Journal of Applied Informatics and Computing (JAIC)
Vol.9, No.6, December 2025, pp. 3012~3024
e-ISSN: 2548-6861

Performance Comparison of Multilayer Perceptron (MLP) and Random Forest for Early Detection of Cardiovascular Disease

Dita Widayanti Setiawan [1]*, Nouval Trezandy Lapatta [2], Amriana [3]*, Deny Wiria Nugraha [4]*, Chairunnisa Ar Lamasitudju [5]***
* Study Program Information Technology, Department of the Faculty of Engineering, Tadulako University
** Master of Information Technology
widayantidita021@gmail.com [1], nouval@untad.ac.id [2], amriana@untad.ac.id [3] , deny.wiria.nugraha@untad.ac.id [4], nisalamasitudju@untad.ac.id [5]

Article Info

Article history:

Received 2025-08-18
Revised 2025-10-27
Accepted 2025-11-05

Keyword:

Cardiovascular Disease,
Early Detection,
Multilayer Perceptron,
Random Forest,
Comparison.

ABSTRACT

Cardiovascular disease is a disorder of the heart and blood vessels that can lead to heart attacks, strokes, and heart failure, so early detection is essential. This study compares Multilayer Perceptron (MLP) and Random Forest for risk classification in a Kaggle dataset containing 70,000 samples with balanced targets. Pre-processing included age conversion, outlier cleaning, standardization, and feature selection based on feature importance. Both models were optimized using RandomizedSearchCV and evaluated using accuracy, precision, recall, F1-score, AUC-ROC, confusion matrix, and k-fold cross-validation. The results show that the accuracy of MLP is 73.90% and Random Forest is 74.23% with an AUC of 0.80 for both. Random Forest is more stable across all folds and performs better on the negative class, while MLP is slightly more sensitive to the positive class. Independent t-test and Mann-Whitney U tests show p>0.05, indicating that the difference in performance is not significant. The most influential features were diastolic blood pressure, age, cholesterol, and systolic blood pressure. The non-clinical Streamlit prototype demonstrated the model's potential for education and initial decision support.

I. INTRODUCTION

Cardiovascular disease is a disorder of the heart and blood vessels that can cause severe conditions such as heart attack, stroke, and heart failure [1]. The WHO notes that 27% of global deaths in 2019 were caused by this disease (WHO, 2020), and in Southeast Asia, the mortality rate reached 124.9 to 421.6 per 100,000 population in 2021 [2]. The latest data from the WHO and the World Heart Federation reports that approximately 19.8 million people died from cardiovascular disease in 2022, which is equivalent to nearly one-third of all global deaths, with more than 85% of cases being heart attacks and strokes. These figures confirm that the burden of cardiovascular disease continues to increase year after year and is a major global health threat [3][4]. In Indonesia, this disease is the leading cause of death, including in Central Sulawesi, with a prevalence of 1.9% and 394 deaths, the highest being in Palu City with 56 deaths [5].

The high incidence rate and lack of adequate treatment, especially in people aged 45 and above, are a serious problem. The different symptoms between men and women, where men more often experience symptoms such as chest pain, while women tend to experience non-specific symptoms, such as shortness of breath, extreme fatigue, or pain in the back and jaw, also cause the risk of misdiagnosis, especially in women who tend to experience non-specific symptoms [6][7].

One solution is early detection using machine learning technology, which can analyze large amounts of data and accurately detect disease patterns [8][9]. In this case, classification algorithms such as Multilayer Perceptron (MLP) and Random Forest can be effective alternatives.

MLP is known to be capable of recognizing complex non-linear patterns through the backpropagation process [9]. At the same time, Random Forest combines multiple decision trees and trains them with random subsets to improve detection accuracy [10].

The selection of MLP and Random Forest represents two complementary approaches in analyzing tabular clinical data. MLP excels at learning complex non-linear relationships between variables. At the same time, Random Forest is more reliable for mixed-scale data because it is less dependent on the scaling process and relatively resistant to outliers and correlations between features. Therefore, comparing the two is relevant for early clinical detection needs using patient data.

Previous studies have tested algorithms such as Random Forest, KNN, and several other algorithms [11][12][13][14][15]. Still, none have specifically compared MLP and Random Forest using a cardiovascular disease dataset of 70,000 data points, accompanied by the development of a Streamlit-based UI, hyperparameter tuning using the RandomizedSearchCV method, feature importance analysis, and comprehensive metric evaluation.

This study aims to compare the performance of Multilayer Perceptron (MLP) and Random Forest algorithms in the early detection of cardiovascular disease, identify the most influential medical features, and determine the best model based on evaluation metrics such as Accuracy, Precision, Recall, F1-Score, AUC-ROC, Confusion Matrix, and Cross-validation.

This study offers three main contributions:

(i) Comparative evaluation of MLP and Random Forest using a large dataset with balanced class distribution.
(ii) Application of cross-validation and hyperparameter tuning to obtain a stable and generalizable model.
(iii) Implementation of an interactive interface prototype to demonstrate the use of the model on patient input.

This study is distinguished from previous ones, which were generally limited to smaller datasets or lacked practical implementation.

This research has the potential to assist general practitioners in early detection, reduce the burden of follow-up examinations, and provide tools for patients to independently understand their disease risks.

This research is expected to contribute to developing an accurate and easy-to-use machine learning-based early detection system for cardiovascular disease and to provide a deeper understanding of risk factors through the visualization and analysis of relevant features.

II. Literature Review

A. Algorithm Comparison Analysis

Algorithm comparison analysis is the process of evaluating two or more machine learning methods to determine the best model based on specific criteria such as accuracy, precision, recall, f1-score, AUC-ROC, confusion matrix, and cross-validation. The use of machine learning in the health sector has been proven to provide practical solutions, ranging from disease detection to the classification of medical data such as heart disease, diabetes, kidney failure, and other diseases [16]. In addition to speeding up diagnosis, machine learning can also improve accuracy and support clinical decision-making by medical personnel. Its ability to process large and complex health data makes it very promising for application in early disease detection systems [17]. Therefore, comparative algorithm analysis is critical to determine the optimal algorithm for early detection of cardiovascular disease [18].

B. Multilayer Perceptron

A Multilayer Perceptron is a type of artificial neural network (ANN) consisting of several layers, namely the input layer, hidden layers, and output layer [19]. The MLP architecture consists of nodes that are directly connected from the input layer to the hidden layer and to the output layer, with weights that are initially randomized and then updated through a learning process [20]. An example of the MLP architecture can be seen in Figure 1.

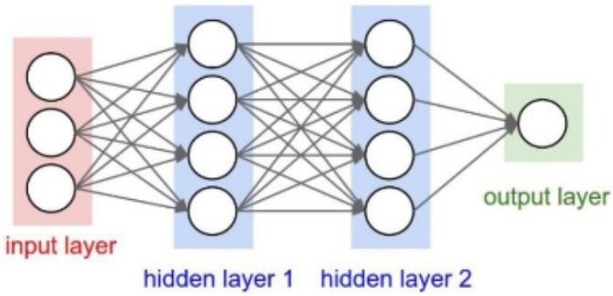

Figure 1. Overview of MLP Architecture
Source: Achmad Reza Fachruroji, 2024

Each layer consists of neurons that are fully connected to the next layer. Each connection between neurons has a weight that determines how strongly the input influences the output [21].

Learning in MLP networks uses the backpropagation algorithm, which consists of three stages: feed-forward, backward, and weight update.

1) The feed-forward stage, where each neuron performs calculations in the form of summing the results of multiplying the input value by its weight, and adding the bias value (v0j) [22]. As formulated in equation (1) below:

$$net_j = v_{0j} + \sum_{i=1}^{n} x_i v_{ij} \quad (1)$$

2) Backward pass, calculating the error value in the neurons in the output layer. This error is obtained from the difference between the target output (tk) and the actual output (yk), then multiplied by the derivative of the activation function of the output neuron. This value is calculated using equation (2). Based on this error value, the weight change (Δwjk) is calculated using equation (3),

which is the product of the error, the output of the hidden neuron (zj), and the learning rate value (α) [22].

$$\delta_k = (t_k - y_k)y_k(1 - y_k) \quad (2)$$

$$\Delta w_{jk} = \alpha\delta_k z_j \quad (3)$$

3) The weight update stage, where the weight values of neurons in the hidden layer to the output layer are updated by summing the old weight values and the weight changes (Δwjk), as shown in equation (4). A similar step is also applied to update the weights from the input layer to the hidden layer, using the weight change (Δvij) as formulated in equation (5)[22].

$$w_{jk}(new) = w_{jk}(old)\Delta w_{jk} \quad (4)$$

$$v_{ij}(new) = v_{ij}(old) + \Delta v_{ij} \quad (5)$$

Next, several vital hyperparameters in the MLP architecture need to be configured to obtain optimal model training results.

1) Activation Function, which converts linear calculation results into a non-linear form [23].
2) Number of Hidden Layers, which processes input through the activation function and connects it to the output layer. The number of neurons in the hidden layer determines the model's capacity to learn patterns from data, which affects the output results [24].
3) Solver, which determines the appropriate algorithm to solve optimization problems [25].
4) Learning Rate, which regulates the magnitude of changes made to the model weights each time training runs (Reyvan Maulid, The Concept of Learning Rate in Machine Learning Algorithms).
5) Regularization, which controls the strength of regularization on the weights [25].

C. Random Forest

Random Forest is a collection of several decision trees formed using random samples, where each tree has a different node separation method. The final prediction of Random Forest can be determined through two approaches.

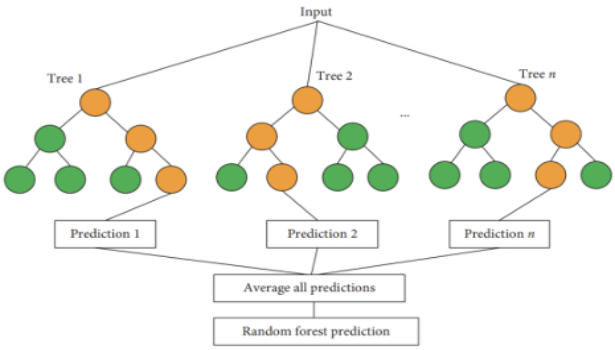

Figure 2. General Structure of the Random Forest Model
Source: Galih Ashari Rakhmat, 2023

The most common approach is to use the majority voting method [26]. The Random Forest algorithm is widely applied in various fields, including healthcare. This is because the process involves creating several decision trees, which are combined to produce more consistent and accurate detection [27]. The following is the general structure of Random Forest, as shown in Figure 2. A decision tree consists of three types of nodes:

1) Root node, which is the topmost part of the tree and is the starting point for data processing.
2) An internal node, which is a branching point that receives at least one input and leads to two or more outputs.
3) A leaf node, which is the final node that only receives one input without producing additional outputs [28].

The decision tree creation process begins with calculating the entropy value to determine how impure an attribute is and the information gain value. The entropy value is calculated using the formula in equation (6), while the information gain value is calculated in equation (7) [29].

$$Entropy\ (Y) = -\sum_i p\,(c|Y)log^2 p(c|Y) \quad (6)$$

Explanation:

Y = Set of cases

P(c|Y) = Proportion of *Y* values in class *c*

$$Information\ Gain(Y, a) = Entropy\ (Y) - \sum_{v\varepsilon Values(a)} \frac{|Y_v|}{|Y_a|} Entropy\ (Y_v) \quad (7)$$

Explanation:

Values (a) = Possible values in case set *a*

Y_v = Subclass of *Y* with class *v* associated with class *a*

Y_a = Semua nilai yang sesuai dengan *a*

To obtain optimal model training results, there are several essential hyperparameters in the Random Forest architecture that need to be configured, namely:

1) *n_estimators*, which determines the number of decision trees [30].
2) max_depth, which determines the maximum depth of a tree [31].
3) *min_samples_split*, which sets the minimum number of samples required to split a node [32].
4) *min_samples_leaf,* the minimum number of samples required to split a node [33].
5) *Bootstrap,* which is a random sampling technique with replacement from the training dataset [34].

In clinical data presented in tables with varying feature scales, Random Forest often shows better stability than feed-forward neural networks due to its ensemble mechanism and insensitivity to scale.

D. Early Detection

Early detection of cardiovascular disease is critical, especially in the productive age population, to prevent more serious complications in the future. By detecting the disease early, individuals have the opportunity to immediately make lifestyle changes, such as improving their diet, increasing physical activity, and undergoing medical treatment as recommended if necessary [35].

Additionally, early detection enables earlier intervention, whether through behavioral changes or medical interventions, aimed at preventing or delaying the progression of cardiovascular disease. This detection also plays a vital role in raising public awareness of the importance of maintaining cholesterol levels, adopting a healthy lifestyle such as consuming a balanced diet, exercising regularly, managing stress, getting enough sleep, and avoiding smoking and excessive alcohol consumption [36]. Thus, the implementation of early detection is not only beneficial for individuals but also helps prevent cardiovascular disease more broadly in the community.

E. Cardiovascular Disease

Cardiovascular disease (CVD) is a group of medical disorders that affect the heart and blood vessels. This type of disease includes various severe conditions such as heart attack, stroke, and heart failure, which can generally interfere with vital bodily functions and cause severe complications [1].

The cardiovascular system itself consists of the heart, blood vessels, and blood components, all of which work together to ensure the distribution of oxygen and nutrients throughout the body. If one of these components is disrupted, the supply of oxygen and essential substances to body tissues will be disrupted, which risks causing various serious diseases [37].

Globally, cardiovascular disease remains one of the leading causes of death. Many factors can increase the risk of this disease, including unhealthy lifestyles such as lack of physical activity, smoking, excessive alcohol consumption, high-fat diets, and obesity. In addition to lifestyle factors, risk can also increase due to non-modifiable factors such as age, gender, and family history of disease [38]. Therefore, understanding cardiovascular disease and its risk factors is crucial as a first step in preventing and detecting this disease early on.

III. RESEARCH METHODOLOGY

The following is a flowchart illustrating the research process.

A. Dataset

The dataset for this study was obtained from Kaggle - Cardiovascular Disease Dataset (https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset). The dataset consists of 70,000 samples with 12 predictor features and 1 target feature. The target distribution is balanced (class 0: 35,021 and class 1: 34,979). The gender composition is 45,530 females and 24,470 males. The average age is 52 years (range 29-64 years). The feature summary is presented in Table I, and the target distribution is shown in Figure 4.

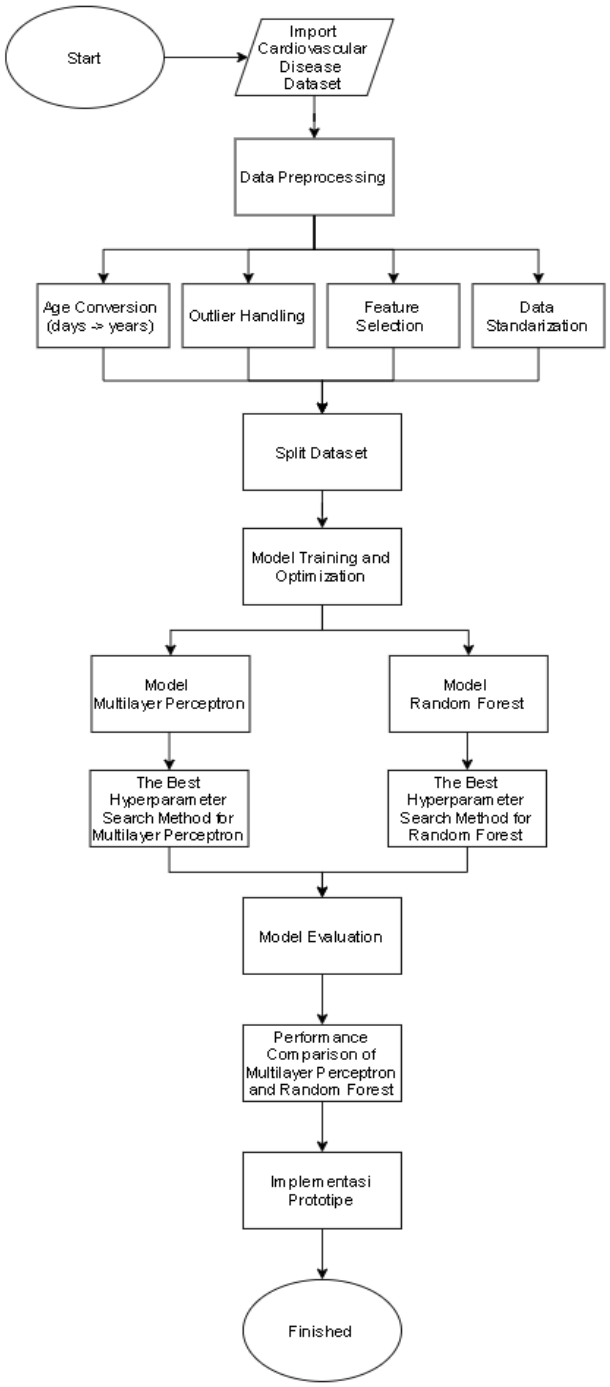

Figure 3. Research Flowchart

TABLE I
FEATURES IN THE CARDIOVASCULAR DISEASE DATASET

NO.	Feature	Description
1.	ID	Unique value for each individual
2.	AGE	Age (in days)
3.	GENDER	Sex (1 = Female, 2 = Male)
4.	HEIGHT	Height (cm)
5.	WEIGHT	Weight (kg)
6.	AP_HI	Systolic Blood Pressure (mmHg)

7.	AP_LO	Diastolic Blood Pressure (mmHg)
8.	CHOLESTEROL	Cholesterol Level (1 = Normal, 2 = Moderate/Above normal, 3 = High/Well above)
9.	GLUC	Glucose Level (1 = Normal, 2 = Moderate/Above normal, 3 = High/Well above)
10.	SMOKE	Smoking habit (1 = Yes, 0 = No)
11.	ALCO	Alcohol consumption (1 = Yes, 0 = No)
12.	ACTIVE	Physical activity (1 = Active, 0 = Inactive)
13.	CARDIO	Target (1 = Cardiovascular Disease, 0 = No Cardiovascular Disease)

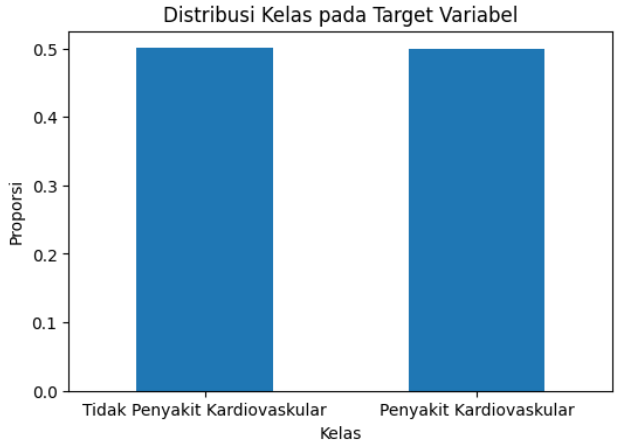

Figure 4. Distribution of "Cardio" Feature Classes

Since the target distribution is balanced, resampling (SMOTE technique) is not performed. Pre-processing focuses on age conversion (days to years), outlier cleaning (IQR/median), standardization, and feature selection (based on feature importance). These characteristics make the dataset suitable for machine learning-based classification tasks.

B. Data Pre-processing

Data pre-processing is carried out in several stages to improve data quality and prepare the data for model training. The following are the steps in data pre-processing:

1) Age Conversion

Convert the "Age" feature from days to years to facilitate data analysis, and age is always input in years as in equation (8).

$$Age\ (years) = \left(\frac{Age\ (days)}{365}\right) \quad (8)$$

2) Outlier Handling

Outliers in the data are detected using the Interquartile Range (IQR) method. This is a descriptive statistical method for measuring the spread or variation of data. IQR helps to understand how far the data is spread from the median value. IQR serves to identify outliers, which are data points that fall outside the lower and upper limits of the interquartile range [39] using formulas (9) and (10).

Lower bound:

$$Lower\ Bound = Q1 - 1.5 \times IQR \quad (9)$$

Values below this bound are considered lower outliers (small extremes).

Upper bound:

$$Upper\ Bound = Q3 + 1.5 \times IQR \quad (10)$$

Values above this bound are considered upper outliers (large extremes).

Explanation:

- Q1 (first quartile) = Middle value of the lower half of the data (25th percentile)
- Q3 (third quartile) = Middle value of the upper half of the data (75th percentile)
- IQR = Difference between Q3 and Q1, i.e., IQR = Q3-Q1

Thus, the values considered outliers are then replaced with the median value to reduce the influence of abnormal data on the model. The results can be seen in Figures 5 and 6..

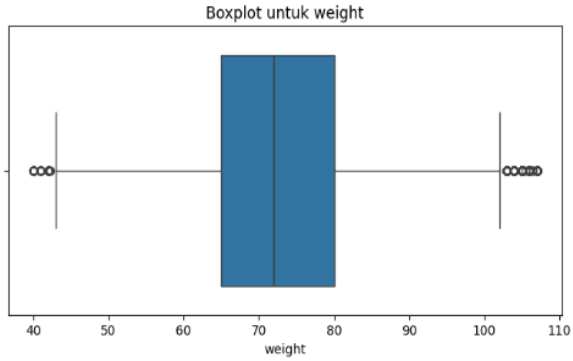

Figure 5. Outlier in the "Weight" Feature

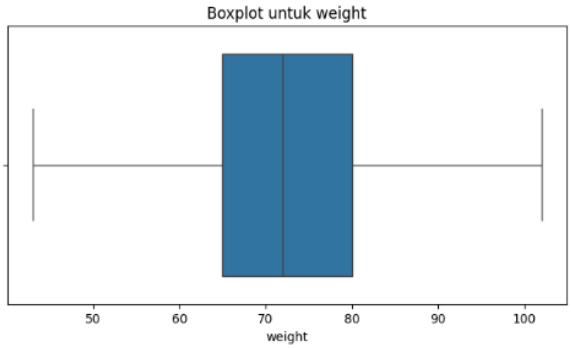

Figure 6. Outliers in the "Weight" Feature After Cleaning

3) Feature Selection

The most relevant features for cardiovascular disease detection were selected using the Feature Importance method calculated with Random Forest. This method was used to identify the features that most influenced the model by removing features that were considered less relevant [29]. Feature Importance analysis with Random Forest, the results of the study can be the basis for more targeted policy recommendations [40]. The results of applying feature importance are shown in Figure 7.

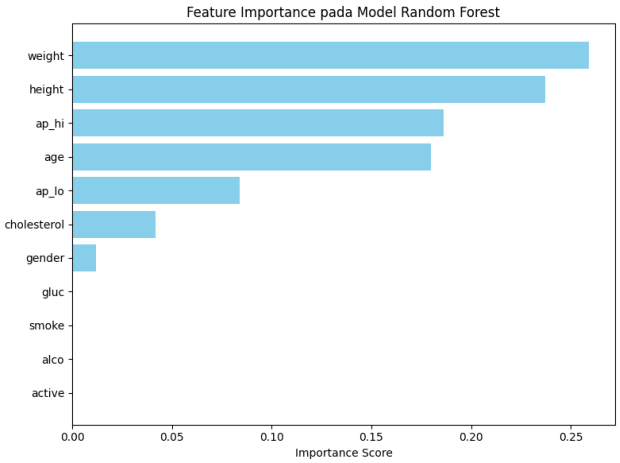

Figure 7. Feature Importance Results

In the Gluc, Smoke, Alco, and Active features, the feature importance results show that they do not provide the most influential feature variation values for the model. This is because these features are less varied or only dominant with 1 value.

4) Data Standardization

Standardization is essential for cardiovascular disease detection because features, such as weight and height, blood pressure, cholesterol, and gender, have different scales, which can disproportionately affect the model. With standardization, machine learning models can work more efficiently, accelerate convergence, and produce more accurate detections [41]. The standardization process is performed using equation (11).

$$z = \frac{x - \mu}{\sigma} \quad (11)$$

Explanation:
x = Original feature value
μ = Mean
σ = Feature standard deviation

C. Split Data

The dataset is divided into two parts, namely 90% (63,000) for training data and 10% (7,000) for test data. This division aims to ensure that the model can generalize well on data that has never been seen before.

D. Model Training and Optimization

After data partitioning, the models were trained and optimized, and hyperparameters were tuned for both models. Hyperparameter tuning was performed using the RandomizedSearchCV method, which selects values for each hyperparameter independently using a probability distribution and takes random sample values. The RandomizedSearchCV method is suitable for hyperparameter tuning because it is more efficient in terms of computation time. RandomizedSearchCV is also more effective when the number of hyperparameters to be tested is quite large or when the search space is high-dimensional [42].

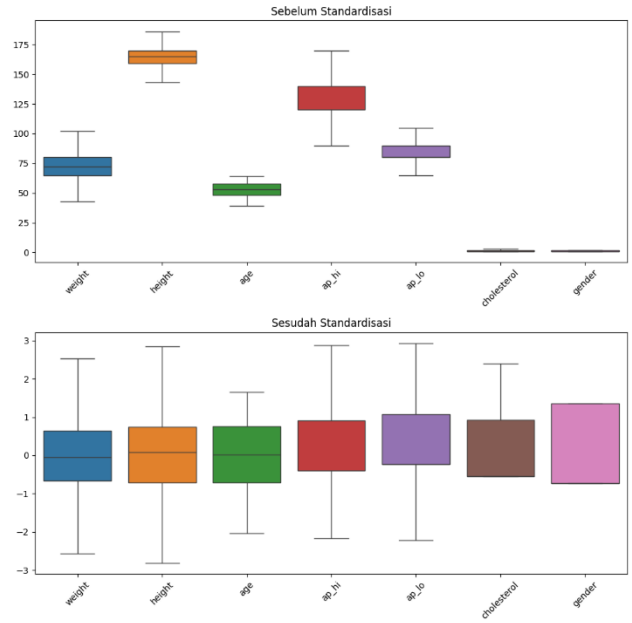

Figure 8. Data Standardization

The following are the values for hyperparameter tuning that will be tested using the RandomizedSearchCV method in Multilayer Perceptron (MLP) model training:

TABLE II
HYPERPARAMETER VALUES TESTED ON MULTILAYER PERCEPTRON

Hyperparameter	Tested Value	Description
Hidden Layer Size	(50,), (100,), (64, 64)	Increases the model's capacity to capture more complex patterns.
Activation Function	RELU, TANH	Assesses its effect on model accuracy and convergence speed.
Learning Rate	ADAPTIVE, CONSTANT	Observing its effect on convergence speed and model stability.
Solver	ADAM, SGD	Observing the effect on convergence speed and stability in the model during training, as well as the model's ability to capture more complex data patterns.
Regularization (Alpha)	0.0001, 0.001, 0.01	Observing whether regularization can reduce overfitting without reducing the model's capacity to learn data patterns.

TABLE III
HYPERPARAMETER VALUES TESTED IN RANDOM FOREST

Hyperparameter	Tested Value	Description
NUMBER OF TREES (N_ESTIMATORS)	100, 200, 300	Increasing the number of trees generally improves accuracy, but also slows down training time.
TREE DEPTH (MAX_DEPTH)	UNLIMITED, 10, 20, 30	A greater tree depth allows the model to capture more complex patterns, but it can also cause overfitting if the model is too deep.
MINIMUM SAMPLES FOR NODE SPLIT (MIN_SAMPLES_SPLIT)	2, 5, 10	Larger values reduce the model's tendency to learn irrelevant details from the training data (overfitting), while smaller values improve the model's ability to capture more detailed patterns.
MINIMUM SAMPLES FOR LEAF (MIN_SAMPLES_LEAF)	1, 2, 4	Prevents the model from splitting the data too small at the bottom of the tree, which can lead to overfitting.
BOOTSTRAP	TRUE, FALSE	Bootstrap sampling enhances tree diversity and mitigates overfitting

E. Model Evaluation

The Multilayer Perceptron and Random Forest models were evaluated using separate test data. The models were evaluated using several metrics to ensure their performance in detecting cardiovascular disease [43]. The evaluation metrics used to measure model performance were:

1) Accuracy is the ratio of the number of correct predictions to the total test data.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (12)$$

2) Precision is the ratio of the number of correct optimistic predictions to all positive predictions generated by the model.

$$precision = \frac{TP}{TP + FP} \times 100\% \quad (13)$$

3) Recall measures how well the model finds all positive examples in the dataset. A high recall value means that the model can identify most of the positive data.

$$recall = \frac{TP}{TP + FN} \times 100\% \quad (14)$$

4) F1-score is the harmonic mean of Precision and Recall, which is used when we want to find a balance between the two metrics.

$$f1 - score = 2 \times \frac{precision \times recall}{precision + recall} \quad (15)$$

5) AUC-ROC Curve, evaluated using the metrics module in scikit-learn; the X-axis is False Positive Rate and the Y-axis is True Positive Rate. A higher AUC value indicates better class separation [44].
6) The confusion matrix presents the number of correct/incorrect predictions for each class, comparing actual labels to predictions, summarized as TP, TN, FP, and FN [44].
7) Cross-validation is a technique for resampling data to evaluate the predictive ability of a model and avoid overfitting. It employs the K-fold Cross-validation technique, which measures the extent to which a machine learning model can be generalized to different test data. The data is divided into k sets of approximately equal size. In the context of classification, the model is trained and tested k times. In each iteration, one of the sets is used as training data, and the others are used as testing data [45].

F. Comparison of MLP and Random Forest Performance

The performance of both models was compared based on accuracy, precision, recall, F1-score, AUC-ROC, and k-fold cross-validation metrics. For clinical tabular data such as this, Random Forest is generally more stable (robust against scaling and outliers). At the same time, MLP is more sensitive in detecting positive classes (potentially reducing false negatives).

G. Prototype Implementation

As a practical contribution, this study developed a simple interactive interface prototype using Streamlit for demonstration and educational purposes (not as a clinical tool). Users input seven features via sliders for weight (kg), height (cm), age (years), ap_hi (mmHg), ap_lo (mmHg), cholesterol (1–3), gender (1=female, 2=male), select a model (MLP or Random Forest), then press the Detect button. The results are class labels (No/Yes cardiovascular disease) and probabilities (%).

IV. RESULTS AND DISCUSSION

Based on the test results conducted in this study, the dataset used is the cardiovascular disease dataset from Kaggle (70,000 samples). Seven relevant features and one target were selected based on the feature importance method from Random Forest (Table IV). In addition to evaluating performance metrics, this study highlights novelty through stability testing via cross-validation and an interactive interface prototype.

TABLE IV
CARDIOVASCULAR DISEASE DATASET FEATURES USED

NO.	FEATURE	DESCRIPTION
1.	WEIGHT	Weight (kg)
2.	HEIGHT	Height (cm)
3.	AP_HI	Systolic Blood Pressure (mmHg)
4.	AGE	Age (in years)

5.	AP_LO	Diastolic Blood Pressure (mmHg)
6.	CHOLESTEROL	Cholesterol Level (1 = Normal, 2 = Moderate/Above normal, 3 = High/Well above)
7.	GENDER	Gender (1 = Female, 2 = Male)
8.	CARDIO	Target (1 = Cardiovascular Disease, 0 = No Cardiovascular Disease)

A. Hyperparameter Optimization and Tuning Results

MLP and Random Forest optimization were performed using the RandomizedSearchCV method:

1) MLP: 10 iterations with 5-fold cross-validation.
2) Random Forest: n_iter = 30 (30 kombinasi) dengan 3-fold cross-validation.

The complete results of the best configurations for both models are presented in Tables V and VI.

TABLE V
RESULTS OF MULTILAYER PERCEPTRON HYPERPARAMETER OPTIMIZATION AND TUNING

HYPERPARAMETER	TESTED VALUE	BEST VALUE
HIDDEN LAYERS SIZE	(50,), (100,), (64, 64)	(64, 64)
ACTIVATION FUNCTION	RELU, TANH	RELU
LEARNING RATE	ADAPTIVE, CONSTANT	CONSTANT
SOLVER	ADAM, SGD	ADAM
REGULARIZATION (ALPHA)	0.0001, 0.001, 0.01	0.01

TABLE VI
RESULTS OF OPTIMIZATION AND TUNING OF RANDOM FOREST HYPERPARAMETERS

HYPERPARAMETER	TESTED VALUE	BEST VALUE
NUMBER OF TREES (N_ESTIMATORS)	100, 200, 300	100
TREE DEPTH (MAX_DEPTH)	UNLIMITED, 10, 20, 30	10
MINIMUM SAMPLES FOR NODE SPLIT (MIN_SAMPLES_SPLIT)	2, 5, 10	2
MINIMUM SAMPLES FOR LEAF (MIN_SAMPLES_LEAF)	1, 2, 4	4
BOOTSTRAP	TRUE, FALSE	TRUE

B. Baseline Testing

For the initial benchmark without pre-processing, a Dummy Classifier (most_frequent) was used, which always predicts the majority class. The results are used as a comparison for the performance of MLP and Random Forest, as shown in Table VII.

TABLE VII
RESULTS OF BASELINE TESTING EVALUATION

MODEL	CLASS	EVALUATION METRIC
BASELINE	0	PRECISION = 50.03%
		RECALL = 100.00%
		F1-SCORE = 66.69%
	1	PRECISION = 0.00%
		RECALL= 0.00%
		F1-SCORE = 0.00%
ACCURACY		50.03%

C. Model Evaluation Results

Both models were evaluated using test data. The evaluation results used several metrics that provided a broad overview of each model's performance in predicting cardiovascular disease.

TABLE VIII
RESULTS OF THE ACCURACY EVALUATION OF MULTILAYER PERCEPTRON AND RANDOM FOREST ON TEST DATA

MODEL	CLASS	EVALUATION METRIC
MULTILAYER PERCEPTRON	0	PRECISION = 72.76%
		RECALL = 76.44%
		F1-SCORE = 74.56%
	1	PRECISION = 75.16%
		RECALL= 71.36%
		F1-SCORE = 73.21%
ACCURACY		73.90%
RANDOM FOREST	0	PRECISION = 71.93%
		RECALL= 79.53%
		F1-SCORE = 75.54%
	1	PRECISION = 77.08%
		RECALL = 68.93%
		F1-SCORE = 72.77%
ACCURACY		74.23%

Based on Table VIII, the accuracy of both models is almost the same (MLP 73.90%, RF 74.23%; a difference of 0.33 points). Random Forest excels in class 0 recall (79.53% vs. 76.44%), while MLP excels in class 1 recall (71.36% vs. 68.93%). The AUC of both is the same (0.80), indicating equivalent class separation ability; the practical difference lies in the trade-off between classes.

To clarify the comparison of metrics, Table IX presents a summary of the accuracy, AUC, macro precision, macro recall, and macro F1 of both models.

TABLE IX
COMPARISON OF MULTILAYER PERCEPTRON AND RANDOM FOREST METRICS

METRICS	MULTILAYER PERCEPTRON	RANDOM FOREST
ACCURACY	73.90%	74.23%
AUC	0.80	0.80
PRECISION MACRO	73.96%	74.50%
RECALL MACRO	73.90%	74.23%
F1 MACRO	73.88%	74.15%

The research data is tabular clinical data (encoded numerical and categorical features) with non-uniform scales and residual outliers. Under such conditions, tree-based ensembles are more robust due to threshold-based separation, scale insensitivity, and stability thanks to the aggregation of many trees. In contrast, MLP is sensitive to scaling and regularization settings, causing its performance to vary more between folds. This variation explains Random Forest's slight but consistent advantage in accuracy and stability.

Practically, Random Forest is more suitable for clinical decision support systems because it is stable and resistant to overfitting. MLP is useful when suppressing false negatives is a priority (positive class sensitivity). The consistency of Random Forest cross-validation (0.73 in all folds) supports the claim of stability, while MLP is slightly more volatile. With a slight difference in accuracy, Random Forest excels in stability and generalization, while MLP is more balanced in detecting positive and negative cases. In addition, the interface prototype enables rapid self-detection and increases health awareness.

In addition, when compared to several previous studies using similar datasets, the results of this study show competitive performance. In a study conducted by I Ketut Adian Jayaditya (2023) using Random Forest, an accuracy of 73.06% was obtained, while Cindy Putri Azizan's (2019) study reported an accuracy of 73.20% with an AUC of 80.30%. The latest study by Peri Hidayat (2025), which used the KNN algorithm, obtained an accuracy of 71.16%. This figure is in line with the results of this study (Random Forest: 74.23% and MLP: 73.90%), which shows that optimal hyperparameter tuning and pre-processing can produce better and more stable performance.

D. Confusion Matrix

Visualize the number of correct and incorrect predictions per category (positive and negative) to assess the error patterns of both models.

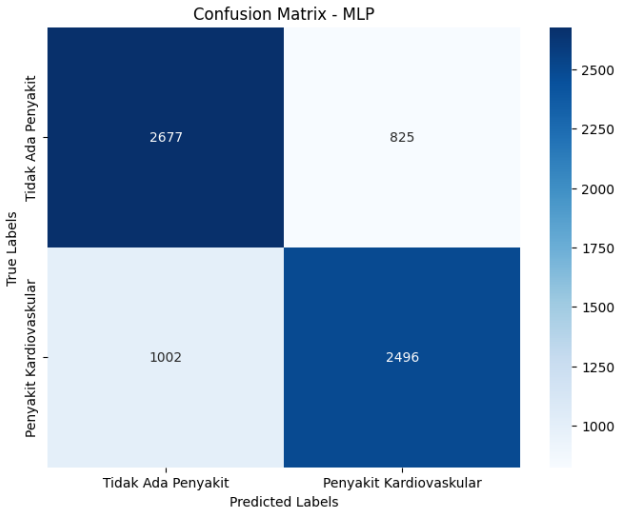

Figure 9. Confusion Matrix MLP

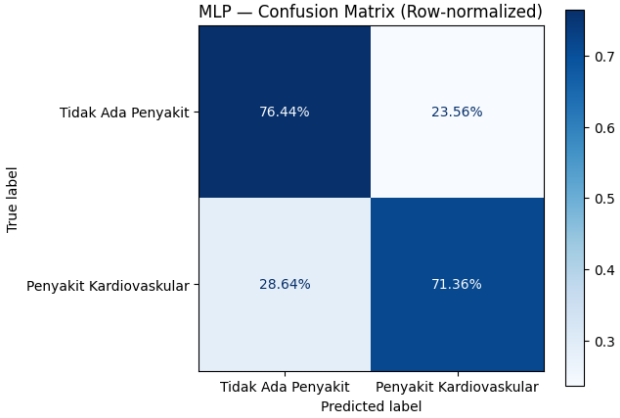

Figure 10. Confusion Matrix Visualization of MLP Prediction Distribution

Keterangan:

1) True Negative: 2,677 (76.44%).
2) True Positive: 2,496 (71.36%).
3) False Positive: 825 (23.56%) - negative predicted positive.
4) False Negative: 1,002 (28.64%) - positive predicted as negative.

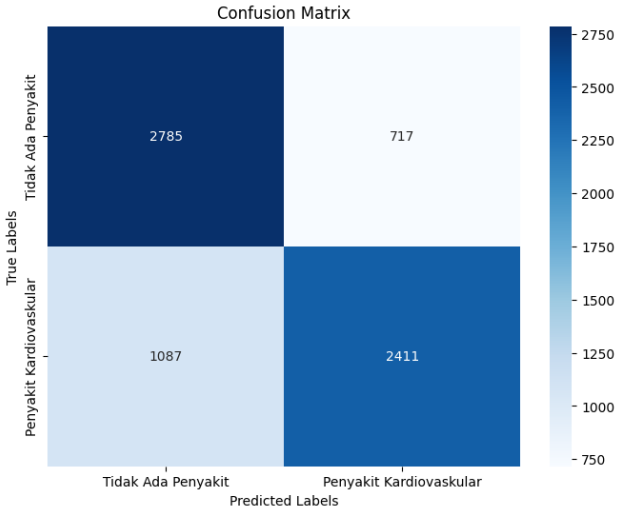

Figure 11. *Confusion Matrix Random Forest*

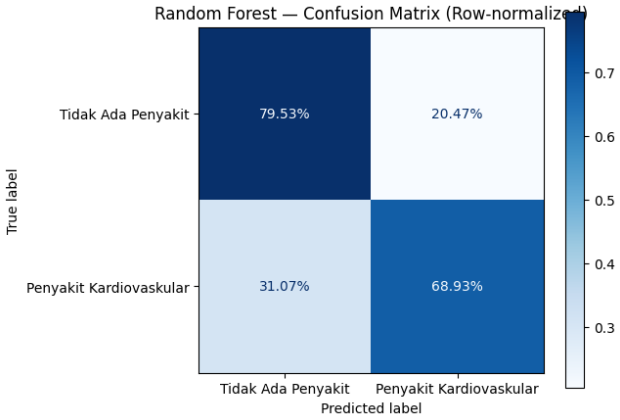

Figure 12. Confusion Matrix Visualization of Random Forest Prediction Distribution

Keterangan:
1) True Negative: 2,785 (79.53%).
2) True Positive: 2,411 (68.93%).
3) False Positive: 717 (20.47%) - negative predicted positive.
4) False Negative: 1,087 (31.07%) - positive predicted negative.

E. Cross-Validation

K-fold cross-validation is used to maintain model generalization. MLP uses 5-fold, Random Forest uses 3-fold. The results are shown in Figures 13-14.

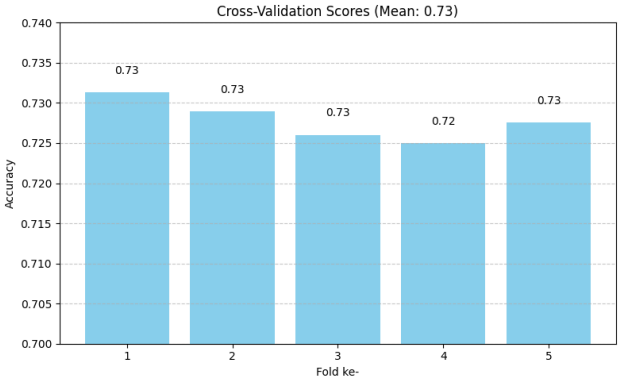

Figure 13. Cross-validation MLP

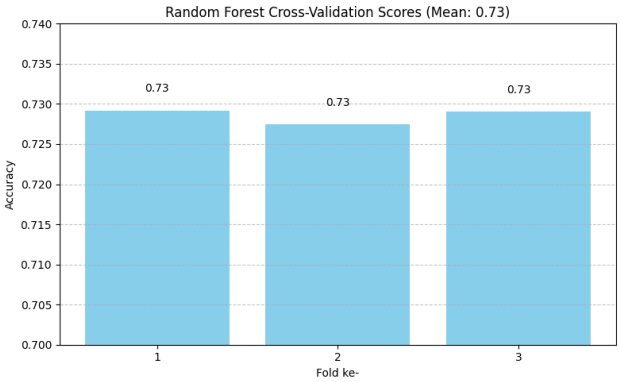

Figure 14. Cross-validation Random Forest

Based on the cross-validation evaluation results of both models, the average accuracy obtained was 0.73. MLP showed slight variations in accuracy between folds (0.72–0.73), while Random Forest had consistent accuracy across all folds (0.73). This indicates that the performance of both models is relatively similar, but Random Forest is slightly more stable.

F. Significance Test of Model Performance Differences

The performance differences between MLP and Random Forest were tested using an independent t-test and Mann-Whitney U test on the accuracy of cross-validation results. The data was independent because the folds were different, with MLP using 5-fold and Random Forest using 3-fold. The results were $p = 0.5495$ (t-test) and $p = 0.5714$ (Mann-Whitney), $p > 0.05$, which means that there was no statistically significant difference between the performance of the two models. Although the accuracy of Random Forest was 74.23%, slightly above the accuracy of MLP at 73.90%, both were statistically equivalent. Model stability is more relevant than a slight difference in accuracy, so both are suitable for early detection of cardiovascular disease.

G. AUC-ROC Curve

The ROC (Receiver Operating Characteristic) curve shows the relationship between the True Positive Rate and False Positive Rate at various thresholds. AUC (Area Under the Curve) is the area under the curve. The larger the AUC, the better the model distinguishes between the two classes. The curve for MLP is shown in Figure 15, and for Random Forest in Figure 16.

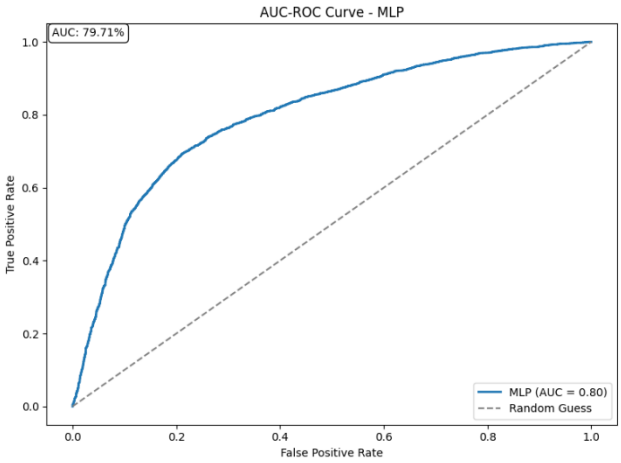

Figure 15. AUC-ROC Curve Multilayer Perceptron

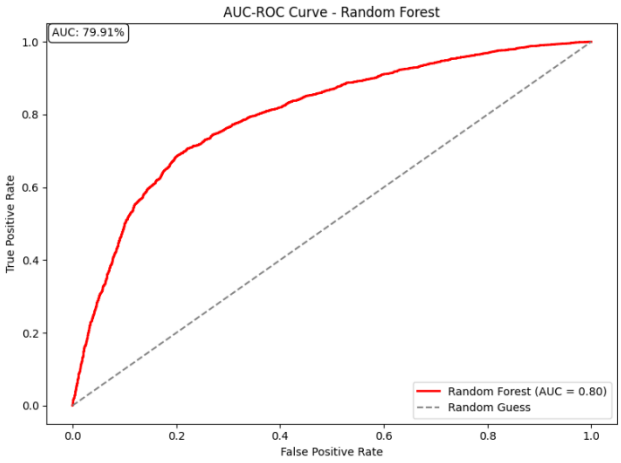

Figure 16. AUC-ROC Curve Random Forest

At low False Positive Rate ranges, Random Forest tends to achieve a higher True Positive Rate, while MLP is slightly more even across thresholds, but the AUC of both is the same at 0.80 (MLP 79.71%, and Random Forest 79.91%).

H. Detection Flow Through Decision Trees

A tree from the Random Forest model is visualized to explain the decision rules. The tree was trained on 63,000 training samples with key features such as ap_lo, age, and cholesterol. Each node represents a split based on feature values until it reaches the leaf (final class: No Cardiovascular Disease or Cardiovascular Disease).

Node descriptions:

1) ap_lo (root): the most informative feature at the root.
2) Gini: node purity (0 = pure and 0.5 = mixed).
3) Samples: number of data points at the node.
4) Value: distribution of the two classes (No Cardiovascular Disease and Cardiovascular Disease).
5) Class: the majority class in the node.

The following is the primary separation flow:

1) Root: ap_lo ≤ 0.488 to the left branch (predominantly No Cardiovascular Disease), ap_lo > 0.488 to the right branch (predominantly Cardiovascular Disease).
2) Left branch (ap_lo ≤ 0.488)
   a. age ≤ 0.245 splits to cholesterol.
   b. Cholesterol ≤ 1.666 tends toward No Cardiovascular Disease.
   c. Further splitting on age, height (still predominantly No Cardiovascular Disease).
   d. Cholesterol> 1.666 tends to indicate Cardiovascular Disease.
   e. Further separation on weight, ap_hi (still predominantly Cardiovascular Disease).
   f. Age> 0.245 also evaluates cholesterol with a similar pattern (reinforcing the decision through further splitting).
3) Right branch (ap_lo > 0.488)
   a. ap_lo ≤ 2.071 separate to ap_hi.
   b. ap_hi ≤ 0.815 then ap_hi ≤ 0.193 or ≤1.471 generally Cardiovascular Disease.
   c. ap_lo > 2.071 to another right node (ap_lo ≤ 2.335) still leads to Cardiovascular Disease, with further breakdown on cholesterol and height.

Interpretation summary:

1) ap_lo (diastolic blood pressure) is the first and most informative discriminator.
2) The cholesterol and ap_hi values reinforce the decision, particularly for the Cardiovascular Disease class on the right branch.
3) In the left branch (lower diastolic), age and cholesterol help distinguish between No Cardiovascular Disease and Cardiovascular Disease.

I. Prototype Implementation Results

To demonstrate the use of the model in user scenarios, a simple interface prototype was created (Figure 17). This interface allows users to enter patient data and obtain model detection results directly. This prototype is intended only as a demonstration and is not a clinical application.

In the example scenario (BB = 70 kg, TB = 170 cm, age = 55 years, ap_hi (systolic blood pressure) = 120, ap_lo (diastolic blood pressure) = 110, cholesterol level = 2 (high), gender = 1 (female), the interface displays the MLP detection of Cardiovascular Disease (90.13%). The output for Random Forest is similar in the form of labels and probabilities, but the values may differ slightly depending on the model's threshold and probability settings.

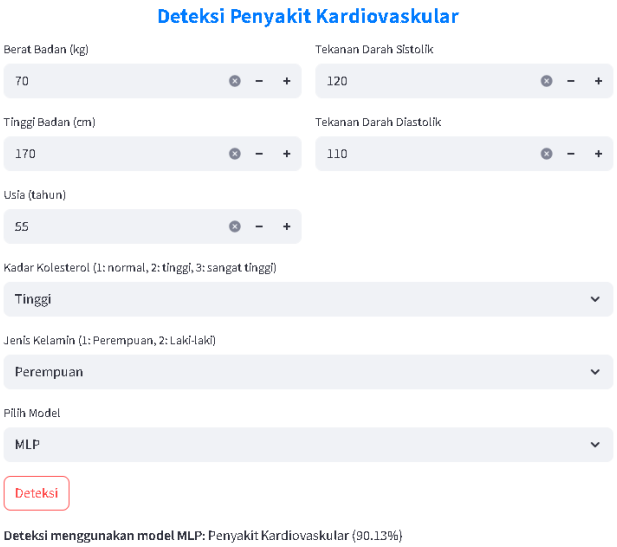

Figure 17. Prototype Interface for Entering Input Features and Displaying Detection of Both Models (Probability and Class)

V. CONCLUSION

This study compares the performance of Multilayer Perceptron (MLP) and Random Forest for early detection of cardiovascular disease in a balanced dataset containing 70,000 samples. After pre-processing, standardization, and tuning with the RandomizedSearchCV method, both yielded nearly identical results, with MLP accuracy at 73.90% and Random Forest at 74.23%, both with an AUC of 0.80. Random Forest was more stable in cross-validation (0.73 in all folds) and better at capturing negative classes, while MLP was slightly superior in positive classes, thus potentially reducing false negatives. Independent t-tests and Mann–Whitney U tests showed p > 0.05, meaning that the difference in performance was not significant. The most influential features are diastolic blood pressure (ap_lo), age, cholesterol, and systolic blood pressure (ap_hi). Practically, Random Forest is more suitable as a stable base model for clinical decision support systems, while MLP is chosen when the main objective is to improve sensitivity. The Streamlit prototype shows that the model can be used for non-clinical education and demonstration. In the future, external validation and the addition of clinical variables are recommended to enhance the model's reliability in the field.

DAFTAR PUSTAKA

[1] I. Naryanti, N. Dwi Astuti, M. Hani Baity Jannaty, and P. Kemenkes Semarang, "Literature Review Faktor Determinan Penyakit Kardiovaskuler di Indonesia," *J. Mandalika Lit.*, vol. 6, no. 1, pp. 2745–5963, 2024.

[2] T. Apriliyani, N. Ainun Oktavia Pusparini, Z. Rohmah, W. Anindito Sri Tunjung, and dan Ardaning Nuriliani, "Mekanisme Penyakit Kardiovaskular Terkait Penuaan (Mechanisms of Cardiovascular Diseases Related to Aging)," *Bioma Berk. Ilm. Biol.*, vol. 26, no. 2, pp. 2598–2370, 2024.

[3] "World Health Organization."

[4] "World Heart Federation."

[5] "TATURA.ID."

[6] dkk Dzulfian Syafrian, "Analisis Struktur Kovarians Indikator Kesehatan pada Lansia di Rumah dengan Fokus pada Persepsi Kesehatan Subjektif," *Sustain.*, vol. 11, no. 1, pp. 1–14, 2025

[7] I. Wahyuniari, D. Ratnayanti, Mayun, S. Wiryawan, Linawati, and Sugiritama, "Deteksi Dini Dan Penanganan Faktor Risiko Penyakit Kardiovaskular," vol. 9, no. 2, pp. 72–74, 2007.

[8] R. D. Muhammad, "Prediksi Penyakit Jantung dengan menggunakan Machine Learning Autogluon," 2024, [Online]. Available: https://dspace.uii.ac.id/handle/123456789/48775

[9] I. Daniel, A. F. Limas Ptr, and A. Ichsan, "Klasifikasi Risiko Penyakit Jantung Dengan Multilayer Perceptron," *Data Sci. Indones.*, vol. 4, no. 1, pp. 78–82, Sep. 2024, doi: 10.47709/dsi.v4i1.4667.

[10] N. H. Alfajr and S. Defiyanti, "Prediksi Penyakit Jantung Menggunakan Metode Random Forest Dan Penerapan Principal Component Analysis (Pca)," *J. Inform. Dan Tek. Elektro Terap.*, vol. 12, no. 3S1, 2024, doi: 10.23960/jitet.v12i3s1.5055.

[11] Y. Amelia, "Perbandingan Metode Machine Learning Untuk Mendeteksi Penyakit Jantung," 2023. [Online]. Available: http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index

[12] I. Ketut *et al.*, "Implementasi Random Forest pada Klasifikasi Penyakit Kardiovaskular dengan Hyperparameter Tuning Grid Search," *Jnatia*, vol. 2, no. 1, pp. 219–226, 2023.

[13] C. Putri and A. Maritza, "Analisis Perbandingan Metode Klasifikasi Menggunakan Regresi Logistik Biner , Random Forest , dan Support Vector Machine pada Cardiovascular Disease Dataset," *Dep. Stat.*, pp. 1–12, 2019.

[14] Y. E. Windarto, "Analisis Penyakit Kardiovaskular Menggunakan Metode Korelasi Pearson, Spearman Dan Kendall," *J. SAINTEKOM*, vol. 10, no. 2, p. 119, 2020, doi: 10.33020/saintekom.v10i2.149.

[15] P. Hidayat, R. Kurniawan, Y. A. Wijaya, and T. Suprapti, "Optimasi Algoritma K-Nearest (Knn) Neighbors Pada Prediksi Risiko Penyakit Kardiovaskular," *J. Inform. Dan Tek. Elektro Terap.*, vol. 13, no. 1, 2025, doi: 10.23960/jitet.v13i1.5864.

[16] I. Akbar, F. Supriadi, and D. I. Junaedi, "Pemanfaatan machine learning di bidang kesehatan," vol. 9, no. 1, pp. 1744–1749, 2025.

[17] F. D. Telaumbanua, P. Hulu, T. Z. Nadeak, R. R. Lumbantong, and A. Dharma, "Penggunaan Machine Learning Di Bidang Kesehatan," *J. Teknol. Dan Ilmu Komput. Prima*, vol. 2, no. 2, pp. 391–399, 2020, doi: 10.34012/jutikomp.v2i2.657.

[18] A. R. Isnain, H. Sulistiani, B. M. Hurohman, A. Nurkholis, and S. Styawati, "Analisis Perbandingan Algoritma LSTM dan Naive Bayes untuk Analisis Sentimen," *J. Edukasi dan Penelit. Inform.*, vol. 8, no. 2, p. 299, 2022, doi: 10.26418/jp.v8i2.54704.

[19] D. Pardede, B. H. Hayadi, and Iskandar, "Kajian Literatur Multi Layer Perceptron Seberapa Baik Performa Algoritma Ini," *J. Ict Apl. Syst.*, vol. 1, no. 1, pp. 23–35, 2022, doi: 10.56313/jictas.v1i1.127.

[20] A. Prasetya Wibawa, W. Lestar, A. Bella Putra Utama, I. Tri Saputra, and Z. Nabila Izdihar, "Multilayer Perceptron untuk Prediksi Sessions pada Sebuah Website Journal Elektronik," *Indones. J. Data Sci.*, vol. 1, no. 3, pp. 57–67, 2020, doi: 10.33096/ijodas.v1i3.15.

[21] I. Firmansyah and B. H. Hayadi, "Komparasi Fungsi Aktivasi Relu Dan Tanh Pada Multilayer Perceptron," *JIKO (Jurnal Inform. dan Komputer)*, vol. 6, no. 2, p. 200, 2022, doi: 10.26798/jiko.v6i2.600.

[22] S. H. Gulo and A. H. Lubis, "Penerapan Multi-Layer Perceptron untuk Mengklasifikasi Penduduk Kurang Mampu," *Explorer (Hayward).*, vol. 4, no. 2, pp. 51–59, 2024.

[23] M. Wahyuni, "Klasifikasi Penyakit Daun Tomat dengan Perbandingan Fungsi Aktivasi Multi Layer Perceptron," vol. 13, pp. 1988–1998, 2024.

[24] S. Sen, D. Sugiarto, and A. Rochman, "Prediksi Harga Beras Menggunakan Metode Multilayer Perceptron (MLP) dan Long Short Term Memory (LSTM)," *Ultim. J. Tek. Inform.*, vol. 12, no. 1, pp. 35–41, 2020, doi: 10.31937/ti.v12i1.1572.

[25] N. K. C. PRATIWI, N. IBRAHIM, and S. SAIDAH, "Prediksi Kanker Paru menggunakan Grid search untuk Optimasi Hyperparameter pada Algoritma MLP dan Logistic Regression," *ELKOMIKA J. Tek. Energi Elektr. Tek. Telekomün. Tek. Elektron.*, vol. 12, no. 3, p. 556, 2024, doi: 10.26760/elkomika.v12i3.556.

[26] D. H. Depari, Y. Widiastiwi, and M. M. Santoni, "Perbandingan Model Decision Tree, Naive Bayes dan Random Forest untuk Prediksi Klasifikasi Penyakit Jantung," *Inform. J. Ilmu Komput.*, vol. 18, no. 3, p. 239, 2022, doi: 10.52958/iftk.v18i3.4694.

[27] D. Sudrajat, A. I. Purnamasari, A. R. Dikananda, D. A. Kurnia, and A. Bahtiar, "Klasifikasi Mutu Pembelajaran Hybrid berdasarkan Algoritma C.45, Random Forest dan Naïve Bayes dengan Optimasi Bootsrap Areggating (Bagging) pada masa COVID-19," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 6, p. 2227, 2022, doi: 10.30865/jurikom.v9i6.5179.

[28] G. Ashari Rakhmat and W. Mutohar, "MIND (Multimedia Artificial Intelligent Networking Database Prakiraan Hujan menggunakan Metode Random Forest dan Cross Validation," *J. MIND J. | ISSN*, vol. 8, no. 2, pp. 173–187, 2023, [Online]. Available: https://doi.org/10.26760/mindjournal.v8i2.173-187

[29] G. Arther Sandag, "Prediksi Rating Aplikasi App Store Menggunakan Algoritma Random Forest," *Cogito Smart J. |*, vol. 6, no. 2, pp. 167–178, 2020.

[30] T. Latifah and G. D. Anggitha, "Implementasi Metode Random Forest , KNN (K-Nearest Neighbour), Decision Tre e Classification menggunakan Machine Learning untuk Stroke Prediction," pp. 1–18.

[31] B. H. Mahendra, Adiwijaya, and U. N. Wisesty, "Kategorisasi Berita Multi-Label Berbahasa Indonesia Menggunakan Algoritma Random Forest," *e-Proceeding Eng.*, vol. 6, no. 2, pp. 9030–9041, 2019.

[32] R. Hidayat *et al.*, "Implementasi Algoritma Random Forest Regression Untuk Memprediksi Penjualan Produk Di Supermarket," vol. 10, no. 1, pp. 101–109, 2025.

[33] A. Rasyid, A. B. Muharam, and A. Solichin, "Prediksi harga saham syariah indonesia berdasarkan analisis fundamental, teknikal dan bandarmology menggunakan metode random forest," vol. 10, no. 2, pp. 1663–1677, 2025.

[34] A. M. A. Rahim, Inggrid Yanuar Risca Pratiwi, and Muhammad Ainul Fikri, "Klasifikasi Penyakit Jantung Menggunakan Metode Synthetic Minority Over-Sampling Technique Dan Random Forest Clasifier," *Indones. J. Comput. Sci.*, vol. 12, no. 5, pp. 2995–3011, 2023, doi: 10.33022/ijcs.v12i5.3413.

[35] G. Royong, "MELALUI SKRINING TRIGLISERIDA PADA USIA PRODUKTIF," pp. 150–156, 2025.

[36] J. Barat, "Article History:" no. April, 2025.

[37] Mk. dr Fika Tri Anggraini, dr Novi Aryanti, M. Dian Paramita Kartikasari, S. apt Wahyu Hendrarti, and Mk. Wa Ode Rahmadania, *Penyakit Kardiovaskuler Penerbit Cv.Eureka Media Aksara*. 2024.

[38] D. Nurina Hafila, K. Wisudawan, S. Darma, and Dahlia, "Fakumi Medical Journal Prevalensi Penyakit Kardiovaskular pada Masa Pandemic Tahun 2020-2021 di RS Arifin Nu'mang Kabupaten Sidrap," vol. 3, no. 10, pp. 710–719, 2023.

[39] R. Efendi, A. Junaidi, and A. M. Rizki, "Penentuan Pusat Klaster Secara Otomatis Pada Algoritma Density Peaks Clustering Berbasis Metode Inter Quartile Range," *J. Inform. Dan Tek. Elektro Terap.*, vol. 12, no. 3, 2024, doi: 10.23960/jitet.v12i3.4997.

[40] D. N. Handayani and S. Qutub, "Penerapan Random Forest Untuk Prediksi Dan Analisis Kemiskinan," *RIGGS J. Artif. Intell. Digit. Bus.*, vol. 4, no. 2, pp. 405–412, 2025, doi:

10.31004/riggs.v4i2.512.
[41] R. R. Hallan and I. N. Fajri, "Prediksi Harga Rumah menggunakan Machine Learning Algoritma Regresi Linier," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 7, no. 1, pp. 57–62, 2025, doi: 10.47233/jteksis.v7i1.1732.
[42] M. A. Abubakar, M. Muliadi, A. Farmadi, R. Herteno, and R. Ramadhani, "Random Forest Dengan Random Search Terhadap Ketidakseimbangan Kelas Pada Prediksi Gagal Jantung," *J. Inform.*, vol. 10, no. 1, pp. 13–18, 2023, doi: 10.31294/inf.v10i1.14531.
[43] A. Sentimen, P. Terhadap, and P. Iphone, "Analisis sentimen publik terhadap penjualan iphone 16 dan kebijakan tkdn di indonesia," vol. 11, no. 1, pp. 73–80, 2025.
[44] W. A. Aziz, *Implementasi metode random forest pada klasifikasi data ulasan konsumen perusahaan (studi kasus: aplikasi kai access)*. 2021.
[45] Y. Mulia, "Perbandingan Metode Machine Learning Untuk Mendeteksi Penyakit Jantung," pp. 1–23, 2016.