JTAM (Jurnal Teori dan Aplikasi Matematika)
http://journal.ummat.ac.id/index.php/jtam
p-ISSN 2597-7512 | e-ISSN 2614-1175
Vol. 6, No. 3, July 2022, pp. 629-637

Best Architecture Recommendations of ANN Backpropagation Based on Combination of Learning Rate, Momentum, and Number of Hidden Layers

Syaharuddin[1], Fatmawati[2]*, Herry Suprajitno[3]
[1,2,3]Department of Mathematics, Universitas Airlangga, Surabaya, Indonesia
[1]Department of Mathematics Education, Universitas Muhammadiyah Mataram, Mataram, Indonesia
syaharuddin-2020@fst.unair.ac.id[1], fatmawati@fst.unair.ac.id[2], herry-s@fst.unair.ac.id[3]

Article History:
Received : 25-04-2022
Revised : 24-05-2022
Accepted: 06-06-2022
Online : 16-07-2022

Keywords:
Neural Network;
Backpropagation;
Learning rate;
Momentum;
Number of Neuron;
Layer hidden.

ABSTRACT

This article discusses the results of research on the combination of learning rate values, momentum, and the number of neurons in the hidden layer of the ANN Backpropagation (ANN-BP) architecture using meta-analysis. This study aims to find out the most recommended values at each learning rate and momentum interval, namely [0.1], as well as the number of neurons in the hidden layer used during the data training process. We conducted a meta-analysis of the use of learning rate, momentum, and number of neurons in the hidden layer of ANN-BP. The eligibility data criteria of 63 data include a learning rate of 44 complete data, the momentum of 30 complete data, and the number of neurons in the hidden layer of 45 complete data. The results of the data analysis showed that the learning rate value was recommended at intervals of 0.1-0.2 with a RE model value of 0.938 (very high), the momentum at intervals of 0.7-0.9 with RE model values of 0.925 (very high), and the number of neurons in the input layer that was smaller than the number of neurons in the hidden layer with a RE model value of 0.932 (very high). This recommendation is obtained from the results of data analysis using JASP by looking at the effect size of the accuracy level of research sample data.

Crossref
https://doi.org/10.31764/jtam.v6i3.8524

———————— ◆ ————————

A. INTRODUCTION

The process of building the ANN Backpropagation architecture to obtain the results of training, testing, and good data prediction requires a process that is not easy. Because the architecture built will affect the level of accuracy of the architecture networks. The selection of learning rate and momentum values serves in the process of accelerating network performance in knowing large amounts of data (Sutskever et al., 2013). This is in accordance with Moreira & Fiesler (1995) and Yu & Chen (1997) that learning rate and momentum can reduce data training time. Thus, Hao et al. (2021) also explained that the selection of learning rate and momentum affects the speed of the training process and testing of data. Learning rate is one of the important parameters in the training process that aims to calculate the value of weight correction between layers in the architecture. This learning rate is at intervals of 0 to 1 (Smith & Topin, 2019). Meanwhile, the momentum parameter aims to update the weight so that the training process is quickly completed.

The researchers recommended learning rate and momentum that vary according to the type and pattern of data. A small learning rate value will certainly slow down the training process, and vice versa, a large learning rate (close to 1) will speed up the training process. The combination of learning rate and momentum needs to be careful because it will affect the performance speed of the data training process and has implications for the number of iterations carried out by the network (Ch & Mathur, 2012). Research on the combination of learning rate and momentum has been done a lot. Learning rates of more than 0.5 are used, for example, for climate change prediction (Syaharuddin et al., 2021), oil production prediction (Aizenberg et al., 2016), prediction of diabetes mellitus (Jayalakshmi & Santhakumaran, 2011), and card classification problem (Abdul Hamid et al., 2011). When making climate change predictions, Syaharuddin et al. (2021) used a learning rate of 0.9, so that the training process took place quickly with an MSE value of 7.48539. This value is certainly not good because the resulting error value is relatively high. This is also seen from the research conducted by Jayalakshmi & Santhakumaran (2011) when they made predictions of diabetes mellitus and obtained a coefficient of determination value of 0.726.

On the other hand, a momentum value of 0.9 was also used to predict river flow (Ghorbani et al., 2016), and classify breast tumors in ultrasound imaging (Singh et al., 2015). A momentum value of 0.8 is used to predict graduation success (Lesinski et al., 2016), thyroid disease classification problem (Rehman & Nawi, 2012), and earthquake (Moustra et al., 2011). The accuracy rate of these experiments with a momentum value of more than 0.6 reached 92.3%. In addition to the combination of learning rate and momentum, the use of hidden layers is also a necessity in the data training process so that the architecture is able to recognize data patterns and reduce errors that occur. The addition of hidden layers will certainly affect many iterations or maximum epochs that occur. However, the more epochs are, the better the data recognition rate becomes (Solanki & Jethva, 2013). Therefore, there needs to be a broader study related to the use of a combination of learning rate, momentum, and number of neurons in the hidden layer both one layer and more. The use of architecture with one hidden layer has been widely done (Zhang et al., 2020; Nawi et al., 2017; Bai et al., 2016; Gowda & Mayya, 2014). Meanwhile, architecture with two hidden layers has been used by Irawan et al. (2013) in predicting hydro-climatology data with architecture 744-100-10-1, and Singh et al. (2015) in the classification of breast tumors in ultrasound imaging with architecture 50-20-1-1.

From this explanation, it is important of conduct an in-depth analysis of the results of research that conducts experiments combining learning rate, momentum, and the number of neurons in the input layer and hidden layer. Because every research result always claims that the architecture is built in the optimal architecture with the highest degree of accuracy. Hence, the purpose of this study is to find out the level of accuracy of architecture based on the learning rate and momentum value used, knowing to comparison of the accuracy rate of architecture based on the number of neurons in the hidden layer. The results of this study are expected to be able to provide recommendations related to good parameters when conducting the process of training, testing, or predicting data.

B. METHODS

1. Architecture of ANN Backpropagation

Generally, the ANN Backpropagation architecture consists of three layers, namely the input layer, hidden layer, and output layer (Haviluddin & Alfred, 2016; Karsoliya, 2012). **Figure 1** shows that output $y_1, y_2, ..., y_k, ..., y_m$ are affected by the magnitude of the weight $v_{11}, v_{1j}, ..., v_{ij}, ..., v_{np}$ from input layer to hidden layer and weight $w_{11}, w_{j1}, ..., w_{jk}, ..., w_{pm}$ from the hidden layer to the output layer. Weight performance in each layer is determined by the learning rate value and momentum through the hidden layer $z_1, ..., z_j, ..., z_p$ to calculate together the input value by the activation function and the training function at each layer to produce an output close to the target value. In this study, there will be an analysis of the value of learning rate, momentum, and hidden layer based on the results of research that has been done, as shown in Figure 1.

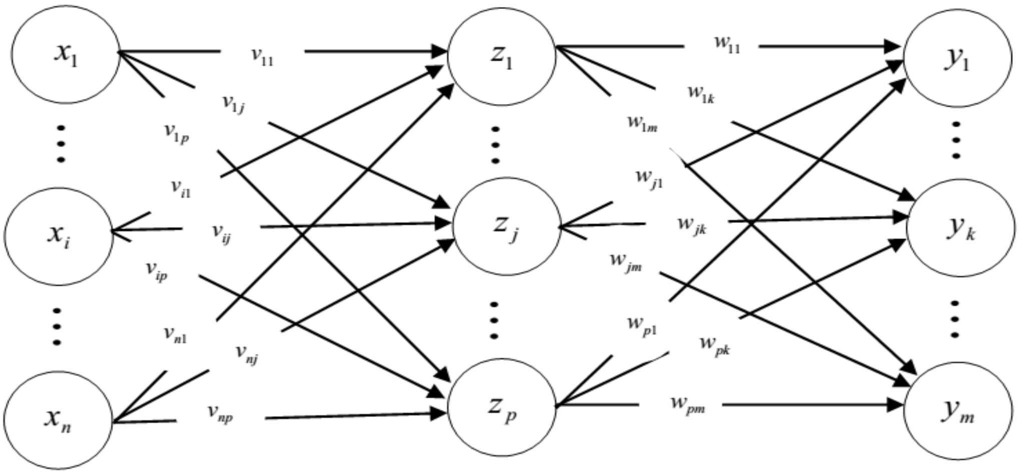
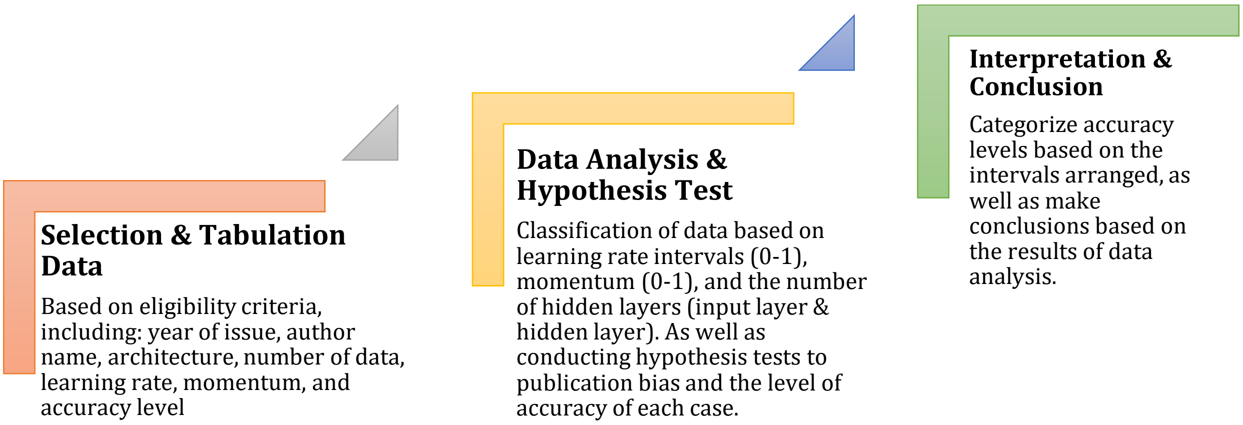


Figure 1. Backpropagation Network Architecture (Fausett, 1994)

2. Research Procedures

This study is a meta-analysis study that aims to analyze more deeply the various intervals of learning rate value, momentum, and the number of hidden layers in the ANN Backpropagation architecture recommended by the researchers. We compile and use simple research procedures according to research needs are presented in Figure 2.



Selection & Tabulation Data

Based on eligibility criteria, including: year of issue, author name, architecture, number of data, learning rate, momentum, and accuracy level

Data Analysis & Hypothesis Test

Classification of data based on learning rate intervals (0-1), momentum (0-1), and the number of hidden layers (input layer & hidden layer). As well as conducting hypothesis tests to publication bias and the level of accuracy of each case.

Interpretation & Conclusion

Categorize accuracy levels based on the intervals arranged, as well as make conclusions based on the results of data analysis.

Figure 2. Research Procedures

a. Data Selection and Tabulation

Data sources are articles published in 2011-2021 from indexing databases such as Scopus, Science Direct, and Google Scholar with forecasting topics using ANN Backpropagation. The components searched and tabulated from the filtered articles include the year of issue, author name, country, type of prediction data, architecture (input-hidden-output layer), amount of data (N), coefficient of determination value (R^2), learning rate (LR), and momentum value. The data that matches the criteria specified the value of effect size (ES) and summary effect (SE) using the equation:

$$ES_i = R_i^2 \qquad (1)$$

$$SE_i = \sqrt{\frac{ES_i(1-ES_i)}{N_i}} \qquad (2)$$

with i is a sequence of data (1,2,3..., N) and N is the amount of data in each case.

b. Hypothesis Test

This study will test two things, namely (1) publication bias from the amount of data used in this study, and (2) difference in accuracy levels based on the combination of the learning rate, momentum, and the number of hidden layers. Publication bias testing was determined by the criteria that if the p-value Rank test is greater than 0.001 (p-value > 0.001), then the data used in this study do not indicated bias. In addition, it can also be determined by the Rosemthal (1979) namely: $5k + 10 < N_R$, with k is the amount of data and N_R is the value of File-Safe N. Furthermore, the coefficient of determination value that correlates with the value of random effect (RE) was categorized according to intervals, namely very weak (0.00-0.199), weak (0.20-0.399), moderately high (0.40-0.599), high (0.60-0.799), and very high (0.80-1.00).

C. RESULT AND DISCUSSION

1. Data Selection Results

Search on the indexing database namely Scopus, ScienceDirect, and Google Scholar found as many as 79 data. From this amount of data, a re-examination was carried out according to the eligibility criteria (Figure 2) so that 45 data that met the criteria of architecture of the number of hidden layers, 44 data that met the learning rate criteria, and 30 data that met the momentum criteria were obtained. More complete details of the data can be seen in Figure 3.

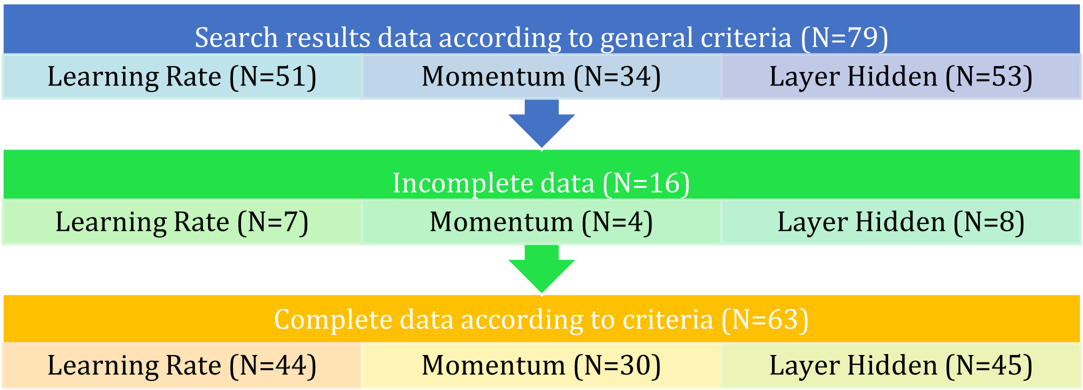


Figure 3. Data collection results

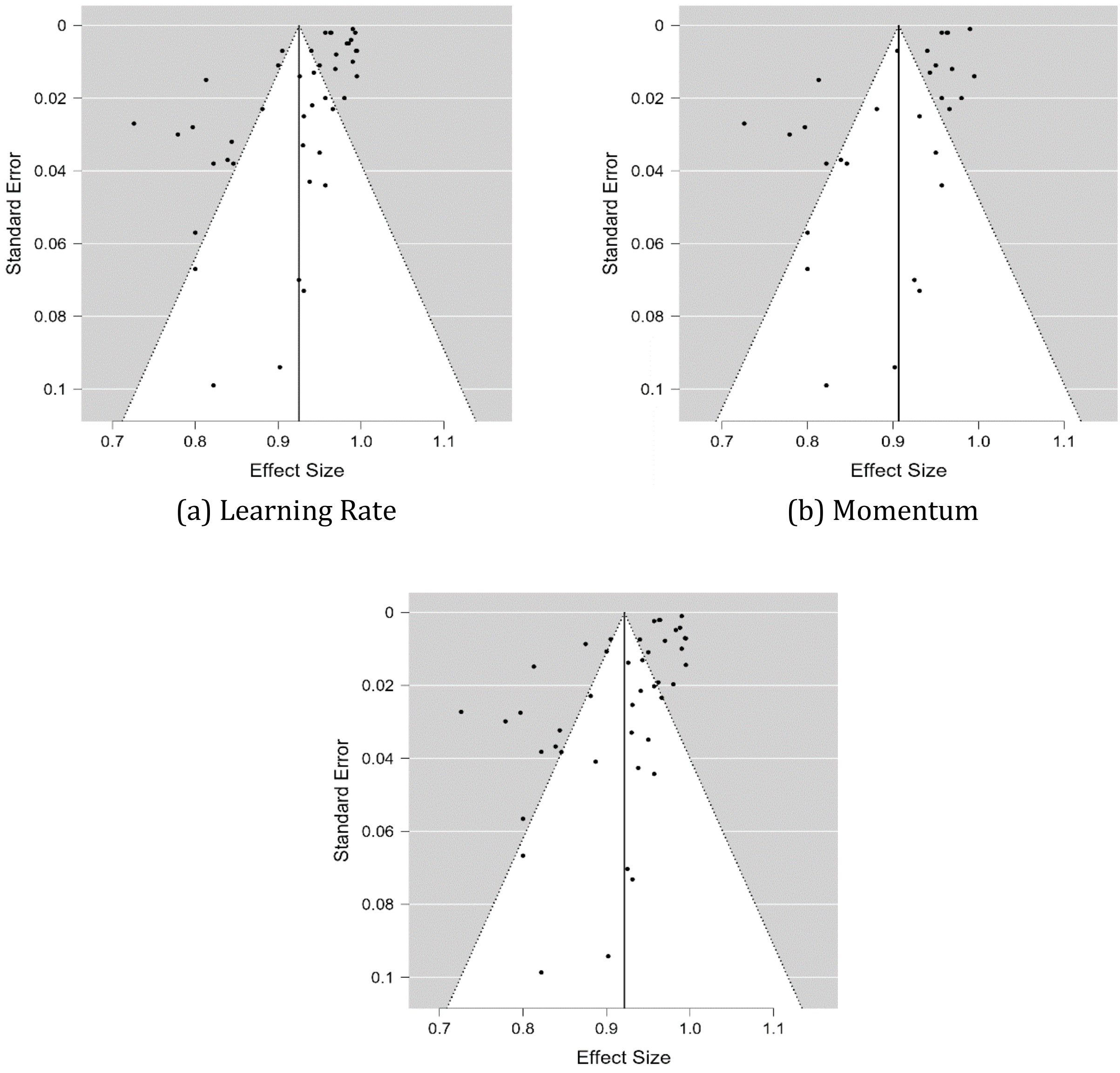
Incomplete data is due to the absence of the amount of data (N) and the coefficient of determination (R2) that does not exist. Incomplete data is not used in the next stage, namely the stage of converting accuracy rate into *ES* and *SE* values. Furthermore, the selection results data is tabulated based on their respective test criteria. The complete learning rate (LR) data is divided into five intervals, namely LR < 0.1; 0.1–0.2; 0.3–0.4; 0.5–0.6; > 0.6. The complete momentum data is divided into three intervals, namely 0.1–0.3; 0.4–0.6; 0.7–0.9. Finally, the number of hidden layers is divided into two types, namely the number of input layers that is greater than the number of hidden layers (LI > LH) and the number of input layers that is smaller than the number of hidden layers (LI < LH).

2. Publication Bias Test (H1)

In this study, researchers used JASP software to perform data analysis. The publication bias test was carried out to see the adequacy of the amount of data used so that the results of the study can be generalized. The JASP output as shown in Table 1 and Figure 4.

Table 1. Publication Bias Test

Parameter	N	Q_{RE}	I^2 (%)	p-Rank Test	Fail-safe N	RE Model
Learning Rate	44	1060.780	99.409	0.491	1098e+7	0.93 [0.90, 0.95]
Momentum	30	878.416	99.501	0.198	4636e+6	0.91 [0.88, 0.93]
Layer Hidden	45	1095.384	99.180	0.822	8605e+6	0.92 [0.90, 0.94]



(a) Learning Rate

(b) Momentum

(c) Layer Hidden

Figure 4. Funnel plot data for learning rate, momentum, and layer hidden

Table 1 shows that in general, ANN Backpropagation architectures that use learning rate parameters during the data training process are able to increase the accuracy or performance of the use of architecture by 93%, the use of momentum by 91%, and the use of hidden layers up to 92%. This shows that the use of learning rate, momentum, and hidden layers, both one layer and two layers, is able to improve the performance of the architecture well. These results are in accordance with the results of previous studies. Furthermore, based on the p-Rank Test value on each data, a learning rate of 0.491 > 0.001, a momentum of 0.198 > 0.001, and the number of hidden layers of 0.822 > 0.001 were obtained. So it can be said that the data used in this study is sufficient according to minimum standards and can be generalized. Figure 4 also shows that the pattern of data distribution is good and there is no evidence of publication bias.

3. Accuracy Level of Each Parameter (H2)

At this stage, the researcher divided the data based on the value intervals of each parameter. This was done to see the accuracy level of the architecture according to the learning rate, momentum, and number of neurons in the hidden layer recommended by other researchers. As for the results of data analysis, they are presented in Table 2.

Table 2. Data analysis results

Parameters	Interval	N	RE Model	Estimate Coefficient	Standard Error
Learning Rate	< 0,1	9	0.94 [0.90, 0.98]	0.937	0.021
	0,1 – 0,2	12	0.94 [0.91, 0.96]	0.938	0.013
	0,3 – 0,4	8	089 [0.83, 0.96]	0.892	0.034
	0,5 – 0,6	10	0.93 [0.90, 0.97]	0.932	0.018
	> 0,6	5	0.92 [0.82, 1.02]	0.919	0.049
Momentum	0,1 – 0,3	9	0.88 [0.82, 0.94]	0.882	0.030
	0,4 – 0,6	9	0.91 [0.88, 0.93]	0.905	0.014
	0,7 – 0,9	12	0.93 [0.88, 0.97]	0.925	0.023
Number of Neurons in the Hidden Layer	LI > LH	33	0.92 [0.89, 0.94]	0.917	0.013
	LI < LH	12	0.93 [0.85, 0.96]	0.932	0.017

Table 2 shows that the learning rate at intervals of 0.1-0.2 provides a higher degree of accuracy than the learning rate at intervals smaller than 0.1 and greater than 0.2 with an estimated coefficient of correlation value of 0.938 (very high category). These results show that the greater the learning rate value used, the data training process will run faster, but the accuracy of the architectural network will be reduced. Conversely, if the learning rate used is getting smaller, then the accuracy of the network will be greater or increased, but consequently, the training process will take longer. The momentum value at intervals of 0.7-0.9 provides the highest degree of accuracy compared to momentum values at intervals smaller than 0.7 with an estimate of the coefficient of correlation value of 0.925 (very high category). This result is in accordance with the research of Mislan et al. (2015) when they conducted rainfall predictions with a hidden layer architecture of 50-20-1 and obtained by MSE of 0.00096341. Furthermore, the results of the use of the number of neurons in the hidden layer obtained the result of an estimate of the coefficient of correlation of 0.932 for the category of the number of neurons in the input layer that is smaller than the number of neurons in the hidden layer. This is in accordance with the results of research from Hamid et al. (2011) when they conducted the introduction of fragments of glass encountered in forensic work. So, the use of a learning rate of 0.1, the momentum of 0.9, and the number of hidden layer neurons (LH) greater than the input layer (LI) can improve architectural performance by 88%-99% (Sun & Huang, 2020), (Baldi et al., 2018), (Tarigan et al., 2017), (Lesinski et al., 2016), (Mislan et al., 2015).

D. CONCLUSION AND SUGGESTIONS

A good architecture cannot be separated from the optimal combination of parameters. Therefore, researchers continue to experiment with the combination of learning rate, momentum, and number of hidden layers to find a reliable architecture for various cases. The results of the data analysis showed that the selection of learning rates at intervals of 0.1-0.2 and momentum greater than 0.7 was able to provide a faster training process and high accuracy. Furthermore, the authors recommend each architecture using a hidden layer, where the number of neurons in the input layer is greater than the number of neurons in the hidden layer. The use of the number of hidden layers of more than one also has a good impact, because it is able to recognize data patterns better when the data training process.

REFERENCES

Abdul Hamid, N., Mohd Nawi, N., Ghazali, R., & Mohd Salleh, M. N. (2011). Accelerating learning performance of back propagation algorithm by using adaptive gain together with adaptive momentum and adaptive learning rate on classification problems. *Communications in Computer and Information Science*, *151(*2), 559–570. https://doi.org/10.1007/978-3-642-20998-7_62

Aizenberg, I., Sheremetov, L., Villa-Vargas, L., & Martinez-Muñoz, J. (2016). Multilayer Neural Network with Multi-Valued Neurons in time series forecasting of oil production. *Neurocomputing*, *175*, 980–989. https://doi.org/10.1016/j.neucom.2015.06.092

Bai, Y., Li, Y., Wang, X., Xie, J., & Li, C. (2016). Air pollutants concentrations forecasting using back propagation neural network based on wavelet decomposition with meteorological conditions. *Atmospheric Pollution Research*, *7*(3), 557–566. https://doi.org/10.1016/j.apr.2016.01.004

Baldi, P., Sadowski, P., & Lu, Z. (2018). Learning in the machine: Random backpropagation and the deep learning channel. *Artificial Intelligence*, *260*, 1–35. https://doi.org/10.1016/j.artint.2018.03.003

Ch, S., & Mathur, S. (2012). Particle swarm optimization trained neural network for aquifer parameter estimation. *KSCE Journal of Civil Engineering*, *16*(3), 298–307. https://doi.org/10.1007/s12205-012-1452-5

Fausett, L. (1994). *Fundamentals of Neural Network*. Prentice Hall, Hoboken.

Ghorbani, M. A., Zadeh, H. A., Isazadeh, M., & Terzi, O. (2016). A comparative study of artificial neural network (MLP, RBF) and support vector machine models for river flow prediction. *Environmental Earth Sciences*, *75*(6). https://doi.org/10.1007/s12665-015-5096-x

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*, 1–6. https://doi.org/10.1155/2014/290127

Hao, Z., Jiang, Y., Yu, H., & Chiang, H. D. (2021). Adaptive Learning Rate and Momentum for Training Deep Neural Networks. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12977 LNAI*, 381–396. https://doi.org/10.1007/978-3-030-86523-8_23

Haviluddin, & Alfred, R. (2016). A genetic-based backpropagation neural network for forecasting in time-series data. *Proceedings - 2015 International Conference on Science in Information Technology: Big Data Spectrum for Future Information Economy*, 158–163. https://doi.org/10.1109/ICSITech.2015.7407796

Irawan, M. I., Syaharuddin, Utomo, D. B., & Rukmi, A. M. (2013). Intelligent irrigation water requirement system based on artificial neural networks and profit optimization for planting time decision making of crops in Lombok Island. *Journal of Theoretical and Applied Information Technology*, *58*(3), 657–671.

Jayalakshmi, T., & Santhakumaran, A. (2011). Statistical Normalization and Back Propagationfor Classification. *International Journal of Computer Theory and Engineering*, 89–93. https://doi.org/10.7763/ijcte.2011.v3.288

Karsoliya, S. (2012). Approximating Number of Hidden layer neurons in Multiple Hidden Layer BPNN Architecture. *International Journal of Engineering Trends and Technology*, *3*(6), 714–717.

Lesinski, G., Corns, S., & Dagli, C. (2016). Application of an Artificial Neural Network to Predict Graduation Success at the United States Military Academy. *Procedia Computer Science*, *95*, 375–382. https://doi.org/10.1016/j.procs.2016.09.348

Mislan, Haviluddin, Hardwinarto, S., Sumaryono, & Aipassa, M. (2015). Rainfall Monthly Prediction Based on Artificial Neural Network: A Case Study in Tenggarong Station, East Kalimantan - Indonesia. *Procedia Computer Science*, *59*, 142–151. https://doi.org/10.1016/j.procs.2015.07.528

Moreira, M., & Fiesler, E. (1995). Neural Networks with Adaptive Learning Rate and Momentum Terms. *Technique Report 95*, *4*, 1–29.

Moustra, M., Avraamides, M., & Christodoulou, C. (2011). Artificial neural networks for earthquake prediction using time series magnitude data or Seismic Electric Signals. *Expert Systems with Applications*, *38*(12), 15032–15039. https://doi.org/10.1016/j.eswa.2011.05.043

Nawi, N. M., Hamzah, F., Hamid, N. A., Rehman, M. Z., Aamir, M., & Ramli, A. A. (2017). An optimized back propagation learning algorithm with adaptive learning rate. *International Journal on Advanced Science, Engineering and Information Technology*, *7*(5), 1693–1700. https://doi.org/10.18517/ijaseit.7.5.2972

Rehman, M. Z., & Nawi, N. M. (2012). Studying The Effect Of Adaptive Momentum In Improving The Accuracy Of Gradient Descent Back Propagation Algorithm On Classification Problems. *International Journal of Modern Physics: Conference Series*, *09*, 432–439. https://doi.org/10.1142/s201019451200551x

Singh, B. K., Verma, K., & Thoke, A. S. (2015). Adaptive gradient descent backpropagation for classification of breast tumors in ultrasound imaging. *Procedia Computer Science*, *46*, 1601–1609. https://doi.org/10.1016/j.procs.2015.02.091

Smith, L. N., & Topin, N. (2019). Super-convergence: Very fast training of neural networks using large learning rates. *Proceeding of Artificial Intelligence and Machine Learning for Multi-Domain Operations*, 1–18. https://doi.org/https://doi.org/10.1117/12.2520589

Solanki, S., & Jethva, H. B. (2013). Modified Back Propagation Algorithm of Feed Forward Networks. *International Journal of Innovative Technology and Exploring Engineering (IJITEE), 2*(6), 131-134.

Sun, W., & Huang, C. (2020). A carbon price prediction model based on secondary decomposition algorithm and optimized back propagation neural network. *Journal of Cleaner Production*, *243*. https://doi.org/10.1016/j.jclepro.2019.118671

Sutskever, I., Martens, J., Dahl, G., & Hinton, G. (2013). On the importance of initialization and momentum in deep learning. *30th International Conference on Machine Learning, ICML 2013*, 2176–2184.

Syaharuddin, Sari, D. A., Sabaryati, J., Zonyfar, C., Sihotang, S. F., Fadillah, A., Sari, T. H. N. I., Putra, D. S., Harun, R. R., & Mandailina, V. (2021). Computational based on GUI MATLAB for back propagation method in detecting climate change: Case study of mataram city. *Journal of Physics: Conference Series*, *1816*(1). https://doi.org/10.1088/1742-6596/1816/1/012001

Tarigan, J., Nadia, Diedan, R., & Suryana, Y. (2017). Plate Recognition Using Backpropagation Neural Network and Genetic Algorithm. *Procedia Computer Science*, *116*, 365–372. https://doi.org/10.1016/j.procs.2017.10.068

Yu, X. H., & Chen, G. A. (1997). Efficient backpropagation learning using optimal learning rate and momentum. *Neural Networks*, *10*(3), 517–527. https://doi.org/10.1016/S0893-6080(96)00102-5

Zhang, Z., Yang, P., Ren, X., Su, Q., & Sun, X. (2020). Memorized sparse backpropagation. *Neurocomputing*, *415*, 397–407. https://doi.org/10.1016/j.neucom.2020.08.055