Optimal multi-layer perceptron parameters for early stage diabetes risk prediction

V N Wijayaningrum1*, T H Saragih2 and N N Putriwijaya3

1 Department of Information Technology, Politeknik Negeri Malang, Malang, Indonesia
2 Faculty of Mathematics and Natural Sciences, Lambung Mangkurat University, Banjarmasin, Indonesia
3 Faculty of Information and Communication Technology, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

*vivinurw@polinema.ac.id

Abstract. Diabetes is an alarming threat to people around the world because the number of diabetics is increasing every year. Diabetics with other complications have a very high risk of death. Therefore, the use of technology to predict the risk of early diabetes is needed. Neural Network as one part of artificial intelligence plays a role in solving prediction problems with satisfying results. In this study, a multi-layer perceptron neural network is used to predict the risk of early stage diabetes with optimal parameters from the optimization results using Improved Crow Search Algorithm. The test results prove that the multi-layer perceptron with optimal parameters is able to provide better accuracy compared to other algorithms such as J48, PART, Decision Table, Naïve Bayes, AIRS1, AIRS2, and Single Layer Perceptron with the highest accuracy values of 97.69% and 96.92% for one and two hidden layers, respectively. This proves that the proposed solution can be used to predict the early stage diabetes risk.

1. Introduction

Diabetes is a disease caused by a metabolic disorder so that blood sugar levels increase. Diabetes patients show weight loss and frequent urination. In addition, the effects of long-term diabetes can cause sufferers to experience kidney failure, nerve damage, blurred vision, even death [1]. World Health Organization (WHO) states that 422 million people worldwide suffer from diabetes, with the majority of sufferers living in low and middle income countries. The number of diabetes cases has also continued to increase over the last few decades. Each year, 1.6 million deaths are directly related to diabetes. This shows that diabetes is a dangerous disease. Thus, early identification and prediction of diabetes risk is needed to avoid more serious problems.

Currently, the health care system is one of the popular research fields used for early identification of a disease by applying data mining techniques. Data mining can be applied to extract hidden patterns, so that relationships between different patterns in clinical data can be found [2]. Furthermore, machine learning can play a role in teaching computers to carry out the learning process and understanding the given parameters, which are then used for prediction or decision making. This learning process can be done using various classification or clustering algorithms, one of which is the Support Vector Machine. This algorithm performs classification by forming a multidimensional hyperplane that separates the two
data clusters by maximizing the margin between the two data clusters. By using the data set on diabetes from the National Health and Nutrition Examination Survey (NHANES) with 14 parameters, SVM performance was measured using the area under the curve (AUC) with a value of 83.47% for the detection of diagnosed or undiagnosed diabetes [3].

In other study, a hybridization mechanism between K-Means and J48 was proposed. This mechanism is carried out using K-Means for data reduction then the classification process is carried out using the J48 decision tree. Data reduction aims to transform raw data into a more useful form, in this case the data set is converted into an ordered and simple form. In this study, PIMA Indian Diabetes Data from UCI repository was used. Missing values in the data set are replaced with mean values according to their attributes, while incorrect classification results are removed from the sample using K-Means algorithm, leaving 532 samples out of 768 samples. Furthermore, J48 classifies the sample data set with an accuracy value of 90.04% [4].

Neural Network is a computational algorithm that is often used as a classification method for high-dimensional data [5]. Several Neural Network algorithms have proven successful for disease diagnosis, such as Multilayer Perceptron for heart disease prediction [6,7], Back Propagation for chronic kidney disease prediction [8], Extreme Learning Machine for Jatropha Curcas disease identification [9], and Learning Vector Quantization for leaf disease detection [10]. However, the learning process on Neural Network algorithms is influenced by the learning rate. Algorithms often take a long time to come up with an optimal solution when the learning rate is too small [11]. To solve this problem, several researchers have performed various hybridization techniques between the Neural Network algorithm and other algorithms, such as genetic algorithms and Back Propagation which are used to determine the initial weights and number of neurons in the hidden layer for the Multi-Layer Perceptron [12]. In other study, Extreme Learning Machine and Modified Simulated Annealing were proposed to optimize Extreme Learning Machine weight values [13]. The use of this hybridization technique aims to increase the accuracy value obtained.

In this study, optimization of Multi-Layer Perceptron parameters using Improved Crow Search Algorithm was proposed. Multi-Layer Perceptron has five main parameters that can affect the given results, namely learning rate, momentum, number of epochs, threshold for errors, and number of neurons in hidden layer. The optimal parameters of the search results using the Improved Crow Search Algorithm are expected to provide high accuracy values for predicting early stage diabetes risk.

2. Research methodology
Multi-Layer Perceptron (MLP) is a class of Neural Network algorithms. It consists of one input layer, one output layer, and one or more hidden layers of nodes that are mapped nonlinearly. The nodes in each layer are connected by weight to all nodes in the next layer. The Perceptron Multi-Layer architecture depends on the choice of the number of layers, the number of nodes in the hidden layer, and the objective functions used [14]. The number of neurons in the input layer is determined based on the number of input variables, while the number of neurons in the output layer is based on the number of output classes [15]. In this study, 16 neurons were used in the input layer and two neurons in the output layer. Meanwhile, the number of hidden layers and the number of neurons in the hidden layer are determined through the optimization process. The architecture of Multi-Layer Perceptron used in this study is shown in Figure 1.

In this study, Multi-Layer Perceptron is used to predict the risk of early stage diabetes. When carrying out the learning process, Multi-Layer Perceptron uses five main parameters, consisting of learning rate, momentum, number of epochs, threshold for errors, and number of neurons in hidden layer. In order to obtain high accuracy, Improved Crow Search Algorithm (ICS) is used to find optimal values for the Multi-Layer Perceptron parameters. The improvised mechanism of ICS was adopted from previous research [16]. The parameter optimization mechanism of the Multi-Layer Perceptron using ICS is illustrated in Figure 2.
The parameter optimization mechanism of Multi-Layer Perceptron in Figure 2 begins by inserting the value of each attribute into Multi-Layer Perceptron. To produce a good quality solution, the parameters used for the learning process on Multi-Layer Perceptron are determined based on the search results from Improved Search Crow Algorithm (ICSA). In this search process, ICSA also uses several parameters consisting of flock size, number of iterations, flight length, and awareness probability. The ICSA search results are the optimal parameters used to regulate the learning process on Multi-Layer Perceptron. After carrying out the learning process, Multi-Layer Perceptron classifies the input data to produce a positive or negative class output. From the dataset used, it can be seen the accuracy value obtained at the end.

2.1. Search agent representation
ICSA is a meta-heuristic algorithm that has a population that contains several search agents. Each search agent states a possible solution to the problem. In this study, each agent has six values that represent the parameters of the Multi-Layer Perceptron. Figure 3 shows a representation of a search agent.

In Figure 3, the variables $x_1$ to $x_6$ respectively represent the values for learning rate, momentum, number of epochs, threshold for errors, number of neurons in the first hidden layer, and number of neurons in the second hidden layer. Each search agent has different values but these values are within the range of values specified for each variable.
2.2. **Fitness function**

The quality of the solutions offered by each search agent can be measured using the fitness function. A search agent who has a high fitness value claims that its solution is good. In this case, the fitness function is calculated using accuracy. To calculate accuracy, (1) is used.

\[
\text{Accuracy} = \frac{\text{The number of sample correctly classified}}{\text{The number of all sample}} \tag{1}
\]

To evaluate the performance of the algorithm, this study uses $k$-fold cross validation, which is a resampling procedure by separating the dataset into $k$ folds. Thus, the calculation of accuracy uses (2) as a fitness function.

\[
\text{Accuracy} = \frac{\text{Sum of Accuracy in each fold}}{k} \tag{2}
\]

3. **Results and discussion**

The dataset used in this study is the early stage diabetes risk prediction dataset derived from the UCI dataset. The dataset consists of 520 instances and 17 attributes consisting of 16 inputs and one output. The output class consists of ‘positive’ and ‘negative’ related to early stage diabetes risk prediction.

The performance of the proposed algorithm for identifying early stage diabetes risk is compared with other algorithms such as J48, PART, Decision Table, Naive Bayes, Artificial Immune Recognition System (AIRS1), and AIRS2. In addition, comparisons of various variants of the Perceptron algorithm are also carried out to determine the quality of the given solutions. The ICSA parameter settings used to optimize the Multi-Layer Perceptron parameters are shown in Table 1.

| Parameter               | Value | Explanation of Parameter                     |
|-------------------------|-------|-----------------------------------------------|
| Flock size              | 100   | The number of agents searching for solutions  |
| Number of iterations    | 50    | The number of searches carried out            |
| Flight length           | 2.5   | How far the search is carried out             |
| Awareness probability   | 0.1   | The tendency of direction in finding a solution |

The performance comparison of these algorithms is shown in Table 2. The performance of the proposed algorithm is highlighted in gray. The values shown are the average accuracy using 10-fold cross validation.

| Algorithm                             | Accuracy (%) |
|---------------------------------------|--------------|
| J48                                   | 92.88        |
| PART                                  | 94.42        |
| Decision Table                        | 93.46        |
| Naive Bayes                           | 86.15        |
| AIRS1                                 | 89.62        |
| AIRS2                                 | 90.58        |
| Single Layer Perceptron               | 95.19        |
| Multi-Layer Perceptron without optimization | 95.38   |
| Multi-Layer Perceptron with optimization using ICSA (one hidden layer) | 92.66 (Max: 97.69) |
| Multi-Layer Perceptron with optimization using ICSA (two hidden layer) | 93.18 (Max: 96.92) |
Because ICSA is a stochastic algorithm, the accuracy given to each execution will be different. Therefore in Table 2, the average accuracy obtained from 10 independent experiments is shown, and the highest accuracy obtained from all these experiments is also shown and symbolized by Max. These results indicate that the Multi-Layer Perceptron with optimization provides the highest accuracy value compared to other algorithms using either one hidden layer or two hidden layers. This is because ICSA is able to find the best parameters on the Multi-Layer Perceptron to get a solution with high accuracy. Meanwhile, the average accuracy given by the Multi-Layer Perceptron with optimization is lower than some other algorithms. This is because the ICSA initialization process is carried out using random initial values which are then improved continuously as the learning process progresses. Thus, the use of a random initial value that is not good enough allows the algorithm to be unable to provide a high enough accuracy.

4. Conclusion
The test results show that the use of Multi-Layer Perceptron with optimal parameters from the optimization results using Improved Crow Search Algorithm can be used to predict the risk of early diabetes with high accuracy. The proposed algorithm is proven to provide the best accuracy compared to other algorithms such as J48, PART, Decision Table, Naive Bayes, AIRS1, AIRS2, Perceptron, and Multi-Layer Perceptron without optimization. By using the optimal parameters of the optimization results, it is known that the accuracy provided by Multi-Layer Perceptron with one hidden layer and two hidden layers has a slight difference. Multi-Layer Perceptron with one hidden layer gives an average accuracy of 92.66% with the highest accuracy of 97.69%, while Multi-Layer Perceptron with two hidden layers gives an average accuracy of 93.18% with the highest accuracy of 96.92%.

The hidden layers on Multi-Layer Perceptron perform computations on input and use activation functions for generating output. However learning with more layers will be easier but more training time is required. The weights of each hidden layer also have a very important role to the output produced by Multi-Layer Perceptron. Weights are used to scale inputs separately. In the next study, the use of optimization mechanisms using Particle Swarm Optimization can be used to optimize the weights used in each hidden layer. Thus, the predicted results are expected to be better so that the algorithm produces higher accuracy.

References
[1] Vijayan V V and Anjali C 2015 Prediction and Diagnosis of Diabetes Mellitus - A Machine Learning Approach 2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS) 122–7
[2] Kumari S and Singh A 2012 A Data Mining Approach for the Diagnosis of Diabetes Mellitus 2013 7th International Conference on Intelligent Systems and Control (ISCO) (IEEE) 373–5
[3] Yu W, Liu T, Valdez R, Gwinn M and Khoury M J 2010 Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes BMC Med. Inform. Decis. Mak. 10 16
[4] Chen W, Chen S, Zhang H and Wu T 2017 A Hybrid Prediction Model for Type 2 Diabetes Using K-means and Decision Tree 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS) 386–90
[5] Wójcik P I and Kurdziel M 2019 Training neural networks on high-dimensional data using random projection Pattern Anal. Appl. 22 1221–1231
[6] Yan H, Jiang Y, Zheng J, Peng C and Li Q 2006 A multilayer perceptron-based medical decision support system for heart disease diagnosis Expert Syst. Appl. 30 272–81
[7] Sonawane J S and Patil D R 2014 Prediction of Heart Disease Using Multilayer Perceptron Neural Network International Conference on Information Communication and Embedded Systems (IEEE) 1–6
[8] Borisagar N, Barad D and Raval P 2017 Chronic Kidney Disease Prediction Using Back Propagation Neural Network Algorithm Proceedings of International Conference on
Communication and Networks (Singapore: Springer) p 295–303

[9] Saragih T H, Fajri D M N, Mahmudy W F, Abadi A L and Anggodo Y P 2018 Jatropha Curcas Disease Identification with Extreme Learning Machine *Indones. J. Electr. Eng. Comput. Sci.* **12** 883–888

[10] Upadhyay A, Singh J and Shinde R 2019 Learning Vector Quantization Based Leaf Disease Detection *International Conference on Recent Developments in Science, Engineering and Technology* (Singapore: Springer) p 246–53

[11] Rahmi A, Wijayaningrum V N, Mahmudy W F and Parewe A M A K 2016 Offline Signature Recognition using Back Propagation Neural Network *Indones. J. Electr. Eng. Comput. Sci.* **4** 678–83

[12] Castillo P A, Merelo J J, Prieto A, Rivas V and Romero G 2000 G-Prop: Global optimization of multilayer perceptrons using GAs *Neurocomputing* **35** 149–63

[13] Saragih T H, Mahmudy W F, Abadi A L and Anggodo Y P 2018 Application of extreme learning machine and modified simulated annealing for jatropha curcas disease identification *Int. J. Adv. Soft Comput. its Appl.* **10** 108–19

[14] Ramchoun H, Idrissi M A J, Ghanou Y and Ettaouil M 2017 New Modeling of Multilayer Perceptron Architecture Optimization with Regularization: An Application to Pattern Classification *IAENG Int. J. Comput. Sci.* **44** 261–9

[15] Taşkıran M, Çam Z G and Kahraman N 2015 An Efficient Method to Optimize Multi-Layer Perceptron for Classification of Human Activities *Int. J. Comput. Commun. Instrum. Eng.* **2** 191–5

[16] Wijayaningrum V N and Putriwijaya N N 2020 An Improved Crow Search Algorithm for Data Clustering *Emit. Int. J. Eng. Technol.* **8** 86–101