Investigations on Impact of Feature Normalization Techniques for Prediction of Hydro-Climatology Data Using Neural Network Backpropagation with Three Layer Hidden

Syaharuddin[1,2], Fatmawati[1*], Herry Suprajitno[1]

[1] Department of Mathematics, Faculty of Science and Technology, Universitas Airlangga, Surabaya 60115, Indonesia
[2] Department of Mathematics Education, Faculty of Teacher Training and Education, Universitas Muhammadiyah Mataram, Mataram 83127, Indonesia

Corresponding Author Email: fatmawati@fst.unair.ac.id

ABSTRACT

Data normalization techniques are a very important initial stage to be carried out in order to obtain a good predictive data approach. Many researchers get different prediction and error results in each use of these data normalization techniques. Thereby, in this article discusses the accuracy rate of the seven normalization techniques at the preprocessing stage in the Neural Network Backpropagation (NNBP) architecture including decimal scaling, Z-score, min-max (there are 6 types), sigmoid, tanh estimators, mean-MAD, and median-MAD. We used two data patterns: seasonal data (rainfall) and stationary data (air humidity) that taken over the past 10 years (at 10-day intervals). We use accuracy rate parameters including number of epochs, MAE, and MSE when conducting training, testing, and predictions. The results showed that the Z-score technique was very good for the normalization of rainfall data with epochs of 10, MAE of 0.051, and MSE of 0.004. In the case of air humidity data, mean-MAD and Z-score techniques can be recommended with the number of mean-MAD technique epochs of 8, MAE of 0.013, MSE of 0.0004, while the number of epochs of Z-score techniques of 7, MAE of 0.018, and MSE of 0.0006. Thus, we conclude that when other researchers predict seasonal data or stationary data can use the Z-score technique for data normalization.

1. INTRODUCTION

Input data in the neural network back propagation (NNBP) architecture varies greatly and has different scales including static data patterns, decreased data patterns, increased data patterns, and even data patterns with extreme trends. The use of original data (raw data) to train neural networks can lead to convergence problems [1]. This will have implications for a high number of epochs and difficult networks to recognize data patterns. Therefore, the stage of normalization or standardization of data needs to be done before the data is trained in the NNBP architecture [2]. This normalization technique is included in the data preprocessing stage which will affect the accuracy of predictions significantly. The purpose of data normalization is to maintain all data value relationships appropriately in order to reduce the potential for bias in data [3]. It is in agreement with Nayak et al. [4] stating that data normalization aims (1) to minimize bias in neural networks from one feature to another, and (2) to accelerate training time by initiating the training process for each feature on the same scale.

Research on data normalization techniques as a stage of data preprocessing in NNBP has been widely carried out [4-12]. So, it can be known that normalization techniques include several data normalization techniques, namely, Z-score, min-max, mean-MAD, median-MAD, sigmoid, decimal scaling, tanh estimator, vector, and softmax. Panigrahi [11] has conducted experiments by training data on Wisconsin employment, monthly interest rates government bond yield 2-year securities,

Reserve Bank of Australia, monthly milk production per cow in pounds, and temperature. He normalized the data with min-max, decimal scaling, median, vector, and Z-score techniques. The results showed that decimal scaling and vector techniques provide good accuracy results. This result is different from Tasdemir et al. [9] when calculated wind power performance and obtained that Z-score became the recommended normalization technique.

Nayak et al. [4] when predicting stock markets, conducted data normalization experiments using min-max techniques, decimal scaling, Z-score, median, median-MAD, sigmoid, and tanh estimators. The results showed that the Z-score technique provided a good degree of accuracy and min-max could not be recommended. This result is also supported by Khond [6] when conducting data normalization experiments using min-max, softmax, decimal scaling, and Z-score techniques. He concluded that the Z-score normalization method is an effective and robust technique of data normalization when input data differ largely on the scale. In contrast, Nayak et al. [4] and Singh et al. [13] recommend the min-max technique to be used in the process of data normalization. Lastly, Eesa and Arabo [10] prefer the Mean-MAD and Median-MAD techniques in conducting normalization experiments of some unique client identifier (UCI) datasets.

The results of these studies provide different information related to recommendations for data normalization techniques that we should use when training-testing data. However, broadly speaking, more dominant normalization techniques are recommended, namely decimal scaling, Z-score, min-max,

sigmoid, tanh estimators, median-MAD, and mean-MAD. Furthermore, when it is viewed from the writing of the formulas of these techniques, the researchers wrote with the same formula. However, for the min-max technique, each researcher gave a different formula. As example, Rohman et al. [14], Wang et al. [15], and Pradhan et al. [16] only use normalized data variables (x_i'), data (x_i), maximum datum (x_{i-max}), and minimum datum (x_{i-min}) in data normalization with the formula:

$$x_i' = \frac{x_i - x_{i-\min}}{x_{i-\max} - x_{i-\min}} \qquad (1)$$

This formula is different from the formula used by Melesse [17] which multiplies x_i' by 0.99 plus 0.01. While, Singh et al. [13] multiply x_i' by 2 minus 1. Then, Lesnussa et al. [18] multiply x_i' by 0.8 plus 0.1. It is also different from Kumar et al. [19] that used the average variable ($\bar{x}_i$). Statistically, if we modify a variables and constants will certainly produce different calculation results. The min-max method is also used by Yang et al. [20] for stirling cryocooler predictions. Velasco et al. [21] utilized the method for the day-ahead base, intermediate, and peak load predictions. Elgin et al. [22] employed the method for bio-inspired classification. Lastly.

Danpakdee and Songpan [23] used the method for skin image lesion classification.

Therefore, it is needed to be more in-depth experiments on the results of calculations for these data normalization techniques, especially the min-max technique compiled by researchers. Thus, it is clear that the min-max technique has a high degree of accuracy at intervals of 84.60%-96%. Hence, in this article, we apply statistical calculations to determine the mean square error (MSE) between actual data and normalization data. So that it can be seen simply which technique has the smallest MSE value. Then, we used all types of min-max techniques and other techniques including decimal scaling, Z-score, sigmoid, tanh estimators, median-MAD, and mean-MAD to data normalization. We also utilize two types of data with different patterns, namely rainfall data for seasonal data and air humidity data for stationary data. The results of this study are expected to provide new recommendations about the use of data normalization techniques to predict time series data using NNBP.

2. RESEARCH METHODS

2.1 Formula of normalization techniques

Table 1. Data normalization techniques

Normalization Techniques	Formula	Source
Min-Max	MM1: $x_i' = \frac{x_i - x_{i-\min}}{x_{i-\max} - x_{i-\min}}$	[14-16]
	MM2: $x_i' = 0.5 + 0.5\left(\frac{x_i - \bar{x}_i}{x_{i-\max} - x_{i-\min}}\right)$	[19]
	MM3: $x_i' = 0.01 + 0.99\left(\frac{x_i - x_{i-\min}}{x_{i-\max} - x_{i-\min}}\right)$	[17]
	MM4: $x_i' = 0.1 + 0.8\left(\frac{x_i}{x_{i-\max}}\right)$	[24]
	MM5: $x_i' = -1 + 2\left(\frac{x_i - x_{i-\min}}{x_{i-\max} - x_{i-\min}}\right)$	[13]
	MM6: $x_i' = 0.1 + 0.8\left(\frac{x_i - x_{i-\min}}{x_{i-\max} - x_{i-\min}}\right)$	[18]
Decimal Scaling	$x_i' = \frac{x_i}{10^d}$	[6, 11]
Z-Score	$x_i' = \frac{x_i - \bar{x}_i}{\sigma}$	[25, 26]
Mean-MAD	Mean: $x_i' = \frac{x_i}{\bar{x}_i}$	[10]
	Mean-MAD: $x_i' = \frac{x_i - \bar{x}_i}{MAD(x_i)}$	[4, 10]
Median-MAD	Median: $x_i' = \frac{x_i}{Median(x_i)}$	[4, 10]
	Median-MAD: $x_i' = \frac{x_i - Median(x_i)}{MAD(x_i)}$	[4, 10]
Sigmoid	$x_i' = \frac{1}{1 + e^x}$	[4]
Softmax	$x_i' = \frac{1}{1 + e^{\frac{x_i - \bar{x}_i}{std(x_i)}}}$	[6, 13]
Tanh Estimators	$x_i' = 0.5\left(\tanh\left(\frac{0.01 \cdot (x_i - \bar{x}_i)}{\sigma}\right) + 1\right)$	[5, 7]
Vector	$x_i' = \frac{x_i}{\sqrt{\sum_{i=1}^{n} x_i^2}}$	[11]

Generally, data predicted using NNBP has four patterns, namely static data, seasonal data, cyclical data, and trend data. These four data patterns have different characteristics. Therefore, data normalization techniques are needed when the data training process (preprocessing stage) is in the NNBP architecture. Several studies have shown that there are nine data normalization techniques with varying degrees of accuracy. For example, the min-max technique used varies in each different case. Therefore, these data normalization techniques need to be experimented on data of different types of data to see which data is good to apply. This process will also show which normalization techniques are good to use for all types of data. Some data normalization techniques are presented in Table 1.

In the initial stage, we calculated the MSE between the actual data and the normalization result data. Normalization techniques used are only decimal scaling, Z-score, min-max (there are six types), sigmoid, tanh estimators, mean-MAD, and median-MAD. In the second stage, we selected five techniques with the smallest MSE. In the third stage, we conducted data training using NNBP normalization techniques that have been selected. In the final stage, we tabulated the output of NNBP, interpreted it, and made conclusions.

2.2 Dataset and architecture NNBP 3-layer hidden

This study employed rainfall data (seasonal data) and air humidity (stationary data) for case studies. We used two different types of data to conduct experiments on each normalization technique. Thus, that each technique can be known the accuracy rate when simulated using the type of seasional data and stationary data. Rainfall data were taken from Ampenan station located at latitude -8.618 and longitude: 116,083 (Mataram City, Indonesia) and air humidity data were taken from Kediri station located at latitude: -8.6364 and longitude: 116.1707 (West Lombok Regency, Indonesia). The data taken were the data over the last 10 years (January 2012 to December 2021) at 10-day intervals, so that in a year there are 36 data.

In the data simulation stage, we used the NNBP architecture with three hidden layers, namely 36-73-37-19-1. The number of neurons in the 1st hidden layer has been determined using the Hecht-Nelson formula, while the number of neurons in the second and third hidden layers was determined using the Lawrence-Fredrickson formula. The other parameters set were trainlm function, activation function in the hidden layer and output layer namely logsig-logsig-logsig-purelin, the goal of 0.001, the max epoch of 1000, the learning rate of 0.1, and the momentum of 0.9. Furthermore, parameters for testing the accuracy level of the architecture were number of epoch, mean absolute error (MAE) and mean square error (MSE).

3. RESULTS AND DISCUSSION

3.1 Selection of normalization techniques

At this stage, we calculated descriptive statistical values of rainfall and air humidity data to determine the MSE value between actual data and data resulted from normalization, both rainfall data and air humidity data. The rainfall data obtained a minimum value of 0, a maximum of 401, a mean of 40.29, a median of 21, and a standard deviation of 49.93. Meanwhile, the air humidity data obtained a minimum value of 71, a

maximum of 91, a mean of 82.64, a median of 83, and a standard deviation of 3.44. In the decimal scaling technique, researchers used a value of d of 2 for air humidity data and d of 3 for rainfall data. The results of the calculation of the MSE value between the actual data and the normalization data can be seen in Figure 1 and Figure 2.

Figure 1 provides information that five normalization techniques with small MSE values are MM1, MM2, MM3, MM4, and decimal scaling. Then, in Figure 2, we obtained a normalization technique with a small MSE value, namely MM2, decimal scaling, Z-score, Mean-MAD, and Median MAD. Thus, the five techniques were selected based on the case studies provided.
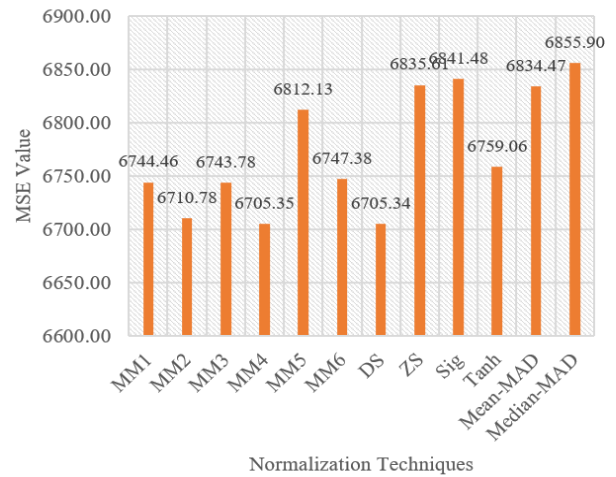

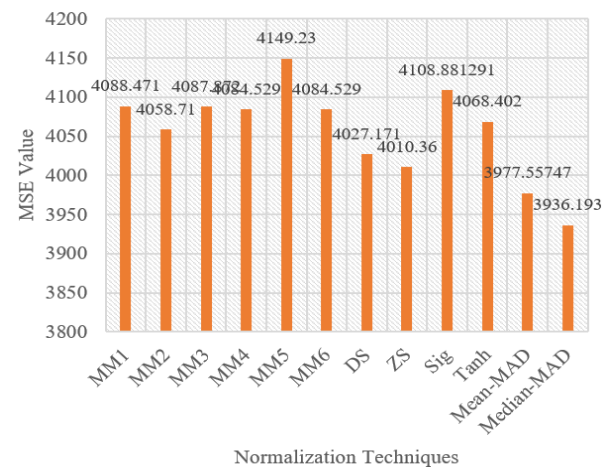
Figure 1. MSE value of air humidity (%)

Figure 2. MSE value of rainfall data (mm)

These five techniques were used in the data training process using a predetermined architecture. Because in computational algorithms the process of normalization is always followed by the process of de-normalization, the equation of each technique is obtained as follows.

- De-normalization for MM2

$$x_i = 2 \cdot (x' - 0.5) \cdot (x_{i-\text{mx}} - x_{i-\text{min}}) + \bar{x}_i \qquad (2)$$

- De-normalization for Decimal scaling

$$x_i = 10^d \cdot x_i' = 10^2 \cdot x_i' = 100 \cdot x_i' \qquad (3)$$

- De-normalization for Z-score

$$x_i = x_i' \cdot \sigma + \bar{x}_i \qquad (4)$$

- De-normalization for Mean-MAD

$$x_i = x_i' \cdot MAD(x_i) + \bar{x}_i \qquad (5)$$

- De-normalization for Median-MAD

$$x_i = x_i' \cdot MAD(x_i) + Median(x_i) \qquad (6)$$

Based on Table 1 and the de-normalization formula, computational scripts can be created using MATLAB as presented in Table 2.

3.2 Result of training data

Because there were two types of data, the data training process was carried out 10 times. Each training result was recorded and tabulated to see the number of epochs, MAE, and MSE. Graphs of actual data approaches and prediction data were also checked. The results of training data can be seen in Table 3.

Table 3 shows that in the case of rainfall data, it can be known that Z-score normalization technique gives good results with epochs of 10, MAE of 0.051, and MSE of 0.004. The next normalization techniques were the median-MAD technique with MAE of 0.111 and MSE of 0.059, the mean-MAD technique with MAE of 0.264 and MSE of 0.107, the min-max technique (MM2) with MAE of 17,519 and MSE of 545.49, and finally decimal scaling technique with MAE of 22,163 and MSE of 977.9. Then, in the case of air humidity data, it can be known that Z-score technique gives good results with epochs of 7, MAE of 0.018, and MSE of 0.006. This result is almost the same as that of the mean-MAD technique with an epoch of 8 and MSE of 0.0004 which is smaller than the MSE value of the Z-score technique. The min-max technique in training rainfall data did not give good results. This was because rainfall data had different patterns in each period so the range and standard deviation values were very large, namely 401 and 49.93, respectively. However, when training air humidity data, MAE and MSE values already looked smaller compared to the normalization of rainfall data, because the range and standard deviation were 20 and 3.44, respectively. The actual data approach and prediction data on the normalization results of the Z-score technique can be seen in Figure 3 and Figure 4.

Table 2. Computational normalization techniques using MATLAB Scripts

Techniques	Normalization Script	De-normalization Script
Min-Max	Norm=0.5+0.5.*((P-mean(data))./(max(data)-min(data)));	Denorm=2.*(Z4-0.5).*(max(data)-min(data))+mean(data);
Decimal Scaling	Norm=0.01.*P;	Denorm=Z4.*100;
Z-score	Norm=(P-mean(data))/std(data);	Denorm=Z4.*std(data)+mean(data);
Mean-MAD	Norm=(P-mean(data))/mad(data);	Denorm=Z4.*mad(data)+mean(data);
Median-MAD	Norm=(P-median(data))/mad(data);	Denorm=Z4.*mad(data)+median(data);

Table 3. Result of data training based on normalization techniques

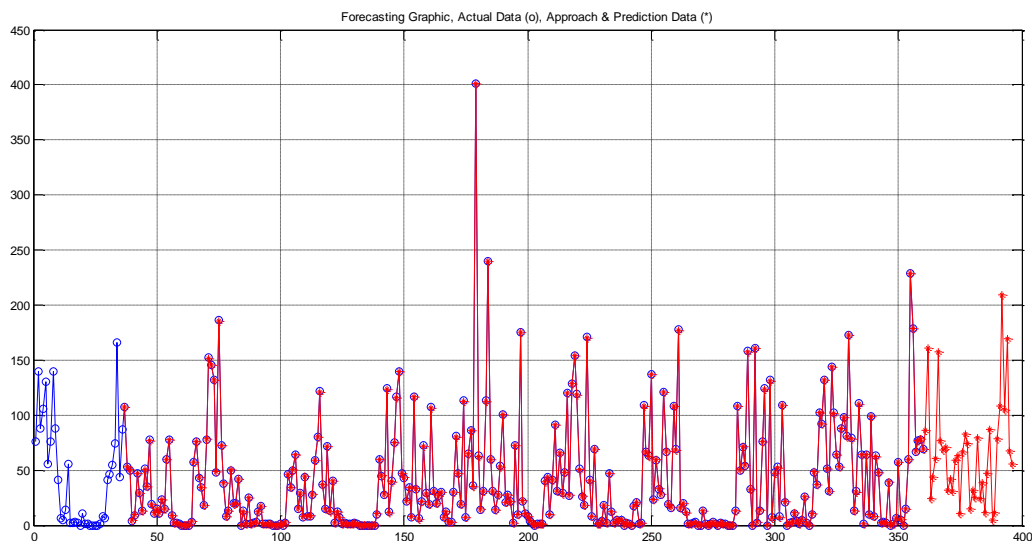| Normalization Techniques	Rainfall			Air Humidity		
	Epoch	MAE	MSE	Epoch	MAE	MSE
Min-Max	10	17.519	545.49	7	0.810	1.163
Decimal Scaling	7	22.163	977.9	3	2.495	9.256
Z-score	10	0.051	0.004	7	0.018	0.0006
Mean-MAD	9	0.264	0.107	8	0.013	0.0004
Median-MAD	8	0.111	0.059	9	0.022	0.0007

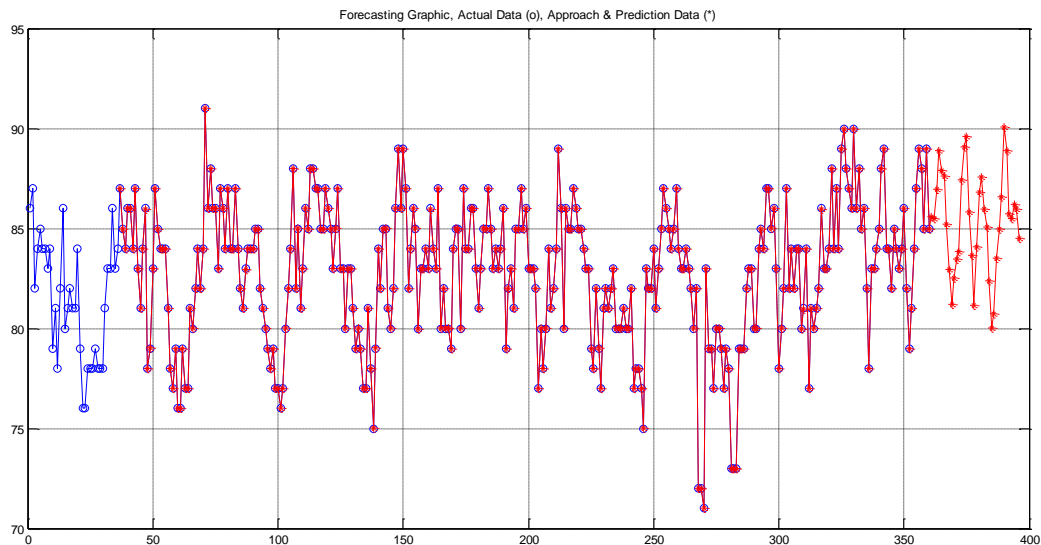**Figure 3.** Actual data approach and prediction data of rainfall data (mm)

Figure 4. Actual data approach and prediction data of air humidity (%)

Figure 3 shows that the maximum rainfall in early November was 209.51 mm, and the minimum rainfall in mid-October was 5.01 mm. Generally, the forecast results show that in 2022 the intensity of rainfall in the Amperan area of Mataram city will be higher than the previous year. Furthermore, Figure 4 shows that the highest humidity occurs in October at 90%, while the lowest humidity occurs in September at 80%. Graphically, the distribution of the predicted data is already very close to the actual data. These results show that the Z-score technique gives good data normalization results before processing the data computationally, because of the Z-score is one of the data normalization techniques that determines how far a data is from its average value in its standard deviation units. It relates the whole to the average and its standard deviation this technique is able to make actual data errors and predictions become small. This is because the Z-score is able to reduce the epoch required by the architecture when the data training process takes place [8].

4. CONCLUSIONS

Rainfall and air humidity are important components for monitoring climate change. Therefore, predictive activities are needed to regulate policy in various fields such as agriculture. Rainfall data have a seasonal pattern, meaning that the data range and standard deviation tend to be high. This is different from air humidity data which have a stationary pattern whose data range and standard deviation are small. Therefore, the results of data training using NNBP with three hidden layers obtained information that the Z-score technique is very well applied to the normalization of rainfall data with a MAE value of 0.051 and MSE of 0.004. While, it is recommended to use the mean-MAD or Z-score technique for air humidity data with MAE value at interval 0.013-0.018 and MSE at interval 0.0004-0.0006. The simulation results showed that these two techniques gave almost the same result, so it would not have a significant effect on the prediction results. These results also open opportunities for new research in the future to learn more in-depth about other modifications of the min-max technique with other numbers or variables, as well as conduct training on more types of time series data. Then compare it with other normalization techniques.

REFERENCES

[1] Gill, J., Singh, B., Singh, S. (2010). Training back propagation neural networks with genetic algorithm for weather forecasting. SIISY 2010 - 8th IEEE International Symposium on Intelligent Systems and Informatics, Subotica, Serbia, pp. 465-469. https://doi.org/10.1109/SISY.2010.5647319

[2] Passalis, N., Tefas, A., Kanniainen, J., Gabbouj, M., Iosifidis, A. (2020). Deep adaptive input normalization for time series forecasting. IEEE Transactions on Neural Networks and Learning Systems, 31(9): 3760-3765. https://doi.org/10.1109/TNNLS.2019.2944933

[3] Li, W., Liu, Z. (2011). A method of SVM with normalization in intrusion detection. Procedia Environmental Sciences, 11(PART A): 256-262. https://doi.org/10.1016/j.proenv.2011.12.040

[4] Nayak, S.C., Misra, B.B., Behera, H.S. (2014). Impact of data normalization on stock index forecasting. International Journal of Computer Information Systems and Industrial Management Applications, 6(2014): 257-269. http://www.mirlabs.org/ijcisim/volume_6.html.

[5] Akilli, A., Atil, H. (2020). Evaluation of normalization techniques on neural networks for the prediction of 305-day milk yield. Turkish Journal of Agricultural Engineering Research, 1(2): 354-367. https://doi.org/10.46592/turkager.2020.v01i02.011

[6] Khond, S.V. (2020). Effect of data normalization on accuracy and error of fault classification for an electrical distribution system. Smart Science, 8(3): 117-124. https://doi.org/10.1080/23080477.2020.1799135

[7] Bhanja, S., Das, A. (2019). Impact of data normalization on deep neural network for time series forecasting. Conference on Advancement in Computation, Communication and Electronics Paradigm, arXiv:1812.05519. https://doi.org/10.48550/arXiv.1812.05519

[8] Al-Faiz, M.Z., Ibrahim, A.A., Hadi, S.M. (2019). The effect of Z-Score standardization (normalization) on binary input due the speed of learning in back-

propagation neural network. Iraqi Journal of Information & Communications Technology, 1(3): 42-48. https://doi.org/10.31987/ijict.1.3.41

[9] Tasdemir, S., Yaniktepe, B., Guher, A.B. (2018). The effect on the wind power performance of different normalization methods by using multilayer feed-forward backpropagation neural network. International Journal of Energy Applications and Technologies, 5(3): 131-139. https://doi.org/10.31593/ijeat.464210

[10] Eesa, A.S., Arabo, W.K. (2017). A normalization methods for backpropagation: A comparative study. Science Journal of University of Zakho, 5(4): 319. https://doi.org/10.25271/2017.5.4.381

[11] Panigrahi. (2013). Effect of normalization techniques on univariate time series forecasting using evolutionary higher order neural network. International Journal of Engineering and Advanced Technology, 3(2): 280-285.

[12] Jayalakshmi, T., Santhakumaran, A. (2011). Statistical normalization and back propagationfor classification. International Journal of Computer Theory and Engineering, 3(1): 89-93. https://doi.org/10.7763/ijcte.2011.v3.288

[13] Singh, B.K., Verma, K., Thoke, A. (2015). Investigations on impact of feature normalization techniques on classifier's performance in breast tumor classification. International Journal of Computer Applications, 116(19): 11-15. https://doi.org/10.5120/20443-2793

[14] Rohman, E.M., Imamah, Rachmad, A. (2016). Predicting medicine-stocks by using multilayer perceptron backpropagation. International Journal of Science and Engineering Applications, 5(3): 188-191. https://doi.org/10.7753/IJSEA0503.1010

[15] Wang, J.Z., Wang, J.J., Zhang, Z.G., Guo, S.P. (2011). Forecasting stock indices with back propagation neural network. Expert Systems with Applications, 38(11): 14346-14355. https://doi.org/10.1016/j.eswa.2011.04.222

[16] Pradhan, B., Lee, S. (2010). Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. Environmental Modelling and Software, 25(6): 747-759. https://doi.org/10.1016/j.envsoft.2009.10.016

[17] Melesse, A.M., Ahmad, S., McClain, M.E., Wang, X., Lim, Y.H. (2011). Suspended sediment load prediction of river systems: An artificial neural network approach. Agricultural Water Management, 98(5): 855-866. https://doi.org/10.1016/j.agwat.2010.12.012

[18] Lesnussa, Y.A., Rumlawang, F.Y., Risamasu, E., Fhilya, C. (2020). Prediction of life expectancy in Maluku province using backpropagation artificial neural networks. Jurnal Matematika Integratif, 16(2): 75. https://doi.org/10.24198/jmi.v16.n2.26606.75-82

[19] Kumar, M., Raghuwanshi, N.S., Singh, R., Wallender, W. W., Pruitt, W.O. (2002). Estimating evapotranspiration using artificial neural network. Journal of Irrigation and Drainage Engineering, 128(4): 224-233. https://doi.org/10.1061/(asce)0733-9437(2002)128:4(224)

[20] Yang, Z., Liu, S., Li, Z., Jiang, Z., Dong, C. (2021). Application of machine learning techniques in operating parameters prediction of Stirling cryocooler. Cryogenics, 113: 103213. https://doi.org/10.1016/j.cryogenics.2020.103213

[21] Velasco, L.C.P., Estoperez, N.R., Jayson, R.J.R., Sabijon, C.J.T., Sayles, V.C. (2020). Performance analysis of artificial neural networks training algorithms and activation functions in day-ahead base, intermediate, and peak load forecasting. In Lecture Notes in Networks and Systems, 70: 284-298. https://doi.org/10.1007/978-3-030-12385-7_23

[22] Elgin Christo, V.R., Khanna Nehemiah, H., Minu, B., Kannan, A. (2019). Correlation-based ensemble feature selection using bioinspired algorithms and classification using backpropagation neural network. Computational and Mathematical Methods in Medicine, 2019: 7398307. https://doi.org/10.1155/2019/7398307

[23] Danpakdee, N., Songpan, W. (2017). Classification model for skin lesion image. Lecture Notes in Electrical Engineering, 424: 553-561. https://doi.org/10.1007/978-981-10-4154-9_64

[24] Gowda, C.C., Mayya, S.G. (2014). Comparison of back propagation neural network and genetic algorithm neural network for stream flow prediction. Journal of Computational Environmental Sciences, 2014: 290127. https://doi.org/10.1155/2014/290127

[25] Abhishek, K., Kumar, A., Ranjan, R., Kumar, S. (2012). A rainfall prediction model using artificial neural network. Proceedings - 2012 IEEE Control and System Graduate Research Colloquium, ICSGRC 2012, Shah Alam, Malaysia, pp. 82-87. https://doi.org/10.1109/ICSGRC.2012.6287140

[26] Ogasawara, E., Martinez, L.C., De Oliveira, D., Zimbrão, G., Pappa, G.L., Mattoso, M. (2010). Adaptive Normalization: A novel data normalization approach for non-stationary time series. Proceedings of the International Joint Conference on Neural Networks, Barcelona, Spain. https://doi.org/10.1109/IJCNN.2010.5596746

NOMENCLATURE

P	training data (Table 2)
Z4	training results data in the output layer (Table 2)
n	amount of data
x_i	actual data
x_i'	normalization data
x_{i-min}	the smallest datum of data
x_{i-max}	the largest datum of data
d	the smallest integer

Greek symbols

e	Euler=2.71828
σ	standard deviation of actual data