Journal of Dinda
Data Science, Information Technology, and Data Analytics
Vol. 5 No. 2 (2025) 208 - 219 E-ISSN: 2809-8064

Heart Failure Classification Using a Hybrid Model Based on SVM and Random Forest

Muh Sajid Abdillah[1*], Harminto Mulyo [2], Gentur Wahyu Nyipto Wibowo [3]
[1*2,3] Informartics Engineering, Sains and Technology, Nahdlatul Ulama Islamic University of Jepara
[1*]211240001176@unisnu.ac.id, [2]minto@unisnu.ac.id, [3]gentur@unisnu.ac.id

Abstract

This study discusses the development of a model to classify heart failure disease by combining two algorithms in the field of data mining: Support Vector Machine (SVM) and Random Forest (RF). The dataset used is the Heart Failure Prediction Dataset, consisting of 918 patient records containing medical information such as blood pressure, cholesterol levels, and heart rate. The research process began with data cleaning, normalization using MinMaxScaler, and data balancing with the SMOTE technique to equalize the number of cases between heart failure patients and non-patients. The data was then split into training and testing sets. Each model (SVM and RF) was tested individually and also combined into a hybrid model. Validation was performed using 5-Fold Cross Validation to ensure consistent results. The results show that SVM performed better in terms of precision for detecting heart failure after applying SMOTE, while RF remained stable even without data balancing. The hybrid model combining both algorithms achieved the best performance, with an accuracy of 91.20%, precision of 90.85%, recall of 92.44%, and an AUC score of 0.961. These results indicate that the hybrid model can detect heart failure more accurately and in a more balanced manner. With its high and consistent performance, this model is suitable for use as a decision support system in the medical field, particularly for early detection of heart failure.

Keywords: Heart Failure, Support Vector Machine (SVM), Random Forest, SMOTE, Hybrid Model

1. Introduction

Heart disease is one of the leading causes of death worldwide, including in Indonesia. One of the most serious forms of this disease is heart failure, a condition in which the heart is unable to pump blood effectively to meet the body's metabolic needs. According to data from the World Health Organization (WHO), cardiovascular diseases, including heart failure, account for approximately 17.9 million deaths each year, with one-third of these occurring in the Asia region[1], [2]. In Indonesia, there were around 15.5 million heart disease cases in 2022, making it the most common illness compared to cancer, stroke, and kidney failure. [1]. This shows that early detection of heart failure is crucial to reduce mortality rates and improve patients' quality of life.

Heart failure is often caused by structural or functional problems in the heart that make it unable to pump blood efficiently. Common symptoms include shortness of breath, fatigue, and swelling in areas such as the legs. If not treated properly, this condition can lead to repeated hospitalizations and even death. [3]. The high incidence rate and the complexity of symptoms make early detection a significant challenge for the medical field. To improve diagnostic accuracy, data- and technology-driven approaches are becoming increasingly important.

With the advancement of technology, Artificial Intelligence (AI) and Machine Learning (ML) have brought significant changes to the healthcare sector, especially in processing medical data. Machine learning has the ability to quickly analyze large and complex datasets, recognize hidden patterns, and generate data-driven decisions that are often more accurate than conventional methods. Various ML algorithms, such as Support Vector Machine (SVM) and Random Forest (RF), have proven to deliver strong performance in classifying cardiovascular diseases [4], [5], [6].

However, each algorithm has its own limitations. SVM excels at handling high-dimensional data and optimally separating classes but is less effective with imbalanced datasets. Conversely, Random Forest can manage large and complex datasets and tends to be more stable with class imbalance, though it is less precise in defining class

Received: 24-07-2025 | Accepted: 13-08-2025 | Published: 15-08-2025

boundaries. Therefore, a hybrid approach combining SVM and RF offers a promising alternative. By leveraging the strengths of both algorithms, the hybrid model is expected to deliver more accurate and robust classification performance [7], [8].

Besides the challenge of choosing the right algorithm, class imbalance in medical datasets is a major issue that can cause models to be biased toward the majority class. In the context of heart failure classification, this is particularly dangerous because misclassifying patients with heart failure can have fatal consequences. To address this problem, the Synthetic Minority Oversampling Technique (SMOTE) is used, which creates synthetic data for the minority class through interpolation. This technique has been proven to improve the sensitivity, precision, and F1-score of models in various previous studies [9], [10], [11]. However, heart failure classification faces challenges such as class imbalance, where heart failure cases are much rarer than normal ones, causing models to favor the majority class. Clinical data are also complex and multidimensional, involving physiological, laboratory, and clinical symptom variables that vary across patients. These issues necessitate advanced machine learning techniques and preprocessing methods like SMOTE to enhance detection of the minority class.[12]

Previous studies have applied algorithms such as SVM, Random Forest, KNN, ANN, and XGBoost for heart failure classification. While XGBoost and ANN often achieve high performance, they require complex tuning and are prone to overfitting on imbalanced data.[6] SVM excels at separating data with optimal margins and is suitable for low-dimensional datasets, while Random Forest handles complex features and missing values well[13]. The hybrid model combining SVM and RF aims to overcome the individual limitations and deliver more stable and balanced classification. This approach is more efficient than complex ensembles like XGBoost stacking, especially for small and imbalanced medical datasets.

This study presents an original contribution by developing a hybrid model that integrates Support Vector Machine (SVM) and Random Forest (RF), complemented by SMOTE to address class imbalance. Unlike previous studies that merely compare algorithms individually, this research leverages the strengths of each algorithm within a unified classification system. The dataset used is the Kaggle Heart Failure Prediction Dataset, consisting of 918 patient records with 11 predictor attributes and 1 target attribute. The main contribution of this study is the development of a hybrid model that combines SVM and Random Forest algorithms for heart failure classification, integrated with SMOTE as a data balancing technique. This approach is designed to overcome the limitations of each individual algorithm and enhance prediction stability for the minority class, which is critical in clinical diagnosis contexts. The main goal of this research is to improve the accuracy and reliability of heart failure classification, while producing a predictive model that is effective, efficient, and capable of supporting more accurate and timely clinical decision-making.

2. Research Methods

This study employs a quantitative experimental approach aimed at testing the performance of a hybrid model combining Support Vector Machine (SVM) and Random Forest (RF) algorithms in classifying heart failure disease. The dataset used was obtained from Kaggle, named the Heart Failure Prediction Dataset, which contains 918 patient records with 11 predictor attributes and 1 target attribute (Heart Disease). Before use, the data were cleaned to remove incomplete values and scaled to ensure all features have comparable ranges. Additionally, the Synthetic Minority Over-sampling Technique (SMOTE) was applied to balance the number of heart failure patients and normal patients, preventing the model from being biased toward any class.

After the data cleaning and balancing process, the dataset was split into two parts: 80% for training and 20% for testing. The hybrid model was then trained using this data and validated through K-Fold Cross Validation to ensure stable results. Model evaluation was carried out using accuracy, precision, recall, F1-score, and ROC-AUC metrics to assess how well the model distinguishes between patients with heart failure and those without. The entire data processing, model training, and evaluation process was conducted using the Python programming language.

Figure 1 shows the flowchart of the entire research process.

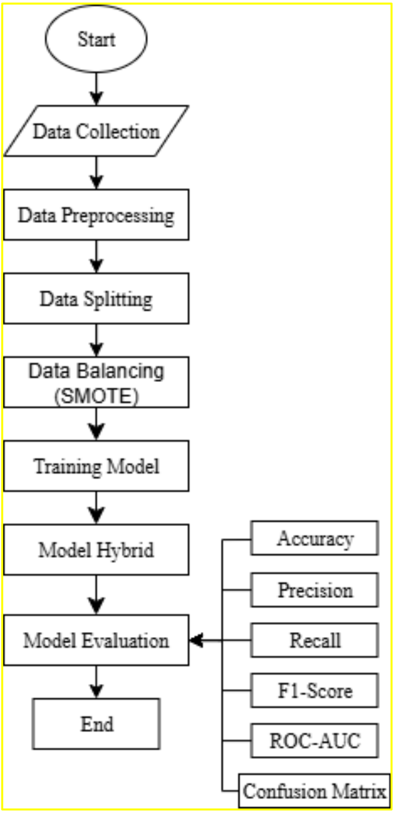


Figure 1. Flowchart Research

2.1 *Data Collection*

The dataset used in this study is the **Heart Failure Prediction Dataset** obtained from the Kaggle platform. This dataset consists of 918 records, with 11 predictor attributes and 1 target attribute. The target attribute indicates heart failure status, labeled as 1 for heart failure patients and 0 for normal conditions. The predictor attributes include Age, Sex, Chest Pain Type, Resting Blood Pressure, Cholesterol, Fasting Blood Sugar, Resting ECG, Max HR, Exercise Angina, Oldpeak, and ST Slope. [13]

Table 1. Description of Dataset Attributes

Attribute Name	Attribute Type	Measurement Scale
Age	Numerical	Ratio
Sex	Categorical	Nominal
ChestPain Type	Categorical	Nominal
RestingBP	Numerical	Ratio
Cholesterol	Numerical	Ratio
FastingBP	Categorical	Nominal
Resting Electrocardiographic Results	Categorical	Nominal
MaxHR	Numerical	Ratio
Exercise Induced Angina	Categorical	Nominal
Oldpeak	Numerical	Ratio
ST Slope	Categorical	Nominal
Heart Disease	Categorical	Nominal

Each attribute in the dataset has clinical relevance to heart failure diagnosis:

- Age and Sex are key risk factors for cardiovascular disease.
- Chest pain, Blood pressure, and Cholesterol reflect vascular and cardiac workload conditions.
- Fasting blood sugar relates to diabetes, a major comorbidity of heart failure.
- ECG, Max heart rate, and Oldpeak help detect arrhythmia and ischemia.
- Exercise-induced angina and ST slope indicate oxygen supply issues during exertion.

2.2 *Data Cleaning*

Before modeling, the dataset goes through several stages, including data cleaning. The data must be cleaned from missing values, duplicates, and irrelevant data. The steps taken include:

1. Remove rows with missing values if their number is small, or fill in missing values with the mean or median if the number is large.
2. Remove duplicate data to prevent the model from learning repetitive information.
3. Remove irrelevant or unimportant data that does not affect the analysis objectives.
4. Check and correct or remove any unreasonable or inconsistent data.

2.3 *Data Scalling*

Normalization was performed using MinMaxScaler from the *Scikit-learn* library, which transforms numerical feature values into the [0, 1] range. This scaling was applied before feeding the data into the model, particularly because algorithms such as SVM are highly sensitive to feature scales. Without normalization, features with large values may dominate distance and margin calculations in SVM, potentially reducing model accuracy.

Standard scaling formula:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

Explanation:

x = raw value,

x_{min} = feature minimum value,
x_{max} = feature maximum value,
x' = feature values that have been scaled into a range [0, 1].

2.4 Data Balancing (SMOTE)

Synthetic Minority Over-sampling Technique (SMOTE) is an oversampling method used to address class imbalance problems in classification datasets, such as in this study predicting heart failure. SMOTE works by creating new synthetic samples for the minority class (heart failure patients) through linear interpolation between existing minority class samples. In this study, SMOTE (Synthetic Minority Over-sampling Technique) was applied using the default parameter k_neighbors=5. SMOTE was used to balance the number of heart failure cases (minority) and normal cases (majority) to a 1:1 ratio. To prevent data leakage, SMOTE was applied only to the training data (x_train, y_train) after splitting, not on the entire dataset. As a result, the model was evaluated on test data that preserved the original class distribution.

This technique does not simply duplicate existing data (as in traditional oversampling methods) but generates new data points based on differences between minority class samples, helping the model to better identify the decision boundary between the two classes. This assists models like Random Forest and Support Vector Machine (SVM) in the hybrid model to avoid being overly biased toward the majority class (normal patients).

SMOTE formula:

$$x_{new} = x_i + \lambda \cdot (x_{zi} - x_i) \qquad (2)$$

Explanation:

x_i = feature vector of a randomly selected minority data point.

x_{zi} = the nearest neighbor of x_i, chosen from the k nearest neighbors.

λ = a random number between 0 and 1.

x_{new} = the generated synthetic sample.

2.5 Data Splitting

The dataset that has undergone preprocessing is then split into two parts: 80% as the training set and 20% as the testing set. This split aims to train the model using the training data so it can recognize patterns within the dataset and to evaluate its performance on previously unseen data. Thus, it ensures that the model not only performs well on the training data but is also capable of effectively generalizing to new data.

2.6 *Model Hybrid*

The model used in this study is a hybrid combination of two popular classification algorithms, namely Support Vector Machine (SVM) and Random Forest (RF). This hybrid approach is designed to combine the strengths of each method to improve the classification performance of heart failure disease, especially in complex and imbalanced data conditions.

2.6.1 Support Vector Machine (SVM)

SVM is an algorithm that works by finding the optimal hyperplane that separates two classes with the maximum margin. The margin is the distance between the hyperplane and the closest data points from each class. The larger the margin, the better the model's ability to generalize to new data.

For linear data, the SVM decision function can be written as follows:

$$f(x) = w^T x + b \qquad (3)$$

Subject to the constraints :

$$y_i(w^T x^i + b) \geq 1 \ untuk\ semua\ i \qquad (4)$$

Explanation:

x^i = input feature vector,

y_i = class label (±1),

w = weight vector,

b = bias,

i = sampel index.

SVM is highly effective in clearly separating minority class data, especially in cases of class imbalance such as in datasets of heart failure patienst.

2.6.2 Random Forest (RF)

Random Forest is an ensemble learning algorithm that builds multiple decision trees randomly and combines their results through majority voting (for classification) or averaging (for regression). RF excels at handling complex features, reducing bias and variance, and is resistant to overfitting thanks to the use of bagging and random feature selection.

The final prediction of Random Forest can be expressed as:

$$\hat{y} = majority\_vote(h_1(x), h_2(x), \dots, h_T(x)) \quad (5)$$

Explanation:

$h_T(x)$ = Prediction from the t-th tree",

T = The total number of trees in the forest,

ŷ = final classification result.

2.7 Model Evaluation

To evaluate the performance of the hybrid model based on Support Vector Machine (SVM) and Random Forest (RF) in classifying heart failure disease, several common machine learning evaluation metrics are used, especially in the context of imbalanced data. The evaluation metrics used are:

- Confusion Matrix
  Used to observe the number of True Positives, True Negatives, False Positives, and False Negatives, helping to understand the types of errors made by the model.

- Accuracy
  Measures the proportion of correct predictions out of the total data..
  $$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$
  High accuracy does not necessarily reflect good model performance on imbalanced data.

- Precision
  Measures the precision of positive predictions.
  $$Precision = \frac{TP}{TP+FP} \quad (7)$$

- Recal (Sensitivity)
  Measures the model's ability to identify all positive cases.
  $$Recall = \frac{TP}{TP+FN} \quad (8)$$

- F1-Score
  The harmonic mean of precision and recall.
  $$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (9)$$

- ROC-AUC
  This metric represents the model's ability to distinguish between the positive class (heart failure) and the negative class. An AUC value close to 1 indicates excellent performance.

- Paired T-Test
  T-test is a method used to compare the means of two groups of data to determine whether the observed difference is statistically significant or merely due to chance. In this study, a paired t-test was employed to compare the performance results of several classification models based on k-fold cross-validation. The paired t-test is appropriate because the measurements in each fold are paired, using the same data splits. This approach helps ensure that the differences in model performance are truly significant and not just the result of data variation across folds.

2.8 Model Validation

This study uses the 5-Fold Cross Validation technique, where the dataset is divided into five parts. Each part is alternately used as the test data, while the other four parts are used for training. This process is repeated five times, and the results are averaged. This technique improves evaluation reliability and helps reduce bias, allowing the hybrid SVM and Random Forest model to be validated more stably and accurately.

2.9 Visualization

Visualization is used as an important part of evaluating the performance of the hybrid SVM and Random Forest model in classifying heart failure disease. Visualization not only helps in understanding data patterns but also supports more intuitive and communicative interpretation of classification results.

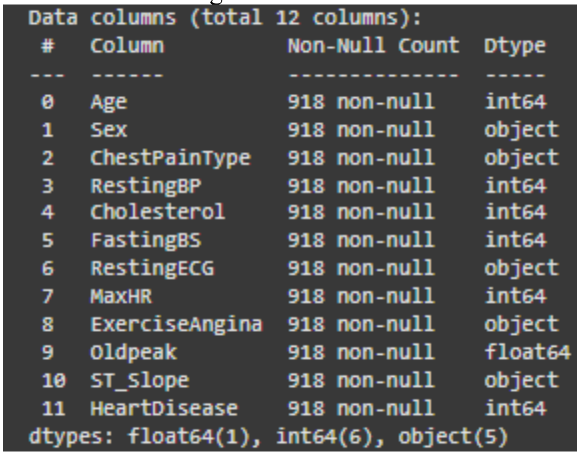
1. Confusion Matrix Plot
   The Confusion Matrix visualization is presented as a heatmap, illustrating the number of correct and incorrect predictions for each class. In this study, the Confusion Matrix heatmap visually demonstrates the model’s performance in identifying heart failure patients, with a focus on maximizing True Positive (TP) values and minimizing False Negative (FN) values as a priority in medical diagnosis
2. Kurva ROC
   The ROC curve is used to compare the trade-off between recall (sensitivity) and the false positive rate (1 - specificity) at various classification thresholds. In this study, the ROC curve illustrates how well the model can distinguish between patients with heart failure

and those without. The larger the area under the curve (AUC), the better the model's performance. The high AUC value obtained from the hybrid model demonstrates its strong capability in binary classification, especially after the data was balanced using SMOTE.

3. Results and Discussion

3.1 *Data Collection*

The data used in this study is sourced from the Heart Failure Prediction Dataset, which is publicly available on the Kaggle platform. This dataset consists of 918 patient records with 11 predictor attributes and 1 target attribute. It was selected due to its diverse features, which include demographic information and clinical examination results, making it suitable for developing machine learning-based classification models.



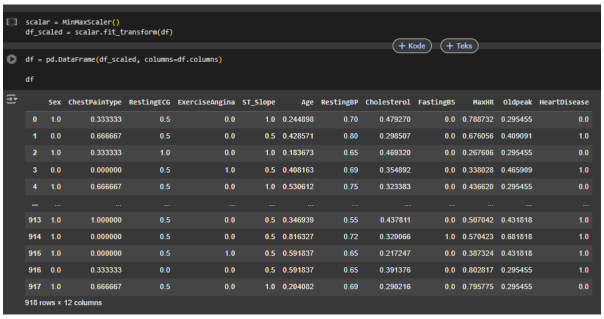
```
Data columns (total 12 columns):
 #   Column          Non-Null Count  Dtype  
---  ------          --------------  -----  
 0   Age             918 non-null    int64  
 1   Sex             918 non-null    object 
 2   ChestPainType   918 non-null    object 
 3   RestingBP       918 non-null    int64  
 4   Cholesterol     918 non-null    int64  
 5   FastingBS       918 non-null    int64  
 6   RestingECG      918 non-null    object 
 7   MaxHR           918 non-null    int64  
 8   ExerciseAngina  918 non-null    object 
 9   Oldpeak         918 non-null    float64
 10  ST_Slope        918 non-null    object 
 11  HeartDisease    918 non-null    int64  
dtypes: float64(1), int64(6), object(5)
```

Figure 1. Result Of Attribute Data Collection

3.2 *Data scalling*

Numerical features in the dataset, such as Age, Resting Blood Pressure, Cholesterol, Max Heart Rate, and Oldpeak, are scaled using MinMaxScaler from the Scikit-Learn library. This process transforms feature values into the range [0, 1], ensuring each feature contributes equally during training. MinMaxScaler is chosen because it effectively preserves the data distribution shape and accelerates the convergence process of machine learning algorithms like SVM.

This is very important because some algorithms, especially distance-based ones like SVM, are highly sensitive to the scale of input features.



```
scalar = MinMaxScaler()
df_scaled = scalar.fit_transform(df)
```

```
df = pd.DataFrame(df_scaled, columns=df.columns)

df
```

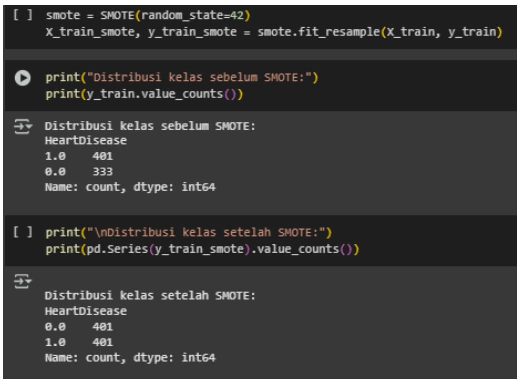
```
     Sex  ChestPainType  RestingECG  ExerciseAngina  ST_Slope       Age  RestingBP  Cholesterol  FastingBS     MaxHR   Oldpeak  HeartDisease
  0  1.0       0.333333         0.5             0.0       1.0  0.244898       0.70     0.479270        0.0  0.788732  0.295455           0.0
  1  0.0       0.666667         0.5             0.0       0.5  0.428571       0.80     0.298507        0.0  0.676056  0.409091           1.0
  2  1.0       0.333333         1.0             0.0       1.0  0.183673       0.65     0.469320        0.0  0.267606  0.295455           0.0
  3  0.0       0.000000         0.5             1.0       0.5  0.408163       0.69     0.354892        0.0  0.338028  0.465909           1.0
  4  1.0       0.666667         0.5             0.0       1.0  0.530612       0.75     0.323383        0.0  0.436620  0.295455           0.0
...  ...            ...         ...             ...       ...       ...        ...          ...        ...       ...       ...           ...
913  1.0       1.000000         0.5             0.0       0.5  0.346939       0.55     0.437811        0.0  0.507042  0.431818           1.0
914  1.0       0.000000         0.5             0.0       0.5  0.816327       0.72     0.320066        1.0  0.570423  0.681818           1.0
915  1.0       0.000000         0.5             1.0       0.5  0.591837       0.65     0.217247        0.0  0.387324  0.431818           1.0
916  0.0       0.333333         0.0             0.0       0.5  0.591837       0.65     0.391376        0.0  0.802817  0.295455           1.0
917  1.0       0.666667         0.5             0.0       1.0  0.204082       0.69     0.290216        0.0  0.795775  0.295455           0.0
918 rows × 12 columns
```

Figure 2. Result Of Scalling Data

3.3 *SMOTE*

The initial dataset had class imbalance, with only 410 out of 918 records (approximately 44.7%) belonging to the positive class (heart failure patients). The class imbalance problem was addressed using the Synthetic Minority Over-sampling Technique (SMOTE). This technique generates synthetic data for the minority class (heart failure patients) through interpolation between nearest neighbors in the feature space.

As a result, the class proportions become balanced, and the model is no longer biased toward the majority class.



```
smote = SMOTE(random_state=42)
X_train_smote, y_train_smote = smote.fit_resample(X_train, y_train)
```

```
print("Distribusi kelas sebelum SMOTE:")
print(y_train.value_counts())
```

```
Distribusi kelas sebelum SMOTE:
HeartDisease
1.0    401
0.0    333
Name: count, dtype: int64
```

```
print("\nDistribusi kelas setelah SMOTE:")
print(pd.Series(y_train_smote).value_counts())
```

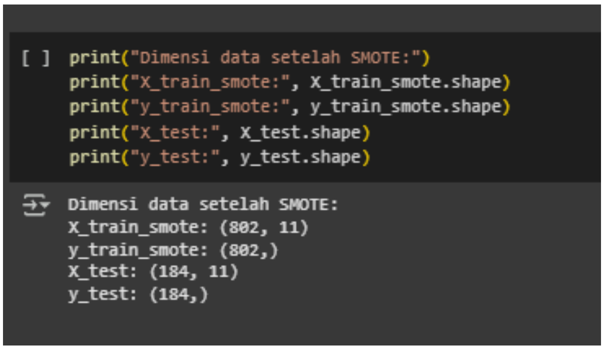
```

Distribusi kelas setelah SMOTE:
HeartDisease
0.0    401
1.0    401
Name: count, dtype: int64
```

Figure 3. Penerapan SMOTE pada Data Heart Disease

Figure 3 shows the class distribution before and after SMOTE was applied. Before SMOTE, the dataset was imbalanced with 401 heart failure cases and 333 normal cases. After SMOTE, the distribution became balanced (401:401). This balance reduces model bias toward the majority class and improves recognition of the minority class. Synthetic data are generated through interpolation between heart failure samples, preserving original patterns while expanding the represented feature space.

3.4 *Data Splitting*

After applying SMOTE, the dataset was split into two parts: 80% for training and 20% for testing. This split aims to allow the model to learn from the majority of the data and be tested on unseen data to measure its generalization capability.



```
print("Dimensi data setelah SMOTE:")
print("X_train_smote:", X_train_smote.shape)
print("y_train_smote:", y_train_smote.shape)
print("X_test:", X_test.shape)
print("y_test:", y_test.shape)
```

```
Dimensi data setelah SMOTE:
X_train_smote: (802, 11)
y_train_smote: (802,)
X_test: (184, 11)
y_test: (184,)
```

Figure 4. Pembagian data

3.5 *Modelling*

In this study, the classification model was built using two machine learning algorithms: Support Vector Machine (SVM) and Random Forest (RF). Each model was first applied separately, then combined into a hybrid model to improve classification performance. *Support Vector Machine (SVM)*

SVM is used to create a hyperplane that separates two classes of data (heart failure and normal) with the maximum margin. SVM performs optimally on scaled data and excels at handling high-dimensional data. The following are the results of applying SVM without SMOTE.

Table 2. Hasil penerapan SVM tanpa SMOTE

	Precision	Recall	f1-score	Support
0.0	0.82	0.90	0.86	77
0.1	0.92	0.86	0.89	107
accuracy			0.88	184
macro avg	0.87	0.88	0.87	184
weighted avg	0.88	0.88	0.88	184

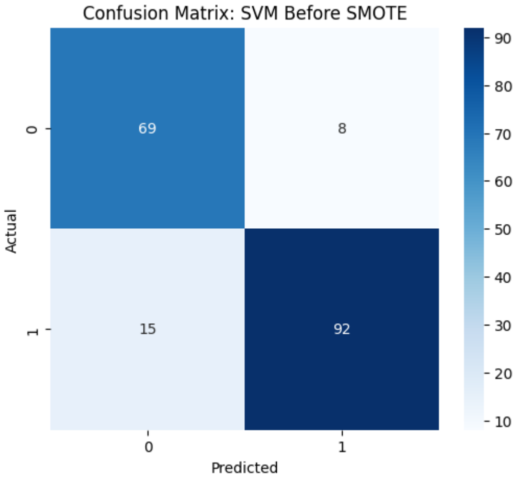


Figure 5. Histogram SVM tanpa SMOTE

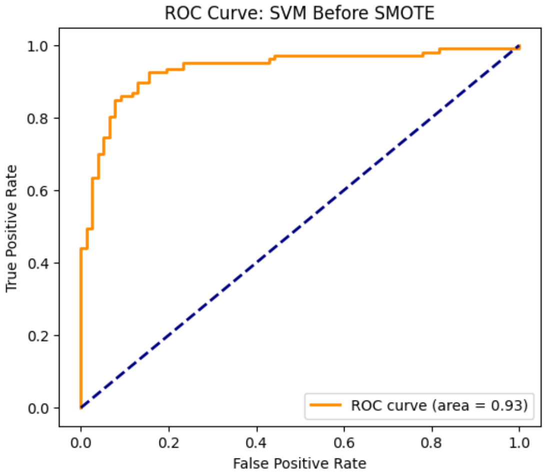


Figure 6. Histogram ROC SVM tanpa SMOTE

3.5.1. Random Forest (RF)

Random Forest builds multiple decision trees randomly on data subsets, then combines their results through voting. RF has advantages in handling complex features and reducing the risk of overfitting. The RF model was first trained on unbalanced data to evaluate its baseline performance before combining and applying balancing techniques.

Table 3. Hasil penerapan RF tanpa SMOTE

	Precision	Recall	f1-score	Support
0.0	0.84	0.88	0.86	77
0.1	0.91	0.88	0.90	107
accuracy			0.88	184
macro avg	0.88	0.88	0.88	184

weighted avg	0.88	0.88	0.88	184

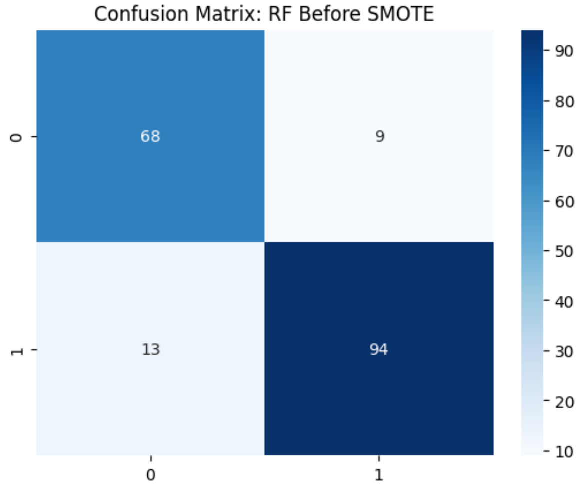


Figure 7. Histogram RF tanpa SMOTE

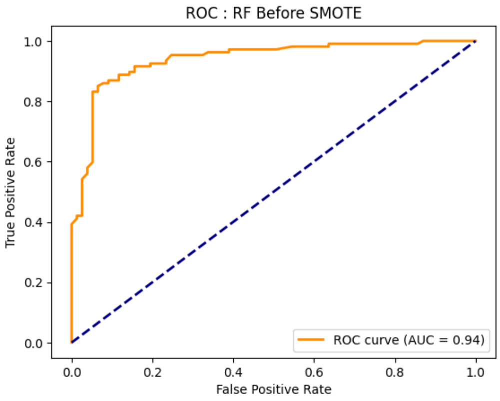


Figure 8. Histogram ROC RF tanpa SMOTE

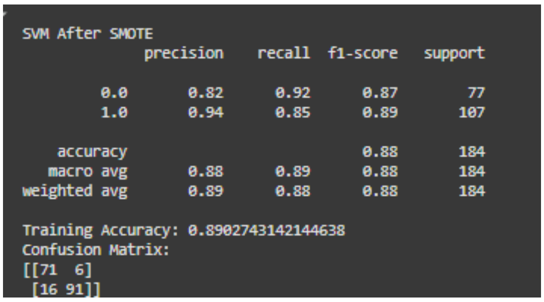
3.6 *Model Evaluation*

After initial training on the unbalanced data, further evaluation was conducted by applying the Synthetic Minority Over-sampling Technique (SMOTE) to balance the class distribution.

The balanced dataset was then used to retrain the three models: SVM, Random Forest, and the Hybrid model. The purpose of applying SMOTE was to improve the models' sensitivity to the minority class, namely heart failure patients.

3.7.1. Support Vector Machine dengan SMOTE

The application of SMOTE to the SVM algorithm showed a significant impact on the model's performance. Recall for the normal class increased from 0.90 to 0.92, while precision for the heart failure class rose from 0.92 to 0.94. However, there was a slight decrease in recall for the heart failure class, from 0.86 to 0.85. Although the overall accuracy remained unchanged (at 0.88), these results indicate that SMOTE helped enhance the model's ability to detect normal patients, while also improving the precision in predicting heart failure cases, despite a slight reduction in sensitivity to that class.



```
SVM After SMOTE
              precision    recall  f1-score   support

         0.0       0.82      0.92      0.87        77
         1.0       0.94      0.85      0.89       107

    accuracy                           0.88       184
   macro avg       0.88      0.89      0.88       184
weighted avg       0.89      0.88      0.88       184

Training Accuracy: 0.8902743142144638
Confusion Matrix:
[[71  6]
 [16 91]]
```

Figure 9. Hasil Penerapan SVM dengan SMOTE

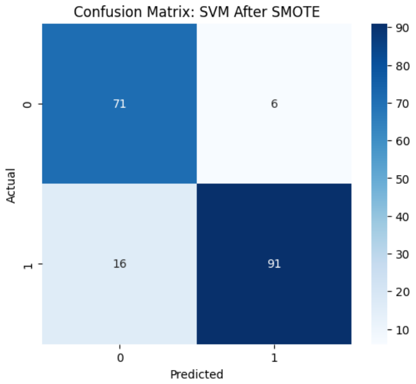


Figure 10. Histogram SVM dengan SMOTE

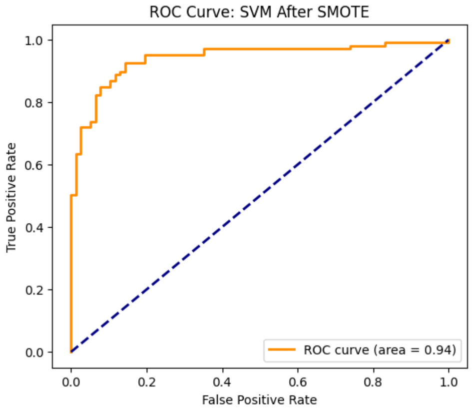
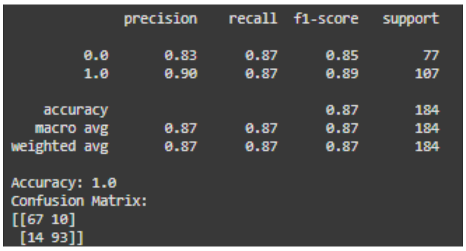


Figure 11. Histogram ROC SVM dengan SMOTE

3.7.2. Random Forest dengan SMOTE

Unlike in SVM, the application of SMOTE to Random Forest did not result in a significant performance improvement. In fact, there was a slight decrease in both precision and recall for both classes. Precision for the heart failure class dropped from 0.91 to 0.90, recall from 0.88 to 0.87, and accuracy from 0.88 to 0.87. This suggests that Random Forest is already robust in handling data imbalance without the need for additional balancing techniques, and in this case, SMOTE did not lead to a noticeable performance enhancement..



```
              precision    recall  f1-score   support

         0.0       0.83      0.87      0.85        77
         1.0       0.90      0.87      0.89       107

    accuracy                           0.87       184
   macro avg       0.87      0.87      0.87       184
weighted avg       0.87      0.87      0.87       184

Accuracy: 1.0
Confusion Matrix:
[[67 10]
 [14 93]]
```

Figure 12. Hasil Penerapan RF dengan SMOTE

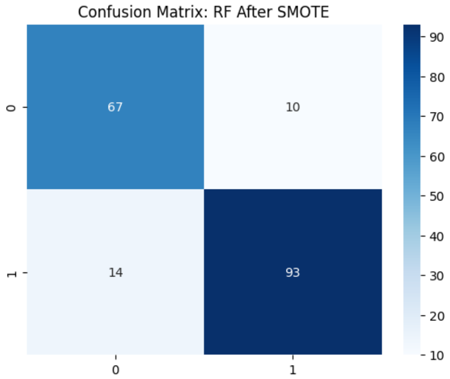


Figure 13. Histogram RF dengan SMOTE

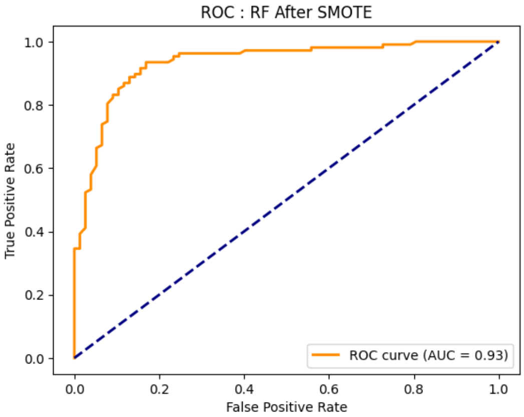


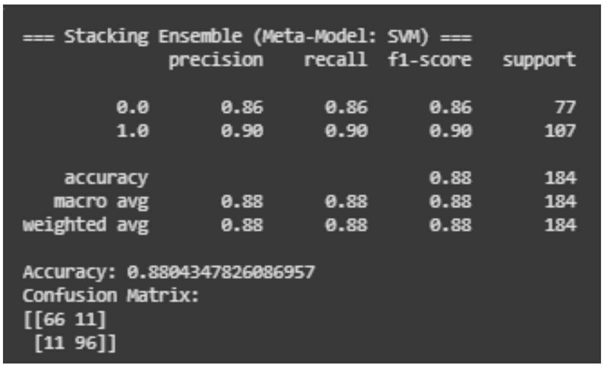
Figure 14. Histogram ROC RF dengan SMOTE

3.7.3. Hybrid SVM dan Random Forest

When SVM and Random Forest were tested on the raw (unbalanced) data, Random Forest demonstrated more stable performance across both classes in terms of precision and recall. The accuracy of Random Forest was slightly higher than that of SVM, at 0.88 compared to 0.867. This indicates that Random Forest is naturally more robust in handling imbalanced data without the need for additional balancing techniques such as SMOTE.

After both algorithms were applied to the balanced data, it was found that Random Forest had slightly more stable recall for the heart failure class (0.87 compared to 0.85 in SVM). However, SVM achieved higher precision for that class (0.94 compared to 0.90). This indicates that SVM with SMOTE is superior in terms of precision for detecting heart failure cases, while Random Forest tends to be more stable in recognizing overall cases.

The hybrid model combining SVM and Random Forest, after the application of SMOTE, demonstrated the most balanced and consistent results. Both precision and recall for the heart failure class reached 0.90, and the overall accuracy remained high at 0.88. This proves that the hybrid approach effectively integrates the optimal margin capability of SVM with the generalization strength of Random Forest. Therefore, the hybrid model emerges as the most optimal choice for heart failure classification, as it maintains a balanced trade-off between detection and precision across both classes.



```
=== Stacking Ensemble (Meta-Model: SVM) ===
              precision    recall  f1-score   support

         0.0       0.86      0.86      0.86        77
         1.0       0.90      0.90      0.90       107

    accuracy                           0.88       184
   macro avg       0.88      0.88      0.88       184
weighted avg       0.88      0.88      0.88       184

Accuracy: 0.8804347826086957
Confusion Matrix:
[[66 11]
 [11 96]]
```

Figure 15. Hasil Penerapan Hybrid SVM dan Random Forest

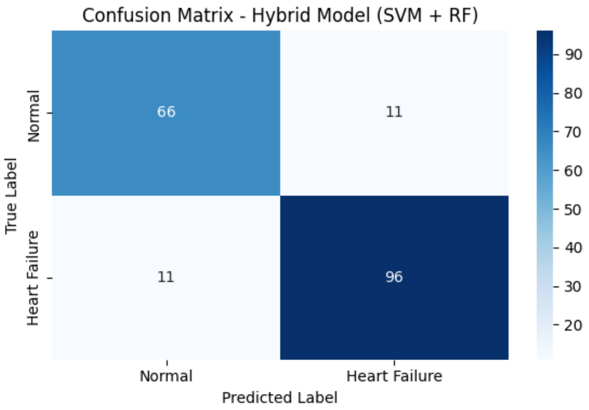


Figure 16. Histogram Penerapan Hybrid SVM dan Random Forest

Table 4. Evaluate Statistical Significance

SVM + SMOTE	RF + SMOTE	Hybrid + SMOTE
0.86956522	0.89440994	0.8707483
0.8447205	0.86956522	0.85034014
0.89375	0.85	0.83673469
0.86875	0.86875	0.87755102
0.83125	0.85	0.87671233

Paired t-test Hybrid vs SVM:
t-statistic = 0.0492, p-value = 0.9631
Paired t-test Hybrid vs RF:
t-statistic = -0.4337, p-value = 0.6868

The results of the paired t-test show that the p-values are greater than 0.05 for the comparison between the hybrid model and SVM + SMOTE (0.9631), and Random Forest + SMOTE (0.6868). This indicates that there is no statistically significant difference in performance among the three models. The observed differences in mean accuracy are likely due to data variation across the folds of cross-validation. Therefore, the three models demonstrate comparable predictive capabilities, and model selection can be based on other factors such as complexity, computational time, or ease of interpretation.

3.7 *Model Validation*

Model validation was carried out using the 5-Fold Cross Validation technique, where the dataset was divided into five parts. Each part was used in turn as the test data, while the remaining parts served as the training data. This process was repeated five times, and the evaluation results were averaged. This technique aims to produce more stable evaluations and to prevent overfitting.

The testing results using 5-Fold Cross Validation showed that the hybrid SVM and Random Forest model produced relatively stable accuracy scores across each fold. The accuracy values obtained for the five folds were 0.8707, 0.8503, 0.8503, 0.8776, and 0.8767, respectively. From these values, the average accuracy (mean accuracy) was calculated to be 0.87, or 87%, indicating a very good classification performance.

In addition, the standard deviation of the accuracy, which was 0.01, indicates that the variation in performance across the folds was very small. This means that the model demonstrated consistent classification performance on different subsets of data in each fold, suggesting that it has good generalization capability and is not overfitting to the training data.

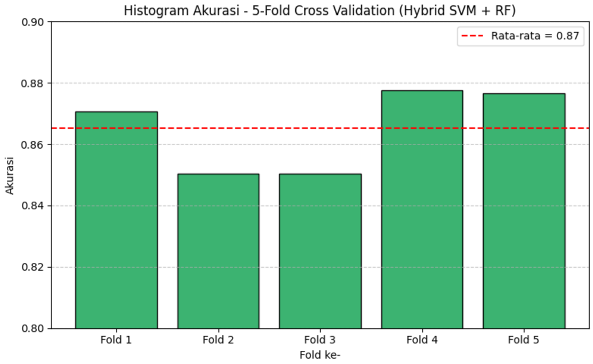


Figure 17. Histogram K-Fold Hybrid

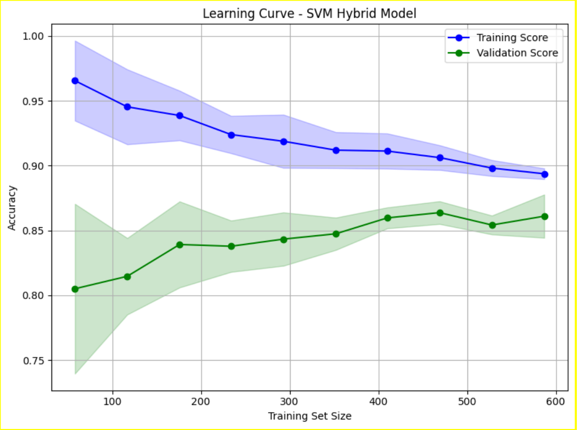


Figure 18. Learning Curve

The learning curve in Figure 18 demonstrates the **robustness** of the hybrid SVM + RF model. The training accuracy (blue line) gradually decreases from approximately 96% to 89% as the training size increases, which is expected due to increased data complexity. Meanwhile, the validation accuracy (green line) shows a consistent upward trend, from about 80% to nearly 86%.

This pattern indicates that the model does not suffer from severe overfitting, as the gap between training and validation curves becomes smaller. Furthermore, the shaded areas (standard deviation) are relatively narrow for larger training sizes, suggesting that the model performs consistently across different data subsets. Thus, the curve supports the claim that the proposed hybrid model is robust and stable.

4. Conclusion

This study successfully developed a heart failure classification model based on a hybrid approach that combines the Support Vector Machine (SVM) and Random Forest (RF) algorithms, supported by the SMOTE data balancing technique. The model was tested using the Heart Failure Prediction dataset from Kaggle, which consists of 918 patient records. Evaluation results revealed that each algorithm has its own strengths: SVM excels in precision for the positive class, while RF demonstrates strong stability in recall.

The application of SMOTE proved effective in enhancing the model's ability to recognize the minority class, particularly in the SVM model. However, for Random Forest, SMOTE did not have a significant impact on performance improvement. The hybrid model, which combines the optimal margin strength of SVM and the generalization capability of RF, delivered the most balanced results, with both precision and recall reaching 0.90 and an accuracy of 88%.

Based on the evaluation using 5-Fold Cross Validation, the hybrid model also demonstrated consistent performance with an average accuracy of 87% and a low standard deviation. This proves that the proposed hybrid model is not only accurate but also reliable and stable, making it highly promising for implementation in medical decision support systems for the early detection of heart failure.

As part of future development, it is recommended that the hybrid model proposed in this study be tested on external datasets, such as the *Cleveland Heart Disease Dataset*. This aims to evaluate the model's generalizability and test its robustness against different data distributions and population characteristics compared to the original dataset. Such testing would also provide insight into the model's broader applicability in clinical contexts beyond the scope of the current study.

References

[1] A. S. Prabowo and F. I. Kurniadi, “Analisis Perbandingan Kinerja Algoritma Klasifikasi dalam Mendeteksi Penyakit Jantung,” *Jurnal SISKOM-KB (Sistem Komputer dan Kecerdasan Buatan)*, vol. 7, no. 1, pp. 56–61, 2023.

[2] S. Adi and A. Wintarti, “Komparasi Metode Support Vector Machine (Svm), K-Nearest Neighbors (Knn), Dan Random Forest (Rf) Untuk Prediksi Penyakit Gagal Jantung,” *MATHunesa: Jurnal Ilmiah Matematika*, vol. 10, no. 2, pp. 258–268, 2022.

[3] H. Munirwan and O. Januaresty, “Penyakit Jantung Hipertensi dan Gagal Jantung,” *Jurnal Kedokteran Nanggroe Medika*, vol. 3, no. 4, pp. 9–17, 2020.

[4] S. Priyadarshinee and M. Panda, “Improving Prediction of Chronic Heart Failure using SMOTE and Machine Learning,” in *2022 Second International Conference on Computer Science, Engineering and Applications (ICCSEA)*, 2022, pp. 1–6.

[5] K. N. V. P. S. S. Nishanth, S. A. Basha, V. Srikanth, K. P. Narendra, and V. Rachapudi, “Forecasting Cardio Vascular Diseases Using Kernel Machine,” in *2024 Second International Conference on Inventive Computing and Informatics (ICICI)*, 2024, pp. 9–14.

[6] B. D. Shivahare *et al.*, “Delving into Machine Learning’s Influence on Disease Diagnosis and Prediction,” *The Open Public Health Journal*, vol. 17, no. 1, pp. 1–13, 2024.

[7] P. Mazumder and D. S. Baruah, “A Hybrid Model for Predicting Classification Dataset based on Random Forest, Support Vector Machine and Artificial Neural Network,” *International Journal of Innovative Technology and Exploring Engineering*, vol. 13, no. 1, pp. 19–25, 2023.

[8] P. P. S. Teja and T. Veeramani, “Improving the Efficiency of Heart Disease Prediction Using Novel Random Forest Classifier Over Support Vector Machine Algorithm,” *Cardiometry*, no. 25, pp. 1468–1476, 2023.

[9] N. V Chawla, K. W. Bowyer, L. O. Hall, and W. S. Philip Kegelmeyer, “Synthetic Minority Over-sampling Technique,” *J Artif Intell Res*,

vol. 16, p. 16, 2018.

[10] N. Jannath, M. A. Masud, M. J. Hossain, and M. Samsuzzaman, “Improving Classification using IS-SMOTE on Imbalanced Heart Failure Data,” in *2023 5th International Conference on Sustainable Technologies for Industry 5.0 (STI)*, 2023, pp. 1–6.

[11] Y.-T. Kim, D.-K. Kim, H. Kim, and D.-J. Kim, “A Comparison of Oversampling Methods for Constructing a Prognostic Model in the Patient with Heart Failure,” in *2020 International Conference on Information and Communication Technology Convergence (ICTC)*, 2020, pp. 379–383.

[12] H. A. Al-Alshaikh *et al.*, “Comprehensive evaluation and performance analysis of machine learning in heart disease prediction,” *Scientific Reports*, vol. 14, no. 1, pp. 1–15, 2024.

[13] L. N. Farida and S. Bahri, “Komputika : Jurnal Sistem Komputer Klasifikasi Gagal Jantung Menggunakan Metode SVM (Support Vector Machine) Classification of Heart Failure using the SVM (Support Vector Machine) Method,” vol. 13, pp. 0–7, 2024.