Paradigma, Vol. 26, No. 1, Maret 2024, P-ISSN 1410-5063, E-ISSN: 2579-3500, Page. 30-37
Published: LPPM Universitas Bina Sarana Informatika
PARADIGMA
website: http://jurnal.bsi.ac.id/index.php/paradigma/

Optimizing Heart Failure Detection: A Comparison between Naive Bayes and Particle Swarm Optimization

Abdul Hamid[1], Ridwansyah[2*]

[1]Universitas Bina Sarana Informatika, Jakarta, Indonesia
[2*]Universitas Nusa Mandiri, Jakarta, Indonesia

ARTICLE INFORMATION

Artikel History:
Received: February 15, 2024
Revised: March 5, 2024
Accepted: March 21, 2024

Keyword:

Heart Failure
Naïve Bayes
Optimizing
Particle Swarm Optimization

ABSTRACT

This research focuses on the importance of early detection of heart failure which is a serious global health problem. Given the variety of symptoms of heart failure, accurate early detection methods are needed with the aim of reducing the impact of this disease. This study uses the Naïve Bayes (NB) method which has been proven effective in classifying heart failure with significant variations in accuracy by integrating Particle Swarm Optimization (PSO) to improve the model. The evaluation model involves a confusion matrix including accuracy, precision, recall, and Area Under the Curve. The research results show that the integration of PSO in NB results in an increase in accuracy of 7.73%, an increase in precision of 6.42%, and an increase in recall of 1.93%. Although there was a small decrease in AUC. This research shows that the success of NB with PSO can help improve the performance of early detection of heart failure. This indicates the importance of this research in developing more accurate and effective detection methods for critical health conditions such as heart failure.

Corresponding Author:
Ridwansyah,
Informatika,
Universitas Nusa Mandiri,
Jln. Jatiwaringin Raya No. 02, Jakarta, (021) 28534236
Email: ridwansyah.rid@nusamandiri.ac.id

INTRODUCTION

Heart failure is one of the types of cardiovascular diseases that have a significant global health impact. This disease affects many individuals worldwide and leads to serious problems with quality of life and mortality rates. Various symptoms can arise due to heart failure, such as difficulty breathing, fluid buildup in the lungs, and excessive fatigue. Therefore, it is important to detect this disease early and manage it effectively to reduce its impact on patients.

Early detection of heart failure is one of the areas receiving significant attention in health data science research, due to its potential in early detection which functions to save lives and can aid in identifying individuals at risk of heart failure since the early stages of the disease. Therefore, in efforts to improve the accuracy of heart failure prediction, various Machine Learning techniques have been explored such as Gradient Boosting Classifier, Extra Tree Classifier, Gaussian Naïve Bayes, and Support Vector Machine. (Ishaq et al., 2021), Random Forest (Pal & Parija, 2021), Logistic Regression, Neural Network (Yaqin, Laksito, & Fatonah, 2021), XGBoost (Sachdeva, Singh, Sharma, Bathla, & Solanki, 2023) dan KNN (Mamun, Farjana, Mamun, Ahammed, & Rahman, 2022). Among these techniques, a comparison of the results in terms of area under the curve (AUC) values and accuracy is conducted (Priyatama & Ridwansyah, 2022). One common approach is to use statistical methods such as Naive Bayes. Various studies in the context of heart failure have been conducted to build predictive models based on clinical data and patient medical information, and they have shown effectiveness in both binary and multiclass classification. Naive Bayes exhibits significant variability in accuracy, as demonstrated by existing research findings indicating that Naive Bayes achieves an accuracy of 84.17%, suggesting that this method is capable of performing classification quite well (Alotaibi, 2019). They also reported that K-nearest neighbour in their study achieved a strong accuracy rate of 96.00% (Alaiad, Najadat, Mohsen, & Balhaf, 2020). Another study reported achieving an accuracy of 86.18% using the Naive Bayes method for heart

DOI: https://doi.org/10.31294/p.v26i1.3284

failure disease (Ridwan, Pratama, Prihandono, & Intakoris, 2021). However, on the other hand, Naive Bayes obtained a lower accuracy of 69.60% (Chicco & Jurman, 2020). The low accuracy results may be attributed to differences in the dataset, data preprocessing techniques, and classification parameters utilized. The Naive Bayes method, being a probabilistic classification method, has been widely adopted due to its ease of implementation and efficiency in handling large-dimensional datasets (Barus, Lauwren, Pangaribuan, & Romindo, 2023). Naive Bayes has demonstrated good results in various studies, but there is still room for improvement. The current issue with the Naive Bayes method lies in its limited ability to handle the complexity of diverse data and cope with dynamic data changes in the risk factors of heart failure diseases, which have not yet achieved maximum accuracy results. Therefore, to enhance the low or suboptimal accuracy results of the Naive Bayes method, a technique is required to address these challenges, such as utilizing the Particle Swarm Optimization (PSO) method.

PSO is an optimization method based on simulated group movement, which can be used to identify optimal parameters in predictive models by adjusting feature weights and model parameters. (Ariyati et al., 2020). Therefore, PSO can enhance prediction accuracy in detecting heart failure (Iqbal et al., 2020). PSO dapat menawarkan cara untuk mengoptimalkan kinerja naive bayes lebih lanjut melalui seleksi fitur yang lebih baik dalam menangani PSO can offer a way to further optimize the performance of Naive Bayes through better feature selection in handling data complexity, thus achieving improved accuracy results obtained by the Naive Bayes method (Nurdin, Sartini, Sumarna, Maulana, & Riyanto, 2023). PSO has been utilized to select features and optimize parameters in various machine learning algorithms, including Naive Bayes (Ridwansyah, Ariyati, & Faizah, 2019).

The issues related to this problem include challenges in collecting and processing accurate medical data, precision in identifying risk factors, as well as comparing the performance between Naive Bayes and PSO methods. This comparison will help determine whether the PSO approach can produce a superior predictive model compared to traditional methods like Naive Bayes. Integrating these two methods may provide a more effective approach for heart failure detection.

By outlining the background of this issue, identifying related issues, and referring to previous research, this study aims to bridge the knowledge gap and make a significant contribution to the development of more effective and accurate early detection methods for heart failure. Additionally, it aims to improve the accuracy of Naive Bayes method in predicting heart failure.

In summary, this research aims to address the challenges in data accuracy and feature selection, compare the performance of Naive Bayes and PSO methods, and contribute to the advancement of early heart failure detection methods.

RESEARCH METHOD

This recent study proposes the use of an innovative classification method specifically designed to optimize the heart failure detection process. This method is unique in that it leverages a dataset consisting of 12 diverse attributes, including patient age, presence of anemia, gender, and various other relevant factors, all of which contribute to improving detection accuracy. To provide a deeper understanding of how this methodology works and its application in practice, a detailed illustration of the classification process is presented in Figure 1.

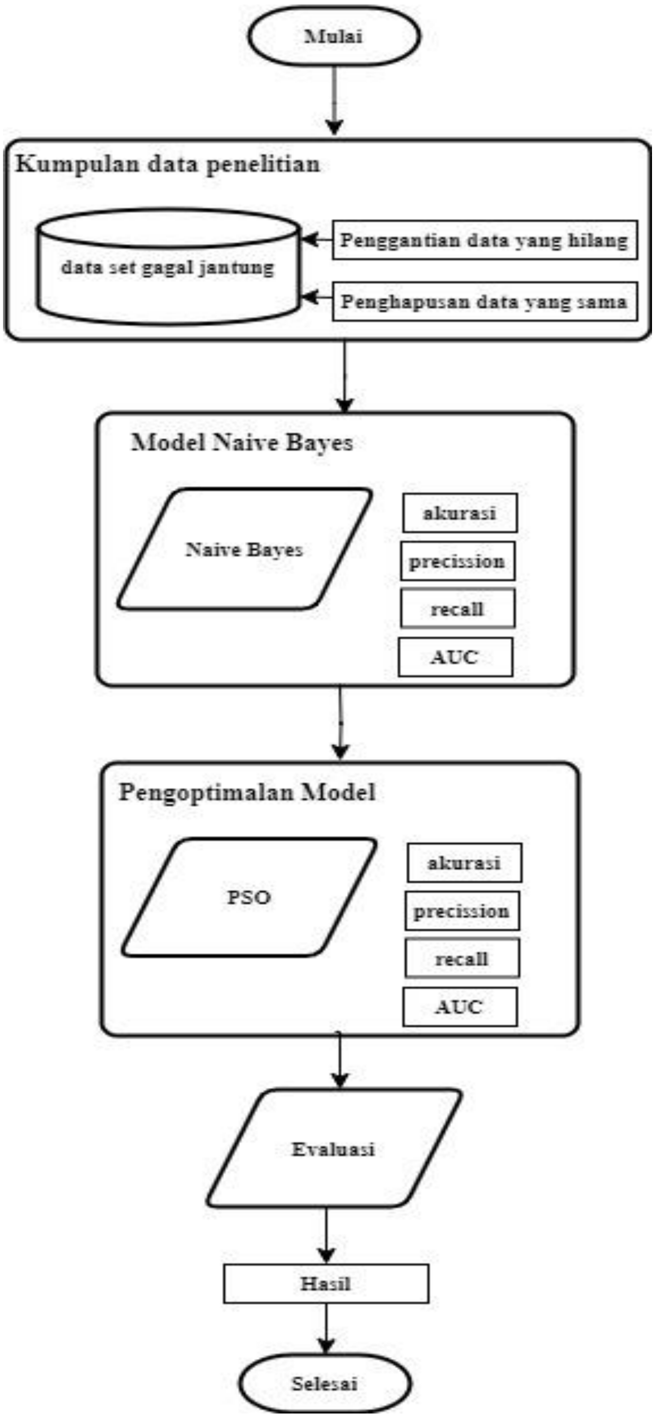

Figure 1. Research Design

Research Data Collection

Figure 1 depicts the initial stage of the research involving the collection of heart failure data consisting of 12 attributes including age, anemia, diabetes, high blood pressure, platelet count, gender, smoking habit, and others. This data, part of a public repository containing heart failure datasets from relevant sources such as the UCI Repository, comprises a total of 299 datasets, each with 12 attributes.

During the data collection phase, adjustments were made to the heart failure dataset. Specific attributes were included in the classification model, and missing data were addressed through data cleansing by identifying and replacing missing data with the mean value. Duplicate data were removed to ensure the uniqueness of each data sample and to optimize the classification process. Additionally, data type conversions were performed, such as converting gender to numerical values (1 for female and 0 for male) and converting nominal data to numerical data. This data is commonly used to train and evaluate the performance of Naive Bayes classification models. The data collection process is further detailed in Table 1.

Table 1 Data Processing

No	Attribute	Type	Value
1	Age	Numerical	Integer
2	Anemia	Strings	Yes, No
3	Creatinine phosphokinase	Numerical	Integer
4	Diabetes	Strings	Yes, No
5	Ejection fraction	Numerical	Integer
6	High blood pressure	Strings	Yes, No
7	Platelets	Numerical	Real
8	Serum creatinine	Numerical	Real
9	Serum_sodium	Numerical	Integer
10	Gender	Strings	Male, Female
11	Smoke	Strings	Yes, No
12	Time	Numerical	Integer

Naïve Bayes Model

In the process of applying the method utilizing heart failure-related data that has undergone initial processing as the main input. After completing this initial stage, the next process to be executed is the classification of heart failure data. Classification is done by applying the Naïve Bayes algorithm in training the model with previously cleaned data sets. Through the implementation of this method, we can generate and measure several important performance parameters in model testing to determine initial metrics such as accuracy, precision, recall, and Area Under the Curve (AUC) related to the application of the Naïve Bayes method in this context.

Model Optimization

In the model optimization process using the Particle Swarm Optimization (PSO) algorithm to optimize Naive Bayes model parameters. By re-evaluating the model with optimized parameters to measure improvements in metrics such as accuracy, precision, recall, and AUC. So that the values of these metrics will be maximal.

Evaluation

The evaluation results are a study of the results obtained from the optimized model. This is done using cross-validation to ensure the stability and confidence of the model. The main objective is to determine the success of the model optimization method by comparing the accuracy levels between optimized and unoptimized Naïve Bayes methods.

RESULTS AND DISCUSSION

Collection of Research Data

The data used as the basis of this research is the result of collection from various secondary data sources originating from the UCI Repository. Specifically, the dataset related to heart failure provides a very detailed and comprehensive insight into various patient health metrics, which will subsequently be analyzed through the application of the naïve Bayes method and also through optimization techniques, namely the Particle Swarm Optimization (PSO) method. Before being used, this dataset has gone through a series of important processes, including imputation to replace missing data by using the average value calculated from each available feature, as well as the elimination process against data identified as duplicates. These steps are taken to ensure that the resulting dataset is accurate and reliable. More detailed information about the results of these processes is presented in Table 2, which provides a sample of the data used in this research.

Table 2. Sample Data after Preprocessing

age	anaemia	creatinpho	diabetes	ejection fraction	high bld_pres	platelets	serum creatin	Serum sodium	sex	smoking	time	Death
75	1	81	0	38	1	368000	4	131	1	1	10	1
62	0	231	0	25	1	253000	0.9	140	1	1	10	1
45	1	981	0	30	0	136000	1.1	137	1	0	11	1
50	1	168	0	38	1	276000	1.1	137	1	0	11	1
49	1	80	0	30	1	427000	1	138	0	0	12	0
82	1	379	0	50	0	47000	1.3	136	1	0	13	1
87	1	149	0	38	0	262000	0.9	140	1	0	14	1
45	0	582	0	14	0	166000	0.8	127	1	0	14	1
…	…	….	….	….	…	…	…	…	…	…	…	…
48	1	582	1	55	0	87000	1.9	121	0	0	15	1

Paradigma, Vol. 26, No. 1, Maret 2024, P-ISSN 1410-5063, E-ISSN: 2579-3500, Page. 30-37
Published: LPPM Universitas Bina Sarana Informatika
PARADIGMA
website: http://jurnal.bsi.ac.id/index.php/paradigma/

Naïve Bayes Model

The data that has been collected and undergone preprocessing will then be used to process the collected and preprocessed data further. This method produces a Confusion Matrix, which is a way to measure the performance of classification models like Naive Bayes. The Confusion Matrix provides information about the actual performance of the model compared to the predictions made by the model. The model's performance is measured by values including accuracy, precision, recall, and Area Under the Curve (AUC) using the Naive Bayes method, the results of which can be seen in Table 3.

Table 3. Confusion Matrix Naïve Bayes

	true hf	true nothf
pred hf	36	17
pred nothf	60	186

From Table 3, which shows the number of correct and incorrect predictions made by the Naive Bayes model, we can understand that True 1 (Actual Positive) represents data that is actually positive, with a total of 36 True Positives (TP), indicating the number of cases correctly predicted as positive by the model. True 0 (Actual Negative) represents data that is actually negative, with a total of 17 False Positives (FP), indicating the number of cases that are actually negative but predicted as positive by the model. Pred 1 (Predicted Positive) represents data predicted as positive by the model, with a total of 60 False Negatives (FN), indicating the number of cases that are actually positive but predicted as negative by the model. Pred 0 (Predicted Negative) represents data predicted as negative by the model, with a total of 186 True Negatives (TN), indicating the number of cases that are truly negative and also predicted as negative by the model. From the results of this confusion matrix, Accuracy, Precision, Recall in Table 4, and AUC as seen in Figure 2 can be generated.

Table 4. The Accuracy, Precision, and Recall results of Naïve Bayes

Method	Accuracy	Precision	Recall
Naïve Bayes	74,24	75,96	91,67

The results from table 4 show an accuracy of 74.24%, indicating that the model is reasonably accurate, although there is still room for improvement as it does not reach precision above 90%. With a precision of 75.96%, it means that when the Naïve Bayes model predicts heart failure as positive, around 76% of those predictions are correct. With a recall value of 91.67%, the Naïve Bayes model is very good at identifying positive cases.

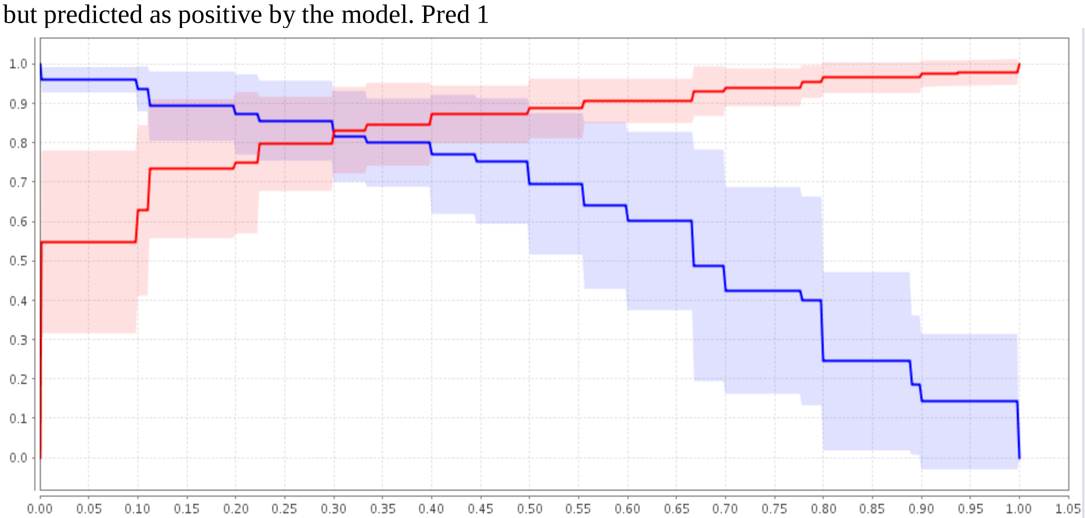

Fig 2. AUC Naïve Bayes

Figure 2 depicts the AUC metric obtained from heart failure data processing, which functions to measure how well the model can differentiate between different classes, with the AUC value of the Naïve Bayes model being 0.847. This indicates that the model has fairly good discriminative ability, where a value of 1.0 indicates perfect prediction and 0.5 indicates no better performance than chance.

Model Optimization

Following the heart failure data testing using Naïve Bayes, model optimization will be performed using the Particle Swarm Optimization method. The performance of the Particle Swarm Optimization method involves eliminating attributes without weights, as shown in Table 5.

DOI: https://doi.org/10.31294/p.v26i1.3284

Table 5: Attribute Weights for Heart Failure Disease

Attribute	Weight
age	0
anaemia	1
Creatinine_phosphokinase	1
diabetes	0
Ejection_fraction	1
High_blood_pressure	0
platelets	1
Serum_creatinine	0
Serum_sodium	0
sex	0.877
smoking	0
time	0.796

In Table 5, it can be explained that there are several relevant factors for the disease or clinical conditions, such as age, anemia, creatine phosphokinase levels, diabetes, ejection fraction, high blood pressure, platelet count, serum creatinine levels, serum sodium levels, gender, smoking, and time. The attribute weight results indicate the relative importance of each attribute by generating values greater than 0. However, some attributes like age and serum sodium levels have a weight of 0, indicating that they may not be important in the model or analysis being optimized using the Naïve Bayes method. On the other hand, attributes with a weight of 1 indicate their utmost importance in the model or analysis. Attribute weights above 0 and below 1 signify a high but not maximal level of importance and will result in a confusion matrix table in Table 6.

Table 6: Confusion Matrix for Model Optimization

	true hf	**true nothf**
pred hf	55	13
pred nothf	41	190

From Table 6, which shows the number of correct and incorrect predictions made by the Particle Swarm Optimization model, we can understand that: There are 55 cases correctly predicted as positive by the model. There are 33 cases actually negative but predicted as positive by the model. There are 60 cases actually positive but predicted as negative by the model. There are 186 cases correctly predicted as negative by the model.

Based on this confusion matrix, we can calculate Accuracy, Precision, and Recall, which are presented in Table 7, along with the AUC depicted in Figure 3.

Table 7: Naïve Bayes + PSO Accuracy, Precision, and Recall Results

Method	**Accuracy**	**Precision**	**Recall**
Naïve Bayes + Particle Swarma Optimization	81,97	82,38	93,60

The results from Table 7 show an accuracy of 81.97%, indicating a good performance of the model in distinguishing between the existing classes. With a precision of 82.38%, it means that when the Naïve Bayes and Particle Swarm Optimization model predicts heart failure as positive, around 82.38% of those predictions are correct. With a recall value of 93.60%, the PSO model is very good at identifying positive cases.

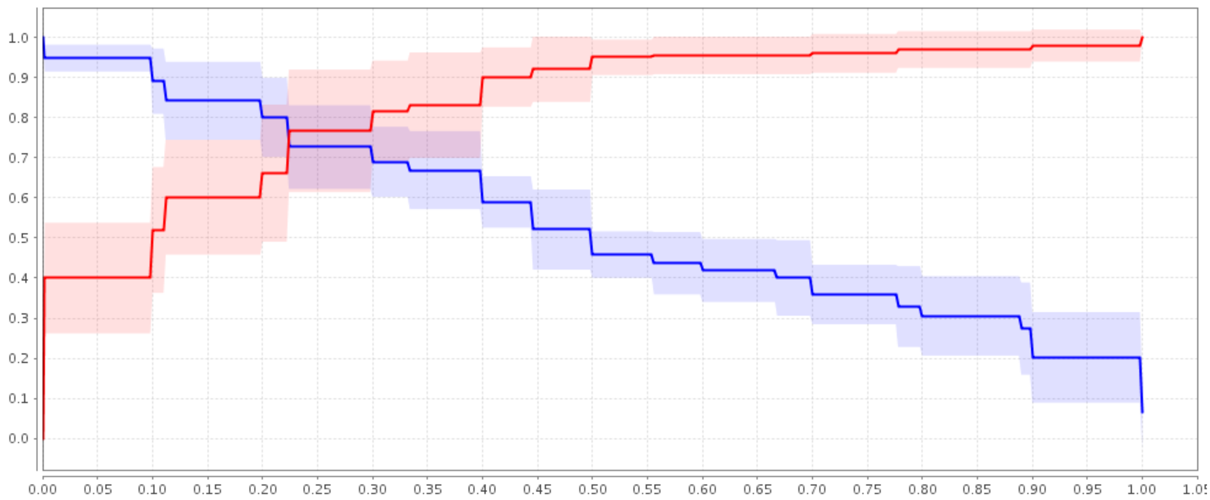

Fig 3. AUC of Naïve Bayes + Particle Swarm Optimization.

Figure 3 depicts the AUC metric obtained from heart failure data processing with model optimization aimed at measuring how well the model can distinguish between different classes. The AUC value of the Naïve Bayes model with Particle Swarm Optimization is 0.830, indicating that the model exhibits fairly good discriminative ability. A value of 1.0 suggests perfect prediction, while 0.5 indicates performance no better than chance.

Evaluation

The following is the research evaluation based on the information available in the graph in Figure 4.

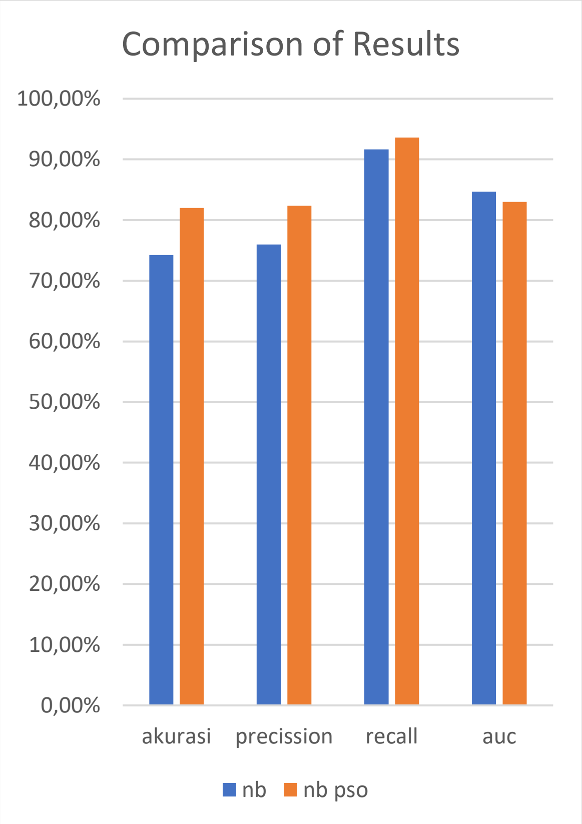

Figure 4: Comparison of Results between NB Method and NB+PSO Method

The accuracy, precision, and recall values of both methods indicate nearly similar performance, with NB+PSO showing slightly higher values. This suggests that for the dataset used in the study, the NB+PSO method is slightly better at correctly classifying data, making accurate predictions, and identifying all positive cases from the relevant class. However, there is a decrease in the AUC value.

CONCLUSION

The final results of optimizing the Naïve Bayes method with Particle Swarm Optimization show that through model optimization, there is an increase of 07.73% in accuracy, an increase of 06.42% in precision, and an increase of 1.93% in recall values. However, there is a slight decrease in the AUC by 0.170%. In the next stage of development, there is an opportunity to conduct deeper research to address challenges related to the decrease in the Area Under the Curve (AUC). One approach to consider is the implementation of the Particle Swarm Optimization (PSO) method. This method has the potential to provide innovative solutions for efficiently selecting features without significantly impacting the AUC value negatively. Additionally, there is a possibility to explore and identify new features that are more relevant and effective in enhancing the accuracy of heart failure prediction. This approach is expected to provide new insights in related research and assist in improving prediction outcomes, thus representing a significant step in enhancing predictive techniques in the medical field, particularly concerning heart failure.

REFERENCES

Alaiad, A., Najadat, H., Mohsen, B., & Balhaf, K. (2020). Classification and Association Rule Mining Technique for Predicting Chronic Kidney Disease. *Journal of Information and Knowledge Management*, *19*(1). https://doi.org/10.1142/S0219649220400158

Alotaibi, F. S. (2019). Implementation of machine learning model to predict heart failure disease. *International Journal of Advanced Computer Science and Applications*, *10*(6), 261–268. https://doi.org/10.14569/ijacsa.2019.0100637

Ariyati, I., Rosyida, S., Ramanda, K., Riyanto, V., Faizah, S., & Ridwansyah. (2020). Optimization of the Decision Tree Algorithm Used Particle Swarm Optimization in the Selection of Digital Payments. *Journal of Physics: Conference Series*, *1641*(1). https://doi.org/10.1088/1742-6596/1641/1/012090

Barus, O. P., Lauwren, K., Pangaribuan, J. J., & Romindo. (2023). Implementation of the Naive Bayes Algorithm to Predict the Safety of Heart Failure Patients. *SEMNASTIK 2023*, *4*(1). https://doi.org/https://doi.org/10.34306/conferenceseries.v4i1.651

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), 1–16. https://doi.org/10.1186/s12911-020-1023-5

Iqbal, M., Herliawan, I., Ridwansyah, Gata, W., Hamid, A., Purnama, J. J., & Yudhistira. (2020). Implementation of Particle Swarm Optimization Based Machine Learning Algorithm for Student Performance Prediction. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, *6*(2), 195–204. https://doi.org/10.33480/jitk.v6i2.1695.IMPLEMENTATION

Ishaq, A., Sadiq, S., Umer, M., Ullah, S., Mirjalili, S., Rupapara, V., & Nappi, M. (2021). Improving the

Prediction of Heart Failure Patients' Survival Using SMOTE and Effective Data Mining Techniques. *IEEE Access*, *9*, 39707–39716. https://doi.org/10.1109/ACCESS.2021.3064084

Mamun, M., Farjana, A., Mamun, M. Al, Ahammed, M. S., & Rahman, M. M. (2022). Heart failure survival prediction using machine learning algorithm: am I safe from heart failure? *2022 IEEE World AI IoT Congress (AIIoT)*. IEEE. https://doi.org/10.1109/AIIoT54504.2022.9817303

Nurdin, H., Sartini, Sumarna, Maulana, Y. I., & Riyanto, V. (2023). Prediction of Student Graduation with the Neural Network Method Based on Particle Swarm Optimization. *Sinkron : Jurnal Dan Penelitian Teknik Informatika*, *8*(4), 2353–2362. https://doi.org/10.33395/sinkron.v8i4.12973

Pal, M., & Parija, S. (2021). Prediction of Heart Diseases using Random Forest. *Journal of Physics: Conference Series*, *1817*(1). https://doi.org/10.1088/1742-6596/1817/1/012009

Priyatama, I. M. D., & Ridwansyah. (2022). Klasifikasi Anak Berkebutuhan Khusus Tunagrahita Menggunakan Metode Algoritma C4.5. *Paradigma*, *24*(1), 90–95. https://doi.org/https://doi.org/10.31294/paradigma.v24i1.1087 Paradigma,

Ridwan, A., Pratama, T. G., Prihandono, A., & Intakoris, S. (2021). Classification of heart failure using the naïve bayes algorithm. *8th International Conference on Public Health 2021,*. https://doi.org/10.26911/ICPHmedicine.FP.08.2021.10

Ridwansyah, Ariyati, I., & Faizah, S. (2019). PARTICLE SWARM OPTIMIZATION BERBASIS CO-EVOLUSIONER DALAM EVALUASI KINERJA ASISTEN DOSEN. *Jurnal SAINTEKOM*, *9*(2), 165–177. https://doi.org/https://doi.org/10.33020/saintekom.v9i2.96

Sachdeva, R. K, Singh, K. D., Sharma, S., Bathla, P., & Solanki, V. (2023). An Organized Method for Heart Failure Classification. *2023 International Conference on Emerging Smart Computing and Informatics (ESCI)*. https://doi.org/10.1109/ESCI56872.2023.10099809

Yaqin, A., Lassito, A. D., & Fatonah, S. (2021). Evaluation of Backpropagation Neural Network Models for Early Prediction of Student's Graduation in XYZ University. *International Journal on Advanced Science Engineering Information Technology*, *11*(2). https://doi.org/10.18517/ijaseit.11.2.11152