Matrik: Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer
Vol. 24, No. 3, July 2025, pp. 507∼520
ISSN: 2476-9843, accredited by Kemenristekdikti, Decree No: 10/C/C3/DT.05.00/2025
DOI: 10.30812/matrik.v24i3.4531

Investigating Liver Disease Machine Learning Prediction Performance through Various Feature Selection Methods

Ahmad Zein Al Wafi , Febry Putra Rochim , Veda Bezaleel
Universitas Negeri Semarang, Semarang, Indonesia

Article Info

Article history:

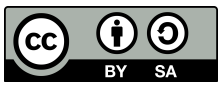
Received October 11, 2024
Revised June 26, 2025
Accepted July 1, 2025

Keywords:

Feature Selection;
Liver Disease Prediction;
Machine Learning.

ABSTRACT

Given the increasing prevalence and significant health burden of liver diseases globally, improving the accuracy of predictive models is essential for early diagnosis and effective treatment. **The purpose of the study** is to systematically analyze how different feature selection methods impact the performance of various machine learning classifiers for liver disease prediction. **The research method** involved evaluating four distinct feature selection techniques—regular, analysis of variance (ANOVA), univariate, and model-based—on a suite of classifiers, including decision forest, decision tree, support vector classifier, multi-layer perceptron, and linear discriminant analysis. The result revealed a significant and variable impact of feature selection on model accuracy. Notably, the ANOVA method paired with the multi-layer perceptron achieved the highest accuracy of 0.801724, while the univariate method was optimal for the decision forest classifier (0.741379). In contrast, model-based selection often degraded performance, particularly for the decision tree classifier, likely due to the introduction of noise and overfitting. The support vector classifier, however, demonstrated robust and consistent accuracy across all selection techniques. These findings underscore that there is no universally superior feature selection method; instead, optimal predictive performance hinges on tailoring the selection technique to the specific machine learning model. **This study contributes** practical, evidence-based insights into the critical interplay between feature selection and model choice in medical data analysis, offering a guide for improving classification accuracy in liver disease prediction. Future work should explore more sophisticated and hybrid feature selection methods to enhance model performance further.



Corresponding Author:

Febry Putra Rochim, +6281225777701,
Faculty of Engineering, Computer Engineering Department,
State University of Semarang, Semarang, Indonesia,
Email: febry.putra@mail.unnes.ac.id

How to Cite:
A. Z. A. Wafi, F. P. Rochim, and V. Bezaleel,, "Investigating Liver Disease Machine Learning Prediction Performance through Various Feature Selection Methods", *MATRIK: Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer*, Vol. 24, No. 3, pp. 505-518, July, 2025, doi:10.30812/matrik.v24i3.4531.

1. INTRODUCTION

Liver disease poses a significant health burden globally, with its prevalence steadily rising over the past decades [1]. Cirrhosis or liver damage is a significant contributor to both death and illness globally, ranking as the 11th major cause of death and the 15th largest cause of morbidity in 2016 [2]. Early detection and accurate prediction of liver disease play pivotal roles in effective patient management, treatment planning, and ultimately, in reducing mortality rates associated with liver-related complications [3]. In recent years, the integration of machine learning techniques in medical decision-making has shown promising results in improving predictive accuracy and patient outcomes [4–7]. Liver disease diagnosis often involves a multifaceted approach, with liver enzyme analysis serving as a key diagnostic tool. Elevated levels of specific liver enzymes, such as alanine aminotransferase (ALT), aspartate aminotransferase (AST), and alkaline phosphatase (ALP), can indicate liver damage or dysfunction [8]. These biomarkers provide valuable insights into the physiological status of the liver and are routinely measured in clinical settings to aid in the diagnosis and monitoring of liver-related conditions [9]. Studies have highlighted the potential of machine learning algorithms to complement traditional diagnostic methods by leveraging a diverse array of clinical and biochemical data for liver disease prediction [10]. Previous studies have demonstrated the versatility and efficacy of various machine learning algorithms in accurately predicting liver disease based on diverse sets of clinical and demographic features [11]. These algorithms, ranging from traditional classifiers like logistic regression and decision trees to more sophisticated approaches such as support vector machines and neural networks, have demonstrated the capacity to analyze complex data patterns and generate reliable predictions [12–14].

While machine learning models have indeed demonstrated substantial success in predicting liver disease, their performance can be further enhanced by employing effective feature selection techniques. Their predictive accuracy can be hindered by the inclusion of irrelevant or redundant features within the input data [15, 16]. Such features not only increase computational overhead but also introduce noise, potentially leading to overfitting and suboptimal model performance [17, 18]. Therefore, the process of feature selection becomes indispensable in identifying the most informative variables that significantly contribute to the predictive task at hand [19, 20]. Feature selection is essential to create a model that effectively generalizes to samples that have not been previously observed [21]. Reducing the dimensionality of the medical information can lead to an increase in the accuracy of the prediction model, while also reducing computing complexity [22]. Previous studies in the realm of liver disease prediction have predominantly focused on employing single-feature selection methods to enhance model performance. For instance, one study deployed the Least Absolute Shrinkage and Selection Operator (LASSO) technique to identify highly correlated attributes of liver disease (LD) [23], while another utilized the Random Under Sampling boosting algorithm to assess feature importance [24]. The landscape of feature selection methods is diverse, encompassing approaches such as Analysis of Variance (ANOVA), Univariate, and Model-Based techniques, each with unique strengths and limitations [25, 26].

Previous research on the dataset we employ has focused on developing different machine-learning models to forecast liver illness, but with a notable omission of feature selection techniques in their methodology [27] While other studies in the broader field have recognized the importance of feature selection, they predominantly rely on a singular methodology, such as deploying LASSO [23], or a boosting algorithm [24] in isolation. This reliance on a single feature selection approach presents a significant, unaddressed limitation [28]. Feature selection methods are not one-size-fits-all [29, 30]; techniques like filter-based (e.g., ANOVA), wrapper-based (e.g., Recursive Feature Elimination), and embedded methods operate on fundamentally different principles [31, 32]. Consequently, a feature subset deemed optimal by one method may be suboptimal when paired with a different machine learning algorithm, potentially leading to biased conclusions about feature importance and constrained predictive performance.

Therefore, the critical research gap lies in the absence of a systematic and comparative analysis that investigates how different classes of feature selection methods perform across various machine learning models for liver disease prediction. The novelty of this research is rooted in its two-dimensional approach, moving beyond the linear focus of previous studies. While prior work has concentrated on either optimizing a single machine learning algorithm or applying one specific feature selection technique [23, 24]. Our study is the first to systematically investigate the crucial interaction between these two components. It remains unclear whether the choice of feature selection technique has a more significant impact on prediction performance than the choice of the learning algorithm itself. This paper seeks to bridge this gap by directly addressing this uncertainty. Our central purpose is to test the hypothesis that optimal predictive performance is not achieved by a single ”best” method, but rather by identifying an optimal pairing of a feature selection technique with a compatible machine learning algorithm. By doing so, we aim to establish an evidence-based methodological framework for researchers in this field.

The contributions of this research extend significantly beyond its academic novelty. For further research, our findings will provide a crucial benchmark and a methodological roadmap, enabling subsequent studies to build upon a more robust evidence base instead of relying on ad-hoc choices for feature selection. This will save valuable computational resources and improve the reproducibility of results in the field. From a clinical perspective, a more nuanced understanding of these model-feature selection interactions is the first step toward developing more accurate and reliable Clinical Decision Support Systems (CDSS). This directly

benefits community organizations, such as patient advocacy groups and professional medical societies, by providing the evidence needed to champion the adoption of superior diagnostic tools that can lead to earlier detection and improved patient outcomes. For government and public health bodies, this research offers tangible insights for policy-making. By identifying more precise predictive models, our work can inform the design of cost-effective national screening strategies for at-risk populations, optimize the allocation of healthcare resources, and ultimately help reduce the significant economic and social burden imposed by advanced liver disease.

2. RESEARCH METHOD

2.1. Research Design and Approach

This study employs a quantitative, experimental research design. The approach is quantitative, as it centers on analyzing numerical data to measure and compare model performance. The research process involves the systematic application of various feature selection techniques and machine learning algorithms to the Indian Liver Patient Dataset. The primary outcome is measured using a single and clear metric named model accuracy.

The experimental nature of this research is defined by its controlled, comparative framework. We establish a baseline performance by training each machine learning algorithm on the full set of features. Subsequently, we systematically introduce different feature selection methods (the independent variable) to create multiple, distinct feature subsets. We then measure the impact of these subsets on the accuracy of each machine learning model (the dependent variable). The core analysis focuses on the percentage of accuracy improvement and the average variance of this improvement relative to the baseline, allowing for a rigorous examination of the interaction and impact of each feature selection and algorithm pairing.

The empirical investigation presented herein follows a rigorous, multi-stage methodological framework, visually delineated in the subsequent flowchart (Figure 1). The process begins with the acquisition of the Indian Liver Patient Dataset (ILPD), followed by a minimal preprocessing step that involves numerically encoding the gender attribute. Subsequently, a suite of diverse feature selection techniques is employed to generate multiple, distinct feature subsets from the training data. Each resulting subset, along with the original full-feature set serving as a baseline, is used to train various machine learning classifiers. The performance of each model is then quantitatively evaluated based on predictive accuracy, culminating in a comparative analysis that assesses the accuracy improvement and systematically investigates the interaction between each feature selection method and learning algorithm.

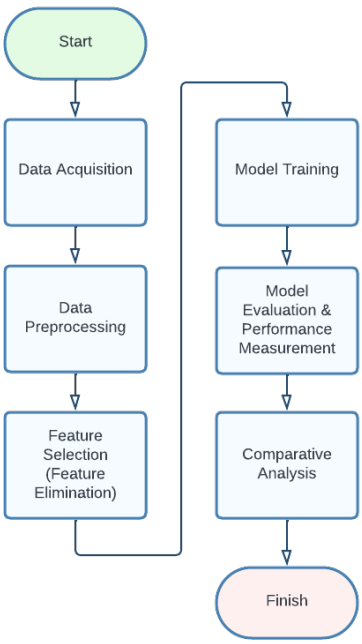


Figure 1. Research Flowchart

2.2. Dataset

The dataset utilized in this study, extracted from the UCI Machine Learning Repository, constitutes the Indian Liver Patient Dataset (ILPD). This dataset has already been studied with various machine-learning techniques. Comprising 416 liver patient records and 167 non-liver patient records, it reflects a comprehensive collection derived from India country [33–36]. The dataset features a gender distribution of 441 male patient records and 142 female patient records, offering a diverse demographic perspective. The dataset encompasses ten discernible features crucial for analysis. The dataset contains several variables, as shown in Table 1.

Table 1. Pembagian data untuk Training dan Testing

Feature Name	Description
Age	The age of the patient.
Gender	The gender of the patient (male/female).
TB (Total Bilirubin)	A compound indicating the total amount of bilirubin in the blood is often associated with liver function.
DB (Direct Bilirubin)	The portion of bilirubin that is directly conjugated in the liver is typically indicative of liver function.
ALP (Alkaline Phosphatase)	An enzyme found primarily in the liver, bone, and placenta, which is often used as a marker for liver health.
ALT (SGPT Alanine Aminotransferase)	An enzyme found predominantly in the liver is often used as a marker for liver damage or disease.
AST (SGOT Aspartate Aminotransferase)	An enzyme found in various tissues including the liver and heart, is often elevated in cases of liver damage.
TP (Total Protein)	The total amount of protein in the blood, including albumin and globulin, is used as a marker for overall health and nutrition.
(ALB) Albumin	The total amount of protein in the blood, including albumin and globulin, is used as a marker for overall health and nutrition.
(A/G) Albumin and Globulin Ratio	The ratio of albumin to globulin in the blood is often used as an indicator of liver or kidney function.
Status	A label indicating whether the patient has liver disorder or not, is assigned by experts for classification purposes.

2.3. Feature Selection Methods

The accuracy of predictive models is highly contingent on the quality of the features used. In this study, we employed three primary feature selection methods: ANOVA, Univariate, and Model-Based methods. These techniques were meticulously chosen to enhance the prediction performance for liver disease by identifying the most relevant features from our dataset. Below, we delineate the principles, implementation specifics, and application of each method.

ANOVA (Analysis of Variance): ANOVA is a statistical method used to determine the significance of individual features by analyzing the differences between the means of different groups. In the context of our study, ANOVA was applied to compare the variance between the liver disease status groups (positive vs. negative) for each feature. The underlying principle of ANOVA is to assess whether the means of different groups are statistically significantly different from each other, thereby identifying features that have a strong discriminative power [37]. The ANOVA F-test was performed for each feature to calculate the F-statistic and the corresponding p-value. Features with p-values below 0.05 were selected, indicating significant differences in their means across the groups.

Univariate Feature Selection: Univariate feature selection involves evaluating each feature individually to determine its relevance to the target variable. This method ranks features based on their univariate statistical tests, selecting the top-ranking features for model-building [38]. The rationale behind univariate selection is that it simplifies the feature space while retaining the most informative features. We utilized the chi-square test for categorical features and the ANOVA F-test for continuous features. Features were ranked according to their test scores. The top-ranked features, based on a predetermined count or percentile, were selected for inclusion in the predictive model.

Model-Based Feature Selection: Model-based feature selection utilizes machine learning models to assess the significance of features. This approach integrates the predictive power of features directly into the model training process, identifying those that most significantly enhance the model's performance. For this study, we utilized the Extra Trees Classifier. This ensemble learning method is particularly effective for feature selection due to its ability to handle large datasets with numerous features.

This method constructs a multitude of decision trees and outputs the mean prediction of the individual trees, thus reducing overfitting and enhancing predictive accuracy [39]. We configured the Extra Trees Classifier with 50 estimators and trained it on the full set of features from the liver disease dataset. After training, the classifier assigned importance scores to each feature based on its contribution to reducing the impurity in the trees. Features with higher scores were deemed more relevant. Features with the highest importance scores were selected. Specifically, we identified a threshold score to include the top-ranked features, ensuring that only the most informative features were retained.

2.4. Model Training

The predictive models used in this study were trained using a diverse set of algorithms, each with specific configurations designed to optimize performance for liver disease prediction. Below, we provide detailed descriptions of the models, their respective training parameters, and the underlying algorithms.

Random Forest Classifier: The Random Forest Classifier is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes for classification tasks. This method reduces overfitting and improves generalization. We employed the Random Forest Classifier with a range of estimators (n_estimators), varying from 50 to 500 in increments of 50 (i.e., 50, 100, 150, . . ., 500). This allowed us to assess the impact of ensemble size on the model's accuracy and robustness.

Decision Tree Classifier: The Decision Tree Classifier is a non-parametric supervised learning method used for classification and regression. It splits the data into subsets based on the value of input features, forming a tree structure. We explored different depths to determine the optimal complexity of the trees. The maximum depth was varied from 50 to 500 in increments of 50 (i.e., 50, 100, 150, . . ., 500), to find the depth that provided the best predictive performance.

Support Vector Classifier (SVC): The Support Vector Classifier (SVC) is a powerful classification method that finds the hyperplane that best separates classes in the feature space. SVC can handle high-dimensional spaces and is effective when the number of dimensions exceeds the number of samples. We evaluated SVC using multiple kernel functions to capture different patterns in the data, testing the linear, polynomial (poly), radial basis function (RBF), sigmoid, and precomputed kernels. Each kernel was assessed for its ability to handle the dataset's complexities and contribute to accurate liver disease prediction.

Multi-Layer Perceptron (MLP): The Multi-Layer Perceptron (MLP) is a type of artificial neural network comprising at least three layers of nodes: an input layer, one or more hidden layers, and an output layer. Each node (neuron) in one layer connects to every node in the following layer with a specific weight. We configured the MLP with two hidden layers and experimented with three different configurations for the number of neuron units per layer: 8, 32, and 128. The activation function used was LeakyReLU, which helps mitigate the vanishing gradient problem. Each model was trained for 100 epochs to ensure convergence and stability in the learning process.

Linear Discriminant Analysis: Linear Discriminant Analysis (LDA) is a statistical and machine learning method that identifies the linear combination of features that best separates two or more classes of objects or events. LDA reduces dimensionality while preserving as much of the class discriminatory information as possible. We implemented LDA using three different solvers: singular value decomposition (SVD), least squares (LSQR), and eigenvalue decomposition (EVD). These solvers offered varied approaches to dimension reduction and classification, enabling us to identify the most effective method for our dataset.

2.5. Experimental Setup

In our experimental setup, we began by preprocessing the dataset to ensure its quality and consistency. All data entries with null values were removed to prevent any biases or inaccuracies in the model training process. This cleaning step was crucial to maintaining the integrity of the dataset and ensuring reliable model performance.

For the implementation of traditional machine learning models, we utilized the Scikit-learn library. This library provided robust and efficient tools for deploying our Random Forest Classifier, Decision Tree Classifier, Support Vector Classifier, and Linear Discriminant Analysis models. The Multi-Layer Perceptron (MLP) was implemented using the TensorFlow library, which offers extensive support for deep learning architectures. Each model underwent a comprehensive grid search to optimize its hyperparameters, allowing us to explore a wide range of parameter configurations systematically. The goal was to identify the set of parameters that yielded the highest accuracy on the validation set.

3. RESULT AND ANALYSIS

This section will analyze the outcomes of our research and provide the findings. We aim to provide insights into the efficiency of various feature selection approaches combined with machine learning models for predicting liver disease through a thorough analysis and interpretation. We examine how model performance varies with different techniques, with a focus on the impact of feature selection on prediction accuracy.

3.1. Selected Features

The ANOVA approach determined that the following variables are significant predictors: Age, Gender, TB, DB, ALP, SGPT, SGOT, ALB, and AG. This approach highlights the statistical disparities across various groups, leading to a thorough selection that encompasses demographic, biochemical, and enzymatic indicators. The Univariate technique, which assesses each feature separately, identified a slightly distinct group of predictors: Age, Gender, TB, DB, ALP, SGPT, SGOT, and ALB. Notably, this technique did not include the Albumin-Globulin Ratio (AG) in its selection, suggesting that AG may have lower predictive power when evaluated independently of other variables. Finally, the Tree-based method, which uses decision trees to determine the relevance of features, identified Age, TB, DB, ALP, SGPT, SGOT, and ALB as the primary predictors. This strategy also omitted Gender and AG, indicating that these characteristics may have a lesser significance in the hierarchical organization of decision tree models.

Upon comparing the chosen characteristics, it becomes apparent that Age, TB, DB, ALP, SGPT, SGOT, and ALB consistently emerge in all techniques, suggesting their strong predictive value for liver disease. Both the ANOVA and Univariate approaches identified gender as a significant factor, but the Tree-based method did not. Additionally, the ANOVA method identified AG as the only significant factor. In subsequent investigations, we will evaluate the performance of predictive models utilizing the chosen features to establish the most optimal combination for precise diagnosis of liver illness. All symbols that have not been mentioned in the equation should be explained in the following text.

3.2. Performance of Feature Selection Methods

In this subsection, we compare the performance of different feature selection methods, with the primary metric for evaluation being the best validation accuracy achieved through grid search optimization. This table presents the results of implementing various machine learning models with different feature selection methods. Table 2 presents the accuracy scores obtained from different machine learning models trained with features selected using various methods, including ANOVA, Univariate, and Model-Based techniques. The ”Regular” column represents models trained without feature selection, serving as a baseline for comparison.

The Decision Forest Classifier (DFC) achieves the maximum accuracy using the Univariate approach, with a score of 0.741379. It is closely followed by the Regular method, which achieves a score of 0.732759. The ANOVA and Model-Based approaches had accuracies of 0.698276 and 0.724138, respectively, which are marginally lower. This suggests that DFC benefits from a more straightforward and statistically significant feature selection process, facilitated by the Univariate technique. The Regular approach achieves the maximum accuracy of 0.715517 for the Decision Tree Classifier (DTC). The ANOVA and Univariate techniques exhibit similar performance with accuracies of 0.698276 and 0.681034, respectively. In contrast, the Model-Based method demonstrates a notable decrease in accuracy, dropping to 0.629310. This suggests that DTC is prone to overfitting when features are chosen using model-based criteria.

The Support Vector Classifier (SVC) consistently performs well with all feature selection approaches, obtaining an accuracy of 0.767241 in each case. The consistent nature of SVC implies that the choice of feature selection method does not significantly influence its resilience. The Multi-Layer Perceptron (MLP) model reaches its greatest accuracy of 0.801724 using the ANOVA approach. The Regular, Univariate, and Model-Based approaches yield accuracies of 0.784483, 0.775862, and 0.775862, respectively, which are slightly lower than the optimal value. MLP benefits from the variance analysis offered by the ANOVA method, potentially because it enhances the comprehension of feature interactions. The Linear Discriminant Analysis (LDA) model attains the greatest accuracy of 0.741379 using both the Regular and Univariate approaches. The ANOVA and Model-Based approaches exhibit a marginal decrease in performance, achieving accuracies of 0.732759 and 0.741379, respectively. The consistency shown across different approaches indicates that LDA is useful in contexts where the characteristics follow a normal distribution.

Table 2. Accuracy Comparison of Machine Learning Models with Feature Selection

♯Model	Regular	Anova	Univariate	ModelBased
Random Forest Classifier	73.275	69.827	**74.137**	72.413
Decision Tree Classifier	**71.551**	69.827	68.103	62.931
Support Vector Classifier	**76.724**	**76.724**	**76.724**	**76.724**
Multi Layer Perceptron	78.448	**80.172**	77.586	77.586
Linear Discriminant Analysis	**74.137**	73.275	**74.137**	**74.137**
Average	74.827	73.965	74.139	72.758

To contextualize our findings, a comparative analysis was conducted against previously published research on liver disease prediction using similar datasets.

Table 3. Pembagian data untuk Training dan Testing

Research	Accuracy Performance	Use Feature Selection
Our Research	80.1%	Yes
[40]	78.6%	Yes
[40]	78.1%	No
[36]	71.8%	No
[41]	71.8%	No
[42]	71.3%	No

As delineated in Table 3, the optimal model developed in this study achieved a predictive accuracy of 80.1%. This result marks a significant performance gain over prior studies that did not employ feature selection, which reported accuracies ranging from 71.3% to 71.8% [36, 41, 42]. Notably, the performance of our model also exceeds that of recent work by [40], which reported accuracies of 78.1% without feature selection and 78.6% with it. This comparison highlights that while the inclusion of feature selection generally yields better results. The multi-method approach adopted in our research provides an additional performance advantage, establishing a new benchmark for the dataset.

3.3. Impact of Feature Selection on Model Performance

The analysis of feature selection methods revealed variations in their impact on the performance of machine learning models for predicting liver disease. To understand the impact of feature selection on model performance, we analyzed the average accuracy, variance, and improvement percentage for each classifier across different feature selection methods. Table 3 presents statistics of validation accuracy results.

The impact of feature selection on model performance reveals distinct patterns and insights when examining the average accuracy, variance, improvement, and improvement percentage for each classification model. Among the observed trends, it is evident that the choice of feature selection method exerted a discernible influence on the predictive accuracy of the models.

The Univariate feature selection approach achieves the highest average accuracy of 0.741379 for the Decision Forest Classifier (DFC), indicating its efficiency in improving the model's performance. This approach also ensures a low variance of 0.000347, which signifies consistent and dependable outcomes. The increase in accuracy is 0.008621, equivalent to a 1.176471% improvement, highlighting the advantage of utilizing statistically significant features identified by the Univariate technique. On the other hand, the ANOVA approach yields the lowest mean accuracy of 0.698276, suggesting that it may not be as suitable for DFC. The Decision Tree Classifier (DTC) achieves the highest mean accuracy of 0.715517 using the Regular feature selection approach. Nevertheless, this model demonstrates a significant decline in performance when using the Model-Based approach, obtaining an average accuracy of merely 0.629310. The substantial reduction, together with the largest variance of 0.001387, demonstrates the Model-Based method's susceptibility to overfitting or introducing noise. The Regular technique has resulted in a negative improvement percentage of -2.409639% for DTC, indicating an adverse impact on performance.

The Support Vector Classifier (SVC) has exceptional consistency, attaining a consistent average accuracy of 0.767241 across all feature selection methods. The fact that SVC exhibits invariance and zero variance indicates that it is resilient and not influenced by the chosen feature selection method. It consistently performs well regardless of the specific characteristics that are selected.

The ANOVA approach yields the highest average accuracy of 0.801724 for the Multi-Layer Perceptron (MLP). The approach exhibits a low variance of 0.000149, indicating consistent and dependable performance. The gain in accuracy is substantial, with a considerable improvement of 0.017241, equal to a 2.197800% increase. The observed pattern suggests that MLP leverages the ANOVA method's ability to efficiently analyze variations in features, leading to improved model performance. Linear Discriminant Analysis (LDA) exhibits consistent performance across various feature selection approaches, with minimal fluctuation. The Regular and Univariate techniques both achieve the highest average accuracy, which is measured at 0.741379. The variance is insignificant, especially for the Regular approach, which has a variance of 0.000019. The improvement % for LDA remains at zero, suggesting that feature selection approaches have no meaningful impact on its performance. The consistency observed indicates that LDA is highly resilient in settings where the characteristics follow a normal distribution. The observed consistency in MLP performance suggests that certain machine learning algorithms may be inherently more resilient to variations in feature selection. This resilience underscores the stability and reliability of MLP models in the context of liver disease prediction, regardless of the specific feature subset used for training search optimization, as shown in Table 4.

Table 4. Validation Accuracy Statistics

♯ Model	Average	Variance(10^{-6})	Improvement	Improvement(%)
Random Forest Classifier	72.413	347	0.862	1.176
Decision Tree Classifier	68.103	1387	-1.724	-2.409
Support Vector Classifier	76.724	0	0	0
Multi Layer Perceptron	78.448	149	1.724	21.978
Linear Discriminant Analysis	73.922	19	0	0

3.4. Discussion and Analysis

This section will analyze the outcomes of our research and provide the findings. We aim to provide insights into the efficiency of various feature selection approaches combined with machine learning models for predicting liver disease through a thorough analysis and interpretation. The analysis revealed instances where the implementation of feature selection did not uniformly enhance model performance. While some models demonstrated notable improvements with feature selection, others exhibited marginal or negligible changes in accuracy compared to the regular model.

The Regular feature which commonly entails utilizing all accessible features without any filtering, shows diverse performance among models. The Regular approach for DFC yielded a notably high average accuracy of 0.732759, suggesting that DFC is capable of effectively managing a bigger feature collection. Similarly, in the case of DTC, the Regular approach exhibited the highest level of accuracy, measuring at 0.715517. Nevertheless, the Regular technique did not demonstrate any notable superiority over other methods, such as SVC, MLP, and LDA. The accuracies of these models were quite similar. This implies that although Regular feature selection may be effective for certain models, it does not constantly provide optimal performance and can result in overfitting in more intricate models [43].

The ANOVA feature selection approach, which assesses the statistical significance of each feature, demonstrated significant advantages for specific models. The MLP model achieved the maximum accuracy of 0.801724 using ANOVA, with a low variance of 0.000149. This suggests that ANOVA is well-suited for handling feature interactions and variation. Study [44, 45] emphasized that ANOVA can enhance the learning process in neural networks by optimizing feature input. However, ANOVA did not yield satisfactory results for models such as DFC and DTC, as their average accuracies were both 0.698276. This implies that whereas ANOVA can be quite useful for models such as MLP, its advantages may not be universally applicable to other classification models.

The Univariate feature selection strategy, which chooses features based on their unique statistical importance, demonstrated great efficacy across various models. The DFC model attained its peak average accuracy of 0.741379 using the Univariate approach while exhibiting a minimal variance of 0.000347. This finding is consistent with the research conducted by [46, 47], which highlighted the efficacy and efficiency of Univariate selection in enhancing model performance by eliminating irrelevant features. The LDA algorithm has consistently demonstrated excellent performance when used in conjunction with the Univariate approach, achieving an accuracy of 0.741379. This result indicates that LDA is highly resilient to variations in feature selection. However, in the case of DTC, the Univariate technique did not produce the maximum level of accuracy, indicating that this strategy may be more advantageous for models that are less susceptible to overfitting.

The Model-Based feature selection strategy, which chooses features based on their impact on a particular model, yielded inconsistent outcomes. The DFC approach achieved an accuracy of 0.724138, which was surpassed by the Univariate method. The Model-Based technique had a substantial negative impact on the performance of DTC, resulting in an average accuracy of 0.629310, which was the lowest among all methods. Additionally, it had the highest variance of 0.001387. This discovery aligns with the findings [48] who observed that the utilization of model-based selection has the potential for higher accuracy than some other embedding methods. The Model-Based technique could not provide any notable advantages or disadvantages for models such as SVC, MLP, and LDA. This implies that its effectiveness may vary depending on the specific model being used.

The observed disparities in performance across different feature selection methods underscore the importance of tailoring selection approaches to the specific characteristics of the dataset and the requirements of the predictive task. Poor selection of feature selection techniques can result in the selection of poor feature subsets, which can hurt the overall performance of the classifier models [49]. Increasing the utilization and enhancement of feature selection methods in disease datasets is crucial, as it significantly contributes to the precise classification of diseases [50].

3.5. Limitations and Considerations

While this study offers valuable insights into the impact of different feature selection methods on the performance of machine learning models for liver disease prediction, several limitations and considerations warrant acknowledgment. Firstly, the generalizability of the findings may be constrained by the specific characteristics of the dataset used in this study. The dataset's size, composition, and representativeness of the target population could influence the observed performance of feature selection methods and machine learning models. Therefore, caution should be exercised when extrapolating the results to broader clinical contexts or diverse patient populations.

Secondly, the evaluation of feature selection methods and machine learning models relied primarily on validation accuracy as the performance metric. While validation accuracy provides a straightforward measure of predictive performance, it may not capture nuances such as model interpretability, sensitivity to class imbalance, or generalization to unseen data. Future studies should consider incorporating additional performance metrics to provide a more comprehensive assessment of model effectiveness. Exploring the underlying mechanisms that drive these variations and identifying strategies for optimizing model performance across diverse clinical settings will also be valuable.

Furthermore, the choice of feature selection methods and machine learning algorithms in this study represents only a subset of the available techniques. Alternative approaches, such as recursive feature elimination, ensemble methods, or deep learning architectures, could yield different results and deserve further exploration. Incorporating more sophisticated feature selection techniques, such as embedded methods and hybrid approaches, could provide further insights into optimizing model performance.

Additionally, the study's focus on predictive performance may overlook other important considerations, such as computational efficiency, scalability, and clinical interpretability. Future research should strive to balance predictive accuracy with these practical considerations to facilitate the real-world application of predictive models in clinical settings. Lastly, while efforts were made to mitigate biases and confounding factors in the analysis, the potential for residual biases or unaccounted variables cannot be entirely ruled out. Sensitivity analyses and robustness checks could provide additional insights into the stability of the findings and the robustness of the conclusions. Longitudinal studies that evaluate the impact of feature selection over time, considering model updates and changes in data patterns, would also be valuable.

4. CONCLUSION

This study demonstrates that feature selection methods have a significant impact on the performance of classification models in predicting liver disease. The Univariate and ANOVA methods proved most effective, particularly for DFC and MLP, respectively, by enhancing accuracy and maintaining low variance. While the Regular method showed variable effectiveness, the Model-Based method was found to introduce noise and overfitting in certain models, such as DTC. These findings underscore the importance of selecting suitable feature selection techniques that are tailored to the specific model and dataset, providing valuable insights for enhancing predictive performance in medical data analysis.

Moving forward, several avenues for future research warrant exploration. Firstly, further investigation is needed to understand the underlying mechanisms driving the observed variations in model performance across different feature selection methods. Additionally, incorporating performance metrics such as interpretability and computational efficiency could provide a more comprehensive assessment of model effectiveness.

5. ACKNOWLEDGEMENTS

The Acknowledgments section is optional. Research sources can be included in this section.

6. DECLARATIONS

AUTHOR CONTIBUTION

Ahmad Zein Al Wafi was responsible for the conceptualization, methodology, investigation, software, and drafting the original manuscript. Ferby Putra Rochim contributed to supervision, validation, review, and editing. Veda Bezaleel contributed to the administration, co-authoring the original draft and reviewing and editing it.

FUNDING STATEMENT

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

[1] P.-L. Gan, S. Huang, X. Pan, H.-F. Xia, X.-Y. Zeng, W.-S. Ren, X. Zhou, M.-H. Lv, and X.-W. Tang, "Global research trends in the field of liver cirrhosis from 2011 to 2020: A visualised and bibliometric study," *World Journal of Gastroenterology*, vol. 28, no. 33, pp. 4909–4919, Sep. 2022, https://doi.org/10.3748/wjg.v28.i33.4909.

[2] S. Cheemerla and M. Balakrishnan, "Global Epidemiology of Chronic Liver Disease," *Clinical Liver Disease*, vol. 17, no. 5, pp. 365–370, May 2021, https://doi.org/10.1002/cld.1061.

[3] R. Williams, C. Alessi, G. Alexander, M. Allison, R. Aspinall, R. L. Batterham, N. Bhala, N. Day, A. Dhawan, C. Drummond, J. Ferguson, G. Foster, I. Gilmore, R. Goldacre, H. Gordon, C. Henn, D. Kelly, A. MacGilchrist, R. McCorry, N. McDougall, Z. Mirza, K. Moriarty, P. Newsome, R. Pinder, S. Roberts, H. Rutter, S. Ryder, M. Samyn, K. Severi, N. Sheron, D. Thorburn, J. Verne, J. Williams, and A. Yeoman, "New dimensions for hospital services and early detection of disease: A Review from the Lancet Commission into liver disease in the UK," *The Lancet*, vol. 397, no. 10286, pp. 1770–1780, May 2021, https://doi.org/10.1016/s0140-6736(20)32396-5.

[4] H. S. R. Rajula, G. Verlato, M. Manchia, N. Antonucci, and V. Fanos, "Comparison of Conventional Statistical Methods with Machine Learning in Medicine: Diagnosis, Drug Development, and Treatment," *Medicina*, vol. 56, no. 9, pp. 1–10, Sep. 2020, https://doi.org/10.3390/medicina56090455.

[5] M. Pourhomayoun and M. Shakibi, "Predicting mortality risk in patients with COVID-19 using machine learning to help medical decision-making," *Smart Health*, vol. 20, pp. 1–8, Apr. 2021, https://doi.org/10.1016/j.smhl.2020.100178.

[6] C. Giordano, M. Brennan, B. Mohamed, P. Rashidi, F. Modave, and P. Tighe, "Accessing Artificial Intelligence for Clinical Decision-Making," *Frontiers in Digital Health*, vol. 3, pp. 1–9, Jun. 2021, https://doi.org/10.3389/fdgth.2021.645232.

[7] S. M. D. A. C. Jayatilake and G. U. Ganegoda, "Involvement of Machine Learning Tools in Healthcare Decision Making," *Journal of Healthcare Engineering*, vol. 2021, pp. 1–20, Jan. 2021, https://doi.org/10.1155/2021/6679512.

[8] M. A. Kalas, L. Chavez, M. Leon, P. T. Taweesedt, and S. Surani, "Abnormal liver enzymes: A review for clinicians," *World Journal of Hepatology*, vol. 13, no. 11, pp. 1688–1698, Nov. 2021, https://doi.org/10.4254/wjh.v13.i11.1688.

[9] K. Pafili and M. Roden, "Nonalcoholic fatty liver disease (NAFLD) from pathogenesis to treatment concepts in humans," *Molecular Metabolism*, vol. 50, pp. 1–20, Aug. 2021, https://doi.org/10.1016/j.molmet.2020.101122.

[10] F. Mostafa, E. Hasan, M. Williamson, and H. Khan, "Statistical Machine Learning Approaches to Liver Disease Prediction," *Livers*, vol. 1, no. 4, pp. 294–312, Dec. 2021, https://doi.org/10.3390/livers1040023.

[11] A. Spann, A. Yasodhara, J. Kang, K. Watt, B. Wang, A. Goldenberg, and M. Bhat, "Applying Machine Learning in Liver Disease and Transplantation: A Comprehensive Review," *Hepatology (Baltimore, Md.)*, vol. 71, no. 3, pp. 1093–1105, Mar. 2020, https://doi.org/10.1002/hep.31103.

[12] M. Ghosh, Md. Mohsin Sarker Raihan, M. Raihan, L. Akter, A. Kumar Bairagi, S. S. Alshamrani, and M. Masud, "A Comparative Analysis of Machine Learning Algorithms to Predict Liver Disease," *Intelligent Automation & Soft Computing*, vol. 30, no. 3, pp. 917–928, 2021, https://doi.org/10.32604/iasc.2021.017989.

[13] R. Choudhary, T. Gopalakrishnan, D. Ruby, A. Gayathri, V. S. Murthy, and R. Shekhar, "An Efficient Model for Predicting Liver Disease Using Machine Learning," in *Data Analytics in Bioinformatics*, 1st ed. Wiley, Feb. 2021, pp. 443–457, https://doi.org/10.1002/9781119785620.ch18.

[14] J. Singh, S. Bagga, and R. Kaur, "Software-based Prediction of Liver Disease with Feature Selection and Classification Techniques," *Procedia Computer Science*, vol. 167, pp. 1970–1980, 2020, https://doi.org/10.1016/j.procs.2020.03.226.

[15] U. M. Khaire and R. Dhanalakshmi, "Stability of feature selection algorithm: A review," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 4, pp. 1060–1073, Apr. 2022, https://doi.org/10.1016/j.jksuci.2019.06.012.

[16] S. Alabdulwahab and B. Moon, "Feature Selection Methods Simultaneously Improve the Detection Accuracy and Model Building Time of Machine Learning Classifiers," *Symmetry*, vol. 12, no. 9, pp. 14–24, Aug. 2020, https://doi.org/10.3390/sym12091424.

[17] G. Kou, P. Yang, Y. Peng, F. Xiao, Y. Chen, and F. E. Alsaadi, "Evaluation of feature selection methods for text classification with small datasets using multiple criteria decision-making methods," *Applied Soft Computing*, vol. 86, pp. 1–14, Jan. 2020, https://doi.org/10.1016/j.asoc.2019.105836.

[18] R.-C. Chen, C. Dewi, S.-W. Huang, and R. E. Caraka, "Selecting critical features for data classification based on machine learning methods," *Journal of Big Data*, vol. 7, no. 1, Dec. 2020, https://doi.org/10.1186/s40537-020-00327-4.

[19] R. Zebari, A. Abdulazeez, D. Zeebaree, D. Zebari, and J. Saeed, "A Comprehensive Review of Dimensionality Reduction Techniques for Feature Selection and Feature Extraction," *Journal of Applied Science and Technology Trends*, vol. 1, no. 1, pp. 56–70, May 2020, https://doi.org/10.38094/jastt1224.

[20] A. Bommert, X. Sun, B. Bischl, J. Rahnenführer, and M. Lang, "Benchmark for filter methods for feature selection in high-dimensional classification data," *Computational Statistics & Data Analysis*, vol. 143, pp. 1–19, Mar. 2020, https://doi.org/10.1016/j.csda.2019.106839.

[21] N. Pudjihartono, T. Fadason, A. W. Kempa-Liehr, and J. M. O'Sullivan, "A Review of Feature Selection Methods for Machine Learning-Based Disease Risk Prediction," *Frontiers in Bioinformatics*, vol. 2, Jun. 2022, https://doi.org/10.3389/fbinf.2022.927312.

[22] M. Rostami, K. Berahmand, E. Nasiri, and S. Forouzandeh, "Review of swarm intelligence-based feature selection methods," *Engineering Applications of Artificial Intelligence*, vol. 100, p. 104210, Apr. 2021, https://doi.org/10.1016/j.engappai.2021.104210.

[23] S. Afrin, F. M. J. M. Shamrat, T. I. Nibir, M. F. Muntasim, M. S. Moharram, M. M. Imran, and M. Abdulla, "Supervised machine learning based liver disease prediction approach with LASSO feature selection," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 6, pp. 3369–3376, Dec. 2021, https://doi.org/10.11591/eei.v10i6.3242.

[24] J. S. Ko, J. Byun, S. Park, and J. Y. Woo, "Prediction of insufficient hepatic enhancement during the Hepatobiliary phase of Gd-EOB DTPA-enhanced MRI using machine learning classifier and feature selection algorithms," *Abdominal Radiology*, vol. 47, no. 1, pp. 161–173, Jan. 2022, https://doi.org/10.1007/s00261-021-03308-0.

[25] N. Biswas, M. M. Ali, M. A. Rahaman, M. Islam, M. R. Mia, S. Azam, K. Ahmed, F. M. Bui, F. A. Al-Zahrani, and M. A. Moni, "Machine Learning-Based Model to Predict Heart Disease in Early Stage Employing Different Feature Selection Techniques," *BioMed Research International*, vol. 2023, no. 1, p. 6864343, Jan. 2023, https://doi.org/10.1155/2023/6864343.

[26] R. Spencer, F. Thabtah, N. Abdelhamid, and M. Thompson, "Exploring feature selection and classification methods for predicting heart disease," *Dgital Healt*, vol. 6, Jan. 2020, https://doi.org/10.1177/2055207620914777.

[27] A. Soni, "Performance Analysis of Classification Algorithms on Liver Disease Detection," in *2021 IEEE Mysore Sub Section International Conference (MysuruCon)*. Hassan, India: IEEE, Oct. 2021, pp. 1–5, https://doi.org/10.1109/mysurucon52639.2021.9641684.

[28] X. Zhang, L. Yu, H. Yin, and K. K. Lai, "Integrating data augmentation and hybrid feature selection for small sample credit risk assessment with high dimensionality," *Computers & Operations Research*, vol. 146, p. 105937, Oct. 2022, https://doi.org/10.1016/j.cor.2022.105937.

[29] O. A. Akinola, J. O. Agushaka, and A. E. Ezugwu, "Binary dwarf mongoose optimizer for solving high-dimensional feature selection problems," *PLOS ONE*, vol. 17, no. 10, p. e0274850, Oct. 2022, https://doi.org/10.1371/journal.pone.0274850.

[30] D. P. M. Abellana and D. M. Lao, "A new univariate feature selection algorithm based on the best–worst multi-attribute decision-making method," *Decision Analytics Journal*, vol. 7, p. 100240, Jun. 2023.

[31] M. Amiriebrahimabadi and N. Mansouri, "A comprehensive survey of feature selection techniques based on whale optimization algorithm," *Multimedia Tools and Applications*, vol. 83, no. 16, pp. 47 775–47 846, Oct. 2023, https://doi.org/10.1007/s11042-023-17329-y.

[32] D. Theng and K. K. Bhoyar, "Feature selection techniques for machine learning: A survey of more than two decades of research," *Knowledge and Information Systems*, vol. 66, no. 3, pp. 1575–1637, Mar. 2024, https://doi.org/10.1007/s10115-023-02010-5.

[33] R. Amin, R. Yasmin, S. Ruhi, M. H. Rahman, and M. S. Reza, "Prediction of chronic liver disease patients using integrated projection based statistical feature extraction with machine learning algorithms," *Informatics in Medicine Unlocked*, vol. 36, p. 101155, 2023, https://doi.org/10.1016/j.imu.2022.101155.

[34] M. P. Behera, A. Sarangi, D. Mishra, and S. K. Sarangi, "A Hybrid Machine Learning algorithm for Heart and Liver Disease Prediction Using Modified Particle Swarm Optimization with Support Vector Machine," *Procedia Computer Science*, vol. 218, pp. 818–827, 2023, https://doi.org/10.1016/j.procs.2023.01.062.

[35] R. K. Sachdeva, P. Bathla, P. Rani, V. Solanki, and R. Ahuja, "A systematic method for diagnosis of hepatitis disease using machine learning," *Innovations in Systems and Software Engineering*, vol. 19, no. 1, pp. 71–80, Mar. 2023, https://doi.org/10.1007/s11334-022-00509-8.

[36] T. A. Assegie, "Support Vector Machine And K-Nearest Neighbor Based Liver Disease Classification Model," *Indonesian Journal of electronics, electromedical engineering, and medical informatics*, vol. 3, no. 1, pp. 9–14, Feb. 2021, https://doi.org/10.35882/ijeeemi.v3i1.2.

[37] M. A. Jamil and S. Khanam, "Influence of One-Way ANOVA and Kruskal–Wallis Based Feature Ranking on the Performance of ML Classifiers for Bearing Fault Diagnosis," *Journal of Vibration Engineering & Technologies*, vol. 12, no. 3, pp. 3101–3132, Mar. 2024, https://doi.org/10.1007/s42417-023-01036-x.

[38] J. Tripathy, R. Dash, B. K. Pattanayak, S. K. Mishra, T. K. Mishra, and D. Puthal, "Combination of Reduction Detection Using TOPSIS for Gene Expression Data Analysis," *Big Data and Cognitive Computing*, vol. 6, no. 1, Feb. 2022, https://doi.org/10.3390/bdcc6010024.

[39] C. Jacob, C. Gopakumar, and F. Nazarudeen, "Optimized Radiomics-Based Machine Learning Approach for Lung Cancer Subtype Classification," *Biomedical Engineering: Applications, Basis and Communications*, vol. 35, no. 5, Oct. 2023, https://doi.org/10.4015/s1016237223500230.

[40] A. Ganji, D. Usha, and P. Rajakumar, "Enhanced Early Diagnosis of Liver Diseases Using Feature Selection and Machine Learning Techniques on the Indian Liver Patient Dataset," *Scalable Computing: Practice and Experience*, vol. 26, no. 3, pp. 1104–1115, Apr. 2025, https://doi.org/10.12694/scpe.v26i3.4254.

[41] I. R. Hikmah and R. N. Yasa, "Perbandingan Hasil Prediksi Diagnosis pada Indian Liver Patient Dataset (ILPD) dengan Teknik Supervised Learning Menggunakan Software Orange," *Jurnal Telematika*, vol. 16, no. 2, pp. 69–76, 2021, https://doi.org/10.61769/telematika.v16i2.402.

[42] A. Gulia, R. Vohra, and P. Rani, "Liver Patient Classification Using Intelligent Techniques," *International Journal of Computer Science and Information Technologies*, vol. 5, no. 4, pp. 5110–5115, 2014.

[43] E. O. Abiodun, A. Alabdulatif, O. I. Abiodun, M. Alawida, A. Alabdulatif, and R. S. Alkhawaldeh, "A systematic review of emerging feature selection optimization methods for optimal text classification: The present state and prospective opportunities," *Neural Computing and Applications*, vol. 33, no. 22, pp. 15 091–15 118, Nov. 2021, https://doi.org/10.1007/s00521-021-06406-8.

[44] H. Nasiri and S. A. Alavi, "A Novel Framework Based on Deep Learning and ANOVA Feature Selection Method for Diagnosis of COVID-19 Cases from Chest X-Ray Images," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–11, Jan. 2022, https://doi.org/10.1155/2022/4694567.

[45] M. Alassaf and A. M. Qamar, "Improving Sentiment Analysis of Arabic Tweets by One-way ANOVA," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 6, pp. 2849–2859, Jun. 2022, https://doi.org/10.1016/j.jksuci.2020.10.023.

[46] S. Williamson, K. Vijayakumar, and V. J. Kadam, "Predicting breast cancer biopsy outcomes from BI-RADS findings using random forests with chi-square and MI features," *Multimedia Tools and Applications*, vol. 81, no. 26, pp. 36 869–36 889, Nov. 2022, https://doi.org/10.1007/s11042-021-11114-5.

[47] A. S. Sumant and D. Patil, "Ensemble Feature Subset Selection: Integration of Symmetric Uncertainty and Chi-Square techniques with RReliefF," *Journal of The Institution of Engineers (India): Series B*, vol. 103, no. 3, pp. 831–844, Jun. 2022, https://doi.org/10.1007/s40031-021-00684-5.

[48] Y. E. Isik, Y. Gormez, Z. Aydin, and B. Bakir-Gungor, "The Determination of Distinctive Single Nucleotide Polymorphism Sets for the Diagnosis of Behçet's Disease," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 19, no. 3, pp. 1909–1918, May 2022, https://doi.org/10.1109/tcbb.2021.3053429.

[49] J. Azimjonov and T. Kim, "Stochastic gradient descent classifier-based lightweight intrusion detection systems using the efficient feature subsets of datasets," *Expert Systems with Applications*, vol. 237, p. 121493, Mar. 2024, https://doi.org/10.1016/j.eswa.2023.121493.

[50] P. Dhal and C. Azad, "A comprehensive survey on feature selection in the various fields of machine learning," *Applied Intelligence*, vol. 52, no. 4, pp. 4543–4581, Mar. 2022, https://doi.org/10.1007/s10489-021-02550-9.

[This page intentionally left blank.]