Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024
DOI : https://doi.org/10.33395/sinkron.v9i1.13017
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Improving the Accuracy of Heart Failure Prediction Using the Particle Swarm Optimization Method

Dewi Yuliandari[1)*], Yudhistira[2)], Anus Wuryanto[3)], Sidik[4)], Findi Ayu Sariasih[5)]
[1,2,3)] Universitas Bina Sarana Informatika, Indonesia
[4,5)] Universitas Nusa Mandiri, Indonesia
[1)]dewi.dwy@bsi.ac.id, [2)]yudhistira.yht@bsi.ac.id, [3)]anus.awu@bsi.ac.id,
[4)]sidik.sdk@nusamandiri.ac.id, [5)]findi.fav@nusamandiri.ac.id

Submitted: Sep 16, 2023 | **Accepted** : Oct 23, 2023 | **Published** : Jan 1, 2024

Abstract: Malfunction of the body's organs, heart failure, makes a person and those closest to them feel worried, which will result in the person's death. There are too many deaths caused by this deadly disease every year. The main problem throughout the world is heart disease, the death rate of which is getting higher and higher and is uncontrollable for various parties due to many factors, especially in terms of knowing it early or not being able to predict it accurately. Therefore, the aim of this research on heart failure problems is to improve heart failure predictions with optimal accuracy, namely the neural network method with the particle swarm optimization method. Previously, heart failure prediction research had been carried out using several methods, but here we will increase the accuracy of the methods that have been carried out. After the testing process on the neural network method and after being optimized and getting results from the particle swarm optimization method, the accuracy increased with an increase of 08.35%. As well as increasing AUC results with an increase of 0.067%. From the results of increasing the accuracy of the neural network method, testing the particle swarm optimization method on heart failure disease data can be used as a reference for stakeholders.

Keywords: Data Mining; Heart Failure; Improve Accuracy; Neural Network; Particle Swarm Optimization.

INTRODUCTION

One of the most serious health problems throughout the world is heart failure. Because heart failure causes people to experience impaired heart function which is caused by the heart's ability to pump blood throughout the body not working properly. Heart failure can be a deadly disease, with high mortality rates worldwide. If not diagnosed early and treated appropriately, this condition can be fatal. One of the main obstacles in treating heart failure is the lack of accuracy in predicting the possible risk of heart failure in the general public, one of the factors being the lack of health professionals in predicting heart failure. Therefore, having the ability for accurate prediction of heart failure risk is essential in modern medical practice. With the hope that in dealing with heart failure, stakeholders can act more quickly and increase the accuracy of heart failure risk predictions through the application of existing methods in datamining techniques. In recent years in data mining, the use of computational techniques and artificial intelligence has become increasingly important in the world of medicine to assist in the prediction and diagnosis of heart failure. And there have been several previous studies on heart failure that have been carried out, here are the studies that have been carried out.

* Dewi Yuliandari

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024
DOI : https://doi.org/10.33395/sinkron.v9i1.13017
e-ISSN : 2541-2019
p-ISSN : 2541-044X

In predicting heart failure, several methods have been used, including Regression tree, Bagged regression tree, Random forest, Boosted regression tree (depth 1 to depth 4), Logistic regression (simple model or full model), Support vector machines by generating and comparing AUC, Sensitivity and Specificity values (Austin, Tu, Ho, Levy, & Lee, 2013). Apart from producing these criteria, there are also those who produce and compare accuracy values in research that has been carried out using many methods and with different accuracies, therefore in the current research we will use the Neural Network (NN) method. However, in previous research the neural network method was still lacking in accuracy. Therefore, to optimize the neural network method, an optimization method, namely the particle swarm optimization method, will be used (I. Ariyati et al., 2020).

One method that has attracted the attention of researchers is Particle Swarm Optimization (PSO), which is an optimization technique inspired by the behavior of animal groups in the wild (Ridwansyah & Purwaningsih, 2018). PSO is an optimization method that can overcome the complexity and multidimensionality of medical data used in heart failure risk modeling (Indah Ariyati, Ridwansyah, & Suhardjono, 2018). The PSO method will be used to optimize the parameters in the prediction model, resulting in a more accurate and reliable model (Ridwansyah, Ariyati, & Faizah, 2018). PSO has been successfully applied in various fields, including disease prediction. By implementing PSO, we hope to identify more precise risk factors and enable earlier and more effective treatment.

Objectives This research has two objectives. First, to increase the accuracy of the method that has been used, namely NN, by improving it using the PSO method. Second, to compare the accuracy of heart failure predictions using comparison methods from previous datamining literature which have been carried out with the PSO optimization method against NN. In this research, we will develop a prediction model that can utilize clinical data and relevant risk factors for heart failure. It is hoped that the results of this study will help health professionals to provide more timely interventions to high-risk patients, with the ultimate goal of reducing the death rate from heart failure and improving the quality of life of affected patients. As a potential solution, it is hoped that this research can make a significant contribution in treating this disease that claims many lives and can make a significant contribution to efforts to prevent and manage heart failure.

LITERATURE REVIEW

Data mining is a data analysis process carried out with the aim of finding unexpected relationships and summarizing and describing data with a different approach than before, so that the information can be described clearly and is useful for those who own the data (Witten, 2017). The data mining process can run automatically or semi-automatically and the patterns found must have significant meaning (Larose & Larose, 2015). Data mining has data collection techniques in a way that can be used to use data that will be processed later (Suhardjono, Wijaya, & Hamid, 2019).

Previous research on heart failure prediction produced and compared accuracy values which can be seen in table 1.

Table 1. Heart failure research

Author	Method	Accuracy
Fahd Saleh Alotaibi (Alotaibi, 2019)	Decision Tree	82.22%
	Logistic Regression	82.56%
	Random Forest	84.17%
	Naïve Baye	84.24%
	SVM	84.85%
Abid Ishaq(Ishaq et al., 2021)	Decision Tree (DT)	78.89%
	Adaptive boosting classifier (AdaBoost)	82.23%
	Logistic Regression (LR)	85.56%
	Stochastic Gradient classifier (SGD)	66.67%

* Dewi Yuliandari

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024
DOI : https://doi.org/10.33395/sinkron.v9i1.13017
e-ISSN : 2541-2019
p-ISSN : 2541-044X

	Random Forest (RF)	88.89%
	Gradient Boosting classifier (GBM)	84.44%
	Extra Tree Classifier (ETC)	83.34%
	Gaussian Naive Bayes classifier (G-NB)	86.67%
	Support Vector Machine (SVM).	86.67%
Madhumita Pal(Pal & Parija, 2021)		
	Random Forest (RF)	86.90%
Priyanka(Priyanka & Ravikumar, 2017)	Naïve Baye	82.35%
	Decision Tree (DT)	98.03%
	Random forests	74.00%
	Decision tree	73.70%
	Gradient boosting	73.80%
	Linear regression	73.00%
Davide Chicco(Chicco & Jurman, 2020)	One rule	72.90%
	Artificial neural network	68.00%
	Naïve bayes	69.60%
	SVM radial	69.00%
	SVM linear	68.40%
	k-nearest neighbors	62.40%
	Logistic Regression	80.00%
	Neural Network	84.80%
Muhammad Saqlain(Saqlain, Hussain, Saqib, & Khan, 2016)	SVM	83.80%
	Random forests	68.60%
	Decision tree	86.60%
	Naïve bayes	86.67%

From table 1, research that has been carried out Neural networks are very suitable for hospital data which has very large and complex data (Riyanto, Hamid, & Ridwansyah, 2019). A neural network is a group of input/output units connected to each other, where each connection has its own weight (Ahmed & Hannan, 2012). During the learning process, these weights are adjusted to allow the network to predict the correct class labels. Neural networks are renowned for their extraordinary ability to recognize patterns and are suitable for use in predictive purposes. In addition, neural networks are very flexible and can be used as universal models capable of modeling data based on past experience. The way neural networks work is biologically inspired by the way the human brain processes information with programs that have been designed by the brain (Yaqin, Laksito, & Fatonah, 2021).

Particle swarm optimization is an optimization model that works by continuously calculating the possible positions of particles in a problem space (Iqbal et al., 2020). These particles move based on certain rules involving their position and speed (Susanto, Hariyanto, & Surateno, 2018) to find the optimal solution (Bumbungan, Kusrini, & Kusnawi, 2023).

Data mining is a process of processing data and the process of finding patterns from databases in large data to obtain knowledge information stored in large data with heart failure data (Larose & Larose, 2015). Data on heart failure is very large data, therefore this data will be tested using experimental research. The following is a picture of the research stages used.

* Dewi Yuliandari

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v9i1.13017 p-ISSN : 2541-044X

METHOD

In this research phase, a model was used that was created using the PSO (Particle Swarm Optimization) optimization algorithm to optimize the Neural Network (NN) method. This test aims to compare the best model between those that use PSO optimization in NN and those that don't. The dataset used in testing comes from heart failure, the data collection of which comes from secondary data sources or public data that can be seen in the UCI Repository. An explanation of the NN and PSO methods and research flow can be found in Fig. 1

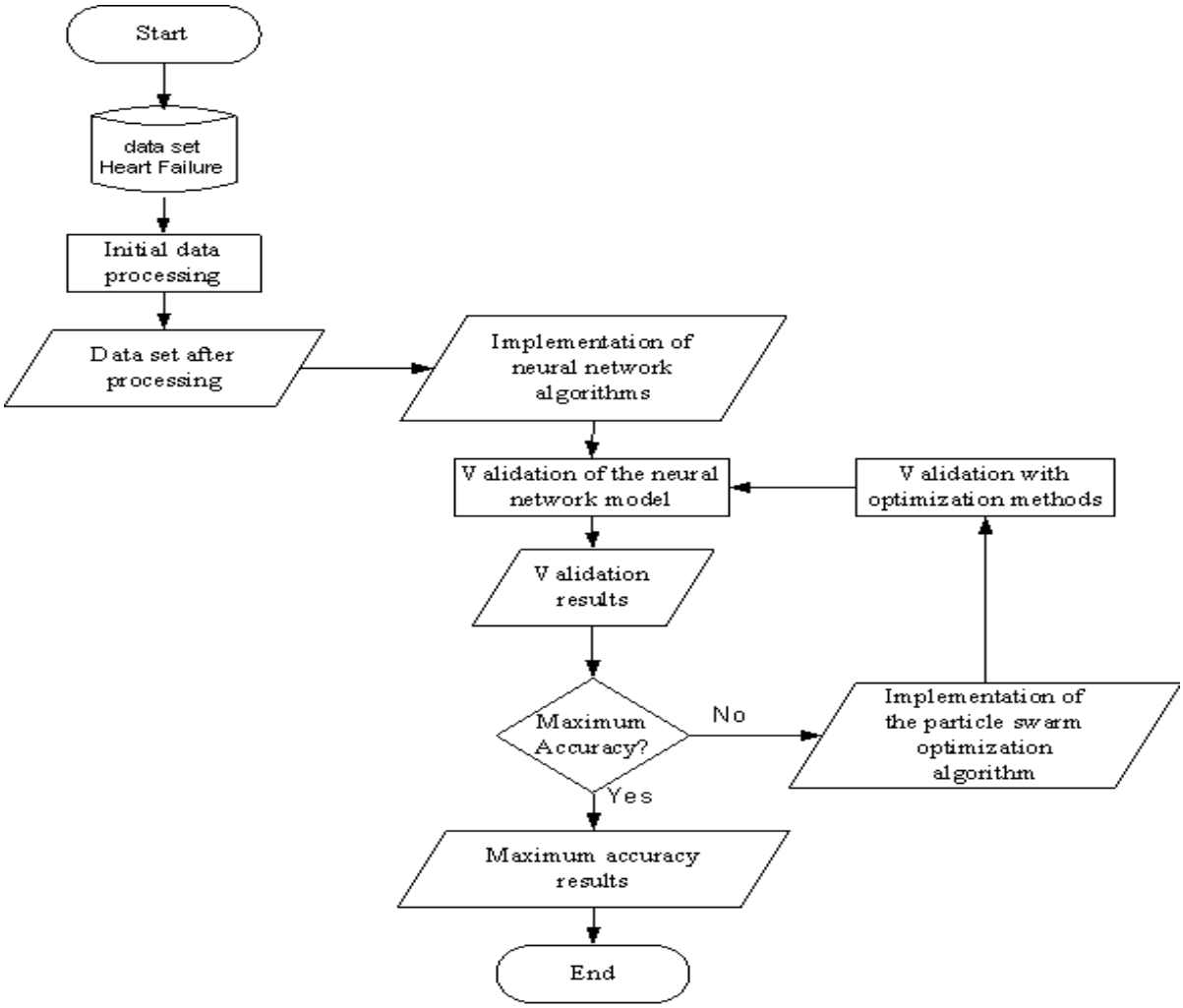

Fig. 1 Stages of heart failure prediction research

From Figure 1, the stages of the research method used to predict heart failure are:

Collection of heart failure data sets

This stage is where the heart failure data set is collected based on secondary data taken from heart failure data contained in a repository, namely the UCI repository. This dataset contains medical records of 299 heart failure patients, and was collected during the follow-up period of these patients. The medical record data shows that each patient profile has 13 clinical features

Data set after processing

The data set after processing after carrying out the initial data processing stage. This processing is carried out after collecting the heart failure data set by first analyzing the data. Data analysis includes data cleaning carried out on incomplete data or noisy data and data that is not meaningful for predicting heart failure. After the data is processed, it will produce a data set that is valid and ready to be tested using the neural network method.

* Dewi Yuliandari

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024
DOI : https://doi.org/10.33395/sinkron.v9i1.13017
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Implementation of the neural network algorithm

After processing the data set, it will be implemented in this research using the neural network method to validate the neural network on the heart failure data set

Validation results

After implementing the neural network algorithm, it will produce validation of the algorithm. Where the results of this validation include the accuracy values obtained from the method. And if the validation results of the neural network method are not optimally accurate, it will be tested again by implementing particle swarm optimization.

Implementation of the particle swarm optimization algorithm

At this stage, if the validation of the neural network is not optimal, it will be tested again using the particle swarm optimization method to increase the accuracy of the neural network model.

Maximum accuracy results

At this stage where accuracy has achieved maximum results with the neural network method. And the results of this accuracy will be used to predict heart failure.

RESULT

Based on the research flow diagram shown in Figure 1 where data is collected and obtained from secondary data sourced from public data, namely the UCI repository, the data will later be experimented with the NN method and the NN method which is optimized by PSO after going through a process of removing data that is not used, so that the data is clean of unused data, after deleting the data. Where the data after deleting unused data can be seen in table 1 for the sample data obtained.

Table 1. Sample heart failure data

age	anaemia	creatinine phosphokinase	diabetes	ejection fraction	high_blood pressure	platelets	serum creatinine	serum sodium	sex	smoking	time	DEATH EVENT
75	0	582	0	20	1	265000	1.9	130	1	0	4	1
55	0	7861	0	38	0	263358.03	1.1	136	1	0	6	1
65	0	146	0	20	0	162000	1.3	129	1	1	7	1
50	1	111	0	20	0	210000	1.9	137	1	0	7	1
65	1	160	1	20	0	327000	2.7	116	0	0	8	1
90	1	47	0	40	1	204000	2.1	132	1	1	8	1
75	1	246	0	15	0	127000	1.2	137	1	0	10	1
60	1	315	1	60	0	454000	1.1	131	1	1	10	1
65	0	157	0	65	0	263358.03	1.5	138	0	0	10	1
80	1	123	0	35	1	388000	9.4	133	1	1	10	1
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

* Dewi Yuliandari

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024
DOI : https://doi.org/10.33395/sinkron.v9i1.13017
e-ISSN : 2541-2019
p-ISSN : 2541-044X

From this data set, the data will be tested and validated with a neural network model using rapid miner software with step 10 cross validation. And you will get the confusion matrix results which can be seen in table 2.

Table 2. Confusion Matrix NN

	true YES	true NO	class precision
Pred YES	55	27	67.07%
pred. NO	41	176	81.11%
class recall	57.29%	86.70%	

From the confusion matrix table, it can be concluded that 55 students were predicted to have heart failure as predicted, however 27 people who were predicted to have heart failure turned out not to have heart failure. Likewise, 176 people were predicted not to suffer from heart failure and in accordance with the prediction results, however, 41 people were predicted not to suffer from heart failure and it turned out that these people had heart failure. So the confusion matrix table produces an accuracy value of 77.26% and produces a bar diagram of heart failure which can be seen in Fig 2.

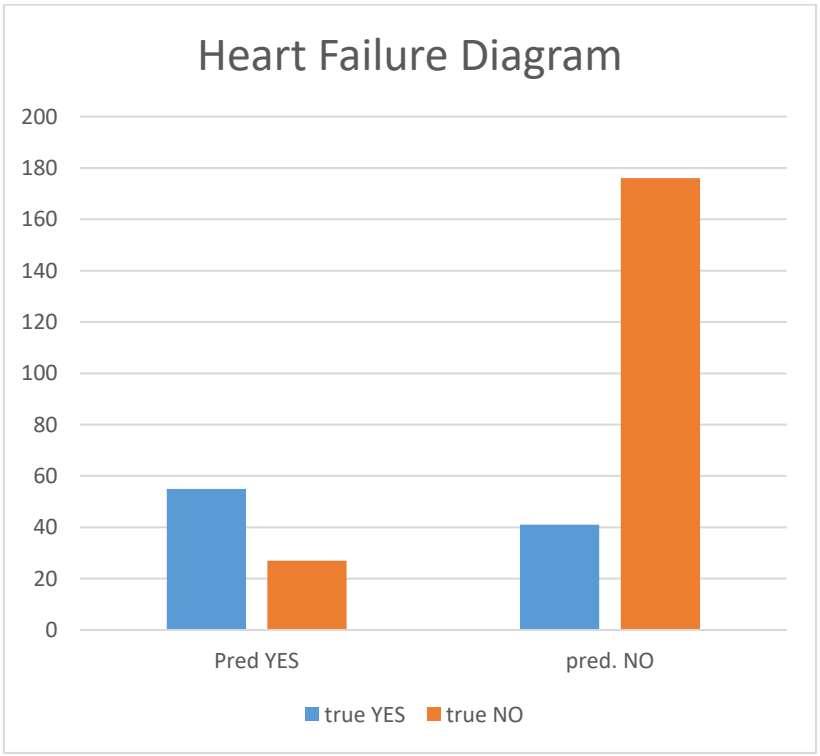

Fig. 2 Bar diagram of heart failure prediction with NN

Figure 2 explains that the blue diagram is the result of a yes prediction in terms of having heart failure, and the orange diagram depicts the result of the prediction of not having heart failure. From table 2 it also produces the area under curve or abbreviated (AUC) which can be seen in Fig 3.

* Dewi Yuliandari

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024
DOI : https://doi.org/10.33395/sinkron.v9i1.13017
e-ISSN : 2541-2019
p-ISSN : 2541-044X

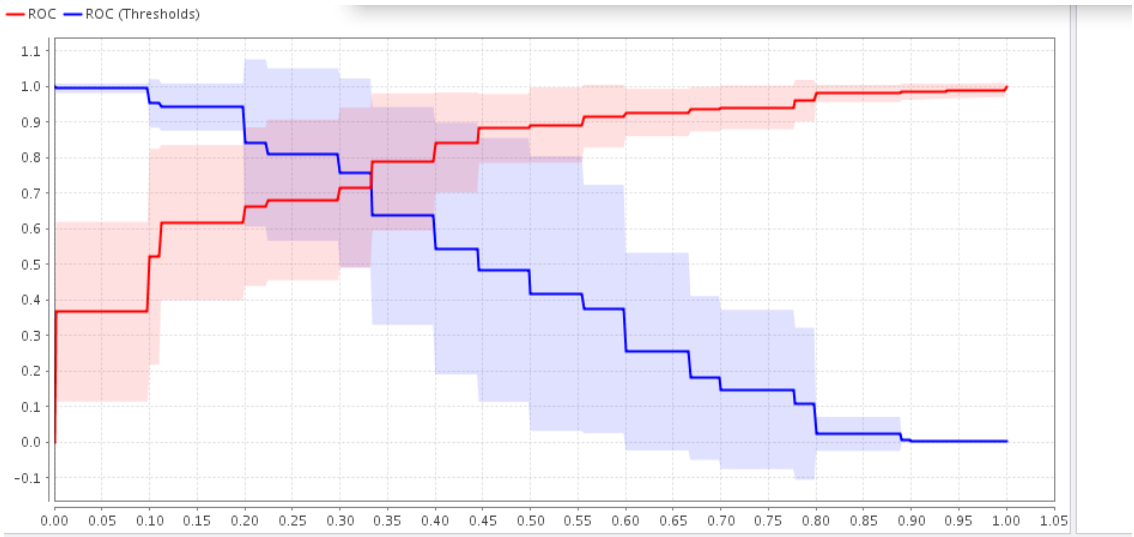

Fig. 3 Area Under Curve (AUC) with NN

In Figure 3 it can produce an AUC of 0.802% or it can be interpreted that the AUC obtained very good results. And by testing data on heart failure using the NN method, the neural network model architecture shown in Figure 4 will be produced

.

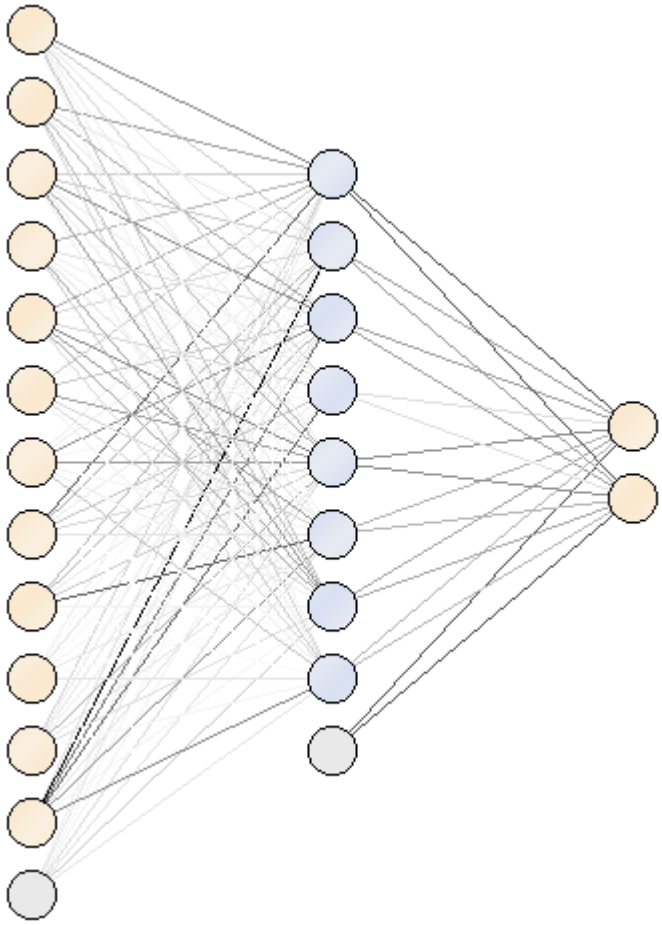

Fig 4 Architecture of NN models in heart failure

* Dewi Yuliandari

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024
DOI : https://doi.org/10.33395/sinkron.v9i1.13017
e-ISSN : 2541-2019
p-ISSN : 2541-044X

In Figure 4 it can be seen that in the input layer there are 13 where there are age, anemia, creatinine phosphokinase, diabetes, ejection fraction, high blood pressure, platelets, serum creatinine, serum sodium, sex, smoking, time and 1 threshold. The hidden layer uses 1 hidden layer with the weight value of each attribute in the input layer, as well as 2 resulting outputs, namely heart failure or no heart failure.

After testing the heart failure data using a neural network, the data will be optimized using the particle swarm optimization method with the performance of the PSO method, namely attributes that have no effect are removed which can be seen in table 3.

Table 3. Heart failure weights attributes

attribute	weight
age	1
anaemia	0
creatinine_phosphokinase	0
diabetes	0.172
ejection_fraction	0.781
high_blood_pressure	0
platelets	0
serum_ creatinine	0.568
serum_sodium	1
sex	0
smoking	1
time	1

From table 3 it can be concluded that the attributes that have no effect are the attributes with a value of 0 where there are three attributes with a value of 5, namely anemia, creatinine_phosphokinase, high_blood_pressure, platelets and sex. From the results of NN with PSO, you will get the confusion matrix results which can be seen in table 4.

Table 4. Confusion Matrix NN PSO Optimization

	true YES	true NO	class precision
Pred YES	72	19	79.12%
pred. NO	24	184	88.46%
class recall	75.00%	90.64%	

From the confusion matrix table, it can be concluded that 72 people who were predicted to have heart failure matched predictions, however 19 people who were predicted to have heart failure turned out to not have heart failure. Likewise, 184 people were predicted not to have heart failure and it was in accordance with the prediction results, but 24 people were predicted not to have heart failure and it turned out that those people had heart failure. So the confusion matrix table produces an accuracy value of 85.61% and produces a heart failure. bar diagram which can be seen in Figure 5.

* Dewi Yuliandari

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024
DOI : https://doi.org/10.33395/sinkron.v9i1.13017
e-ISSN : 2541-2019
p-ISSN : 2541-044X

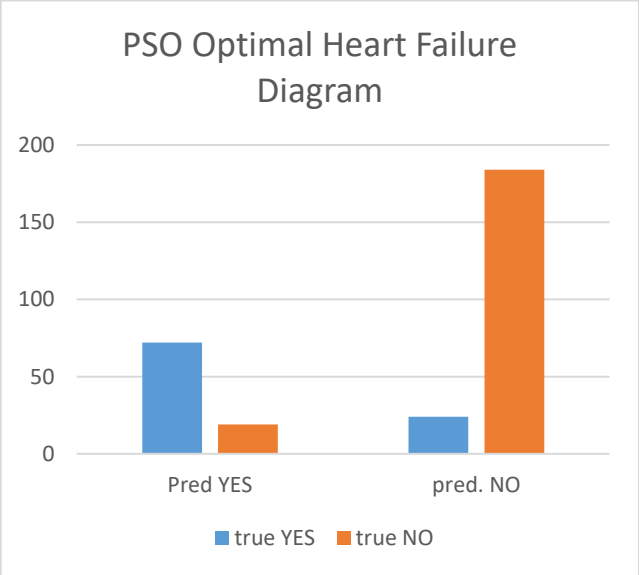

Fig. 5 Bar diagram of heart failure prediction with PSO Optimal NN

Figure 5 explains that the blue diagram is the result of a yes prediction in terms of heart failure, and the orange diagram depicts the result of a no heart failure prediction. From table 4 it also produces the area under curve or abbreviated (AUC) which can be seen in Figure 6.

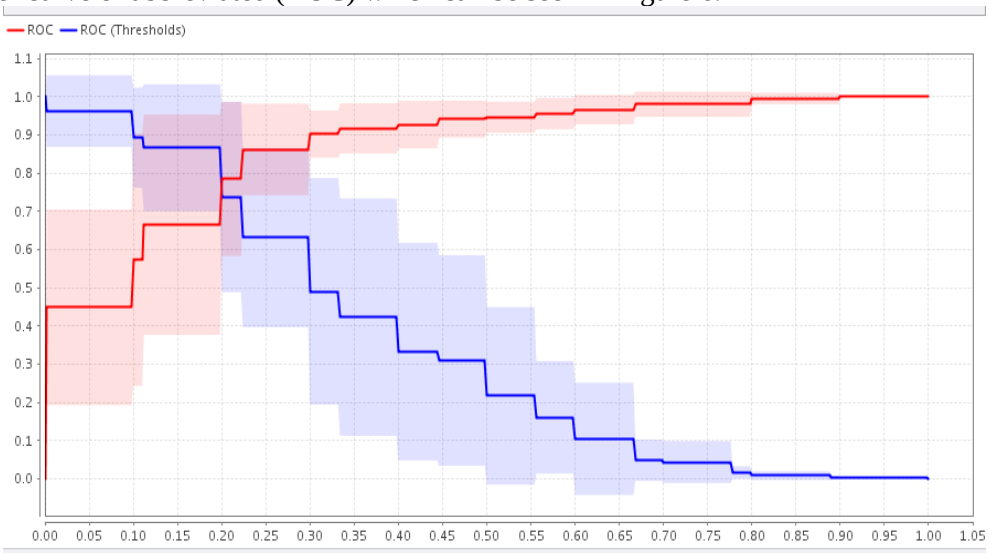

Fig. 6 Area Under Curve (AUC) NN PSO Optimization

In Figure 6 it can produce an AUC of 0.869% or it can be interpreted that the AUC obtained very good results.

DISCUSSIONS

Based on the results obtained from the testing process carried out on the heart failure dataset when implementing the neural network method with the application of particle swarm optimization techniques, it was found that the proposed method is the most precise and efficient method variant that has been applied to build neural network models. Initially, the use of a neural network resulted in an accuracy rate of around 77.26% and an Area Under the Curve (AUC) of 0.802%. However, when the particle swarm optimization method was used to optimize the process, the level of accuracy increased significantly, reaching the optimal peak with an accuracy of 85.61% and an AUC of 0.869%. This improvement reflects an 8.35% increase in accuracy rate and a 0.067% increase in AUC. Therefore, the results obtained from analyzing heart failure data have great potential to become a valuable resource for

* Dewi Yuliandari

hospitals in carrying out data analysis of patients suffering from heart failure at this time. In addition, this method also has great potential to increase accuracy in the prediction of heart failure, thereby allowing more effective treatment for patients affected by this disease.

CONCLUSION

The process of testing heart failure data involves the use of neural network methods together with particle swarm optimization. This process begins by collecting heart failure data from secondary data sources. The data consists of 12 parameters as predictor variables and one variable as the prediction result. Next, the data undergoes a filtering stage to eliminate incomplete data or what is commonly referred to as noise data. After the noise data has been successfully removed, the next step is to test the data using the neural network method. After the testing process using the neural network method is complete, the quality of the data is then improved through optimization using the particle swarm optimization method. The result of this optimization is determining optimal attribute weights, so that attributes that do not make a significant contribution can be removed.

The evaluation results of these two methods, which are evaluated using the confusion matrix table, show that the neural network method that has been optimized with the help of particle swarm optimization has experienced a substantial increase in the level of accuracy.

REFERENCES

Ahmed, A., & Hannan, S. A. (2012). Data Mining Techniques to Find Out Heart Diseases : An Overview. *(Sem Qualis) International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, *1*(4), 18–23.

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

Ariyati, Indah, Ridwansyah, & Suhardjono. (2018). Implementasi Particle Swarm Optimization untuk Optimalisasi Data Mining Dalam Evaluasi Kinerja Asisten Dosen. *JIKO (Jurnal Informatika Dan Komputer) STMIK AKAKOM*, *3*(2), 70–75.

Austin, P. C., Tu, J. V., Ho, J. E., Levy, D., & Lee, D. S. (2013). Using methods from the data-mining and machine-learning literature for disease classification and prediction: A case study examining classification of heart failure subtypes. *Journal of Clinical Epidemiology*, *66*(4), 398–407. https://doi.org/10.1016/j.jclinepi.2012.11.008

Bumbungan, S., Kusrini, & Kusnawi. (2023). Penerapan Particle Swarm Optimization(PSO)dalam Pemilihan Parameter Secara Otomatis padaSupport Vector Machine(SVM)untuk Prediksi Kelulusan Mahasiswa Politeknik Amamapare Timika. *Jurnal Teknik AMATA*, *4*(1), 81–93.

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

Larose, D. T., & Larose, C. D. (2015). *Data Mining And Predictive Analytics*. John Wiley and Sons.

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

Priyanka, N., & Ravikumar, P. (2017). Usage of data mining techniques in predicting the heart diseases

* Dewi Yuliandari

- Naïve Bayes & decision tree. *Proceedings of IEEE International Conference on Circuit, Power and Computing Technologies, ICCPCT 2017*. https://doi.org/10.1109/ICCPCT.2017.8074215

Ridwansyah, Ariyati, I., & Faizah, S. (2018). PARTICLE SWARM OPTIMIZATION BERBASIS CO-EVOLUSIONER DALAM EVALUASI KINERJA ASISTEN DOSEN. *Jurnal SAINTEKOM*, *9*(2), 166–177. https://doi.org/https://doi.org/10.33020/saintekom.v9i2.96

Ridwansyah, & Purwaningsih, E. (2018). Particle Swarm Optimization Untuk Meningkatkan Akurasi Prediksi Pemasaran Bank. *Jurnal PILAR Nusa Mandiri*, *14*(1), 83–88.

Riyanto, V., Hamid, A., & Ridwansyah. (2019). Prediction of Student Graduation Time Using the Best Algorithm. *Indonesian Journal of Artificial Intelligence and Data Mining*, *2*(2), 1–9. https://doi.org/http://dx.doi.org/10.24014/ijaidm.v2i1.6424

Saqlain, M., Hussain, W., Saqib, N. A., & Khan, M. A. (2016). Identification of Heart Failure by Using Unstructured Data of Cardiac Patients. *Proceedings of the International Conference on Parallel Processing Workshops*, *2016-Septe*, 426–431. https://doi.org/10.1109/ICPPW.2016.66

Suhardjono, Wijaya, G., & Hamid, A. (2019). PREDIKSI WAKTU KELULUSAN MAHASISWA MENGGUNAKAN SVM BERBASIS PSO. *Bianglala Informatika*, *7*(2), 97–101.

Susanto, B. M., Hariyanto, A., & Surateno. (2018). Performance comparison of proactive and reactive routing protocol in mobile ad hoc network. *J. Commun*, *13*(5), 218–224. https://doi.org/10.12720/jcm.13.5.218-224

Witten, I. H. (2017). *DATA MINING (Fourth Edition)*.

Yaqin, A., Laksito, 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).

* Dewi Yuliandari