



## Enhanced Predictive Modeling for Non-Invasive Liver Disease Diagnosis

Donni Prabowo<sup>1\*</sup>, Bety Wulan Sari<sup>2</sup>, Yoga Pristyanto<sup>3</sup>, Afrig Aminuddin<sup>4</sup>

<sup>1,2,3,4</sup>Department of Information System, Faculty of Computer Science, Universitas AMIKOM Yogyakarta, Yogyakarta, Indonesia

<sup>1</sup>donniprabowo@amikom.ac.id, <sup>2</sup>bety@amikom.ac.id, <sup>3</sup>yoga.pristyanto@amikom.ac.id, <sup>4</sup>afrig@amikom.ac.id

### Abstract

*Liver diseases (e.g. cirrhosis, hepatitis, and fatty liver disease) are globally one of the leading causes of mortality and are typically diagnosed in advanced stages due to vague symptoms and the difficulty involved in existing diagnostic techniques (e.g. biopsies). To optimize the early diagnosis of liver disease, this study proposes an enhanced, non-invasive approach using machine learning techniques. The research is enriched with a full pipeline, from exploratory data analysis and imputation of the dataset, treatment of the outlier, encoding of labels and scaling using ILPD (Indian Liver Patient Dataset). The classification models compared were RandomForest, XGBoost, LGBM, and CatBoost. The CatBoost algorithm fine-tuned with RandomizedSearchCV showed the highest performance with a test accuracy of 93%. The performance was again better than any already published methods showing that advanced ensembling and hyperparameter optimization worked. The proposed model is suitable for incorporation into clinical decision support systems and provides reliable and accurate diagnostic assistance. In addition to its high accuracy, the model is robust for missing and categorical data, which is a challenge in any real-world clinical scenario. These findings add to the growing body of evidence supporting AI-based medical diagnostics and suggest that CatBoost is a highly promising tool for facilitating timely screening and diagnosis of liver disease. Furthermore, the study stresses the need for thorough preprocessing and cross-validation, which serve to reduce biases that are present in widely applied datasets. Ongoing future efforts may involve the integration of multi-source data and implementation of explainable AI techniques to allow for wider clinical trust and use.*

**Keywords:** catboost; liver disease classification; machine learning; predictive modeling

*How to Cite:* D. Prabowo, Bety Wulan Sari, Yoga Pristyanto, and Afrig Aminuddin, "Enhanced Predictive Modeling for Non-Invasive Liver Disease Diagnosis", *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 9, no. 4, pp. 745 - 753, Aug. 2025.  
*Permalink/DOI:* <https://doi.org/10.29207/resti.v9i4.6449>

*Received:* March 6, 2025

*Accepted:* July 23, 2025

*Available Online:* August 13, 2025

*This is an open-access article under the CC BY 4.0 License  
Published by Ikatan Ahli Informatika Indonesia*

### 1. Introduction

Liver diseases, such as cirrhosis, hepatitis, and fatty liver disease, represent a major global health concern, accounting for millions of deaths annually. Early diagnosis and timely treatment are critical for preventing complications like liver failure and liver cancer [1], [2]. Therefore, the liver is very compensatory, and only 30% of the liver is sufficient to meet the body's normal functioning. Symptoms are extremely uncommon and typically do not indicate the initial phases of liver damage. In most patients, liver disease is diagnosed at an advanced stage [3].

However, conventional diagnostic methods like liver biopsies are costly, painful, and invasive to the patient. These challenges have resulted in increasing efforts toward creating non-invasive diagnostic tools to aid in early diagnosis and improve patient treatment. In the past few years, ML has received much attention as a

disruptive innovation in medical diagnosis. Machine learning (ML) is an emerging method for medical diagnosis that is quick, accurate, and automated disease classification. ML algorithms tend to perform well in analyzing big data, a specialty of human clinicians, but not to the same extent. ML algorithms are capable of helping healthcare systems improve the clinical diagnosis processes, strengthen the systems by decreasing the possibility of errors, and lessen the ferocity of procedures that are known to be invasive [4]–[6]. Smarter machine learning models can achieve superior liver disease diagnosis by simultaneously capturing and analyzing clinical exam and lab test findings [7]–[9]. Liver disease's clinical symptoms are usually used for diagnosing diseases while using machine learning algorithms to recognize the associated patterns of the disease in the dataset [10].

The use of machine learning with ILPD as it relates to connecting clinical features to liver disease indicators like hemoglobin, total bilirubin, SGOT, and SGPT is noteworthy. Previous research has pointed out weaknesses in the dataset's limitations such as the small sample size and suspected geographical bias, which undoubtedly paints a picture of a defined population group. These considerations may affect the applicability of ML algorithms trained with ILPD data, which emphasizes the need to validate such models against more diverse data for broader clinical endorsement. This study attends to these issues by employing effective cross-validation methodologies while proposing more representative and larger datasets for enhanced model performance [11].

However, although ML models offer important benefits, such as high levels of predictive accuracy and effectiveness in classifying large amounts of data, they are not without drawbacks, such as interpretability, possibility of bias, and the necessity for large training data sets. Counteracting these frailties is essential for establishing trustworthy and clinically relevant ML-based diagnostic tools.

Tanwar et. al. [12] discussed the comparison between ML-based and conventional non-ML techniques for liver disease diagnosis indicated that the traditional methods included, among others, liver function tests, imaging by ultrasonographic, CT, MRI elastographic, and biopsy, which were dependent on manual interpretation, time-consuming, and often required an expert involvement. The ML approaches automatically analyze the data, enhancing diagnostic accuracy and enabling earlier disease detection by identifying patterns in humongous datasets. The CNN model techniques exhibited increased performance potency in processing medical imaging and predicting disease pathways in specific cases. Additionally, ML enabled predictive analytics, in which doctors could assess patient risk and tailor treatment decisions accordingly. However, ML models did not have large-scale real-world validation, making them less reliable in the clinical milieu. Furthermore, their integration called for quality medical data, the computational structure to support wide-scale implementation, and tackling regulatory challenges before they can be entered into normal operations.

Minnor and Baths [13] evaluated the proficiency of many supervised ML algorithms in identifying liver disease using blood enzyme levels, such as Logistic Regression, etc. According to the study, the Extra Trees classifier was the most effective technique for prompt and precise diagnosis, with a maximum accuracy of 88.94% and an F1 score of 0.88. The study showed that using machine learning in the medical industry to supplement human experience and lower clinical diagnosis errors is feasible. However, the paper's limitations include a geographically biased dataset, which may not generalize well to diverse populations, and a relatively small sample size of 583 data points,

potentially affecting the robustness of the model training.

Ghazal et. al. [14] presented an advanced predictive model for the early detection of liver disease by leveraging ML methodologies. The researchers emphasized the critical importance of the timely identification of liver disease a worldwide major health issue. The model applied machine learning approaches to incorporating Artificial Neural Networks (ANN), which led to an estimated predictive accuracy of 88.4% and an error rate of 0.1164. The authors constructed and trained a model and tested it with a machine learning dataset from UCI Machine Learning Repository to verify the model's efficiency compared to the conventional liver disease diagnosis methods in terms of costs and time. The authors erroneously specified the tool that can be used as (ANN) and focused only on the lack of advantages of integrating multiple algorithms or hybrid models for better accuracies and robustness in liver disease prediction without examining the scenario from multiple facets.

Geetha and Arunachalam [15] evaluated methods for utilizing ML algorithms to predict liver disease. It aimed to enhance early detection of liver disease, which is essential for enhancing patient survival rates. The study compared various classification algorithms, including Linear Regression and SVM, to classify healthy individuals from liver datasets. According to the study, Logistic Regression had an accuracy of 73.23%, while Support Vector Machine (SVM) reached the maximum accuracy of 75.04%. The findings suggest that machine learning techniques hold substantial potential in facilitating the early diagnosis of liver diseases, thereby enhancing treatment effectiveness. However, the paper notes that the dataset used for training and testing was inadequate, leading to underperformance in diagnosing liver diseases.

Afreen et. al. [16] presented a ML system to categorize patients with liver illness using gradient boosting algorithms, particularly CatBoost and LightGBM, based on the Indian Liver Patient Dataset (ILPD). The data were subject to extensive pre-processing, comprising missing value imputation and outlier detection utilizing the Isolation Forest technique. Random Forest and Extra Trees techniques were used for feature selection to enhance model performance. Comparative analysis dealt with the performance evaluation of CatBoost and LightGBM through precision, accuracy, recall, and F1-score. The experiments exhibited that the CatBoost model attains a high classification accuracy of 86.8%, exceeding the performance of LightGBM and dominating prior approaches documented in the literature. Yet, the key focus of the paper was improving classification accuracy and did not address alternative datasets or techniques for balancing data and selecting features.

Nigatu et. al. [17] presented several supervised ML algorithms for liver disease prediction using the ILPD.

Eight algorithms were used such as Random Forest, Artificial Neural Network (ANN), etc. ANN had the best accuracy (87%) among all methods. Feature scaling was done on the data and hyperparameter tuning was done via grid search. The top-performing model can, therefore, be integrated into a decision support system to facilitate early detection of liver disease, especially in regions where there are limited resources. However, The study focused the majority centered on the Indian Liver Patient Dataset and did not investigate the performance gains that data like genetics or personalized treatment suggestions could make for the predictions offered.

Ghosh et. al. [18] presented seven predictive models composed of Logistic Regression alongside AdaBoost, etc to anticipate chronic liver disease. The data came from UCI Machine Learning Repository and contained 583 instances with ten attributes, and liver disease affected 71.36% of patient samples. The predictive algorithm that performed best among all tested systems was Random Forest due to its accuracy reaching 83.76%. The effectiveness of Random Forest exceeds other algorithms for detecting liver disease at early stages based on the provided dataset. The study used only a solitary dataset and omitted testing various feature engineering methods and multiple model fusion methods to enhance performance.

Hartatik et. al. [19] evaluated liver disease prediction using K-Nearest Neighbors (KNN) and Naïve Bayes algorithms. This study utilized the ILPD Indian Liver Patient Dataset, which included 583 instances between liver and non-liver patients. The research showed Naïve Bayes performed better than KNN, with AUC scores reaching 72.5% and 63.19%, respectively, for this particular classification task. The study showed limited data collection as its main weakness since it failed to investigate what advances would result from combining multiple algorithms or implementing sophisticated feature selection techniques.

Sadhasivam et. al. [20] used Support Vector Machine (SVM) algorithms to predict liver diseases. The authors employed SVM through Weka and Python tools for analyzing patient datasets, including liver disease markers. The team prepared the data by trimming unordinary values and eliminating repeated entries before constructing their model. Analysis indicated that SVM functions within Python surpassed the Weka software, resulting in an 87.09% accuracy rate, while Weka demonstrated 69.10% accuracy. The research investigation concentrated on a restricted collection of characteristics without evaluating how multiple classifying systems might aid in better results.

Existing studies on the prediction of liver disease prediction using ML techniques have focused on various models such as SVM, Naïve Bayes, Random Forest, and Neural Networks, demonstrating their potential for improving diagnosis accuracy. However, many of these works rely on small datasets,

predominantly the Indian Liver Patient Dataset, which limits their generalizability across diverse populations. Furthermore, while individual algorithms have shown promising results, integrating multiple models or applying hybrid approaches remains underexplored. Few studies have also addressed the challenges of imbalanced datasets, missing values, and the potential to incorporate advanced feature engineering or additional data sources like genetic information to improve model robustness.

This research makes several key contributions to the field of liver disease diagnosis and machine learning, including: (1). The development of a highly accurate machine learning model, using CatBoost, for liver disease diagnosis, achieving an accuracy of 91.47%. (2). Advanced hyperparameter tuning techniques, such as RandomizedSearchCV, were applied to enhance the performance of models like XGBoost and CatBoost. (3). A comprehensive evaluation of multiple machine learning models was conducted, with detailed comparisons across performance metrics. (4). The proposed model is designed for integration into clinical decision support systems, improving early liver disease detection and diagnosis.

This research proposed developing predictive models to distinguish liver disease patients from healthy individuals, utilizing algorithms like RandomForest, XGBoost, LGBM, and CatBoost. The goal is to recognize the most suitable model by evaluating accuracy, precision, recall, and F1-score, while also optimizing performance through hyperparameter tuning. This research supports the enhancement of ML in medical diagnostics, specifically for liver disease classification.

This paper is structured as follows: Section 1 outlines the motivation for applying machine learning in non-invasive liver disease diagnosis, relevant literature, comparing models and identifying research gaps. Section 2 describes the dataset, preprocessing, model selection, and hyperparameter tuning. Section 3 presents and analyzes results, highlighting the optimal model. Section 4 concludes with key findings and future research directions, such as using larger datasets and exploring hybrid models.

## 2. Methods

The method proposed in this research consists of eight stages: data acquisition, exploratory data analysis (EDA), data preprocessing, data splitting, model selection, hyperparameter tuning, model training, and model evaluation. The dataset used in this research is the Indian Liver Patient Dataset (ILPD) data. This dataset has 11 columns and 1700 rows. The research flowchart is presented in Fig. 1.

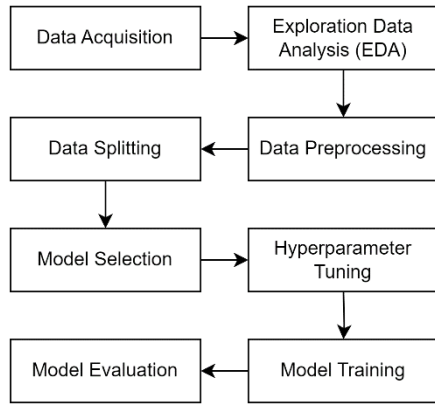


Figure 1. Flowchart of the research experiment

### 2.1 Data Acquisition

Due to its high popularity concerning liver disease classification and the availability of labeled clinical attributes, ILPD was chosen with high consideration for this study. In contrast to other liver disease datasets, ILPD is unique in providing a good balance of features, such as hemoglobin levels, total bilirubin, SGOT, and SGPT, all of which are paramount in the diagnosis of liver conditions. In addition, it has previously served as a benchmark dataset, allowing a direct comparison with existing research.

The dataset, which contains clinical data such as patient demographics and biochemical markers, is loaded using the `pd.read_csv()` function. The dataset  $D$  can be represented as Equation 1.

$$D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\} \quad (1)$$

where  $x_i$  denotes the feature vector associated with the  $i$ -th patient and  $y_i$  represents the corresponding label, indicating the presence  $y=1$  or absence  $y=0$  of liver disease.

### 2.2 Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is fundamental to comprehensively examining the underlying structure and distribution of the data [20].

EDA allows for identifying missing and inconsistent data that needs proper handling through methods like imputation or removal to maintain data quality for machine learning. A clean dataset appears for model building after conducting this process.

### 2.3 Data Preprocessing

Data preprocessing involves transforming categorical variables while applying scale techniques to numerical features to prepare the dataset for machine learning applications [21], [22]. The present study has used multiple imputation techniques to address missing values. Categorical variables, like gender, were imputed using mode imputation, while numerical variables like SGOT and SGPT levels used imputation by the median to avoid skewing of the data distribution. Outliers in biochemical markers are identified and treated by

applying the IQR method for their information to be consistent before training the models.

**Categorical Variable Encoding:** Categorical variables such as gender are encoded using `LabelEncoder()`. For a categorical variable with  $m$  distinct categories, the label encoding function assigns each category  $c_k$  a unique integer as shown in Equation 2.

$$\text{Encoded}(c_k) = k, k \in \{0, 1, \dots, m-1\} \quad (2)$$

**Feature Scaling:** Numerical features are scaled to a fixed range using the `MinMaxScaler`. The transformation for each feature  $x_j$  is defined as Equation 3.

$$x_j' = \frac{x_j - \min(x_j)}{\max(x_j) - \min(x_j)} \quad (3)$$

This ensures that all features lie within the range  $[0,1]$ , which improves the convergence of gradient-based algorithms.

### 2.4 Data Splitting

The generalization ability evaluation method divides the available dataset into two partitions where the training set contains 80% of data and the testing set consists of the remaining 20%. When splitting the data through the `train_test_split()` function the program reserves 80% for training purposes and dedicates 20% for testing purposes. Using `random_state=42` as a seed parameter enables reproducibility, and the testing set receives 20% of the data for performance testing evaluation. Model learning and parameter optimization occur using the training dataset, but the testing dataset functions to determine generalization capabilities and prediction outcomes from new data points [23]. The split is expressed as Equation 4.

$$\begin{aligned} X_{\text{train}}, X_{\text{test}}, y_{\text{train}}, y_{\text{test}} = \\ \text{train\_test\_split}(X, y, \text{test\_size} = \\ 0.2, \text{random\_state} = 42) \end{aligned} \quad (4)$$

### 2.5 Model Selection

ML models were chosen for this study due to their effectiveness in medical classification tasks, particularly in liver disease diagnosis. Gradient boosting algorithms like XGBoost, LGBM, and CatBoost are state-of-the-art techniques for handling tabular medical data, optimizing performance with limited data, and providing feature-importance insights. Previous studies, such as those by Ghazal et al. [14] and Afreen et al. [16], have shown that these models outperform traditional classifiers like Logistic Regression and SVM, making them suitable choices for this research.

The subsequent phase involves selecting a suitable ML model through cross-validation. This technique partitions the training dataset into  $k$  distinct subsets, wherein the model is trained on  $k-1$  subsets and examined on the unused subset. This procedure is iteratively performed  $k$  times, and the resulting

performance metrics are averaged for robust estimation. The cross-validation accuracy is mathematically expressed as Equation 5.

$$CV\ accuracy = \frac{1}{k} \sum_{i=1}^k accuracy_i \quad (5)$$

where  $k$  is the number of folds (in this case, 5), and  $accuracy_i$  represents the accuracy of the  $i$ -th fold.

### 2.6 Hyperparameter Tuning

Hyperparameter tuning is conducted through RandomizedSearchCV. The goal is to identify the optimal hyperparameters  $\theta$  that maximize model performance. Formally, the objective of hyperparameter tuning can be expressed as Equation 6.

$$\hat{\theta} = \arg \max_{\theta \in \Theta} \frac{1}{k} \sum_{i=1}^k accuracy_i(\theta) \quad (6)$$

where  $\theta$  is the hyperparameter space, and  $accuracy_i(\theta)$  is the accuracy of the  $i$ -th fold using hyperparameter configuration  $\theta$ . For the CatBoost model, the hyperparameters tuned include depth as the depth of the trees, iterations as the quantity of boosting iterations, and learning rate as the step size in gradient descent.

This research employed CatBoost because this model appropriate for medical classification problems due to its automated dealing with missing values as well as handling of categorical data. Moreover, with CatBoost, users are guaranteed high accuracy as well as prevention of overfitting.

To achieve improvements in the model's performance, hyperparameter optimization was performed using RandomizedSearchCV. This technique of optimization is better for finding the optimal solution because it randomly searches through a specified set of options. Due to having many parameters to test, it's faster and more efficient than GridSearchCV.

### 2.7 Model Training

After selecting the best hyperparameters, the model is trained on the full training set  $X_{train}, Y_{train}$ . The training process seeks to minimize a loss function  $L(y, \hat{y})$  where  $\hat{y}$  is the predicted value. The log-loss function is widely utilized for binary classification tasks shown in Equation 7.

$$L(y, \hat{y}) = - \frac{1}{n} \sum_{i=1}^n [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (7)$$

### 2.8 Model Evaluation

The efficacy of the trained model is rigorously evaluated on the test dataset through the application of multiple performance metrics to facilitate a comprehensive assessment. The primary assessment metrics include accuracy, precision, recall, and the F1-score. The F1-score, which measures the balance between precision and recall using their harmonic mean, is particularly useful in addressing class

imbalances, as it effectively accounts for disparities in the significance of false positive and false negative errors [24].

## 3. Results and Discussions

The efficacy of the trained model is rigorously evaluated on the test dataset through the application of multiple performance metrics to facilitate a comprehensive assessment. The primary assessment metrics include accuracy, precision, recall, and the F1-score. The F1-score, which measures the balance between precision and recall using their harmonic mean, is particularly useful in addressing class imbalances, as it effectively accounts for disparities in the significance of false positive and false negative errors [24].

### 3.1 Dataset and Preprocessing Results

The dataset, *Liver\_disease\_data.csv*, contains 11 columns with 1700 complete entries featuring demographic and health-related attributes such as Age, Gender, SegmentBMI, AlcoholConsumption, Smoking, GeneticRisk, PhysicalActivity, Diabetes, Hypertension, Diagnosis, and LiverFunctionTest. The analysis begins with exploratory data analysis (EDA), which includes calculating descriptive statistics, checking for duplicates, and creating visualizations to understand feature relationships. Data preprocessing encompasses transforming categorical features through LabelEncoder, normalizing numerical features using MinMaxScaler, and examining feature relationships via a heatmap analysis. The dataset is subsequently divided into training and testing subsets following an 80:20 proportion, with stratification for target variable proportions. Fig. 2 displays the histogram of liver disease patient distributions.

This research uses correlation heatmaps shown in Fig. 3 to visualize the relationships between different features in the dataset. In addition, it also provides Fig. 4 for outliers for numeric columns of some categories.

Fig. 3 and 4 provide insights into liver disease risk factors and numerical data distributions. The correlation heatmap shows weak correlations among most variables, with moderate correlations for smoking, genetic risk, and liver function tests. Boxplots indicate that alcohol consumption has many outliers, while physical activity has fewer, concentrated around the median. Conversely, liver function tests and age display symmetrical distributions with minimal extremes, highlighting the dataset's overall consistency.

### 3.2 Model Selection and Cross-Validation Result

The process of model selection and cross-validation follows several steps. First, a set of machine learning models is defined for evaluation, including RandomForest [25], XGBoost [26], LGBM [27], CatBoost [28], GradientBoosting [28], AdaBoost [28], ExtraTrees [29], MLP [29], Support Vector Classifier (SVC) [30], LogisticRegression [31], DecisionTree

[32], and Kneighbors [33]. Each model undergoes training and evaluation using 5-fold cross-validation, in which the dataset is divided into five subsets, with four used for training and one for validation in rotation. Performance is measured by accuracy, and the mean accuracy over all folds is computed for each model. The model achieving the highest mean accuracy, as presented in Table 1, is selected as the best performer.



Figure 2. Histogram for 10 distributions of liver disease patients : (a) Age, (b) Gender, (c) LiverFunctionTest, (d) Smoking, (e) GeneticRisk, (f) PhysicalActivity, (g) AlcoholConsumption, (h) Diabetes, (i) Hypertension, and (j) SegmentBMI

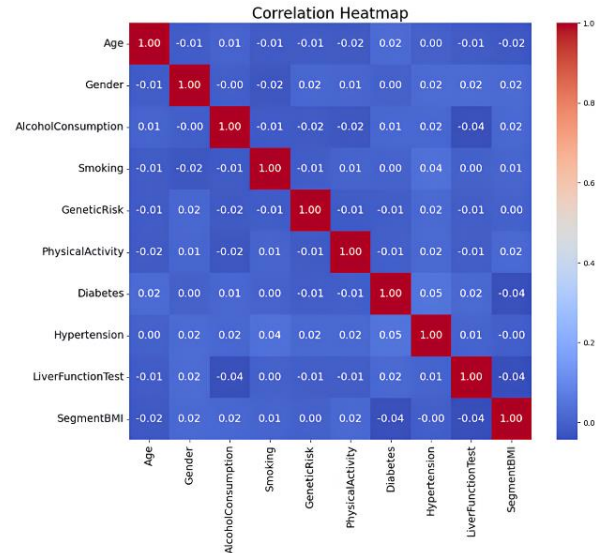


Figure 3. Correlation heatmap

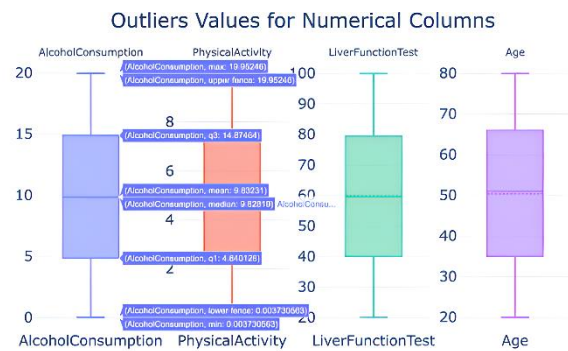


Figure 4. Outliers for numerical columns

Table 1. Comparison of cross-validation results

| Model              | Mean Accuracy |
|--------------------|---------------|
| RandomForest       | 0.8662        |
| XGBoost            | 0.8838        |
| LGBM               | 0.8890        |
| CatBoost           | 0.9125        |
| GradientBoosting   | 0.8941        |
| AdaBoost           | 0.8816        |
| ExtraTrees         | 0.8588        |
| MLP                | 0.8176        |
| SVC                | 0.8162        |
| LogisticRegression | 0.8081        |
| DecisionTree       | 0.7926        |
| KNeighbors         | 0.7765        |

Table 1 indicates that ensemble methods like CatBoost (91.25%), GradientBoosting (89.41%), and LGBM (88.90%) outperformed other models due to their ability to combine multiple weak learners, improving accuracy. In contrast, simpler models like KNeighbors and LogisticRegression showed lower performance, likely attributed to their restricted proficiency in modeling complex data patterns.

### 3.3 Hyperparameter Tuning

To optimize the hyperparameters of various machine learning models, RandomizedSearchCV defines parameter grids for each model, outlining potential hyperparameter ranges. This approach conducts a

randomized search to identify the best hyperparameters that maximize model accuracy through cross-validation. The CatBoost model emerged as the best performer, with optimal hyperparameters of a learning rate of 0.05, l2\_leaf\_reg of 3, 200 iterations, early\_stopping\_rounds of 10, a depth of 4, and a border\_count of 128. This process enhances the model's predictive performance for classification tasks. Table 2 presents a comparison of results before and after tuning.

Table 2 showcases how hyperparameter tuning affects the accuracy of the classification model. XGBoost saw the most significant improvement, increasing from 88.38% to 91.03%. MLP's accuracy rose slightly from 81.76% to 82.72%, while CatBoost improved marginally from 91.25% to 91.47%. Conversely, ExtraTrees decreased from 85.88% to 85.81%, and KNeighbors improved slightly from 77.65% to 80.59%. These results indicate that hyperparameter tuning

enhances model performance, but the degree of improvement varies by model, highlighting the need for tailored optimization strategies. Fig. 5 presents a comparison based on Table 2.

Table 2. Comparison of before and after tuning

| Model              | Score before tuning | Score after tuning |
|--------------------|---------------------|--------------------|
| RandomForest       | 0.8662              | 0.8735             |
| XGBoost            | 0.8838              | 0.9103             |
| LGBM               | 0.8890              | 0.8978             |
| CatBoost           | 0.9125              | 0.9147             |
| GradientBoosting   | 0.8941              | 0.8868             |
| AdaBoost           | 0.8816              | 0.8860             |
| ExtraTrees         | 0.8588              | 0.8581             |
| MLP                | 0.8176              | 0.8272             |
| SVC                | 0.8162              | 0.8265             |
| LogisticRegression | 0.8081              | 0.8103             |
| DecisionTree       | 0.7926              | 0.8140             |
| KNeighbors         | 0.7765              | 0.8059             |

Comparison of Mean CV Accuracy and Best CV Accuracy

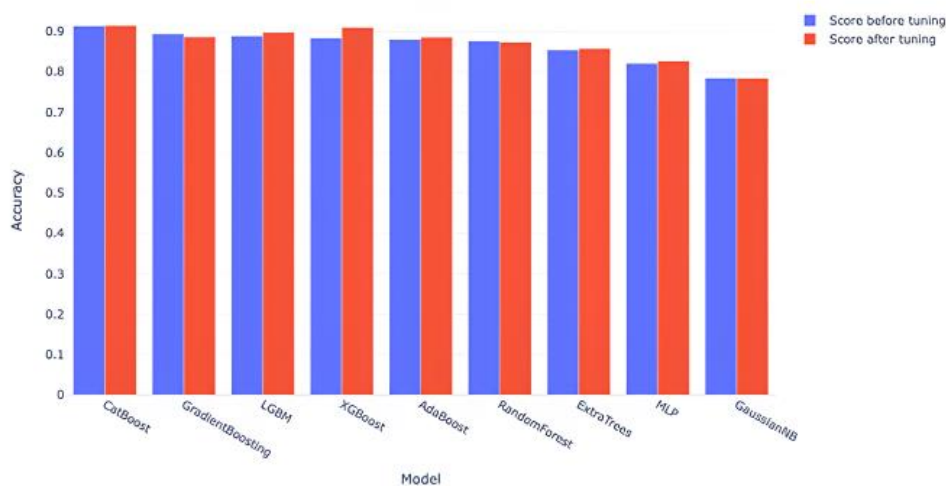


Figure 5. Comparison Bar

### 3.4 Model Evaluation on Test Set

This research uses a confusion matrix and classification report to provide a detailed breakdown of the model's performance on the test set. Fig. 6 shows the confusion matrix for model evaluation.

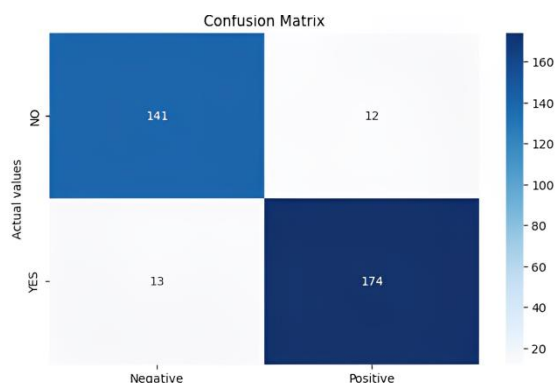


Figure 6. Confusion matrix

The classification report of Fig. 6 is shown in Table 3.

Table 3. Classification report

| Values    | Precision | Recall | F1-Score | Support |
|-----------|-----------|--------|----------|---------|
| Negative  | 0.92      | 0.92   | 0.92     | 153     |
| Positive  | 0.94      | 0.93   | 0.93     | 187     |
| Accuracy  |           |        | 0.93     | 340     |
| Macro avg | 0.93      | 0.93   | 0.93     | 340     |
| Weigh Avg | 0.93      | 0.93   | 0.93     | 340     |

Table 3 shows the demonstrates high effectiveness, attaining a training accuracy of 95% with precision and recall at 95% and 96%, respectively. On the test set, the accuracy slightly decreases to 93%, with precision, recall, and F1-score also showing slight reductions at 94%, 93%, and 93%, respectively. The classification report confirms these results, indicating consistent and effective performance with an overall accuracy of 93%, demonstrating the model's robustness and generalizability.

### 3.5 Comparative Analysis

Numerous ML algorithms have been applied to liver disease prediction, each with varying degrees of success. Models such as RandomForest, XGBoost, LGBM, and CatBoost have shown promise in accurately predicting liver disease. However, there is a need for a detailed comparative analysis to identify which models perform best and under what conditions. This analysis aims to evaluate the effectiveness of various ML models in terms of accuracy using the Liver Patient Dataset as a case study. Table 4 provides a comparative analysis of several supervised ML models for liver disease detection.

Table 4. Comparative analysis

| Author                      | Model                            | Accuracy |
|-----------------------------|----------------------------------|----------|
| Minnor and Baths [13]       | Extra Trees                      | 88.94%   |
| Ghazal et al. [14]          | Artificial Neural Networks (ANN) | 88.4%    |
| Geetha and Arunachalam [15] | Support Vector Machine (SVM)     | 75.04%   |
| Afreen et al. [16]          | CatBoost                         | 86.8%    |
| Nigatu et al. [17]          | ANN                              | 87%      |
| Ghosh et al. [18]           | Random Forest                    | 83.76%   |
| Hartatik et al. [19]        | Naïve Bayes                      | 72.5%    |
| Sadhasivam et al. [20]      | Support Vector Machine (SVM)     | 87.09%   |
| Proposed Method             | CatBoost with RandomizedsearchCV | 93.00%   |

Table 4 provides the performance of different models used in various studies for classification tasks alongside the proposed method employing CatBoost with RandomizedSearchCV. Each study's model is listed with its respective accuracy, providing insight into how different approaches fare in predictive performance. The Extra Trees model by Minnor and Baths [13] achieved an accuracy of 88.94%, while Ghazal et al. [14] reported an accuracy of 88.4% using Artificial Neural Networks (ANN). Support Vector Machine (SVM) models showed varying results, with Geetha and Arunachalam [15] obtaining 75.04% accuracy and Sadhasivam et al. [20] reaching 87.09%. Afreen et al. [16] and Nigatu et al. [17] also applied ANN and CatBoost models, resulting in 86.8% and 87% accuracy, respectively. Ghosh et al. [18] achieved 83.76% using Random Forest, while Hartatik et al. [19] obtained the lowest performance with Naïve Bayes at 72.5%. In contrast, compared to the earlier techniques, this study using CatBoost with RandomizedSearchCV outperformed all other models with an accuracy of 93%, illustrating the effectiveness of hyperparameter tuning in enhancing model performance. This highlights the importance of using optimization techniques like RandomizedSearchCV to achieve higher accuracy and better results than existing approaches.

### 4. Conclusions

This study demonstrates the usefulness of machine learning in classifying liver diseases according to several clinical and biochemical attributes of the Indian Liver Patient Dataset (ILPD). Based on an accuracy of 93%, the CatBoost model shows excellent results for

non-invasive and efficient diagnosis. Nevertheless, challenges such as data privacy, clinical interpretation, and regulatory compliance exist during the adoption process. The other barriers are the geographical and demographic biases inherent in the dataset, which limit the generalization capability of the model. Thus, more extensive validation is required. Future studies should emphasize methodologies that would enhance validation, such as cross-institutional testing and longitudinal studies, to ensure the reliability of the model. Employing explainable AI methods, such as SHAP and hybrid modeling, will provide clinicians with extra insight into the application of such tools.

### Acknowledgments

The authors would like to acknowledge the valuable funding provided by the Department of Research and Community Service, Universitas Amikom Yogyakarta.

### References

- [1] M.-C. Guan *et al.*, "Early diagnosis and therapeutic strategies for hepatocellular carcinoma: From bench to bedside," *World J. Gastrointest. Oncol.*, vol. 13, no. 4, pp. 197–215, Apr. 2021, doi: 10.4251/wjgo.v13.i4.197.
- [2] H. Wazir *et al.*, "Diagnosis and Treatment of Liver Disease: Current Trends and Future Directions," *Cureus*, Dec. 2023, doi: 10.7759/cureus.49920.
- [3] Z. Yao, J. Li, Z. Guan, Y. Ye, and Y. Chen, "Liver disease screening based on densely connected deep neural networks," *Neural Networks*, vol. 123, 2020, doi: 10.1016/j.neunet.2019.11.005.
- [4] S. A. Alowais *et al.*, "Revolutionizing healthcare: the role of artificial intelligence in clinical practice," *BMC Med. Educ.*, vol. 23, no. 1, p. 689, Sep. 2023, doi: 10.1186/s12909-023-04698-z.
- [5] S. Maleki Varnosfaderani and M. Forouzanfar, "The Role of AI in Hospitals and Clinics: Transforming Healthcare in the 21st Century," *Bioengineering*, vol. 11, no. 4, p. 337, Mar. 2024, doi: 10.3390/bioengineering11040337.
- [6] D. A. Hashimoto, G. Rosman, D. Rus, and O. R. Meireles, "Artificial Intelligence in Surgery: Promises and Perils," *Ann. Surg.*, vol. 268, no. 1, pp. 70–76, Jul. 2018, doi: 10.1097/SLA.0000000000002693.
- [7] G. Chongo and J. Soldera, "Use of machine learning models for the prognostication of liver transplantation: A systematic review," *World J. Transplant.*, vol. 14, no. 1, Mar. 2024, doi: 10.5500/wjt.v14.i1.88891.
- [8] D. Nam, J. Chapiro, V. Paradis, T. P. Seraphin, and J. N. Kather, "Artificial intelligence in liver diseases: Improving diagnostics, prognostics and response prediction," *JHEP Reports*, vol. 4, no. 4, 2022, doi: 10.1016/j.jhepr.2022.100443.
- [9] A. Alizargar, Y.-L. Chang, M. Alkhaleefah, and T.-H. Tan, "Precision Non-Alcoholic Fatty Liver Disease (NAFLD) Diagnosis: Leveraging Ensemble Machine Learning and Gender Insights for Cost-Effective Detection," *Bioengineering*, vol. 11, no. 6, p. 600, Jun. 2024, doi: 10.3390/bioengineering11060600.
- [10] T. A. Assegie, R. Subhashni, N. K. Kumar, J. P. Manivannan, P. Duraisamy, and M. F. Engidaye, "Random forest and support vector machine-based hybrid liver disease detection," *Bull. Electr. Eng. Informatics*, vol. 11, no. 3, 2022, doi: 10.11591/eei.v11i3.3787.
- [11] A. Spann *et al.*, "Applying Machine Learning in Liver Disease and Transplantation: A Comprehensive Review," *Hepatology*, vol. 71, no. 3, pp. 1093–1105, Mar. 2020, doi: 10.1002/hep.31103.
- [12] N. Tanwar and K. F. Rahman, "Machine learning in liver disease diagnosis: Current progress and future opportunities," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1022, no. 1, 2021, doi: 10.1088/1757-899X/1022/1/012029.
- [13] M. Minnoor, "Liver Disease Diagnosis Using Machine

- Learning”, doi: 10.1109/aic.2022.08.
- [14] T. M. Ghazal, A. U. Rehman, M. Saleem, M. Ahmad, S. Ahmad, and F. Mehmood, “Intelligent Model to Predict Early Liver Disease using Machine Learning Technique,” 2022. doi: 10.1109/ICBATS54253.2022.9758929.
- [15] C. Geetha and A. R. Arunachalam, “Evaluation based Approaches for Liver Disease Prediction using Machine Learning Algorithms,” Jan. 2021. doi: 10.1109/ICCCI50826.2021.9402463.
- [16] N. Afreen, R. Patel, M. Ahmed, and M. Sameer, “A Novel Machine Learning Approach Using Boosting Algorithm for Liver Disease Classification,” 2021. doi: 10.1109/ISCONS2037.2021.9702488.
- [17] S. S. Nigatu, P. C. R. Alla, R. N. Ravikumar, K. Mishra, G. Komala, and G. R. Chami, “A Comparative Study on Liver Disease Prediction using Supervised Learning Algorithms with Hyperparameter Tuning,” in *2023 International Conference on Advancement in Computation and Computer Technologies, InCACCT 2023*, 2023, pp. 353–357. doi: 10.1109/InCACCT57535.2023.10141830.
- [18] M. Ghosh et al., “A comparative analysis of machine learning algorithms to predict liver disease,” *Intell. Autom. Soft Comput.*, vol. 30, no. 3, pp. 917–928, 2021, doi: 10.32604/iasc.2021.017989.
- [19] H. Hartatik, M. B. Tamam, and A. Setyanto, “Prediction for Diagnosing Liver Disease in Patients using KNN and Naïve Bayes Algorithms,” in *2020 2nd International Conference on Cybernetics and Intelligent System (ICORIS)*, Oct. 2020, pp. 1–5. doi: 10.1109/ICORIS50180.2020.9320797.
- [20] J. Sadhasivam, S. J. G. R.M, and C. N, “Liver Disease Prediction Using Machine Learning Classification,” *Webology*, vol. 18, no. 02, pp. 441–452, Sep. 2021, doi: 10.14704/web/v18si02/web18293.
- [21] V. Mehta et al., “Machine Learning based Exploratory Data Analysis (EDA) and Diagnosis of Chronic Kidney Disease (CKD),” *EAI Endorsed Trans. Pervasive Heal. Technol.*, vol. 10, Mar. 2024, doi: 10.4108/eetpht.10.5512.
- [22] N. Pandey, P. K. Patnaik, and S. Gupta, “Data Pre Processing for Machine Learning Models using Python Libraries,” *Int. J. Eng. Adv. Technol.*, vol. 9, no. 4, pp. 1995–1999, Apr. 2020, doi: 10.35940/ijeat.D9057.049420.
- [23] K. Maharana, S. Mondal, and B. Nemade, “A review: Data pre-processing and data augmentation techniques,” *Glob. Transitions Proc.*, vol. 3, no. 1, 2022, doi: 10.1016/j.gltp.2022.04.020.
- [24] B. Juba and H. S. Le, “Precision-Recall versus Accuracy and the Role of Large Data Sets,” *Proc. AAAI Conf. Artif. Intell.*, vol. 33, no. 01, pp. 4039–4048, Jul. 2019, doi: 10.1609/aaai.v33i01.33014039.
- [25] C. N. Noviyanti and A. Alamsyah, “Early Detection of Diabetes Using Random Forest Algorithm,” *J. Inf. Syst. Explor. Res.*, vol. 2, no. 1, Jan. 2024, doi: 10.52465/joiser.v2i1.245.
- [26] K. Budholiya, S. K. Shrivastava, and V. Sharma, “An optimized XGBoost based diagnostic system for effective prediction of heart disease,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 7, 2022, doi: 10.1016/j.jksuci.2020.10.013.
- [27] N. M. . R. -, V. S. -, S. T. -, and P. V. S. -, “Heart Disease Risk Assessment by Using Light GBM Technique,” *Int. J. Multidiscip. Res.*, vol. 5, no. 2, 2023, doi: 10.36948/ijfmr.2023.v05i02.2620.
- [28] J. An, “Using CatBoost and Other Supervised Machine Learning Algorithms to Predict Alzheimer’s Disease,” in *2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA)*, Dec. 2022, pp. 1732–1739. doi: 10.1109/ICMLA55696.2022.00265.
- [29] K. S. R. Kundra, B. J. Lakshmi, I. V. S. Venugopal, and V. Guthula, “Flood Prediction using MLP, CATBOOST and Extra-Tree Classifier,” *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 11, no. 7 s, 2023, doi: 10.17762/ijritcc.v11i7s.6974.
- [30] B. H. Cho, Y. H. Kim, K. B. Lee, Y. K. Hong, and K. C. Kim, “Potential of Snapshot-Type Hyperspectral Imagery Using Support Vector Classifier for the Classification of Tomatoes Maturity,” *Sensors*, vol. 22, no. 12, 2022, doi: 10.3390/s22124378.
- [31] A. Abdalrada, Ali Fahem Neamah, and Hayder Murad, “Predicting Diabetes Disease Occurrence Using Logistic Regression: An Early Detection Approach,” *Iraqi J. Comput. Sci. Math.*, vol. 5, no. 1, 2024, doi: 10.52866/ijcsm.2024.05.01.011.
- [32] C. Zhou et al., “A machine learning-based predictor for the identification of the recurrence of patients with gastric cancer after operation,” *Sci. Rep.*, vol. 11, no. 1, 2021, doi: 10.1038/s41598-021-81188-6.
- [33] O. Bardhi and B. G. Zapiirain, “Machine learning techniques applied to electronic healthcare records to predict cancer patient survivability,” *Comput. Mater. Contin.*, vol. 68, no. 2, 2021, doi: 10.32604/cmc.2021.015326.