Informatics and Software Engineering
p-ISSN/e-ISSN: 2988-2818/2988-2222
Homepage: https://sanscientific.com/journal/index.php/ise
4(1) 1-7 (2026)
https://doi.org/10.58777/ise.v4i1.609

Research Article

Improving Survival Prediction For Heart Failure Patients Using Random Forest and Grid Search CV

Sari Susanti[1*], Rui Septiansyah Putra[2]
[1]Information System, Faculty of Information Technology, Universitas Adhirajasa Reswara Sanjaya, Bandung
[2]Information System, Faculty of Information Technology, Universitas Adhirajasa Reswara Sanjaya, Bandung

Received: May 05, 2026; Revision: June 21, 2026;
Accepted: June 24, 2026; Available Online: June 28, 2026

Abstract
Heart failure remains a major cause of mortality worldwide, and predicting patient survival has become a key area where machine learning can support clinical decision-making. This study aims to improve the accuracy of survival prediction for patients with heart failure by applying hyperparameter tuning to the Random Forest algorithm. Using a publicly available dataset from the UCI Machine Learning Repository, a structured machine learning pipeline was developed. This includes data preprocessing, outlier treatment using the capping method, stratified data splitting, and model training. The Random Forest model was first trained with default parameters to establish a baseline, then optimized via Grid Search Cross-Validation to identify the best hyperparameter configuration. Results show that the optimized model achieved higher accuracy (80.83%), recall (66.00%), and F1-score (0.7416) than the baseline. These improvements demonstrate that systematic tuning of machine learning models can significantly enhance their predictive capability in clinical settings. The model demonstrated greater sensitivity in identifying high-risk patients, which is essential for early intervention. Although constrained by the dataset size, this study offers a replicable framework for predictive modeling in healthcare and underscores the potential of machine learning for mortality risk stratification.

Keywords: GridsearchCV, Heart Failure, Hyperparameter Tuning, Machine Learning, Random Forest

*Corresponding author: Sari Susanti (sarisusanti@ars.ac.id)

1. Introduction
Heart failure (HF) is a complex clinical syndrome that arises when the heart is unable to pump blood effectively to meet the body's needs (Bozkurt et al., 2021). It has become a significant global health burden, affecting an estimated 26 million people worldwide and driving rising morbidity, mortality, and healthcare costs (Shahim et al., 2023). In Indonesia, data from the 2018 Basic Health Research (Riskesdas) recorded over one million cases of heart failure, with a consistent annual increase (Adisasmito et al., 2020). The rising prevalence of HF emphasizes the urgent need for accurate tools to assess patient prognosis and support clinical decision-making.

One of the critical challenges in managing heart failure is effectively predicting patient survival. Clinicians must often make high-stakes decisions regarding intensive treatment plans, medication adjustments, or palliative care (Cleland, 2021). In this context, predictive modeling can play a vital role. Over the past decade, machine learning (ML) has emerged as a promising tool for clinical prediction tasks due to its ability to learn complex patterns from large and multidimensional datasets (Chicco & Jurman, 2020). This technology works by analyzing patterns in large datasets and using algorithms to classify or predict outcomes (Janiesch et al., 2021).
Several studies have explored the use of ML algorithms for medical purposes. Chicco and Jurman (Chicco & Jurman, 2020), for instance, demonstrated that machine learning models, particularly Random Forest (RF), can predict patient survival using only two features, serum creatinine and ejection fraction, with a reported accuracy of approximately 74%. Their study marked an important step in demonstrating the effectiveness of minimal-feature models(Chicco & Jurman, 2020). However, despite its merits, their work also revealed limitations, such as the exclusion of other relevant clinical features and the lack of hyperparameter optimization to improve model

performance(Sitanggang & Sitompul, 2024). In cardiology, this technology is used in a variety of ways, from analyzing electrocardiograms (ECGs) to detect arrhythmias to assessing coronary artery stenosis via computed tomography angiography (CTA) (Al'Aref et al., 2019). Furthermore, machine learning models are used to estimate survival rates for patients with cardiovascular diseases using clinical data and medical imaging, thereby facilitating early disease detection and improved medical decision-making.

In subsequent studies, Random Forest has continued to demonstrate strong performance in various medical classification tasks due to its robustness, resistance to overfitting, and interpretability (Abubakar et al., 2023). Nevertheless, many of these models are implemented using default parameters, which may not yield optimal results. The process of tuning hyperparameters that define the model structure and learning process, rather than learning them from data, has been shown to significantly affect the performance of machine learning models (Jalal et al., 2022). One well-established method for tuning is Grid Search Cross Validation (Grid Search CV), which systematically explores a predefined range of hyperparameter values and evaluates performance using cross-validation (Pratiwi et al., 2024).

Despite the recognition of the importance of hyperparameter tuning, its application in heart failure survival prediction remains limited (Shah et al., 2020). Few studies have integrated both outlier handling techniques and systematic hyperparameter optimization to enhance the predictive power of Random Forest classifiers (Alfajr & Defiyanti, 2024). This reveals a crucial gap between theoretical advances in ML and their practical implementation in medical predictive modeling (Garg & Mago, 2021). Additionally, there is limited research utilizing publicly available datasets that apply a full ML pipeline from preprocessing and outlier treatment to hyperparameter tuning and validation in a reproducible manner (Reig et al., 2020).

This study aims to address these gaps by developing an improved Random Forest model optimized via Grid Search CV to predict survival among patients with heart failure. Using the publicly available dataset from the UCI Machine Learning Repository, which includes 299 patient records and 13 clinical attributes, this research implements a complete preprocessing pipeline. The pipeline involves outlier detection and treatment using the capping method, stratified data splitting, and rigorous model validation through k-fold cross-validation. The effectiveness of the optimized model is then evaluated and compared with a baseline RF model using standard classification metrics, including accuracy, precision, recall, and F1-score.

By improving the baseline predictive model through thorough data preprocessing and careful hyperparameter tuning, this research makes meaningful contributions to the field. It provides empirical evidence that optimized machine learning models outperform those with default configurations, particularly in medical prediction tasks. Additionally, the study presents a replicable framework for applying machine learning techniques in clinical settings, even with limited data. These insights are expected to support the practical adoption of AI in healthcare environments, providing valuable guidance for medical professionals, data scientists, and institutions seeking to improve decision-making.

2. Methods

In this chapter, the research methods used will be presented as follows:

2.1 Data Collection

This study used secondary clinical data from the publicly available Heart Failure Clinical Records dataset on the UCI Machine Learning Repository. The dataset comprises 299 patient records collected from the Faisalabad Institute of Cardiology and Allied Hospital in Pakistan. Each record includes 13 clinical features, such as age, sex, ejection fraction, serum creatinine, presence of diabetes, anemia, and high blood pressure. The binary target variable indicates whether a patient died during the follow-up period (1 = death, 0 = survival).

The dataset was selected for its accessibility, completeness, and relevance to clinical prediction modeling. Because it is widely used in machine learning research, it enables benchmarking and reproducibility. The dataset's characteristics are outlined in the original documentation and form the foundation for building the predictive model in this study.

2.2 Research Procedure

The study followed a structured experimental procedure comprising several steps: data cleaning and preprocessing, model development, optimization, and evaluation. All experiments were performed using Python 3.10 in Google Colaboratory, with essential libraries including Pandas, NumPy, Scikit-learn, Matplotlib, and Seaborn.
The steps carried out in this research are summarized as follows:

1. Exploratory Data Analysis (EDA): Basic statistical summaries and visualizations were used to understand the data distribution and identify anomalies.

2. Outlier Detection and Treatment: Numerical features were examined using the Interquartile Range (IQR) method to identify outliers. Any values falling outside the range defined in Equation (1) were considered outliers and treated using the capping method.

$$OutlierRange = [Q1 - 1.5 * IQR, Q3 + 1.5 * IQR] \quad (1)$$

3. Feature and Target Separation: The features (X) and target variable (y) were separated for further modeling.
4. Data Splitting: The dataset was split into 80% for training and 20% for testing, using stratified sampling to preserve class balance.

2.3. Machine Learning Model

The predictive model employed in this study is the Random Forest Classifier, an ensemble-based supervised learning algorithm well known for its robustness and generalization capabilities. A baseline model was first trained using default hyperparameters to establish a performance benchmark.

To improve the model, a Grid Search Cross-Validation (Grid Search CV) technique was used to systematically tune hyperparameters. The hyperparameters explored included:

1. n_estimators: [100, 150, 200]
2. max_depth: [5, 10, 15, none]
3. min_samples_split: [2, 4, 6]
4. min_samples_leaf: [1, 2, 3]

The best hyperparameter combination was selected via 5-fold cross-validation, with accuracy as the evaluation metric.

2.4. Model Evaluation

The trained model was evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score. These metrics provide insights into both overall performance and class-wise behavior, particularly important in imbalanced clinical datasets.

Each metric was computed using the following equations:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

$$F1-Score = 2\,X\,\frac{Precision\ X\ Recall}{Precision + Recall} \quad (5)$$

1. TP = True Positive
2. TN = True Negative
3. FP = False Positive
4. FN = False Negative

2.5. Research Workflow

The overall research workflow is illustrated in Figure 1, showing each major step from data collection to model evaluation and result generation. This flow ensures a systematic, reproducible process for developing a predictive machine learning model.

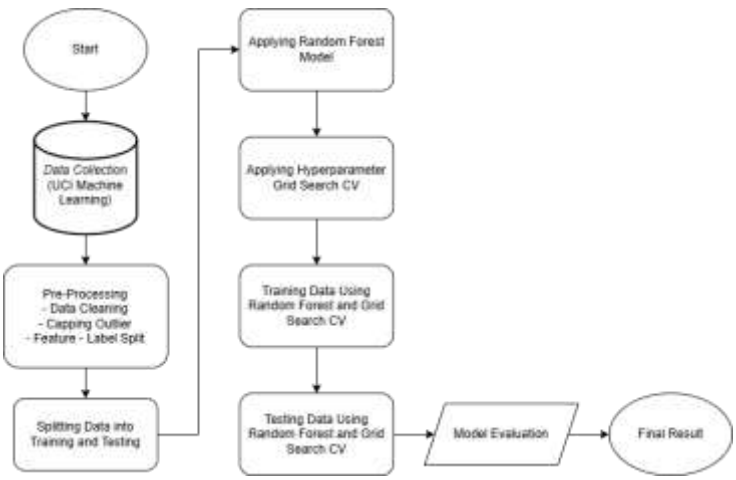

Figure 1. Research Workflow

The flowchart illustrates the sequential steps undertaken in this study, beginning with the acquisition of a publicly available dataset from the UCI Machine Learning Repository. The dataset was preprocessed by data cleaning, outlier treatment using the capping method, and feature-label separation. Subsequently, the data were split into training and test sets using stratified sampling. The baseline Random Forest model was trained and evaluated, followed by hyperparameter tuning using Grid Search Cross Validation to optimize model performance. The optimized model was retrained and tested, and its performance was evaluated using standard classification metrics. The final results were derived from comparative analysis between the baseline and tuned models.

3. Results and Discussion

This section presents the experimental results obtained from both the baseline and optimized Random Forest models. Performance metrics, including accuracy, precision, recall, and F1-score, were used to evaluate each model. The aim was to compare the impact of hyperparameter tuning using Grid Search Cross Validation on predictive performance.

3.1. Baseline Model Performance

The baseline model was developed using the default parameter settings of the Random Forest classifier, without any prior optimization or tuning. This initial model served as a performance benchmark to assess the impact of further enhancements applied later in the study. The model was trained on 60% of the dataset and tested on the remaining 40% to evaluate its generalization capabilities on unseen data.

Table 1 presents the classification results of the baseline model, including accuracy, precision, recall, and F1-score, which are standard metrics for evaluating binary classification performance in clinical prediction tasks.

Table 1. Performance of the baseline Random Forest Model

Metric	Value
Accuracy	74.17%
Precision	77.81%
Recall	74.17%
F1-Score	71.82%

The baseline model achieved an accuracy of 74.17%, indicating that nearly three-quarters of predictions were correct. The recall score of 74.17% indicates that the model was fairly effective in identifying patients who did not survive, which is critical in clinical contexts where false negatives can be costly. A precision score of 77.81% indicates that most patients predicted as deceased were indeed deceased, suggesting a relatively low false-positive rate. The F1-score, which harmonizes precision and recall, was 71.82%, indicating balanced yet modest performance.

3.2. Optimized Model Performance

Following hyperparameter tuning with Grid Search Cross Validation (Grid Search CV), the performance of the Random Forest model improved across all key evaluation metrics. This optimization process involved an exhaustive search over a predefined hyperparameter space to identify the most effective combination that enhances model generalization and accuracy. The refined model was then retrained using the optimal parameters and evaluated on the same test set as the baseline model to ensure a fair comparison.

Table 2 summarizes the optimized model's performance in terms of accuracy, precision, recall, and F1-score.

Table 2. Performance of the optimized Random Forest model

Metric	Value
Accuracy	80.83%
Precision	81.35%
Recall	80.83%
F1-Score	80.35%

The optimized model achieved an accuracy of 80.83%, representing a substantial improvement of over 6 percentage points over the baseline model. This indicates a stronger ability to make correct predictions overall. The precision also increased to 81.35%, suggesting that the model made fewer false-positive predictions and was more reliable at identifying patients who did not survive. Meanwhile, the recall remained high at 80.83%, indicating the model's ability to capture most true positives.

The F1-score, which balances precision and recall, rose from 71.82% to 80.35%, reflecting more stable,

harmonized classification performance. This enhancement confirms that hyperparameter tuning not only improved raw accuracy but also improved the balance between sensitivity and specificity—crucial in high-stakes medical predictions.

3.3. Confusion Matrix Analysis

To gain a deeper insight into the classification behavior of both models, confusion matrices were generated for the baseline and optimized Random Forest classifiers. These matrices provide a breakdown of true and false predictions across the two outcome classes (survived vs. deceased), which is essential for evaluating the models' practical reliability, especially in clinical settings where misclassifications can have serious consequences.

Table 3 and Table 4 present the confusion matrices of the baseline and optimized models

Table 3. Confusion Matrix of the baseline Random Forest Model

	Prediction 0	Prediction 1
Actual 0	67 *True Negative*	3 False Positive
Actual 1	28 *False Negative*	22 *True Positive*

Table 4. Confusion matrix of the optimized Random Forest model

	Prediction 0	Prediction 1
Actual 0	64 *True Negative*	6 False Positive
Actual 1	17 *False Negative*	33 *True Positive*

From the baseline model in Table 3, we observe that the classifier correctly predicted 67 out of 70 negative cases (patients who survived) and 22 out of 50 positive cases (patients who died), while misclassifying 28 actual deaths as survivors (false negatives). This suggests significant underdetection of high-risk cases, a critical limitation in the medical domain where false negatives may lead to missed opportunities for intervention.

In contrast, the optimized model in Table 4 shows an improved classification of actual deaths, with 33 true positives and only 17 false negatives, indicating a notable gain in sensitivity (recall). Although there was a slight increase in false positives (from 3 to 6), this trade-off is generally acceptable in clinical prediction models where minimizing false negatives is often prioritized over false alarms.

3.4. Hyperparameter Selection Results

Table 5 presents the optimal hyperparameter combination obtained via Grid Search Cross Validation. This method systematically evaluated multiple parameter configurations across the training set using five-fold cross-validation to identify the settings that yielded the highest average performance, particularly classification accuracy.

Table 5. Best Hyperparameters for Random Forest Model

Hyperparameter	Selected Value
n_estimators	50
max_depth	none
min_samples_split	2
min_samples_leaf	4
max_features	none
class_weight	balanced

The selected hyperparameters reflect a configuration designed to maximize the model's generalization across clinical data. Using 50 estimators provides sufficient ensemble diversity without excessive computational overhead, while allowing the trees to grow without a depth restriction (max_depth = None), enabling the model to learn detailed hierarchical relationships between features.
Setting min_samples_leaf = 4 reduces the likelihood of overfitting by enforcing a minimum number of data points

at terminal nodes, thereby smoothing decision boundaries. Additionally, using class_weight = balanced addresses class imbalance by assigning higher penalties to the minority class (death events), thereby improving sensitivity to rare but clinically important cases.

3.5. Comparison Summary

To visualize the impact of hyperparameter tuning, a comparative bar chart was created (Figure 2). This figure illustrates the changes in four key evaluation metrics—accuracy, precision, recall, and F1 Score—between the baseline Random Forest model and the tuned version obtained via Grid Search Cross Validation.

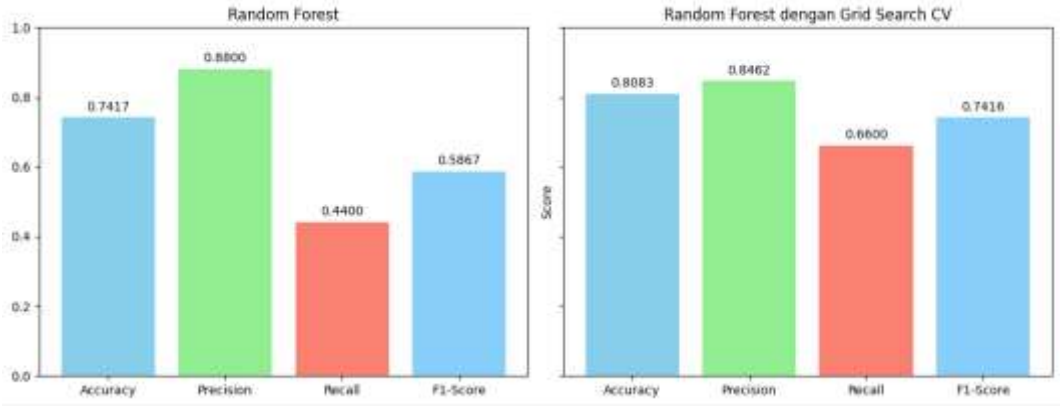

Figure 2. Comparison of baseline and optimized model performance

From Figure 2, it is evident that the optimized model achieved a notable increase in accuracy, rising from 74.17% to 80.83%, demonstrating improved overall predictive reliability. Interestingly, precision decreased slightly from 88.00% to 84.62%, suggesting a minor increase in false positives. However, this trade-off was accompanied by a substantial improvement in recall, rising from 44.00% to 66.00%, indicating a much better ability to identify patients who did not survive correctly.
Most notably, the F1-score, which balances precision and recall, increased from 0.5867 to 0.7416. This significant gain reflects more harmonious, stable model performance, especially in class-imbalanced contexts, where both sensitivity and specificity are critical.

4. Conclusion

The results of this study demonstrate that applying Grid Search Cross-Validation significantly improved model performance. Notably, the optimized model achieved higher accuracy and a substantial increase in recall, reflecting a more sensitive approach to identifying patients at risk of mortality. These enhancements are especially meaningful in a medical context, where accurate identification of high-risk patients can facilitate earlier interventions, targeted treatments, and ultimately, better patient outcomes.

In addition to its technical contributions, this study reinforces the value of hyperparameter tuning as an essential step in machine learning model development. While many prior studies rely on default algorithm settings, this research shows that even well-established algorithms like Random Forests can yield significantly better results when carefully fine-tuned. Furthermore, the use of confusion matrix analysis and metric comparison provided deeper insights into the model's practical behavior, highlighting the trade-offs between precision and recall common in healthcare prediction tasks.

Despite its contributions, the study is not without limitations. The dataset used, while clinically relevant, is relatively small and imbalanced, which may affect the model’s ability to generalize to broader patient populations. The findings, therefore, should be interpreted as a foundational step rather than a conclusive clinical tool. Future work should focus on incorporating additional datasets from diverse sources, exploring more advanced ensemble or deep learning models, and implementing interpretability tools such as SHAP or LIME to ensure transparency and clinical trust.

In summary, this research confirms that the predictive performance of machine learning models in medical applications can be significantly improved through methodological rigor and thoughtful optimization. The Random Forest model, when properly tuned, offers a promising, interpretable solution for mortality prediction in patients with heart failure. These insights contribute to the growing field of machine learning in healthcare and provide a replicable framework for future research aiming to integrate artificial intelligence into clinical decision-making processes.

References

Abubakar, M. A., Muliadi, M., Farmadi, A., Herteno, R., & Ramadhani, R. (2023). Random Forest Dengan Random Search Terhadap Ketidakseimbangan Kelas Pada Prediksi Gagal Jantung. Journal Informatika, 10(1), 13–18. https://doi.org/10.31294/inf.v10i1.14531

Adisasmito, W., Amir, V., Atin, A., Megraini, A., & Kusuma, D. (2020). Geographic and socioeconomic disparity in cardiovascular risk factors in Indonesia: Analysis of the basic health research 2018. BMC Public Health, 20. https://doi.org/10.1186/s12889-020-09099-1

Al'Aref, S. J., Anchouche, K., Singh, G., Slomka, P. J., Kolli, K. K., Kumar, A., Pandey, M., Maliakal, G., Van Rosendael, A. R., Beecy, A. N., Berman, D. S., Leipsic, J., Nieman, K., Andreini, D., Pontone, G., Schoepf, U. J., Shaw, L. J., Chang, H. J., Narula, J., … Min, J. K. (2019). Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging. In European Heart Journal (Vol. 40, Number 24, pp. 1975–1986). Oxford University Press. https://doi.org/10.1093/eurheartj/ehy404

Alfajr, N. H., & Defiyanti, S. (2024). Prediksi Penyakit Jantung Menggunakan Metode Random Forest Dan Penerapan Principal Component Analysis (PCA). Jurnal Informatika Dan Teknik Elektro Terapan, 12(3S1). https://doi.org/10.23960/jitet.v12i3S1.5055

Bozkurt, B., Coats, A. J., Tsutsui, H., Abdelhamid, M., Adamopoulos, S., Albert, N., Anker, S. D., Atherton, J., Böhm, M., Butler, J., Drazner, M. H., Felker, G. M., Filippatos, G., Fonarow, G. C., Fiuzat, M., Gomez-Mesa, J. E., Heidenreich, P., Imamura, T., Januzzi, J., … Zieroth, S. (2021). Universal Definition and Classification of Heart Failure: A Report of the Heart Failure Society of America, Heart Failure Association of the European Society of Cardiology, Japanese Heart Failure Society and Writing Committee of the Universal Definition of Heart Failure. Journal of Cardiac Failure, 27(4), 387–413. https://doi.org/10.1016/j.cardfail.2021.01.022

Chicco, D., & Jurman, G. (2020). Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. BMC Medical Informatics and Decision Making, 20(1). https://doi.org/10.1186/s12911-020-1023-5

Cleland, J. G. F. (2021). The struggle towards a Universal Definition of Heart Failure—how to proceed? European Heart Journal, 42(24), 2331–2332. https://doi.org/10.1093/eurheartj/ehab082

Garg, A., & Mago, V. (2021). Role of machine learning in medical research: A survey. In Computer Science Review (Vol. 40). Elsevier Ireland Ltd. https://doi.org/10.1016/j.cosrev.2021.100370

Jalal, N., Mehmood, A., Choi, G. S., & Ashraf, I. (2022). A novel improved random forest for text classification using feature ranking and optimal number of trees. Journal of King Saud University - Computer and Information Sciences, 34(6), 2733–2742. https://doi.org/10.1016/j.jksuci.2022.03.012

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. https://doi.org/10.1007/s12525-021-00475-2/Published

Pratiwi, N. K. C., Ibrahim, N., & Saidah, S. (2024). Prediksi Kanker Paru menggunakan Grid search untuk Optimasi Hyperparameter pada Algoritma MLP dan Logistic Regression. Elkomika: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika, 12(3), 556. https://doi.org/10.26760/elkomika.v12i3.556

Reig, B., Heacock, L., Geras, K. J., & Moy, L. (2020). Machine learning in breast MRI. In Journal of Magnetic Resonance Imaging (Vol. 52, Number 4, pp. 998–1018). John Wiley and Sons Inc. https://doi.org/10.1002/jmri.26852

Shah, D., Patel, S., & Bharti, S. K. (2020). Heart Disease Prediction using Machine Learning Techniques. SN Computer Science, 1(6). https://doi.org/10.1007/s42979-020-00365-y

Shahim, B., Kapelios, C. J., Savarese, G., & Lund, L. H. (2023). Global Public Health Burden of Heart Failure: An Updated Review. Cardiac Failure Review, 9. https://doi.org/10.15420/cfr.2023.05

Sitanggang, B. F., & Sitompul, P. (2024). Deteksi Awal Kelangsungan Hidup Pasien Gagal Jantung Menggunakan Machine Learning Metode Random Forest. Innovative: Journal Of Social Science Research, 4, 3347–3357.