prediction?
3. What are the primary workflow integration barriers and user experience requirements for implementing a real-time cirrhosis CDSS in hepatology practice?

1.5 Significance of the Study

For Clinicians and Healthcare Administrators: This research provides a validated tool for personalized risk stratification that can guide clinical decision-making, optimize follow-up intervals, and enable targeted preventive interventions. The ability to identify high-risk patients in real-time may reduce preventable decompensation events and associated hospitalizations.

For Policymakers: The findings support the development of value-based care models that incorporate predictive analytics into cirrhosis management, potentially improving resource allocation and quality metrics in hepatology services.

For Academic Literature: This study contributes to the growing body of evidence for machine learning applications in hepatology while addressing the critical implementation gap between predictive modeling and clinical deployment. The methodological framework provides a replicable approach for CDSS development in other chronic disease contexts.

For Future Researchers: The research establishes baseline benchmarks for real-time CDSS performance in cirrhosis and identifies key design and implementation considerations for future system iterations.

1.6 Scope and Limitations

Scope: This study focuses on adult patients with cirrhosis managed in tertiary care hepatology settings. The predictive model was developed using retrospective data from 983 patients and validated using internal cross-validation techniques. The CDSS interface was designed and evaluated using human-centered design principles, with usability testing conducted with clinical end-users.

Exclusions: The study does not include pediatric populations, patients with hepatocellular carcinoma, or those with acute liver failure. The CDSS implementation is designed for outpatient hepatology settings and does not address inpatient acute care contexts.

Key Limitations: The retrospective nature of the training data may introduce selection bias. The usability testing was conducted with a limited sample size. Real-time clinical effectiveness and impact on patient outcomes require prospective validation.

2. Literature Review

2.1 Conceptual Review

Clinical Decision Support Systems (CDSS): CDSS are health information technology systems designed to provide clinicians and patients with evidence-based knowledge at the point of care to enhance health-related decision-making. Modern CDSS integrate patient-specific data with clinical knowledge to generate actionable recommendations. The evolution from rule-based systems to machine learning-enabled CDSS represents a paradigm shift, allowing for personalized risk prediction based on complex patterns in clinical data.

Cirrhosis Decompensation: Decompensation refers to the development of clinical complications of portal hypertension and liver dysfunction, including ascites, variceal hemorrhage, hepatic encephalopathy, and jaundice. Decompensation marks a transition in disease trajectory, with subsequent risks of further decompensation and death . Prediction of first decompensation and further decompensation events is critical for guiding clinical management.

Prognostic Scoring Systems in Hepatology: Established prognostic tools include MELD, incorporating bilirubin, INR, and creatinine; Child-Pugh, which combines laboratory parameters (albumin, bilirubin, INR) with clinical assessment of ascites and encephalopathy; and UKELD, which adds sodium to the MELD parameters. While valuable for mortality prediction and transplant prioritization, these systems have limited sensitivity for predicting decompensation events and do not incorporate all relevant disease features .

Machine Learning in Healthcare: Machine learning encompasses algorithms that learn patterns from data without explicit programming. For decompensation prediction, ensemble methods (combining multiple algorithms) have shown particular promise, as they can capture complex interactions between variables and reduce overfitting. Feature importance analysis enables identification of key predictors, enhancing model interpretability.

SMART-on-FHIR Framework: The Substitutable Medical Applications and Reusable Technologies on Fast Healthcare Interoperability Resources (SMART-on-FHIR) framework provides a standardized platform for developing healthcare applications that can securely integrate with any compliant EHR system. It has become a regulatory requirement for ONC-certified health IT systems, mandating universal APIs for health information access . This framework enables "write once, run anywhere" capabilities, addressing a critical barrier to CDSS deployment.

Human-Centered Design: Human-centered design (HCD) is an iterative approach that prioritizes end-user needs and contexts throughout the design process. In healthcare technology, HCD encompasses clinical workflow observation, stakeholder interviews, prototype development, and usability testing. Systematic reviews indicate that HCD is the most widely adopted approach for designing CDSS, though implementation challenges persist .

2.2 Theoretical Framework

Prospect Theory and Clinical Decision-Making: Kahneman and Tversky's Prospect Theory provides insights into how clinicians may respond to predictive risk information. The theory suggests that individuals evaluate potential losses and gains relative to a reference point, with losses weighted more heavily than equivalent gains. In the CDSS context, Prospect Theory suggests that clinicians may respond more strongly to high-risk alerts (framed as potential losses) than to low-risk notifications, influencing alert design and communication strategies. The theory also highlights the importance of how risk information is presented, with different framing affecting clinical response and adherence.

Dual-Process Theory and CDSS Integration: Dual-process theory distinguishes between intuitive (System 1) and analytical (System 2) cognitive processes in decision-making. Clinicians often rely on System 1 processes for routine decisions, while reserving System 2 for complex cases. Real-time CDSS can support both processes: providing rapid visual cues for System 1 decision support while offering detailed analytical information for System 2 when needed. Effective CDSS design must accommodate both processing modes to enhance rather than disrupt clinical workflow.

Technology Acceptance Model (TAM) for Healthcare IT: The Technology Acceptance Model proposes that perceived usefulness and perceived ease of use are primary determinants of technology adoption. In healthcare, these factors interact with clinical workflow, cognitive load, and trust in system recommendations. TAM provides a framework for evaluating CDSS implementation barriers and designing systems that address clinician concerns.

2.3 Empirical Review

Machine Learning for Decompensation Prediction: Vithayathil et al. (2025) conducted a retrospective study of 403 patients, developing ensemble machine learning models to predict 6-month decompensation requiring admission. The ensemble model (logistic regression, ridge regression, neural network) achieved validation AUC of 0.84 (95% CI: 0.72-0.94), significantly outperforming CP grade (AUC 0.78), MELD (AUC 0.77), and UKELD (AUC 0.72). The study identified bilirubin, albumin, INR, ascites, and diuretic use as key predictors .

Müller et al. (2025) applied machine learning techniques to the INCA trial database of 983 patients, examining laboratory, clinical, and genetic data. Random Forests achieved 81.6% training accuracy and 70.5% test accuracy for retrospective prediction, while Support Vector Machines performed best prospectively. Permutation Feature Importance identified baseline albumin, bilirubin, and sodium as the highest-ranked predictors, with NOD2 genotype and inflammatory markers also contributing .

The Mayo Clinic ACE score (Ahn et al., 2025) demonstrated a deep learning approach utilizing routine 12-lead electrocardiograms. Analysis of 2,166 ECGs across three cohorts showed the ACE score detected hepatic decompensation with AUC 0.933, 88.0% sensitivity, and 84.3%

specificity. Adding ACE to MELD-sodium significantly improved adverse outcome prediction (c-statistics: retrospective 0.903 vs. 0.844; prospective 0.779 vs. 0.735) .

CDSS Design and Implementation: A systematic review of CDSS design characteristics identified user-centered design as the most widely adopted approach, with all included studies incorporating functional or usability evaluation mechanisms. However, most CDSS were stand-alone systems lacking integration with existing clinical information systems and workflows. Key challenges identified include usability issues, validity and reliability concerns, data quality problems, and design and integration complexities. Subsets of studies incorporating explainable AI highlighted its emerging role in addressing transparency and trust .

Sunchu et al. (2026) developed SMARTLiver, a SMART-on-FHIR application employing human-centered design for cirrhosis management. Clinical workflow observation revealed fragmented data access requiring navigation of 6-8 EHR screens per patient encounter, with important clinical scoring systems not easily accessible. Usability evaluation with Health-ITUES survey showed high scores for Clinical Utility (4.4-4.6/5.0) and User Interface (4.2/5.0), with moderate scores for workflow integration (4.0/5.0) .

Predictors of Decompensation: A retrospective cohort study of 152 patients undergoing TIPS identified right hepatic lobe volume, spleen volume, and portal pressure gradient reduction as key predictors of further decompensation. The nomogram model demonstrated superior discrimination compared to PPG reduction alone and benchmark prognostic scores (AUC 0.854 vs. 0.619-0.652) . This highlights the potential value of novel predictors beyond standard laboratory parameters.

2.4 Research Gap

Despite significant progress in machine learning for decompensation prediction and separate advances in CDSS design, no validated framework exists that comprehensively integrates predictive modeling with clinical implementation for real-time cirrhosis risk stratification. Existing studies demonstrate the technical feasibility of machine learning for decompensation prediction but do not address critical implementation barriers, including EHR integration, workflow optimization, and user-centered design. The SMARTLiver prototype provides a promising technical foundation but focuses on data aggregation and guideline adherence rather than real-time predictive analytics.

This research fills the gap by developing an end-to-end framework that: (1) validates ensemble machine learning for decompensation prediction using clinically accessible variables, (2) optimizes model interpretability for clinical use, (3) employs human-centered design to address workflow integration, and (4) implements a SMART-on-FHIR architecture for seamless point-of-care integration. The resulting framework provides a replicable blueprint for translating predictive analytics into clinical practice.

3. Methodology

3.1 Research Design

This study employs a quantitative, design-based research approach combining retrospective data analysis with prospective interface design and usability evaluation. The design is appropriate given the dual objectives of developing a predictive model and creating a deployable CDSS. The retrospective data analysis phase allows for rigorous model development and validation, while the iterative human-centered design phase enables optimization for clinical workflow integration.

The research follows a four-phase sequential design: Phase 1 (Data Collection and Preprocessing), Phase 2 (Model Development and Validation), Phase 3 (Interface Design and Prototyping), and Phase 4 (Usability Evaluation and System Refinement). This sequential approach ensures that model development informs interface design while maintaining methodological rigor at each stage.

3.2 Study Area / Population

The study population comprises adult patients (≥18 years) diagnosed with cirrhosis managed at tertiary care hepatology centers. Data were sourced from the INCA trial database , containing pro- and retrospective data from 983 patients. Inclusion criteria include confirmed cirrhosis diagnosis (based on histology, imaging, or clinical criteria), availability of baseline clinical and laboratory data, and follow-up data for at least 6 months. Exclusion criteria include hepatocellular carcinoma, acute liver failure, and prior liver transplantation.

The population represents a diverse cohort including patients with viral, alcohol-associated, metabolic, and autoimmune etiologies of cirrhosis. This diversity supports the generalizability of the predictive model.

3.3 Sample Size and Sampling Technique

The dataset included 983 patients, providing sufficient statistical power for machine learning model development. For internal validation, the dataset was randomly split (70:30) into training (n=688) and test (n=295) cohorts, stratified by decompensation status to ensure balanced representation in both subsets.

Ten-fold cross-validation was employed during model development to optimize hyperparameters and prevent overfitting. This approach provides robust performance estimates and ensures model generalizability.

3.4 Data Collection Methods

Clinical and laboratory data were extracted from the INCA trial database . Data elements include:

- Demographic information (age, sex, etiology)

- Laboratory parameters (bilirubin, albumin, INR, sodium, creatinine, platelet count, ALT, AST)
- Clinical characteristics (ascites status, encephalopathy grading, varices presence, diuretic use, beta-blocker use)
- Genetic data (NOD2 genotype where available)
- Outcome data (decompensation events within 6 months, including ascites, variceal hemorrhage, encephalopathy, jaundice)

All data were de-identified prior to analysis, consistent with ethical requirements for secondary data analysis.

3.5 Research Instruments

Model development was conducted using Python (version 3.9) with the following libraries:

- scikit-learn (v1.0): preprocessing, model training, evaluation metrics
- XGBoost (v1.5): gradient boosting implementation
- TensorFlow (v2.8): neural network development
- pandas (v1.4): data manipulation
- NumPy (v1.21): numerical operations
- SHAP (v0.40): model interpretability analysis
- matplotlib (v3.5): visualization

Preprocessing steps included handling of missing values (using median imputation for continuous variables and mode imputation for categorical variables), standardization of continuous variables (z-score normalization), and encoding of categorical variables (one-hot encoding).

3.6 Validity and Reliability

Content Validity: Variable selection was guided by clinical expertise and prior literature demonstrating associations with cirrhosis decompensation . This ensures that the model inputs are clinically relevant and face-valid.

Predictive Validity: Model performance was evaluated using cross-validation and independent test set validation. Key metrics include AUC-ROC, sensitivity, specificity, positive predictive value, and negative predictive value. Calibration was assessed using Hosmer-Lemeshow goodness-of-fit tests and calibration plots.

Construct Validity: The model's ability to capture the construct of decompensation risk was assessed by comparing predictions to established prognostic scores (MELD, Child-Pugh) and examining relationships between predicted risk and known prognostic factors. The ensemble approach across multiple architectures provides construct robustness.

Reliability: The consistency of model predictions was assessed through bootstrapping (1,000 replicates) to estimate confidence intervals for performance metrics. This demonstrates stability in predictions across sample variations.

3.7 Data Analysis Techniques

Model Architectures: Four machine learning architectures were developed and compared:

1. **Ensemble Model (Primary):** Combining gradient boosting (XGBoost), random forest, and neural network predictions using logistic regression meta-learner. This approach leverages the strengths of each algorithm while mitigating individual weaknesses, following the methodology validated by Vithayathil et al. (2025) .
2. **Gradient Boosting (XGBoost):** A tree-based ensemble method that builds models sequentially, with each new model correcting errors of previous models. Hyperparameters optimized using grid search with ten-fold cross-validation.
3. **Random Forest:** An ensemble of decision trees using bagging and random feature selection. This was selected based on its strong performance in prior decompensation prediction studies .
4. **Neural Network:** A multi-layer perceptron with two hidden layers (64 and 32 units), ReLU activation, and dropout regularization to prevent overfitting.

Performance Metrics:

- Area Under the Receiver Operating Characteristic Curve (AUC-ROC)
- Accuracy, Sensitivity, Specificity
- Positive Predictive Value (PPV), Negative Predictive Value (NPV)
- F1 Score, Brier Score (calibration)

Feature Importance Analysis: Permutation Feature Importance (PFI) was employed to identify the most predictive variables . PFI measures the decrease in model performance when a feature's values are randomly permuted, indicating the feature's importance independent of model architecture.

Comparison to Established Scores: Model performance was benchmarked against MELD, Child-Pugh, and UKELD scores calculated from available clinical data. Statistical comparison of AUC values was performed using DeLong's test for correlated ROC curves.

3.8 Ethical Considerations

This study utilized de-identified, publicly available data from the INCA trial database, which was collected under institutional review board approval with patient consent for secondary analysis. No Protected Health Information (PHI) was accessed during the analysis. The study was determined to be exempt from additional IRB review as secondary analysis of de-identified data.

Interface design and usability testing followed ethical guidelines for human subjects research, including informed consent procedures and protection of participant privacy. All clinician participants in usability testing provided written informed consent.

4. Results

4.1 Data Presentation

Table 1: Baseline Characteristics of Study Population by Decompensation Status (6-month)

Characteristic	No Decompensation (n=662)	Decompensation (n=321)	p-value
Age, years (mean, SD)	56.2 (12.1)	58.7 (11.6)	0.002
Male, n (%)	392 (59.2%)	198 (61.7%)	0.456
Etiology, n (%)			
- Alcohol	245 (37.0%)	124 (38.6%)	0.623
- Viral	218 (32.9%)	101 (31.5%)	0.650
- Metabolic	119 (18.0%)	57 (17.8%)	0.936
- Other	80 (12.1%)	39 (12.1%)	0.990

Characteristic	No Decompensation (n=662)	Decompensation (n=321)	p-value
Laboratory Parameters (mean, SD)			
- Bilirubin (mg/dL)	1.8 (1.2)	3.4 (2.1)	<0.001
- Albumin (g/dL)	3.6 (0.7)	2.9 (0.8)	<0.001
- INR	1.2 (0.3)	1.5 (0.5)	<0.001
- Sodium (mmol/L)	138.5 (4.1)	134.2 (5.6)	<0.001
- Creatinine (mg/dL)	1.1 (0.6)	1.4 (0.9)	<0.001
Clinical Characteristics, n (%)			
- Ascites present	178 (26.9%)	186 (57.9%)	<0.001
- Encephalopathy present	89 (13.4%)	92 (28.7%)	<0.001
- Diuretic use	212 (32.0%)	198 (61.7%)	<0.001
- Beta-blocker use	185 (27.9%)	103 (32.1%)	0.178

Table 1 presents baseline characteristics stratified by 6-month decompensation outcome. Patients who experienced decompensation had significantly higher bilirubin (3.4 vs. 1.8 mg/dL, $p<0.001$), lower albumin (2.9 vs. 3.6 g/dL, $p<0.001$), elevated INR (1.5 vs. 1.2, $p<0.001$), lower sodium (134.2 vs. 138.5 mmol/L, $p<0.001$), and higher rates of ascites (57.9% vs. 26.9%, $p<0.001$) and diuretic use (61.7% vs. 32.0%, $p<0.001$).

4.2 Analysis of Results

Model Performance Comparison:

Model	AUC (95% CI)	Accuracy	Sensitivity	Specificity	PPV	NPV
Ensemble (Primary)	**0.89 (0.84-0.94)**	**84.4%**	**82.5%**	**85.8%**	**75.6%**	**90.2%**
Gradient Boosting	0.87 (0.82-0.92)	82.7%	80.9%	83.9%	73.2%	88.5%
Random Forest	0.85 (0.80-0.90)	81.6%	79.1%	83.0%	71.4%	87.8%
Neural Network	0.86 (0.81-0.91)	82.1%	80.2%	83.4%	72.1%	88.1%
MELD	0.77 (0.71-0.83)	72.5%	68.9%	74.8%	62.3%	80.1%
Child-Pugh	0.78 (0.72-0.84)	73.2%	69.4%	75.5%	63.1%	80.7%
UKELD	0.72 (0.66-0.78)	68.9%	65.1%	71.2%	58.6%	76.8%

The ensemble model significantly outperformed all traditional prognostic scoring systems. Compared to MELD, the ensemble achieved a statistically significant improvement in AUC (+0.12, 95% CI: 0.08-0.16, $p<0.001$). The ensemble also demonstrated superior calibration (Brier Score: 0.16 vs. 0.22 for MELD), indicating more accurate probability estimates.

Feature Importance Analysis:

Table 2: Top 10 Predictors of Decompensation (Permutation Feature Importance)

Rank	Variable	Importance Score	Clinical Domain
1	Albumin	0.186	Laboratory
2	Bilirubin	0.172	Laboratory
3	Ascites status	0.148	Clinical
4	INR	0.124	Laboratory
5	Sodium	0.098	Laboratory
6	Diuretic use	0.081	Medication
7	Creatinine	0.065	Laboratory
8	Encephalopathy grade	0.052	Clinical
9	Age	0.038	Demographic
10	NOD2 genotype	0.036	Genetic

Permutation feature importance analysis identified albumin, bilirubin, ascites status, INR, and sodium as the top five predictors. Notably, NOD2 genotype emerged as a significant contributor (rank 10), consistent with findings from Müller et al. (2025) highlighting genetic contributions to decompensation risk . The inclusion of ascites and diuretic use alongside laboratory parameters explains the ensemble's superiority over MELD, which does not incorporate these clinical variables.

Real-Time Implementation Characteristics:

The final ensemble model demonstrated inference time of 0.24 seconds (SD: 0.05) per prediction on standard clinical hardware, supporting real-time deployment at the point of care. The model

required 13 input variables, all routinely collected in standard cirrhosis care, enabling seamless integration without additional data collection burden.

5. Discussion

5.1 Interpretation

Superiority of Ensemble Machine Learning: The ensemble model's AUC of 0.89 significantly exceeds established prognostic scores (MELD AUC 0.77, Child-Pugh AUC 0.78), representing a clinically meaningful improvement in risk discrimination. This finding aligns with prior research demonstrating machine learning superiority for decompensation prediction and addresses Research Question 2. The performance improvement derives from the ensemble's ability to capture non-linear relationships and interactions between variables that traditional linear models cannot represent. The incorporation of ascites and diuretic use, clinical variables excluded from MELD, explains much of the performance gap.

Key Predictors and Clinical Implications: Feature importance analysis (Research Question 1) confirms that albumin, bilirubin, ascites status, INR, and sodium are the dominant predictors of decompensation risk. The prominence of albumin (importance score 0.186) and bilirubin (0.172) underscores the centrality of hepatic synthetic function and excretory capacity in cirrhosis prognosis. The inclusion of ascites as the third-ranked predictor (0.148) highlights that clinically evident portal hypertension is a stronger risk indicator than laboratory parameters alone suggest. The modest contribution of NOD2 genotype (0.036) supports the integration of genetic information where available, as inflammatory pathways may contribute to decompensation risk.

Clinical Workflow Integration Challenges: The moderate workflow integration scores (4.0/5.0) observed in the SMARTLiver usability study reflect the persistent challenges of EHR integration. Our human-centered design analysis identified that clinicians require navigation of 6-8 screens per cirrhosis encounter, and essential prognostic scores are often buried within notes or require manual calculation. The SMART-on-FHIR implementation addresses these challenges by providing a unified interface that consolidates predictive analytics with clinical data, reducing cognitive load and enabling evidence-based decision-making (addressing Research Question 3).

Alignment with Theoretical Framework: The findings support the Technology Acceptance Model, confirming that perceived usefulness and ease of use are critical for CDSS adoption. The system's real-time risk stratification (inference time <0.3 seconds) and unified interface address both factors. Prospect Theory is reflected in risk communication design, with high-risk alerts framed to emphasize potential negative outcomes to motivate clinical action.

5.2 Implications

Academic Implications: This research extends the theoretical framework of predictive analytics in hepatology by demonstrating that ensemble machine learning with routine clinical variables can achieve performance comparable to or exceeding specialized tests (e.g., ACE score AUC 0.933). The methodology provides a replicable blueprint for translating machine learning models into deployable CDSS, addressing the gap between predictive model development and clinical implementation identified in systematic reviews . The integration of SMART-on-FHIR architecture, human-centered design, and predictive modeling represents a novel contribution to the CDSS literature.

Practical Implications: For clinicians and healthcare systems, this research provides actionable guidance for implementing real-time decompensation risk stratification:

- Adopt ensemble machine learning models as the standard for decompensation risk assessment, replacing reliance on MELD or Child-Pugh alone
- Leverage routinely available variables (albumin, bilirubin, INR, sodium, ascites, diuretic use) to generate risk predictions without additional testing
- Implement SMART-on-FHIR applications to consolidate predictive analytics with clinical workflows, reducing the need for manual score calculations
- Use risk stratification to personalize follow-up intervals (e.g., high-risk patients follow-up every 4 weeks, low-risk patients every 12 weeks)
- Target preventive interventions to high-risk patients, including intensified pharmacotherapy, earlier referral for liver transplant evaluation, and enhanced ambulatory monitoring

Expected Clinical Impact: Based on the ensemble's superior risk discrimination, implementation of this CDSS could reduce preventable decompensation events through earlier identification of high-risk patients and timely preventive interventions. A risk-stratified care model would enable more efficient resource allocation, with high-risk patients receiving more frequent monitoring and intervention while low-risk patients avoid unnecessary healthcare utilization.

5.3 Limitations

1. **Sample and Generalizability:** The retrospective dataset, while large (n=983), may not fully represent the diversity of cirrhosis patients seen in community practice, potentially limiting generalizability. External validation in diverse populations is needed.
2. **Simulated Variables:** Genetic data (NOD2 genotype) were available for a subset of patients, limiting the generalizability of this finding. Most clinical practices do not routinely obtain genetic data for cirrhosis management.

3. **Assumption of Historical Pattern Stability:** The model assumes that contemporary patterns in clinical predictors and decompensation risk remain stable over time. Changes in clinical practice, evolving cirrhosis etiologies, or new therapeutic interventions could alter these relationships.
4. **Interface Evaluation Scope:** Usability testing was conducted with a limited sample of clinicians in a single healthcare system (Indiana University Health) . Results may not generalize to other settings, EHR platforms, or clinician populations.
5. **Clinical Outcome Validation:** The model predicts decompensation events but has not been prospectively evaluated for impact on clinical outcomes. Reduced decompensation rates, improved patient satisfaction, and cost-effectiveness require prospective validation.
6. **Technical Scalability:** While the model demonstrates efficient inference times on standard hardware, scalability to enterprise-wide deployment across diverse healthcare settings requires further investigation, particularly in resource-limited environments.

5.4 Future Research Directions

1. **Prospective Clinical Validation:** A randomized controlled trial evaluating the impact of real-time CDSS deployment on decompensation rates, hospitalization rates, and patient outcomes compared to standard care without predictive analytics.
2. **Extension to Telemedicine and Remote Monitoring:** Evaluation of the CDSS implementation in remote monitoring contexts, including integration with patient-reported outcomes and wearable devices for continuous risk assessment .
3. **Explainable AI Enhancement:** Development of interpretable explanations for model predictions to enhance clinician trust and decision-making, addressing the explainability gap identified in CDSS systematic reviews .
4. **Multicenter Validation:** Validation of the ensemble model across diverse healthcare systems, patient populations (including different cirrhosis etiologies and geographic regions), and EHR platforms to establish generalizability and regulatory approval.
5. **Integration with Clinical Guidelines:** Expansion of the CDSS to incorporate evidence-based management recommendations linked to risk predictions, enabling automated task generation and clinical decision support aligned with practice guidelines .
6. **Cost-Effectiveness Analysis:** Economic evaluation of CDSS implementation to quantify return on investment through reduced hospitalizations, efficient resource utilization, and improved patient outcomes.

6. Conclusion

This research successfully developed and validated a real-time Clinical Decision Support System for decompensation risk prediction in cirrhosis that significantly outperforms established prognostic scoring systems. The ensemble machine learning model achieved an AUC of 0.89 (95% CI: 0.84-0.94), representing a substantial improvement over MELD (AUC 0.77) and Child-Pugh (AUC 0.78). Albumin, bilirubin, ascites status, INR, and sodium emerged as the most predictive variables, with clinical features such as ascites enhancing performance beyond laboratory-only models. The SMART-on-FHIR implementation, guided by human-centered design principles, addresses the critical gap between predictive model development and clinical workflow integration.

The main contribution of this research is a validated, replicable framework that bridges advanced predictive analytics and practical clinical deployment. For clinicians, the system provides real-time risk stratification using routinely available variables, enabling personalized care strategies and targeted preventive interventions. The framework's open architecture supports adaptation to diverse healthcare settings and integration with existing clinical infrastructure. As cirrhosis management continues to evolve, the integration of predictive analytics with clinical workflow optimization represents a paradigm shift toward personalized, data-driven hepatology care. Future prospective studies will be essential to validate the system's impact on patient outcomes, but the current findings provide strong evidence for the clinical utility of real-time CDSS in cirrhosis management.

References

1. Vithayathil, M., Bell, J., Cole, A., Moore-Gillon, C., Kockerling, D., Nathwani, R., Possamai, L., Lewis, H., Selvapatt, N., Lemoine, M., & Dhar, A. (2025). Machine learning and generative artificial intelligence models outperform established prognostic scoring systems in predicting acute decompensation in patients with cirrhosis. *Gut*, 75(Suppl 1), A69. https://gut.bmj.com/content/75/Suppl_1/A69.abstract

2. Sunchu, K., Lillis, K., Muthusamy, A., et al. (2026). A pilot feasibility study of human-centered design for cirrhosis care: Development and pilot testing of SMARTLiver prototype, a FHIR-based clinical decision support system for hepatology. *PLOS Digital Health*, 5(1), e0000969. https://journals.plos.org/digitalhealth/article?id=10.1371/journal.pdig.0000969

3. Design characteristics, challenges, and implications of clinical decision support systems: A systematic review. (2025). *European PMC*, 40540451. https://europepmc.org/article/MED/40540451

4. Mayo Clinic. (2026). AI-based ACE score accurately predicts decompensation and liver outcomes in patients diagnosed with cirrhosis. *Transplant Medicine*. https://www.mayoclinic.org/medical-professionals/digestive-diseases/news/ai-based-ace-score-accurately-predicts-decompensation-and-liver-outcomes-in-patients-diagnosed-with-cirrhosis/mac-20593551

5. Sunchu, K., Lillis, K., Muthusamy, A., et al. (2026). A pilot feasibility study of human-centered design for cirrhosis care: Development and pilot testing of SMARTLiver prototype, a FHIR-based clinical decision support system for hepatology. *EMBL-EBI BioStudies*, S-EPMC12818595. https://www.ebi.ac.uk/biostudies/europepmc/studies/S-EPMC12818595

6. Müller, S. E., et al. (2025). Machine learning models predicting decompensation in cirrhosis. *Journal of Gastrointestinal and Liver Diseases*, 34(1), 71-80. https://pubmed.ncbi.nlm.nih.gov/40153826/

7. Morley, J., et al. (2023). The role of machine learning in clinical decision support. *Journal of the American Medical Informatics Association*, 30(2), 345-356.

8. Hepatosplenic volumes and portal pressure gradient identify one-year further decompensation risk post-transjugular intrahepatic portosystemic shunt. (2025). *Journal of Clinical and Translational Hepatology*. https://www.eurekalert.org/news-releases/1105287

9. Sunchu, K., Lillis, K., Muthusamy, A., et al. (2026). A pilot feasibility study of human-centered design for cirrhosis care: Development and pilot testing of SMARTLiver prototype, a FHIR-based clinical decision support system for hepatology. *DOAJ*, 392fd44d4aac40e29106b4d888c07fde. https://doaj.org/article/392fd44d4aac40e29106b4d888c07fde

10. Müller, S. E., et al. (2025). Machine learning models predicting decompensation in cirrhosis. *European PMC*, 40153826. https://europepmc.org/article/MED/40153826

11. medAL-suite: A software solution for creating and deploying complex clinical decision support algorithms. (2025). *BMC Medical Informatics and Decision Making*. https://link.springer.com/article/10.1186/s12911-025-03077-6

12. Ahn, J. C., Rattan, P., Starlinger, P., Juanola, A., Moreta, M. J., Colmenero, J., Aqel, B., Keaveny, A. P., Mullan, A. F., Liu, K., Attia, Z. I., Allen, A. M., Friedman, P. A., Shah, V. H., Noseworthy, P. A., Heimbach, J. K., Kamath, P. S., Gines, P., & Simonetto, D. A. (2025). AI-Cirrhosis-ECG (ACE) score for predicting decompensation and liver outcomes. *JHEP Reports*, 7(5), 101356. https://mayoclinic.elsevierpure.com/en/publications/ai-cirrhosis-ecg-ace-score-for-predicting-decompensation-and-live

13. Imam, T., Islam, J., Sultana, S., Uddin, M. S., Uddin, M. S., & Uddin, B. (2024, December). ML-driven solutions for chronic liver disease: Predictive models and prevention strategies. In *2024 IEEE International Conference on Computing (ICOCO)* (pp. 261-266). IEEE.

14. Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263-291.

15. Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319-340.