Recursive Journal of Informatics
Vol. 2, No. 2, September 2024

p-ISSN 2963-5551 https://journal.unnes.ac.id/journals/index.php/rji e-ISSN 2986-6588

Neural Network Optimization Using Hybrid Adaptive Mutation Particle Swarm Optimization and Levenberg-Marquardt in Cases of Cardiovascular Disease

Rima Ayu Cahyani[1], Aji Purwinarko[2]
[1,2]Computer Science Department, Faculty of Mathematics and Natural Sciences,
Universitas Negeri Semarang, Indonesia

Abstract. Cardiovascular disease is a condition generally characterized by the narrowing or blockage of blood vessels, which can lead to heart attacks, chest pain, or strokes. It is the leading cause of death worldwide, accounting for approximately 31% or 17.9 million deaths each year globally. Deaths caused by cardiovascular disease are projected to continue increasing until 2030, with the number of patients reaching 23.3 million. As cases of death due to cardiovascular disease become more prevalent, early detection is crucial to reduce mortality rates.
Purpose: Many previous researchers have conducted studies on predicting cardiovascular disease using neural network methods. This study extends these methods by incorporating feature selection and optimization with Hybrid AMPSO-LMA. The research is designed to explore the implementation and predictive outcomes of Hybrid AMPSO-LMA in optimizing MLP for cases of cardiovascular disease.
Methods/Study design/approach: The first step in conducting this research is to download the Heart Disease Dataset from Kaggle.com. The dataset is processed through preprocessing by removing duplicates and transforming the data. Then, data mining processes are carried out using the MLP algorithm optimized with Hybrid AMPSO-LMA to obtain results and conclusions. This system is designed using the Python programming language and utilizes Flask for website access in HTML.
Result/Findings: The research results demonstrate that the method employed by the author successfully improves the accuracy of predicting cardiovascular disease. Predicting cardiovascular disease using the MLP algorithm yields an accuracy of 86.1%, and after optimization with Hybrid AMPSO-LMA, the accuracy increases to 86.88%.
Novelty/Originality/Value: This effort will contribute to the development of a more reliable and effective cardiovascular disease prediction system, with the goal of early identification of individuals exhibiting symptoms of cardiovascular disease.

Keywords: classification, prediction, cardiovascular, neural network, multilayer perceptron, Levenberg Marquardt, adaptive mutation PSO
Received January 05, 2024 / **Revised** May 02, 2024 / **Accepted** September 17, 2024

INTRODUCTION

Cardiovascular disease is a condition that often occurs with advancing age. It generally involves the narrowing or blocking of blood vessels, leading to heart attacks, chest pain, or strokes [1]. Cardiovascular disease is the leading cause of death worldwide, accounting for approximately 31% or 17,9 million deaths each year globally [2]. Deaths due to cardiovascular disease are projected to continue increasing until 2030, with the number of patients reaching 23,3 million people. Given the increasing vulnerability of deaths caused by cardiovascular disease, early detection is crucial to reduce mortality rates. One effective way to identify and predict cardiovascular disease is by utilizing machine learning (ML) algorithms [3].

ML is a technique for making inferences from data using a mathematical approach [4]. ML can enhance the performance of search engines, making the accuracy of information sought by users more precise [5]. The algorithms used in building ML include supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning [6]. One classification method or supervised learning is neural network (NN). NN is an information system designed to mimic the human brain's functioning to solve problems through a learning process by adjusting the weights of its synapses [7].

[1]*Corresponding author.
Email addresses: rimaayucahyani@students.unnes.ac.id (Cahyani)
DOI: 10.15294/rji.v2i2.78550

Previous researchers have used NN as a predictive method. [8] used NN and naive Bayes (NB) to predict cardiovascular disease. The research results showed that the accuracy value in the NN method was better, at 84,52%, while NB was 79,88%. [9] conducted research on predicting cardiovascular disease using support vector machine (SVM), LR, and artificial neural network (ANN). The research results indicated LR's accuracy at 83%, ANN at 82%, and SVM at 80%. [10] used ANN in their research to predict cardiovascular disease, achieving an accuracy of 81,19%.

One issue in previous research is that ANN and NB have poor area under the curve (AUC) classification values and need additional attributes relevant to cardiovascular disease [11]. The training time of ANN is too long with a large amount of data, especially when performing numerical operations with high precision [11]. Therefore, this research aims to apply the MLP method to determine better weight values and optimize it with Hybrid AMPSO-LMA to increase the efficiency and effectiveness of the MLP method in determining accuracy outcomes. The goal of this research is to improve the accuracy of early detection of cardiovascular disease in individuals in the medical field.

METHODS

In this study, the first step involved the initial dataset acquisition, followed by data preprocessing that began with removing duplicates and data transformation. Afterward, a feature selection process was carried out, and the selected features were used in the classification process. The classification algorithm employed in this study was the Multilayer Perceptron. The flowchart of the method used in this research is illustrated in Figure 1.

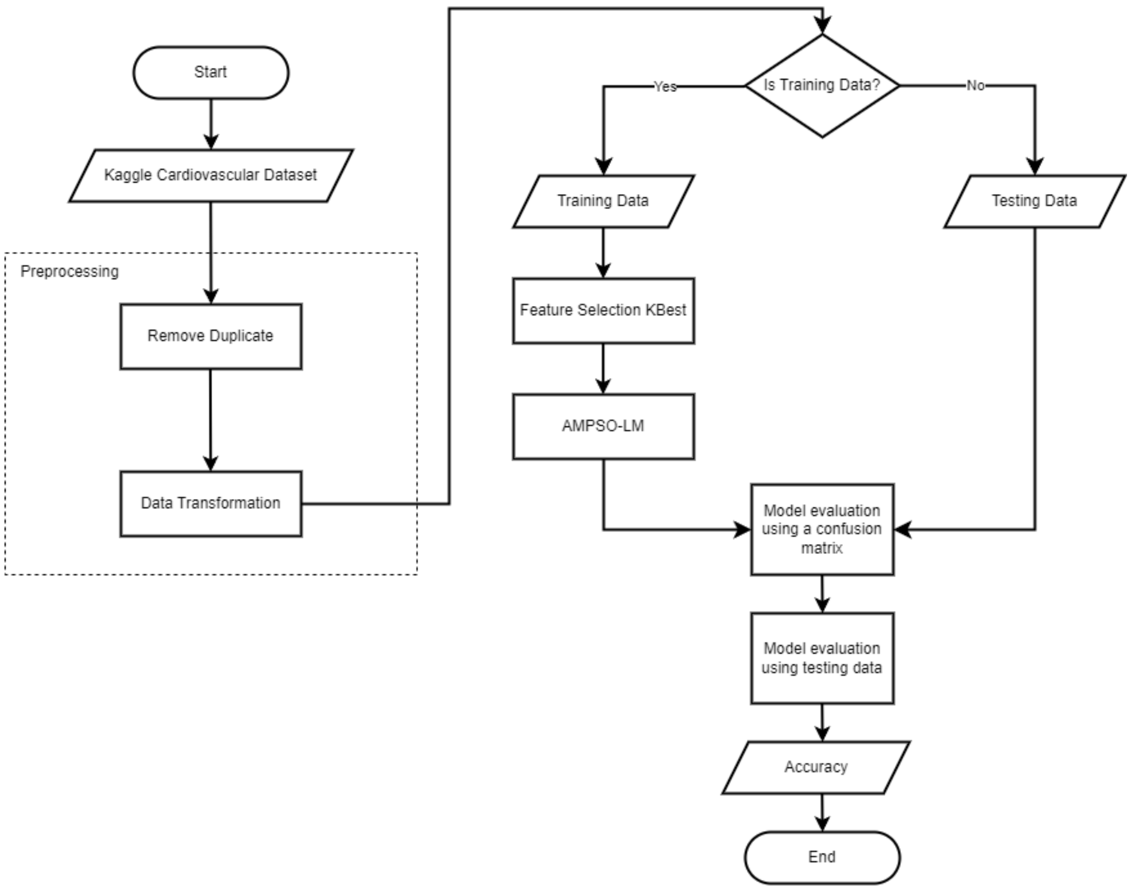

Figure 1. Research Method Flowchart

Data Collection

The data was obtained from the Heart Disease Dataset (HDD), downloaded from the Kaggle Public Health Dataset. This dataset comprises 1024 rows of data with 13 features and 1 class label. The attributes of the dataset and their descriptions can be seen in Table 1.

Table 1. HDD Attributes and Their Descriptions

Attributes	Description	Type
age	Age	Numerik
sex	Gender	Nominal
cp	Chest pain	Nominal
trestpbs	Resting blood pressure	Numerik
chol	Cholesterol	Numerik
fbs	Blood sugar	Nominal
restecg	Result electrocardiographic	Nominal
thalach	Heart rate maximum	Numerik
exang	Induced angina	Nominal
oldpeak	Depression	Numerik
slope	Tilt ST segment	Nominal
ca	Number of vessels main blood colored with fluoroscopy	Numerik
thal	Heart rate type	Nominal
target	Diagnosis heart disease	Nominal

Remove Duplicate
Duplicate removal involves eliminating duplicate or repeated data in each attribute of the dataset. This process is carried out with the aim of eliminating redundancy during the weighting phase.

Transformation Data
The data transformation used includes min-max normalization and standard scaler. Min-max normalization is a data transformation process that works by placing data within a range, with 0 as the minimum value and 1 as the maximum value [12]. This method is employed to achieve the quickest convergence compared to other methods [13]. The Min-max normalization calculation formula in Equation 1.

$$x' = \frac{x - xmin}{xmax - xmin} \quad (1)$$

Standard scaler is a data transformation method based on the mean and standard deviation, aiming to prevent data values from being too large compared to others. Standard scaler produces a data distribution with a standard deviation equal to a unit variance of 1 [14]. This method maintains stability against outliers or values greater than maxA or smaller than minA [13]. The standard scaler calculation formula in Equation 2.

$$Z = \frac{x - \mu}{\sigma} \quad (2)$$

Feature Selection
Feature selection is utilized to create a subset of relevant features while discarding redundant, excessive, unused, or unwanted features [15]. SelectKBest is part of the univariate feature selection stage that chooses the best features based on univariate statistical tests or ANOVA tests. Statistical tests can be employed to identify features with the strongest relationship with the output variable. SelectKBest retains only the features with the highest scores [16].

Adaptive Mutations Particle Swarm Optimization (AMPSO)
PSO is a population-based algorithm that adjusts positions and velocities referencing optimal particles to find solutions [17]. The PSO algorithm combines local search (local exploration) and global search (global exploration) methods [18]. The changes in position and velocity of each particle are determined by the best position obtained from all particles or global best (gbest) and the best position of each particle or personal best (pbest). Mutation in PSO involves adding a random number within the range (0,1) to gbest. The AMPSO flowchart is shown in Figure 2. The use of mutation in PSO still struggles to prevent local optima as it covers the search space too narrowly. A new solution to address the limited reach of mutated particles is to introduce adaptive mutation, where the mutation size is related to the size of the search space. The adaptive mutation equation in PSO can be seen in Equation 3.

$$gbest_j(t+1) = gbest_j(t) + \left[b_j(t) - a_j(t)\right] \& rand() \quad (3)$$

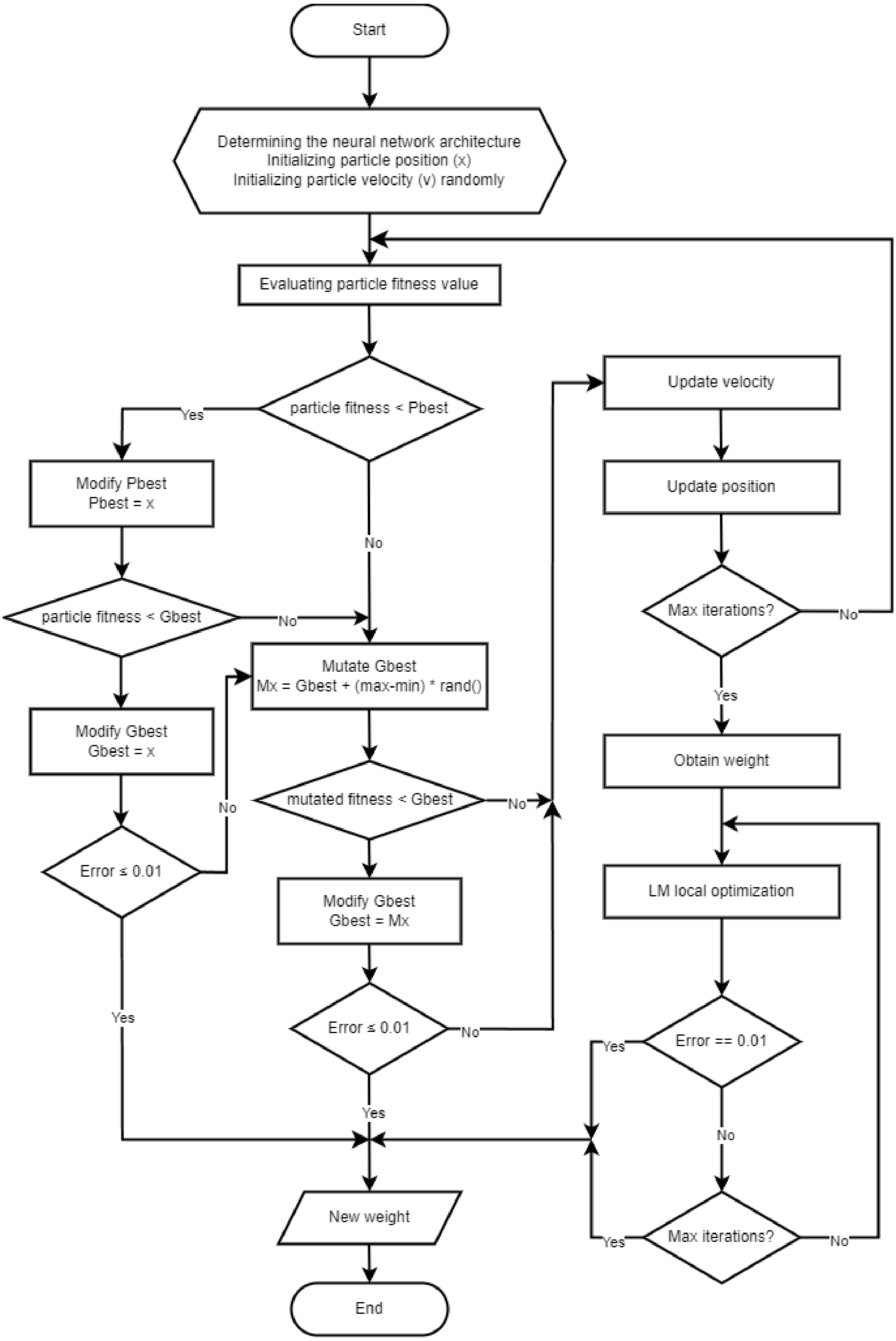

Figure 2. Flowchart AMPSO

Levenberg-Marquardt Algorithm (LMA)

LMA (Levenberg-Marquardt Algorithm) is the fastest supervised training method for training large-sized feedforward neural networks, even with hundreds of weights [19]. The basic operation of LMA involves searching for the minimum value based on the lowest sum of squares. The quadratic approach utilizes the Gauss-Newton algorithm to expedite convergence. The LMA flowchart is shown in Figure 3.

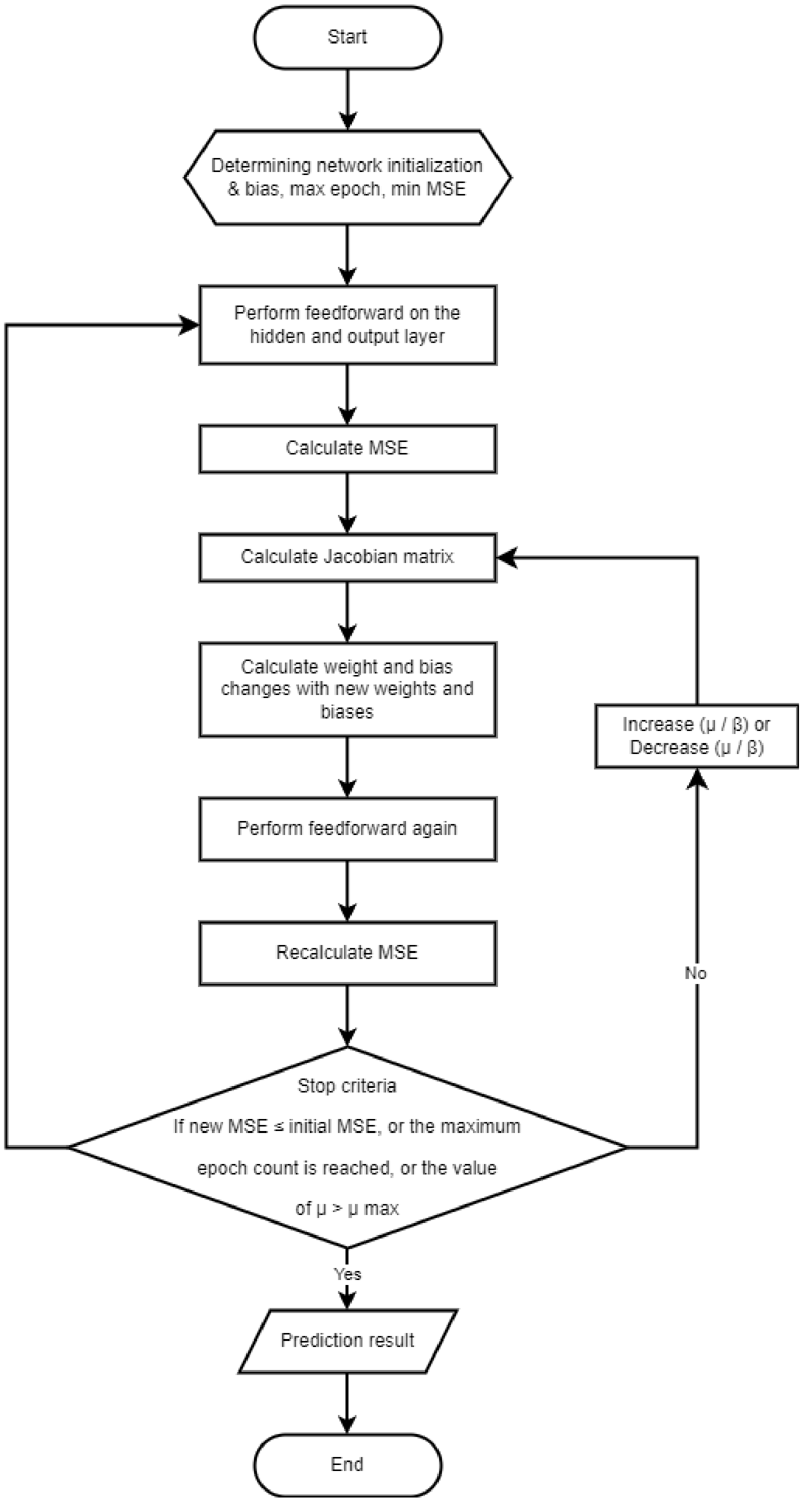

Figure 3. Flowchart LMA

Confusion Matrix

The aim of this step is to assess the classification model's performance. Performance is calculated based on the confusion matrix. The evaluation steps using the confusion matrix are as follows:
1. Classify the test data using the classification model.
2. Input the classification results into the confusion matrix.
3. Calculate performance metrics such as accuracy, sensitivity/recall, specificity, precision, and F1 score based on the data in the confusion matrix.

Table 2. Confusion Matrix

	Positive	Negative	
Positive	True Positive (TP)	False Negative (FN)	Sensitivity $\frac{TP}{(TN+FN)}$
Negative	False Positive (FP)	True Negative (FN)	Specificity $\frac{TN}{(TN+FP)}$
	Precision $\frac{TP}{(TP+FP)}$	Negative Predictive Value $\frac{TN}{(TN+FN)}$	Accuracy $\frac{TP+TN}{(TP+TN+FP+FN)}$

4. Draw conclusions based on the obtained performance metrics.

RESULT AND DISCUSSION

In this study, SelectKBest is applied for feature selection and combined with transformation as the choice for data normalization. The first step carried out is preprocessing. The purpose of preprocessing is to process the data so that it can be used for data mining. The first step in preprocessing is to ensure that the data is numeric. Data that is not in numeric format will be discarded or converted into numeric format with a specific range because feature selection can only process numeric data. Then, it is ensured that there is no empty or missing data. In the remove duplicate stage, the database's quantity is reduced from the initial 1024 to 302 rows of data. These data points are selected because there were identical rows in the dataset.

Next, the data is transformed to equalize the range of attributes that differ significantly using a specific scale. Many data sets have different ranges, so normalization is necessary. The data is transformed using the min-max normalization method by processing the minimum and maximum values of each attribute. The range used in this method is 0-1. In addition, Z-score normalization is also applied. This method has stable values against outliers or values greater than maxA or smaller than minA. Z-score normalization is carried out by processing the mean and standard deviation of the attribute values. With the completion of these steps, the preprocessing stage is finished. Then, the data splitting stage is carried out. Data splitting divides the data with sequential testing data: training data ratios of 90:10, 80:20, 70:30, and 60:40. The next process is the data mining process. In this study, the MLP algorithm is applied, followed by optimization using AMPSO. Subsequently, it is implemented with Hybrid AMPSO-LMA. This analysis process is conducted with various experiments involving combinations of data transformation and data splitting.

In the Multilayer Perceptron classification process, 10 hidden layers, a learning rate of 0,05, and a maximum of 1000 iterations are used. In this stage, SelectKBest feature selection is performed. The output of this process is the weights of each attribute. The output of this process is the weights of each attribute. Table 3 presents the features along with the filtered weights.

Table 3. Filter Weight Coefficient

Atribut	Koefisien Weight
age	0,39804465
sex	-0,136785
cp	0,6496193
trestbps	0,54239704
chol	0,69662968
fbs	0,10508909
Restecg	-0,36279852
Thalach	0,14478061
Exang	0,49629384
Oldpeak	0,12882088
Slope	0,14478061
Ca	0,49629384
thal	0,12882088

The data produces a confusion matrix:

$$\begin{bmatrix} 0 & 61 \\ 0 & 60 \end{bmatrix}$$

Thus, the obtained confusion matrix results in an accuracy of 49,587%, precision of 49,587%, sensitivity of 100%, specificity of 0%, and F1 score of 1,96%. These results provide an overview of how well the model can predict with the actual dataset.

AMPSO optimization using parameters c1: 0,5; c2: 0,3; and w: 0,9. The classification results using AMPSO feature selection are superior to SelectKBest. This is evidenced by the difference in accuracy results obtained. When using MLP and SelectKBest feature selection, the accuracy is 86,1%, while applying AMPSO and LMA increases the accuracy to 86,9%. AMPSO is designed to handle multi-objective problems, meaning it can consider multiple criteria or goals in the feature selection process. AMPSO is more capable of dealing with complex problems compared to the simpler SelectKBest, thus seeking more optimal solutions. Table 4 provides a summary of the data mining process results.

Table 4. Mining Data Table

Algorithm		MLP	AMPSO	Hybrid AMPSO-LMA
No Transformation	9:1	83,9%	80,6%	51,6%
	8:2	80,3%	81,9%	82%
	7:3	84,6%	74,7%	47,3%
	6:4	83,5%	50,4%	80,2%
MinMax Scaler	9:1	86,1%	74,2%	77,4%
	8:2	85,9%	81,9%	83,6%
	7:3	84,6%	82,4%	82,4%
	6:4	81%	78,5%	83,5%
Z-mean Scaler	9:1	86,1%	70,9%	83,9%
	8:2	85,2%	78,7%	86,9%
	7:3	84,6%	76,9%	79,1%
	6:4	82,6%	80,2%	84,3%

Based on Table 4, the best classification result is obtained using Z-mean normalization transformation with a data split ratio of 8:2, resulting in an accuracy of 86,88%.

The results of this study will be compared with three previous studies, namely [8], [9], and [10]. The comparison of results with the proposed method in the three previous studies can be seen in Table 5.

Table 5. Comparison of Results with Previous Studies

Researcher	Dataset	Method	Accuracy
Nawawi *et al*, (2019)	Heart Disease Data Set	NN	84,52%
		NB	79,88%
Handayani (2021)	Heart Disease Data Set	SVM	80%
		LR	83%
		ANN	82%
Kurniawan and Silvanie (2021)	Heart Disease Data Set Cleveland	ANN	81,19%
Proposed method	Heart Disease Data Set	MLP	86,1%
		AMPSO-LMA	86,88%

The study conducted by [8] resulted in an accuracy of 84,52% for the backpropagation neural network algorithm. The current study achieved an accuracy of 86,88%, indicating a difference in accuracy of 2,36% compared to the previous research. Furthermore, the study conducted by [9] reported an accuracy of 82%. A drawback of that study is the absence of feature selection and data transformation. In comparison, the research method used by [10] resulted in an accuracy of 81,19%. A limitation of that study is the lack of optimization with other algorithms. Optimizing MLP has the potential to improve performance and reduce computational costs. In conclusion, the proposed method in this study appears to yield better results.

The instability of accuracy results that constantly fluctuate in each analysis can be a limitation in this research. This may be attributed to fluctuations in weights and other coefficients in the model, influenced by various factors, including initializations and randomness in specific algorithms. To address this limitation, you might want to try various approaches, such as setting a seed for randomization or conducting repeated analyses with parameter variations to obtain a more stable understanding of the model's performance. Additionally, it's crucial to consider the accuracy and representativeness of the dataset used, as well as the stability of the applied algorithms.

CONCLUSION

The implementation of the Hybrid AMPSO-LMA in optimizing the MLP for predicting cardiovascular diseases has yielded excellent results. The best accuracy achieved on the MLP was 86,1%, and when the Hybrid AMPSO-LMA method was applied, the accuracy increased to 86,88%. This improvement of 0,78% indicates that the use of LMA for classification has successfully enhanced the performance and quality of the model in predicting whether defined patients have cardiovascular diseases or not. The research results demonstrate that this implementation can address the complexity and time efficiency issues in predicting cardiovascular diseases that are not optimized. A limitation of this research is the fluctuating accuracy results because the weights and coefficients change every time the analysis is conducted.

REFERENCES

[1] D. Dona, H. Maradona, and M. Masdewi, "Sistem pakar diagnosa penyakit jantung dengan metode case based reasoning (CBR)," *Zo. J. Sist. Inf.*, vol. 3, no. 1, pp. 1–12, 2021, doi: 10.31849/zn.v3i1.6442.

[2] S. S. Virani *et al.*, *Heart disease and stroke statistics—2020 update a report from the American Heart Association*, vol. 141, no. 9. 2020. doi: 10.1161/CIR.0000000000000757.

[3] W. Nugraha, "Prediksi penyakit jantung cardiovascular menggunakan model algoritma klasifikasi," *J. Manag. dan Inform.*, vol. 9, no. 2, pp. 3–8, 2021.

[4] J. W. G. Putra, *Pengenalan Konsep Pembelajaran Mesin dan Deep Learning*, 1.4. Tokyo: wiragotama.github.io, 2020. [Online]. Available: https://www.researchgate.net/publication/323700644_Pengenalan_Pembelajaran_Mesin_dan_Deep_Learning

[5] A. S. Diantika and Y. Firmanto, "Implementasi machine learning pada aplikasi penjualan produk digital (studi pada GrabKios)," *J. Ilm. Mhs. FEB*, vol. 53, no. 9, pp. 1689–1699, 2019.

[6] E. Retnoningsih and R. Pramudita, "Mengenal machine learning dengan teknik supervised dan unsupervised learning menggunakan Python," *Bina Insa. Ict J.*, vol. 7, no. 2, pp. 156–165, 2020, doi: 10.51211/biict.v7i2.1422.

[7] A. F. Achmalia, Walid, and Sugiman, "Peramalan penjualan semen menggunakan backpropagation neural network dan recurrent neural network," *UNNES J. Math.*, vol. 8, no. 1, pp. 92–106, 2019, doi: 10.15294/UJM.V9I1.29970.

[8] H. M. Nawawi, J. J. Purnama, and A. B. Hikmah, "Komparasi algoritma neural network dan naïve Bayes untuk memprediksi penyakit jantung," *J. Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 189–194, 2019, doi: 10.33480/pilar.v15i2.669.

[9] F. Handayani, "Komparasi support vector machine, logistic regression dan artificial neural network dalam prediksi penyakit jantung," *J. Edukasi dan Penelit. Inform.*, vol. 7, no. 3, p. 329, 2021, doi: 10.26418/jp.v7i3.48053.

[10] A. Kurniawan and A. Silvanie, "Prediksi penyakit jantung menggunakan jaringan syaraf tiruan multi layer perceptron dan Python pada basis data penyakit jantung di Ceveland," *JUNIF J. Nas. Inform.*, vol. 2, no. 1, pp. 21–28, 2021.

[11] A. N. Sari and S. Alfionita, "Klasifikasi penyakit jantung menggunakan metode naïve Bayes," *AMRI (Analisa Metod. Rekayasa Inform.*, vol. 1, no. 1, pp. 22–26, 2022, doi: 10.12487/AMRI.v1i1.xxxxx.

[12] F. N. S. Inggih Permana, "The Effect of Data Normalization on the Performance of the Classification Results of the Backpropagation Algorithm," *IJIRSE Indones. J. Inform. Res. Softw. Eng.*, vol. 2, no. 1, pp. 67–72, 2022.

[13] Gde Agung Brahmana Suryanegara, Adiwijaya, and Mahendra Dwifebri Purbolaksono, "Peningkatan Hasil Klasifikasi pada Algoritma Random Forest untuk Deteksi Pasien Penderita Diabetes Menggunakan Metode Normalisasi," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 1, pp. 114–122, 2021, doi: 10.29207/resti.v5i1.2880.

[14] Ainurrochman, D. P. Adi, and A. B. Gumelar, "Deteksi Emosi Wicara pada Media On-Demand menggunakan SVM dan LSTM," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 5, pp. 799–804, 2020, doi: 10.29207/resti.v4i5.2073.

[15] D. ARIYOGA, "Perbandingan Metode Seleksi Fitur Filter, Wrapper, Dan Embedded Pada Klasifikasi Data Nirs Mangga Menggunakan Random Forest Dan Support Vector Machine …," 2022, [Online]. Available: https://dspace.uii.ac.id/handle/123456789/38955

[16] Generosa Lukhayu Pritalia, “Analisis Komparatif Algoritme Machine Learning dan Penanganan Imbalanced Data pada Klasifikasi Kualitas Air Layak Minum,” *KONSTELASI Konvergensi Teknol. dan Sist. Inf.*, vol. 2, no. 1, pp. 43–55, 2022, doi: 10.24002/konstelasi.v2i1.5630.
[17] E. Subiyantoro and Azhari, “Autonomous Cognitive Leveling Game Pada Serious Game Menggunakan Particle Swarm Optimization,” *J. Buana Inform.*, vol. 8, no. 2, pp. 87–98, 2017.
[18] V. Julianto, H. S. Utomo, and M. R. Arrahimi, “Penerapan bat algorithm dalam penyelsaian kasus travelling salesman problem (TSP) pada internship program,” *J. Ilm. Inform.*, vol. 6, no. 2, pp. 111–116, 2021, doi: 10.35316/jimi.v6i2.1485.
[19] J. Bilski, B. Kowalczyk, A. Marchlewska, and J. M. Zurada, “Local Levenberg-Marquardt algorithm for learning feedforwad neural networks,” *J. Artif. Intell. Soft Comput. Res.*, vol. 10, no. 4, pp. 299–316, 2020, doi: 10.2478/jaiscr-2020-0020.