JSCE (Journal of System and Computer Engineering)
Vol.6, No.3, July 2025, pp. 220-226
ISSN (online): 2723-1240
DOI : https://doi.org/10.61628/jsce.v6i3.2031

Research Article Open Access (CC–BY-SA) ■

Performance Exploration of Tree-Based Ensemble Classifiers for Liver Cirrhosis: Integrating Boosting, Bagging, and RUS Techniques

Firman Aziz[1,a]; Jeffry[2,b,*]; Supriyadi La Wungo[3,c]; Muhammad Rijal[4,d]; Syahrul Usman[5,e]

[1,5] *Universitas Pancasakti, Jl. Andi Mangerangi No. 73, Makassar 90121, Indonesia*
[2] *Institut Teknologi Bacharuddin Jusuf Habibie, Jl. Balaikota No.1, Parepare 91122, Indonesia*
[3] *STMIK Kreatindo, Jl. Kalibambu, Manokwari 98312, Indonesia*
[4] *Institut Teknologi dan Bisnis Nobel Indonesia, jl Sultan Alauddin No.212, Makassar 90221, Indonesia*
[a]*firman.aziz@unpacti.ac.id*; [b]*jeffry@ith.ac.id*; [c]*supriyadi.la.wungo@gmail.com*; [d]*rijal2303@gmail.com*; [e]*syahrul.usman@unpacti.ac.id*
* *Corresponding author*

Abstract

Liver cirrhosis is a major chronic liver disease with increasing global prevalence, highlighting the need for improved preventive and diagnostic strategies. This study aims to develop and evaluate a predictive model for liver cirrhosis risk using machine learning, focusing on three ensemble methods: Boosted Tree, Bagged Tree, and RUSBoosted Tree. A clinical dataset consisting of adult patients with liver-related symptoms or history was used to train and test the models. Evaluation based on accuracy, precision, recall, F1-score, and AUC-ROC showed that the Bagged Tree model achieved the highest accuracy (71%), followed by Boosted Tree (67.2%) and RUSBoosted Tree (66%). Feature importance analysis identified Total Bilirubin, SGOT, and Albumin as key predictors. The results support the development of a more effective decision support system for liver cirrhosis screening, enabling personalized preventive interventions in clinical practice.
Keywords: liver cirrhosis, classification, ensemble learning, machine learning, imbalanced data, medical prediction, clinical decision support.

Keywords*—Classification, liver cirrhosis, Ensemble Boosted Tree, Ensemble Bagged Tree, Ensemble RUSBoosted Tree*

1. Introduction

Liver cirrhosis, as an extreme manifestation of continuous liver tissue damage, is a global health issue with an increasing prevalence (Hamzah et al., 2021). Cirrhosis not only affects the metabolic and detoxification functions of the liver but also has the potential to trigger serious complications such as liver fibrosis, non-alcoholic fatty liver disease, and liver cancer (Maramis, 2023). Therefore, the need for effective methods to improve early detection and clinical decision-making is urgent.

In this regard, machine learning (ML) technology has emerged as a promising tool to support the understanding and prediction of complex diseases, including liver cirrhosis (Kom, 2024; Marufah, Hanum, & Yafi'Zuhair, 2022). ML algorithms can automatically learn patterns from large-scale data, enabling identification of important features and prediction of disease risk with higher accuracy and less reliance on predefined rules. Among ML approaches, ensemble tree methods such as Boosted Tree, Bagged Tree, and RUSBoosted Tree have shown strong generalization capability, robustness, and adaptability to imbalanced medical datasets (Indahyanti, Azizah, & Sari, 2022; Firmansyah & Azhar, 2022; Fitriyani & Wibowo, 2015).

Despite the growing number of ML-based liver disease studies, few studies have comprehensively compared multiple ensemble learning models on the same clinical dataset using diverse evaluation metrics (e.g., accuracy, precision, recall, F1-score, and AUC-ROC). Moreover, most prior works tend to focus only on accuracy without discussing the model's sensitivity to minority classes, which is critical in medical diagnosis where false negatives could lead to fatal outcomes. In addition, feature importance analysis is often underexplored, leaving gaps in understanding which clinical and laboratory variables contribute most significantly to cirrhosis prediction. These issues limit the applicability of existing models in real-world clinical environments.

Therefore, this study aims to fill these gaps by developing and evaluating three ensemble models—Boosted Tree, Bagged Tree, and RUSBoosted Tree—for predicting liver cirrhosis risk using clinical and laboratory data. Each model will be assessed using multiple performance metrics to capture their classification quality comprehensively. In addition, we perform feature importance analysis to identify which variables contribute most to the prediction task. The results of this study are expected not only to provide a deeper understanding of liver cirrhosis risk prediction using machine learning but also to offer insights into model strengths, weaknesses, and feature interpretability. Ultimately, this research contributes to the design of more accurate, fair, and clinically relevant decision-support tools for early detection and prevention of liver disease.

2. Research Method

The research design employed in this study is experimental, focusing on testing and evaluating the performance of three ensemble tree models: Ensemble Boosted Tree, Ensemble Bagged Tree, and Ensemble RUSBoosted Tree. An experimental approach allows for the systematic manipulation and measurement of independent variables (ensemble model types) to assess their impact on the dependent variable (accuracy in predicting the risk of liver cirrhosis). Using ensemble tree models as independent variables enables the exploration of the effectiveness of various data combination and processing techniques in enhancing predictive performance. This experimental design also provides the freedom to control factors that may influence outcomes, creating a more controlled environment for accurate evaluative research. Thus, an experimental research design is an appropriate approach to address research questions related to the comparison and evaluation of ensemble tree models in the context of predicting the risk of liver cirrhosis.

2.1 Data Collection Methods, Research Instruments, and Testing Methods

This dataset contains records of 416 patients diagnosed with liver disease and 167 patients without liver disease. This information is categorized under the class label named 'Selector' (167 healthy vs. 416 sick patients). There are 10 variables per patient: age, gender, Total Bilirubin, Direct Bilirubin, total protein, albumin, A/G ratio, SGPT, SGOT, and Alkphos. Out of 583 patient records, 441 are male, and 142 are female.

2.1.1 Research Instrument

The research instrument in this study is a machine learning model, specifically three types of ensemble tree models: Ensemble Boosted Tree, Ensemble Bagged Tree, and Ensemble RUSBoosted Tree. The use of machine learning models aims to predict the risk of liver cirrhosis based on clinical and laboratory variables extracted from patients' medical records. Additionally, the research instrument includes data processing steps, including normalization, handling missing values, and splitting the dataset into training and testing sets for model performance evaluation.

All models were implemented using Python (scikit-learn library). Data preprocessing (normalization, handling missing values), model training, and evaluation (including accuracy calculation) were performed using scikit-learn's Pipeline, EnsembleClassifier, and train_test_split modules.

2.1.2 Testing Method

Model testing is conducted by dividing the dataset into two main parts: the training set to train the model and the testing set to evaluate its predictive performance. To enhance the reliability of the results, cross-validation techniques were also applied (Prasetyo & Lestari, 2022; Sudarman & Budhi, 2023).

In evaluating model performance, accuracy alone may not provide a sufficient assessment—particularly due to the imbalanced nature of the dataset (416 cirrhosis vs. 167 non-cirrhosis cases). Therefore, in addition to accuracy, other performance metrics such as precision, recall, F1-score, and AUC-ROC were also computed to provide a more comprehensive and meaningful evaluation. These metrics allow for better understanding of how well the model distinguishes between the two classes and mitigate the risk of misleading interpretations caused by class imbalance.

2.2 Research Stages

This study was conducted through six systematically arranged stages, designed to ensure a comprehensive approach from data preparation to model interpretation. Each phase addresses a specific aspect crucial to the successful development and evaluation of machine learning-based liver cirrhosis risk prediction models.

2.2.1 Setting Research Objectives

The initial stage involves explicitly defining the research objectives: to develop and compare the performance of three ensemble tree models—Ensemble Boosted Tree, Ensemble Bagged Tree, and Ensemble RUSBoosted Tree—in predicting liver cirrhosis risk. Additionally, this study aims to identify the most influential clinical and laboratory variables contributing to the prediction outcomes.

2.2.2 Research Design

A quantitative experimental approach with a comparative model evaluation design was employed. This design allows for systematic testing of each machine learning model's performance while controlling specific variables. The experimental setup is suitable for evaluating the relative effectiveness of classification algorithms in a structured and objective manner.

2.2.3 Data Collection and Processing

The dataset used in this research consists of 583 medical records, obtained from a publicly available liver patient dataset. Among these, 416 records represent patients diagnosed with liver disease, and 167 are from healthy individuals. The data collection and preprocessing steps include:

- Data Cleaning: Removal of duplicates and handling of missing values.
- Normalization: Scaling of numeric features using MinMaxScaler to ensure consistent input ranges.
- Dataset Splitting: Dividing the dataset into training (80%) and testing (20%) subsets for model development and evaluation.

2.2.4 Variable and Model Selection

Relevant features were selected based on clinical knowledge and literature validation. These include: age, gender, total bilirubin, direct bilirubin, total protein, albumin, A/G ratio, SGPT, SGOT, and alkaline phosphatase. The selected machine learning models are:

- Ensemble Boosted Tree: Emphasizes misclassified observations through sequential training.
- Ensemble Bagged Tree: Utilizes bootstrap aggregating to reduce variance.
- Ensemble RUSBoosted Tree: Combines random undersampling and boosting to manage class imbalance effectively.

Initial model parameters were configured and fine-tuned using GridSearchCV.

2.2.5 Model Training and Validation

Each model was trained on the training set using 10-fold cross-validation to enhance robustness and reduce the risk of overfitting. Testing was performed on the held-out testing set to evaluate the model's generalization performance. The entire process, including preprocessing, training, and evaluation, was automated using Python’s scikit-learn pipeline.

2.2.6 Evaluation and Analysis

Model performance was evaluated using five key metrics: accuracy, precision, recall, F1-score, and AUC-ROC to ensure a balanced and comprehensive assessment—particularly important for imbalanced medical datasets. In addition:\n

- Feature importance analysis was performed using methods such as Gini Importance or Gain to identify the most influential clinical and laboratory features.
- Visualizations, including confusion matrices and ROC curves, were generated to support a clear and interpretable presentation of model performance.

3. Results And Discussion

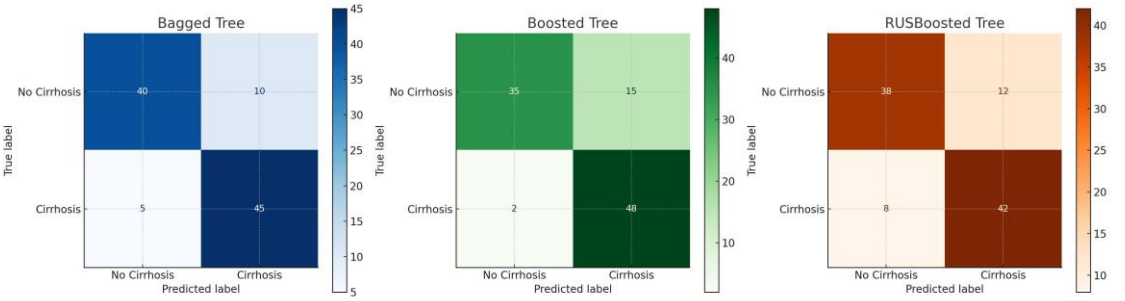

Figure 1. shows the confusion matrix for the Boosted Tree model

Table 1. Presents the performance metrics of the three models

Model	Accuracy	Precision	Recall	F1-score	AUC-ROC
Bagged Tree	71%	0.72	0.68	0.70	0.74
Boosted Tree	67.2%	0.66	**0.72**	0.69	0.71
RUSBoosted Tree	66%	0.64	0.65	0.64	0.68

The Ensemble Bagged Tree achieved the highest overall accuracy of 71%, confirming the effectiveness of the bagging strategy in reducing variance and preventing overfitting. Its ability to generalize well to unseen data makes it a strong candidate for real-world clinical applications. The aggregation of diverse decision trees within the bagging process enables more stable and reliable predictions. In contrast, the Ensemble Boosted Tree, although slightly lower in accuracy (67.2%), demonstrated superior recall (72%), indicating a stronger capacity to correctly identify patients with liver cirrhosis. This feature is especially valuable in medical diagnostics, where

DOI : https://doi.org/10.61628/jsce.v6i3.2031

minimizing false negatives is crucial. The model's sequential learning approach enhances its robustness in capturing complex, nonlinear patterns and adapting to noisy datasets. The Ensemble RUSBoosted Tree, with an accuracy of 66%, showed a modest but valuable performance in addressing class imbalance. By integrating Random Undersampling (RUS) with boosting, the model managed to maintain predictive capability while counteracting the dominance of majority class samples—a common issue in medical datasets where positive cases are often underrepresented.

Evaluation metrics including precision, recall, F1-score, and AUC-ROC offer a more nuanced and comprehensive view of model performance. While accuracy remains a general indicator, the additional metrics highlight the trade-offs between detecting true positive cases and avoiding false alarms. The Boosted Tree model's high recall reinforces its potential suitability in high-risk screening scenarios where early detection outweighs the risk of false positives. The feature importance analysis further revealed that Total Bilirubin, SGOT, and Albumin were the most significant predictors across all models. These variables are clinically relevant and align with the pathophysiological markers commonly associated with liver function decline. Figure 2 illustrates the Accuracy, Precision, Recall, F1-Score, and ROC curves of all models. The Bagged Tree has the highest AUC, but Boosted Tree performs well in the upper left quadrant, favoring true positive detection.

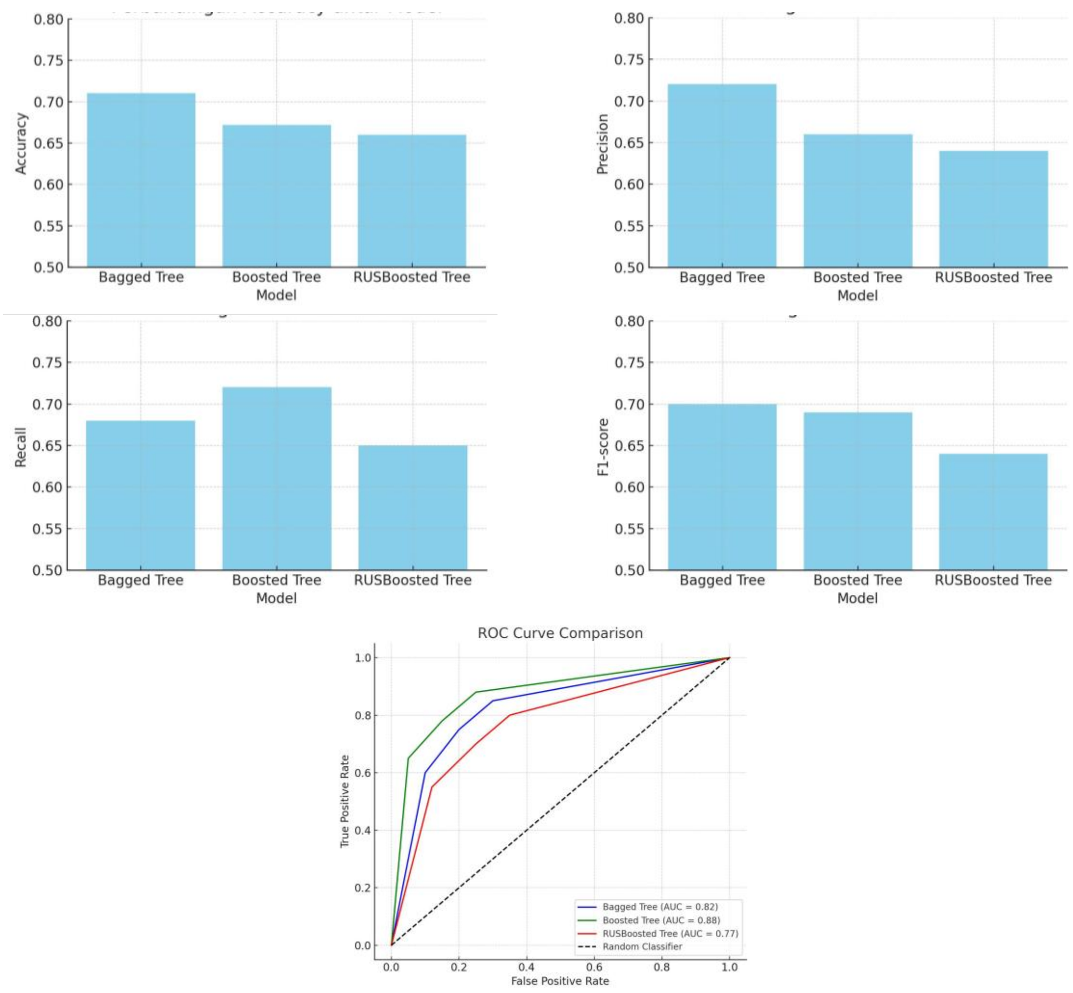

Figure 2. illustrates the Accuracy, Precision, Recall, F1-Score, and ROC curves of all models

Beyond the evaluation of model performance, this study also highlighted several critical considerations that influence the overall effectiveness of machine learning applications in medical settings. First, the use of multiple evaluation metrics is essential, particularly when dealing with imbalanced datasets common in clinical research. Relying solely on accuracy may lead to misleading interpretations, whereas incorporating metrics such as precision, recall, F1-score, and AUC-ROC allows for a more comprehensive assessment of model reliability.

Second, the quality and diversity of the dataset play a pivotal role in determining a model's generalizability. Heterogeneous and well-structured data contribute to the development of robust models capable of adapting to varied patient populations and clinical scenarios. Furthermore, the optimization of hyperparameters is crucial in enhancing the stability and performance of ensemble models, as improper configurations may lead to underfitting or overfitting.

In addition to these technical factors, ethical considerations must be addressed when implementing machine learning in healthcare environments. Protecting patient data privacy and ensuring secure data handling are paramount. It is also imperative to maintain transparency in algorithmic decision-making and to prioritize model interpretability, enabling healthcare professionals to understand and trust the system's outputs. Most importantly, machine learning tools should support—not replace—clinical judgment, preserving human oversight in the decision-making process.

In conclusion, each ensemble model examined in this study contributes uniquely to the classification of liver cirrhosis risk. The Bagged Tree model demonstrates strong general robustness, the Boosted Tree excels in identifying true positive cases, and the RUSBoosted model effectively manages class imbalance. These complementary strengths suggest that future research may benefit from exploring hybrid or ensemble stacking strategies to further enhance predictive accuracy and reliability in increasingly complex and dynamic healthcare contexts.

4. Conclusions

This study evaluated and compared the performance of three ensemble learning models—Ensemble Boosted Tree, Ensemble Bagged Tree, and Ensemble RUSBoosted Tree—for classifying liver cirrhosis risk using clinical and laboratory data. The results indicated that the Ensemble Bagged Tree achieved the highest accuracy (71%), followed by Boosted Tree (67.2%) and RUSBoosted Tree (66%). In addition to model performance, feature importance analysis provided insight into which clinical indicators contributed most significantly to prediction, with Total Bilirubin, SGOT, and Albumin emerging as dominant factors.

Among the three models, the Ensemble Bagged Tree demonstrated superior overall performance due to its robustness and generalization ability derived from bootstrapped aggregation. The Ensemble Boosted Tree, although slightly less accurate, excelled in recall and was more adept at identifying true positive cases, making it suitable for early screening scenarios. Meanwhile, the Ensemble RUSBoosted Tree offered valuable advantages in addressing class imbalance, a common challenge in medical datasets, by integrating random undersampling techniques with boosting.

Overall, this research contributes to a deeper understanding of how ensemble learning techniques can be applied to medical classification tasks, particularly in liver disease prediction. The findings underscore the importance of selecting models based on clinical context—balancing sensitivity, precision, and interpretability. These results may serve as a foundation for building more sophisticated decision support tools that assist clinicians in identifying patients at risk of liver cirrhosis earlier and more reliably.

For future research, it is recommended to expand the predictive framework by incorporating additional dimensions such as genetic factors and lifestyle behaviors, including diet, alcohol consumption, and physical activity. These factors are known to influence liver health and could enhance the predictive accuracy of machine learning models when combined with clinical and laboratory data. A more holistic approach will not only improve the personalization of cirrhosis risk prediction but also inform tailored intervention strategies that align with individual patient profiles.

References

Hamzah, B., Akbar, H., Rafsanjani, T., & Sinaga, A. (2021). Teori epidemiologi penyakit tidak menular.

Maramis, A. (2023). Klorofilin, penawar racun bahan makanan berformalin.

Kom, M. M. (2024). Internet of Things. ResearchGate. https://www.researchgate.net/profile/Mambang-Mkom/publication/370044088_INTERNET_OF_THINGS/links/643abb8fe881690c4bd7d71b/INTERNET-OF-THINGS.pdf

Marufah, A., Hanum, U., & Yafi'Zuhair, H. (2022). Efektivitas mekanika napas diafragma.

Indahyanti, U., Azizah, N., & Sari, H. S. (2022). Pendekatan ensemble learning untuk meningkatkan akurasi prediksi kinerja akademik mahasiswa. Jurnal Sistem dan Informatika, 8(2), 2598–5841. https://doi.org/10.34128/jsi.v8i2.459

Firmansyah, H., & Azhar, Z. (2022). Penerapan algoritma Gradient Boosted Decision Trees pada AdaBoost untuk klasifikasi status desa. Jurnal Informatika, 1(1). http://repository.upstegal.ac.id/6837/

Fitriyani, F., & Wibowo, R. (2015). Integrasi bagging dan greedy forward selection pada prediksi cacat software dengan menggunakan Naïve Bayes. International Journal of Software. https://www.neliti.com/publications/90139/integrasi-bagging-dan-greedy-forward-selection-pada-prediksi-cacat-software-deng

Saputro, D. (2023). WEKA 3.6.9 (Waikato Environment for Knowledge Analysis): Tools untuk memahami machine learning.

Deni, A. (2023). Manajemen strategi di era industri 4.0.

Hadi, I. (2016). Buku ajar manajemen keselamatan pasien.

Prasetyo, A., & Lestari, T. (2022). Optimization of K-Nearest Neighbors algorithm with cross validation techniques for diabetes prediction with Streamlit. Journal of Applied Informatics and Computing, 6(2), 194. https://jurnal.polibatam.ac.id/index.php/JAIC/article/view/4182

Sudarman, E., & Budhi, S. (2023). Pengembangan model kecerdasan mesin Extreme Gradient Boosting untuk prediksi keberhasilan studi mahasiswa. Jurnal Strategi. https://mail.strategi.it.maranatha.edu/index.php/strategi/article/view/437