Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)
Vol. 8, No. 1, March 2025, pp. 229 – 235
p-ISSN: 2614-3372 | e-ISSN: 2614-6150

Prediction of Heart Disease Using K-Nearest Neighbor and Particle Swarm Optimization (PSO) Method

[1*]Elin Panca Saputra, [2]Andi Saryoko, [3]Aryo Tunjung Kusumo, [4]Priyono
[1,4]Department of System and Tecnology Information, Cyber University, Indonesia
[2]Department of Informatic, Universitas Nusa Mandiri, Indonesia
[3]Department of Information System, Universitas Bina Sarana Informatika, Indonesia
Email: [1]elin.epa@bsi.ac.id, [2]andi.asy@nusamandiri.ac.id, [3]aryo.atk@bsi.ac.id, [4]priyono.pyo@bsi.ac.id

Article Info
Article history:
Received Nov 6th, 2024
Revised Dec 23th, 2024
Accepted Feb 14th, 2025

Keyword:
Heart Disease
K-Nearest Neighbor
Particle Swarm Optimization
Prediction

ABSTRACT
One of the leading causes of death nowadays is heart disease, so more has to be done to avoid it, such as by making prediction models work better. Among the machine learning algorithms is K-NN is among the best methods for predicting heart disease based on several risk factors, including smoking, high blood pressure, diabetes, age, and so on. To get accurate values and attribute selection features, we tested them with K-NN, and to improve the results of our research predictions, we combined them using the PSO algorithm. The results are very interesting after we do the calculations, the algorithm that uses PSO-based K-NN gets a higher weight compared to using only the K-NN algorithm. The predicted value of the weight resulting from the PSO-based K-NN is 97.67%. while the results only use K-NN of 64.92%. The advantages of PSO can also select attributes that can affect it, namely age, diabetes, and ejection fraction. So gathering information through data mining. The PSO-Based K-NN method, which is the primary machine learning technique used in this computation, yields the greatest results in terms of accuracy for heart disease when applied to the data assets. Using the K-NN - PSO algorithm can provide promising results for predicting symptoms that cause heart disease with very good accuracy. PSO is Used to choose features and optimize k values on the K-NN dataset, after which the accuracy is output on the K-NN.

Corresponding Author:
Elin Panca Saputra,
Department of System and Tecnology Information, Cyber University, Indonesia,
Jl. TB Simatupang, Tj. Bar., Kec. Jagakarsa, Jakarta Selatan,
Daerah Khusus Ibukota Jakarta, Indonesia.
Email: elin.epa@bsi.ac.id

DOI: http://dx.doi.org/10.24014/ijaidm.v8i1.33373

1. INTRODUCTION

As per data from the World Health Organization (WHO), diseases play a significant role in the rate of fatalities. Heart disease stands as the second leading cause of mortality among all illnesses [1]. Machine Learning (ML) is used as a support system that is so effective in making diagnoses in the field of health sciences that contain large amounts of data [2]. Another Artificial Intelligence approach that most clearly demonstrates its supremacy in the field of study is ML [3]. Globally, heart disease is the leading cause of death. The most prevalent kind of heart disease is coronary heart disease. It happens when plaque builds up in the heart's blood vessels, causing blood flow to become constricted [4]. A high-accuracy identification of cardiac illness is a critical aspect in preventing additional harm to the patient [5].

Predicting cardiac disease is one area where machine learning may be used effectively [6]. Since machine learning can uncover patterns in data and help specialists reduce misdiagnosis in categorizing cardiovascular disease, it plays a crucial role in the medical sector [7]. Because of the greater detection of cardiac disease as well as persons at high risk, employing predictive models is usually suggested for reduced

mortality and enhanced decision-making for future treatment and prevention [8]. To avoid this, we as researchers attempt to categorize the data first. The classifier's accuracy improves noticeably after employing the selecting function [9]. Feature selection can be used most relevantly in selecting variables in heart disease [10].

Our research's primary objective is to precisely identify the cardiac diseases that might develop in a person. In addition, in the contents of our research, aims to investigate the effectiveness of various algorithms contained in machine learning to predict heart disease. To achieve this goal, we use several techniques that we test. The various algorithms in machine learning that we will be testing include the utilization of K-Nearest Neighbor (K-NN) and K-NN-Particle Swarm Optimization (K-NN-PSO). The K-NN approach is a widely used method for classification, known for its lazy nature. The most popular, effective, and successful pattern recognition method is K-NN [11]. The reason we opt for one of the ML algorithms on K-NN is because it represents an Artificial Intelligence (AI) that finds growing applications in cardiovascular medicine [12]. ML can also improve algorithms that provide promising results for predicting heart disease symptoms [13]. From previous research using K-NN is superior to algorithms found in other ML, K-NN is much better at predicting heart disease [14].

It is thus vital to boost accuracy and choose relevant features to anticipate the degree of accuracy since the results of numerous algorithms that we evaluated after calculating the results obtained did not meet expectations and presented challenges in picking attribute selection features. because some of the methods we tried had stochastic features and required a lengthy time to convert [15]. So, we endeavor to compare each of the algorithms we have tested with the Particle Swarm Optimization (PSO) algorithm. The utilization of PSO stems from its effectiveness as a meta-heuristic algorithm that efficiently identifies optimal solutions within a compressed timeframe [2]. Furthermore, PSO has the benefit of being a very efficient optimization approach that can improve algorithm training [16]. to overcome this, the K-NN algorithm can be optimized using feature selection techniques, especially PSO Regarding the categorization of heart disease [1].

However, choosing the best weights to update the particles' position and speed is another issue with the PSO method. To solve these problems, this study provides a unique method to calculate the optimal weight value based on the tuning function and the population diversity function. Therefore we propose a new fitness function for PSO with the help of several algorithms, namely, K-NN and K-NN-PSO, so that it will produce the algorithm with the best performance when it has been compared with the PSO algorithm. In general the technique of using PSO techniques that are so good in its use that many other researchers use the PSO algorithm in various fields [17]. The PSO optimization approach is widely used because of its efficacy in solving optimization problems, ease of implementation, and relative simplicity. [18]. and has been widely used in various fields such as health, education, and social. economy and others. As in previous research, apart from the health sector, the PSO algorithm also obtained superior results which increased accuracy values [19].

2. MATERIAL AND METHOD

Formula thoughts on this research consist of several stages as shown in the image below the problem with this research is the existence of several attributes that will reduce the accuracy and increase the complexity of the Algorithm K-NN support. For that PSO to solve that problem.

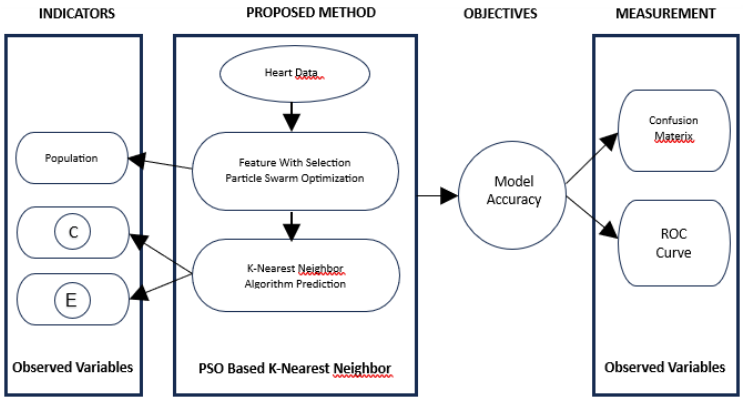

Figure 1. Research Formulate Thoughts

The image 1 is the framework of the idea, of how we manage and calculate data using the Rapid Miner version 5.3 tools. Based on these considerations, the use of K-NN with the Particle Swarm Optimization Approach is expected to improve the performance of heart disease prediction so that it can get a higher accuracy value.

2.1 Collecting Data

The data that we classified, before we got it on the Kaggle site, Kaggle itself is a site that shares ideas that there are community groups on algorithms found in online machine learning which includes some data science. The data that we will process is a sample of 300 data. The following is an example. For more details, it can be shown in Table 1.

Table 1. Heart Variable Data

No	Variable	Information
1	Patient_age	Determining the age of one indication
2	Patient _naemia	Decrease of red blood The CPK level of cells or hemoglobin (boolean)
3	creatinine_phosphokinase	Enzyme in the blood <mcg/L>
4	Diabetes_disease	If the patient has diabetes (boolean)
5	ejection_fraction	Percentage of blood leaving the heart at each contraction (percentage)
6	high_blood_pressure	If the patient has hypertension (boolean)
7	platelets	Platelets in the blood (kilo platelets/mL)
8	serum_creatinine	Level of serum creatinine in the blood (mg/dL)
9	serum_sodium	Level of serum sodium in the blood (mEq/L)
10	sex	Woman or man (binary)
11	Is_smoking	How many people smoke cigarettes each day
12	time	Show time

2.2 K-Nearest Neighbors (K-NN)

This approach to data classification is quite good. The most widely used, straightforward, highly successful, and efficient pattern recognition technique is called K-NN. K-NN is a simple classifier in which samples are categorized according to the Nearest Neighbors class. There are many medical databases. Classification may yield erroneous results if the data collection includes repetitive and unnecessary properties. [20]. K-NN is the most well-liked, successful, and effective pattern recognition technique. Medical data sets may encompass a considerable number of traits. If the data sets contain noisy features, the classifier's performance will deteriorate [16].

2.3 Particle Swarm Optimization (PSO)

Kennedy and Eberhart brought forth the PSO algorithm in the year 1995. PSO is analogous to a group of birds flying around the skies with elegance and purposelessly searching for food. There was only one food item discovered in the searched area. The location of the meal is unknown to the birds. But they are aware of how far the food and their standing among peers go. the best course of action.

A group of particles that can update their relative position from one iteration to another, PSO can improve the algorithm as it should do the search process [17]. To get optimal results, each moving particle looks for the previous personal best_position ($\mathbf{p}_{best}$) and the best global_position ($\mathbf{g}_{best}$) which is included in the swarm [21].

As a unique collective and distributed intelligent approach for tackling problems, particularly in optimization, without centralized control or a global model, One particularly noteworthy idea in the fields of computers and artificial intelligence is particle swarm optimization or PSO. Research in behavioral sciences, social ethology, cognitive psychology, and neuroscience served as inspiration for PSO. [22]. A velocity is computed based on the vector of each particle's personal best performance and the vector of the particle with the greatest performance overall, and these vectors are added up with the updated best-performing vectors of each particle, respectively [23]. It is possible to make precise predictions about the occurrence of heart disease in an individual patient by combining it with the Particle Swarm Optimization (PSO) feature selection technique. Experimental results have demonstrated that the highest accuracy was produced by PSO.[24]. Three control parameters—bobot inersia (w), koefisien percepatan (C1 dan C2), and bilangan acak (R1 dan R2)—are the performance of PSO operations [25].

$$V_{i,m} = W.V_{i,m} + \quad_{ci} * R * \left(pbest_{i,m} - V_{i,m}\right) + C_2 * R * \left(gbest_m - X_{i,m}\right) \tag{1}$$

$$X_{id} = X_{i,m} + V_{i,m} \tag{2}$$

Where: n: is the number of particles in the group, d: is dimensions, $v_{i,m}$: is article velocity i-th on interaction with, w : is inertia weight factor, c_1, c_2:is acceleration constant(learning rate), R: is random numbers (0-1), $x_{i,d}$:is the particles' current location on iteration i-th, $pbest_i$:is the best previous position of the particles i-th, and gbest: is the best particle in a population or group out of all the others.

3. RESULTS AND ANALYSIS

After the results of our calculations have been carried out, we will detail one by one the algorithms that have the highest level of accuracy with the highest performance for classification to determine the level of prediction of heart disease at this point. For the calculations performed, we use as many as 300 data with 12 attributes for attribute selection feature selection. The following is the result of the algorithm that we have implemented.

3.1 Algorithm Testing Experiments Using K-NN

The observation results from calculations with the K-NN algorithm with K-Fold Cross Validation show accuracy with a value of 64.92%, then an AUC value of 0.5%. The next steps we take to calculate the results of matrix calculations using the K-NN method are as follows which can be shown in Table 2 the values of Accuracy, Sensitivity, Specificity, PPF, and NPV using the following computations.

Table 2. K-NN Data Processing Results (2023)

Classification	Predicted Class	
	Class= 1	Class= 2
Class= 1	169	51
Class= 2	56	23

$$\text{Accuracy}= \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} = \frac{169+23}{169+23+56+51} = 0{,}6492$$

$$\text{Sensitivity}= \frac{\text{TP}}{\text{TP+FN}} = \frac{169}{169+51} = 0{,}7681$$

$$\text{Specificity}= \frac{\text{TN}}{\text{TN+FP}} = \frac{23}{23+56} = 0{,}2911$$

$$\text{PPF}= \frac{\text{TP}}{\text{TP+FP}} = \frac{169}{169+56} = 0{,}7511$$

$$\text{NPV}= \frac{\text{TN}}{\text{TN+FN}} = \frac{23}{23+51} = 0{,}310$$

For further explanation, please see graphic pig 2 for the data processing results using only K-NN.

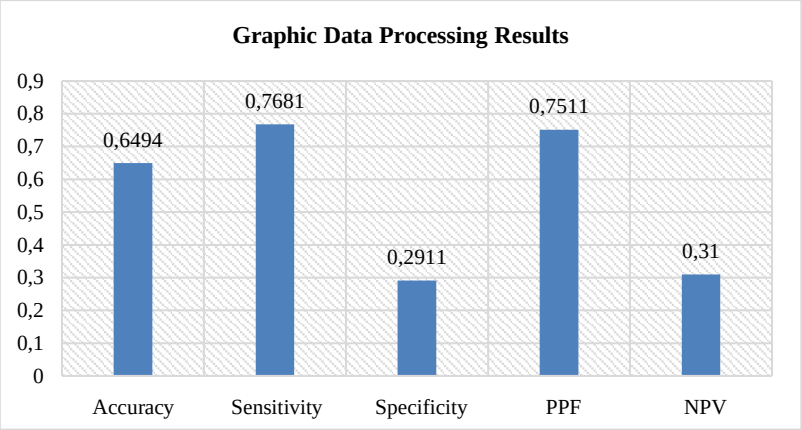

Figure 2. Data Preprocessing Results K-NN

3.2 Optimizer K-NN and K-NN-PSO Algorithm Testing Experiment

The accuracy with the K-NN method is 64.21%, and the AUC value is 0.5%, according to the observation findings from computations using the K-NN and K-NN-PSO algorithms with K-Fold Cross Validation. When we compared the results using the PSO-based K-NN, the significant value was 97.67%, but the AUC value did not grow steadily by 0.5%. The following process is used to determine the results of the confusion matrix using the K-NN-PSO approach and is given in Table 3 and Grafik in Figure 2. The values of Accuracy, Sensitivity, Specificity, PPF, and NPV using the following computations.

Table 3. K-NN-PSO Data Processing Results (2023)

Classification	Predicted Class	
	Class= 1	Class= 2
Class= 1	220	2
Class= 2	5	72

$$\text{Accuracy}= \frac{TP+TN}{TP+TN+FP+FN} = \frac{220+72}{220+72+5+2} = 0{,}9767$$

$$\text{Sensitivity}= \frac{TP}{TP+FN} = \frac{220}{220+2} = 0{,}9909$$

$$\text{Specificity}= \frac{TN}{TN+FP} = \frac{72}{72+5} = 0{,}9350$$

$$\text{PPF}= \frac{TP}{TP+FP} = \frac{220}{220+5} = 0{,}9777$$

$$\text{NPV}= \frac{TN}{TN+FN} = \frac{72}{72+2} = 0{,}9729$$

For further explanation, please see graphic Figure 3 for the data processing results using K-NN-PSO.

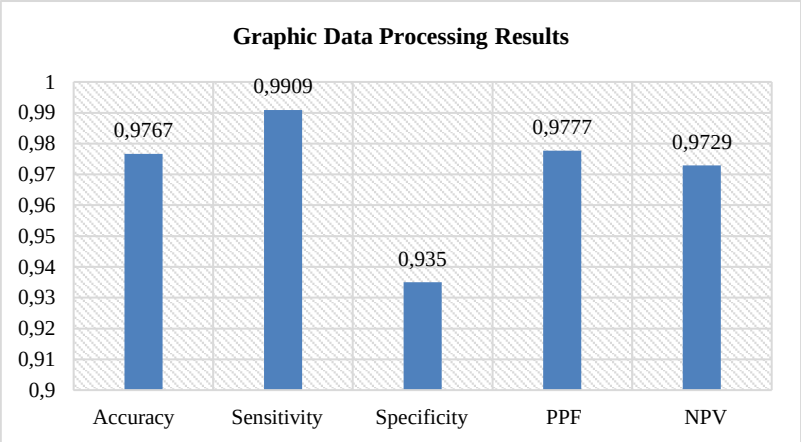

Figure 3. Data Processing Results K-NN-PSO

Table 3. Result from Observasi C, ε dan population

C	ε	Population	Accuracy	AUC
1.0	0.0	10	97.67%	50.00%

The observation results show that the highest value of accuracy is 97.67% and AUC is 0.50% obtained with a C value of 1.0, ε is 0.0, and a population of 10. So these values are used in this study. The next step is to select the attributes used, namely in Table 4.

Table 4. The result of the attribute selection is 12

Attribute	Weight
age	1
anaemia	0
creatinine_phosphokinase	0
diabetes	1
ejection_fraction	1
high_blood_pressure	0
platelets	0
serum_creatinine	0
serum_sodium	0
sex	0
smoking	0
Time	0

3.3 Evaluation between the PSO-Based K-NN and K-NN

The results of carrying out a comparison of the various methods that we have tested can be concluded by comparing the accuracy level values of the various methods tested, for details can be shown in Table 6.

Table 6. Comparison of Accuracy Value and AUC Value

Algorithm	Accuracy	Precision	AUC
K-NN	64.92%	29.45%	0.50%
K-NN and PSO	97.78%	93.85%	0.50%

The results of the PSO-based K-NN algorithm assessment. The Accuracy value of 97.78% is very good and the Area Under Curve (AUC) is 50% with a classification accuracy value of less. whereas if only using the K-NN algorithm alone the accuracy value is only 64.92% which is categorized as sufficient, while the AUC value is 50% with a classification accuracy value of less. The Precision value of PSO-based K-NN has a very high accuracy of 93.85% while with KKN alone it is 29.45%.

4 CONCLUSION

After computing the machine learning algorithm, it is evident that the predicted value had an accuracy rate of 64. 92%, precision of 29. 45%, and an AUC of 0. With a precision value of 93. 85% and an accuracy rate of 97. 67%, the PSO-based K-Nearest Neighbor method achieves an accuracy rate of 0. 50% for the AUC value. From experiments with different machine learning algorithms, it is evident that the PSO-based K-Nearest Neighbor algorithm holds greater predictive value for enhancing yields by 32. 86%, as opposed to using the K-Nearest Neighbor strategy alone. Age, diabetes, and ejection fraction are the qualities that have an impact on making predictions, according to the outcomes of choosing feature traits, out of the 12 attributes that have been determined. Our study doesn't finish here; we'll continue to play around with all the machine learning algorithms and figure out how to evaluate how well they work if the computations approach thousands and include more criteria for predicting heart disease. In addition, additional study is still needed to improve the PSO algorithm's outcomes. In the future, we will try how to process images to detect the disease based on AI (artificial intelligence) to produce better accuracy.

REFERENCES

[1] R. Wahyusari, E. H. Herraprastanti, and H. Gunawan, "Machine Learning Techniques for Heart Disease Classification Using K-Nearest Neighbor Optimization with Particle Swarm Optimization," *E3S Web Conf.*, vol. 500, 2024.

[2] J. Vijayashree and H. P. Sultana, "A Machine Learning Framework for Feature Selection in Heart Disease Classification Using Improved Particle Swarm Optimization with Support Vector Machine Classifier," *Program. Comput. Softw.*, vol. 44, no. 6, pp. 388–397, 2018.

[3] B. S. and R. K. Apurv Garg, "Heart disease prediction using machine learning techniques," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1022, no. 1, 2021.

[4] I. D. Mienye and Y. Sun, "Improved heart disease prediction using particle swarm optimization based stacked sparse autoencoder," *Electron.*, vol. 10, no. 19, 2021.

[5] M. S. Pathan, A. Nag, M. M. Pathan, and S. Dev, "Analyzing the impact of feature selection on the accuracy of heart disease prediction," *Healthc. Anal.*, vol. 2, no. February, p. 100060, 2022.

[6] Y. Khourdifi and M. Bahaj, "Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization," *Int. J. Intell. Eng. Syst.*, vol. 12, no. 1, pp. 242–252, 2019.

[7] T. G. and P. L. M. Chintan M. Bhatt , Parth Patel, "Effective Heart Disease Prediction Using Machine Learning Techniques," *MPDI*, 2023.

[8] U. Nagavelli, D. Samanta, and P. Chakraborty, "Machine Learning Technology-Based Heart Disease Detection Models," *J. Healthc. Eng.*, vol. 2022, 2022.

[9] D. Panda, R. Ray, A. A. Abdullah, and S. R. Dash, "Predictive Systems: Role of Feature Selection in Prediction of Heart Disease," *J. Phys. Conf. Ser.*, vol. 1372, no. 1, 2019.

[10] F. F. Firdaus, H. A. Nugroho, and I. Soesanti, "A Review of Feature Selection and Classification Approaches for Heart Disease Prediction," *IJITEE (International J. Inf. Technol. Electr. Eng.*, vol. 4, no. 3, p. 75, 2021.

[11] M. A. Jabbar, "Prediction of heart disease using k-nearest neighbor and particle swarm optimization," *Biomed. Res.*, vol. 28, no. 9, pp. 4154–4158, 2017.

[12] C. Krittanawong *et al.*, "Machine learning prediction in cardiovascular diseases : a meta - analysis," *Sci. Rep.*, pp. 1–11, 2020.

[13] R. T. Neha Nandal, Lipika Goel, "Machine learning-based heart attack prediction : A symptomatic heart attack prediction method and exploratory analysis [version 1 ; peer review : awaiting peer review] Neha Nandal," *F1000Research*, pp. 1–17, 2023.

[14] I. M. A. Dara DVVNS Saikumar, Uma Priyadarsini P. S, "Prediction of Heart Disease Using K-Nearest Neighbour Algorithm in Comparison with Support Vector Machine Algorithm," *IEEE*, no. International Conference on

Informatics and Computational Sciences (ICICoS), 2022.
[15] M. Amanullah, "A PSO-SVM Feature Selection Method For Improving The Prediction," *Int. J. Comput. Eng. Technol.*, no. September, 2022.
[16] A. Al Bataineh and S. Manacek, "MLP-PSO Hybrid Algorithm for Heart Disease Prediction," *J. Pers. Med.*, vol. 12, no. 8, 2022.
[17] A. G. Gad, *Particle Swarm Optimization Algorithm and Its Applications: A Systematic Review*, vol. 29, no. 5. Springer Netherlands, 2022.
[18] M. P. Behera, A. Sarangi, D. Mishra, and S. K. Sarangi, "A Hybrid Machine Learning algorithm for Heart and Liver Disease Prediction Using Modified Particle Swarm Optimization with Support Vector Machine," *Procedia Comput. Sci.*, vol. 218, no. 2022, pp. 818–827, 2023.
[19] E. P. Saputra, Supriatiningsih, Indriyanti, and Sugiono, "Prediction of Evaluation Result of E-learning Success Based on Student Activity Logs with Selection of Neural Network Attributes Base on PSO," *J. Phys. Conf. Ser.*, vol. 1641, no. 1, 2020.
[20] M. A. jabbar, B. L. Deekshatulu, and P. Chandra, "Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm," *Procedia Technol.*, vol. 10, pp. 85–94, 2013.
[21] Y. Zhang, S. Balochian, P. Agarwal, V. Bhatnagar, and O. J. Housheya, "Artificial Intelligence and Its Applications 2014," *Math. Probl. Eng.*, vol. 2016, 2016.
[22] Okwara Jerry Chizoba and Buba Abba Kyari, "The particle swarm optimization (PSO) algorithm application – A review," *Glob. J. Eng. Technol. Adv.*, vol. 05, no. 02, pp. 047–056, 2020.
[23] T. Eswaran and V. S. Kumar, "Particle swarm optimization (PSO)-based tuning technique for PI controller for management of a distributed static synchronous compensator (DSTATCOM) for improved dynamic response and power quality," *J. Appl. Res. Technol.*, vol. 15, no. 2, pp. 173–189, 2017.
[24] I. Yekkala, S. Dixit, and M. A. Jabbar, "Prediction of heart disease using ensemble learning and Particle Swarm Optimization," *Proc. 2017 Int. Conf. Smart Technol. Smart Nation, SmartTechCon 2017*, pp. 691–698, 2018.
[25] W. Y. Dong and R. R. Zhang, "Order-3 stability analysis of particle swarm optimization," *Inf. Sci. (Ny).*, vol. 503, no. 61672024, pp. 508–520, 2019.

BIBLIOGRAPHY OF AUTHORS

Elin Panca Saputra as a lecturer I have had an ID Scholar, SINTA, and SCOPUS. In addition, I also have various important certificates such as Serdik, Serkom, Askom, and training, especially regarding online learning. My research interests are in Data Mining, Machine Learning, and Online Learning.

Andi Saryoko a lecturer at Numandiri University has an ID Scholar, SINTA and SCOPUS. In addition, I also have various important certificates such as Serdik, Serkom, and Askom, and Andi Saryoko is also the head of LPPM Nusa Mandiri University.

Aryo Tunjung Kusumo a lecturer at Bina Sarana Informatika University is a certified lecturer, who is an expert in the field of information systems and software development, and has many certificates in his scientific field.

Priyono as a lecturer at Cyber University who is a certified lecturer, I have had an ID Scholar, SINTA, and SCOPUS. who is an expert in the field of information systems.