Rainfall prediction using artificial neural network with historical weather data as supporting parameters

Awang Hendrianto Pratomo [a,1,*], Budi Santosa [a,2], Sylvert Prian Tahalea [a,3], Eko Teguh Paripurno [a,4], Johan Ddanu Peasetyo [a,5], Herlina Jayadianti [a,6], Muhammad Fathurrahman Pitayandanu [a,7]

[a] Informatika, Universitas Pembangunan Nasional Veteran Yogyakarta, Indonesia
[1] awang@upnyk.ac.id*; [2] dissan@upnyk.ac.id; [3] sylvert@upnyk.ac.id; [4] paripurno@upnyk.ac.id; [5] johandanu@upnyk.ac.id; [6] herlinajayadianti@gmail.com; [7] fathurdanu@gmail.com;
* Corresponding Author

Received 25 January 2022; accepted 8 April 2022; published 15 May 2022

ABSTRACT

Changing climatic patterns are caused by changes in variables, such as rainfall and air temperature that occur continuously in the long term. Rainfall itself is influenced by several weather factors such as air humidity, wind speed, air pressure, and temperature. This study experimented to test a combination of 9 additional weather parameters such as dew point, wind gusts, cloud cover, humidity, rainfall, air pressure, air temperature, wind direction, and wind speed to predict daily rainfall for one year using the main parameters of the rainfall time series. Prediction is done using Artificial Neural Network (ANN). The ANN architecture used is to use 3 to 11 input parameters, 1 hidden layer totaling 60 neurons with the ReLu activation function, and 1 neuron in the output layer without an activation function. ANN without additional weather parameters obtained an MSE of 0.01654, while prediction using additional weather parameters obtained an MSE of 0.00884. So the combination of rainfall time series parameters with additional weather parameters is proven to provide a smaller MSE value

Check for updates

KEYWORDS
ANN
Rainfall forecasting
ReLU activation function

1. Introduction

Climate change occurs due to changes in variables such as changes in temperature and rainfall continuously in the long term [1]. Climatechange is also caused by unstable weather, such as drastic changes in wind direction, extreme temperatures, frequent storms, and erratic rainfall [2]. The uncertainty of the intensity of this rainy rainfall can be seen in the rainy season where the intensity is always changing and unstable. Rainfall is influenced by several weather factors such as wind speed, air humidity, temperature, and air pressure as evidenced by multiple linear regression [3]. Therefore, it is necessary to predict rainfall with reference to weather data calculations .

There are several methods for conducting forecasting. The majority of predictive research is conducted using artificial intelligence (AI). Several AI models were compared to predict rainfall, the MSE Fast Fourier Transform (FFT) was 14.92, and the MSE Autoregressive Integrated Moving Average (ARIMA) was 17.49. Artificial Neural Network (ANN) MSE value is 0.07. The conclusion of this study is that the MSE value produced by ANN is the smallest MSE [4].

ANN was able to provide better daily forecasting results in terms of identifying rainfall fluctuation patterns and successfully obtaining them using four input parameters (air temperature, humidity, pressure and evaporation with seven hidden layers) obtaining MSE 0. 16 with *hidden layer* architecture [35 (tansig) -30 (tansig) -25 (Sigmoid log) -20 (tansig) -15 (tansig) -10 (tansig) -5 (Sigmoid log)], and using linear activation function on layer output [5]. Then in Mislan's study, rainfall parameters two months and one month before the event were used as input parameters, and MSE resulted from two hidden layers (50-20) and neuron output layers of 0.00096341 to predict monthly rainfall [6]. Velasco also uses 11 input parameters for weekly precipitation forecasts, namely average temperature, minimum temperature,

Jurnal Informatika
Vol. 16, No. 2, May 2022, pp. 63-73

maximum temperature, average wind speed, maximum wind speed, relative humidity, total rainfall/snow melting, visibility, date, bulan and year. The architecture used in the hidden layer [50 (tanh)] and one neuron output layer. The OBTAINED MSE value is 0.0001682209 [7].

Through the previous researches, rainfall time series parameters has a great influence on predictions, and they also reported a small MSE. Therefore, using the main parameters of the precipitation time series and weather parameters as supporting parameters, it is estimated that the MSE value is lower on the rainfall prediction. However, because not all weather parameters have a significant influence on rainfall, with the help of the correlation coefficient formula, we can see the effect of each of these parameters on the results of rainfall prediction. This gave rise to the idea of using the ANN and the weather parameters to be studied (i.e. dew point temperature, gusts, cloud cover, humidity, pressure, temperature, wind direction and wind speed) to find the best combination of precipitation prediction parameters.

2. Method

This research consists of several stages in a row such as data collection, data preprocessing, architectural testing, and parameter testing. These stages can be depicted as in **Fig. 1**

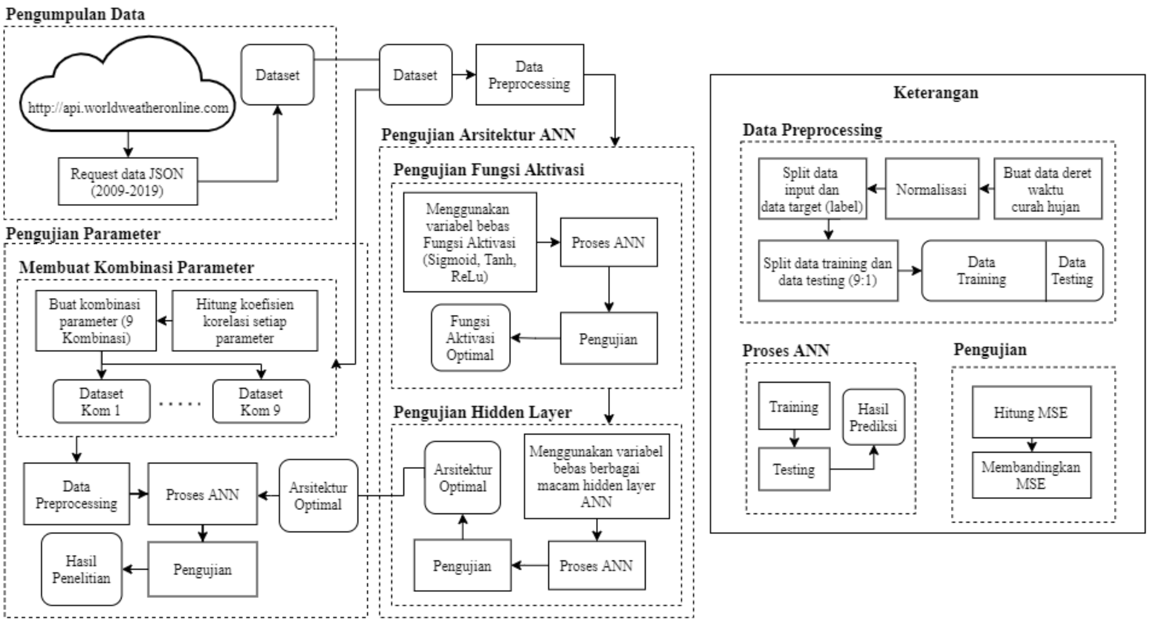

Fig 1. Research Methodology

2.1. Data Collection

This research began with data collection. The data used is secondary data about weather data obtained from World Weather Online by means of an API request. The parameters used in the study were dew point temperature, wind gusts, cloud cover, humidity, precipitation, pressure, temperature, wind direction, and wind speed from 2009 to 2019. The location studied is Caturtunggal Village.

2.2. Creating Parameter Combinations

Combinations are made based on the value of the correlation coefficient. The sequence of combinations starts from the most influential (far from zero) to the least influential (close to zero). A positive value indicates that the variable x is directly proportional to the varibel y. Conversely, a negative value indicates that the variable y is inversely proportional to the variable y. Equation The correlation coefficient is written in Eqaution 1.

$$r = \frac{(n\sum_{i=1}^{n} X_i Y_i) - (\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i)}{\sqrt{n\sum_{i=1}^{n} X_i^2 - (\sum_{i=1}^{n} X_i)^2}\sqrt{n\sum_{i=1}^{n} Y_i^2 - (\sum_{i=1}^{n} Y_i)^2}} \qquad (1)$$

2.3. Data Prepocessing

The preprocessing stage in this study consists of four steps. The preprocessing step sequentially is to add input parameters for the rainfall time series, normalize data, splitting input parameters with target data, splitting training data and testing data. The preprocessing flowchart can be seen in **Fig. 2**

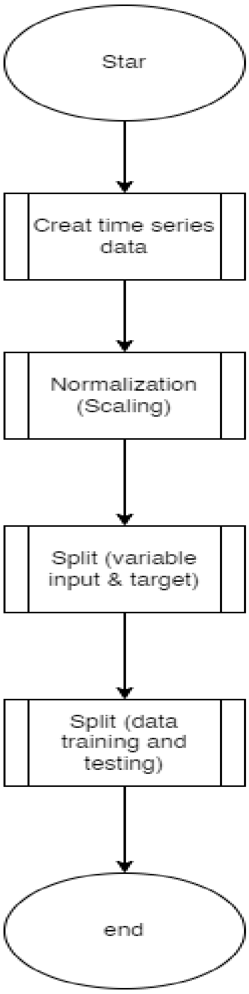

Fig 2. Preprocessing Data Flowchart

The time series in question is rainfall data that is in order based on the time of the event. This study will apply a rainfall time series for 3 days. So that there are new input parameters created, such as rainfall data three days before the event with the symbol t-3, rainfall two days before the event with the symbol t-2, and precipitation the day before the event with the symbol t-1.

The data normalization stage is carried out with the Minmax Scaler, whose function is to equalize the scale with the interval [0,1]. Minmax Scaler is inscribed into Equation 2.

$$\bar{x}_{new} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (2)$$

2.4. Artificial Neural Network

Artificial Neural Network (ANN) is the processing of information systems that are based on how human neural networks work. ANN can be implemented in various fields in human life [8]. The ANN training algorithm can be seen in **Fig. 3**

Jurnal Informatika
Vol. 16, No. 2, May 2022, pp. 63-73

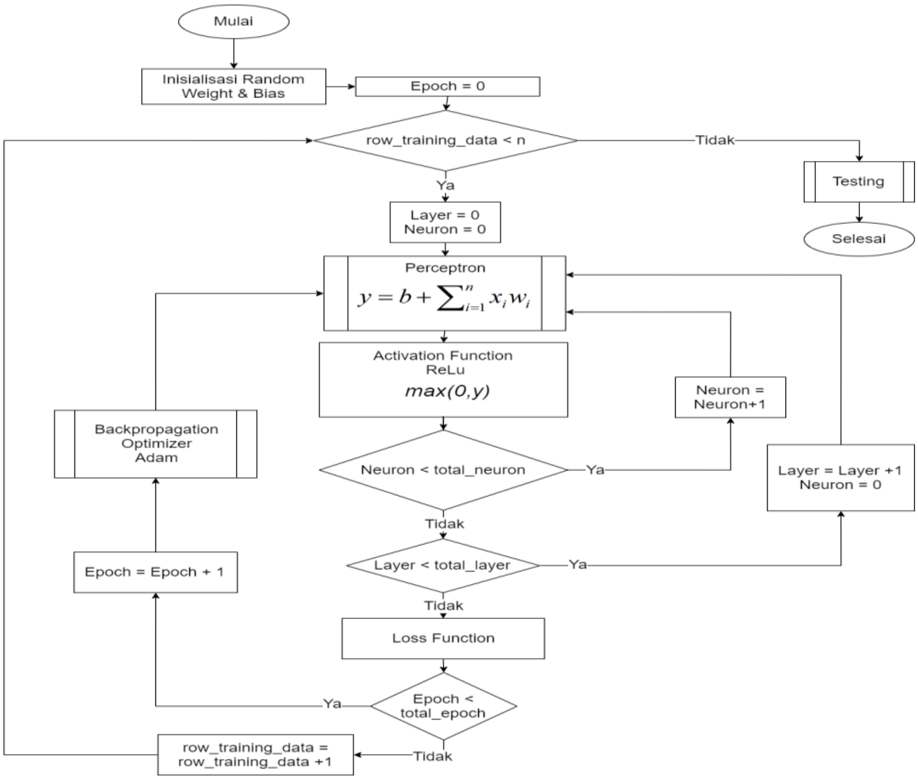

Fig 3. ANN Training scenario

Unlike an ANN training flowchart, ANN testing does not use an optimizer for weight and bias improvement [9]. ANN testing only runs once using weight and bias obtained from training. Flowchart testing can be seen in **Fig. 4**

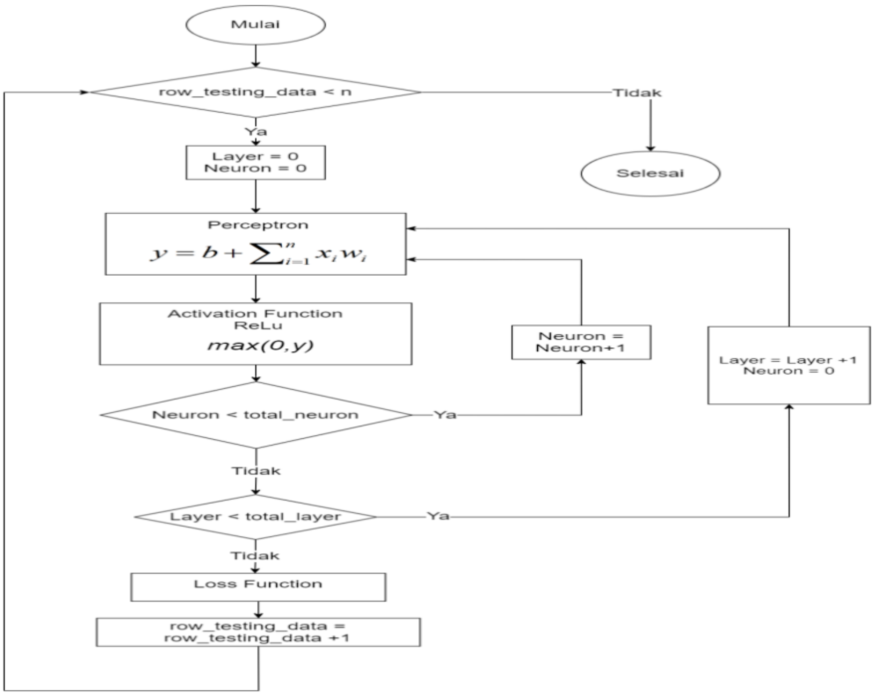

Fig 4. ANN Testing scenario

The ANN structure consists of at least two layers, namely the input layer and the output layer [10]. While a hidden layer can be added between the input layer and the output layer. Each layer consists of at least one neuron. When data passes through a neuron, then the perceptron shown in Equation 3 will be active.

$$y = b + \sum_{i=1}^{n} x_i w_i \tag{3}$$

Neurons on the hidden layer are activated using ReLu. ReLu activation will generate a value with an interval of [0.1]. The ReLu activation function is written in Equation 4.

$$g(x) = max\ (0, x) \tag{4}$$

The ANN architecture that was tested for performance before being used in this study was nine input parameters, 60 hidden layers with ReLu activation function, and one neuron output without activation function [11]. This architecture can be described as in **Fig. 5**.

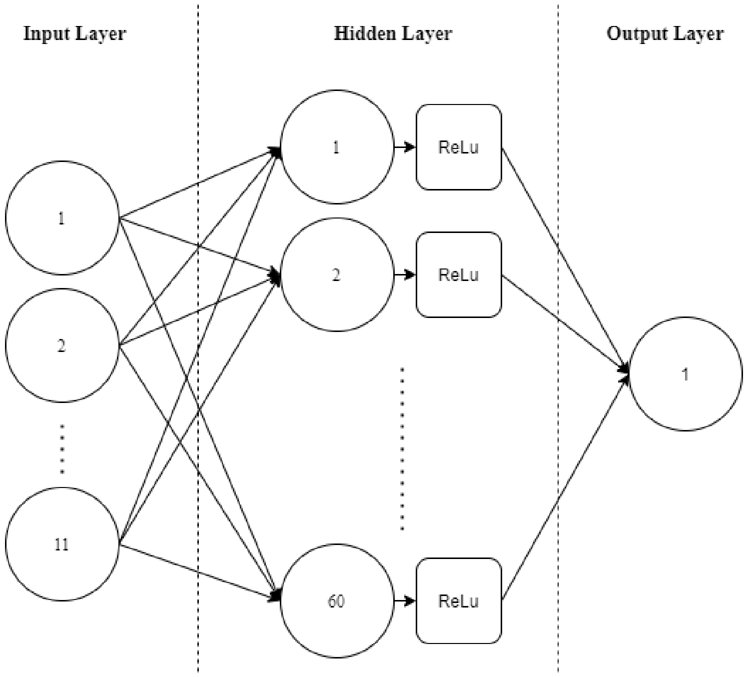

Fig 5. ANN Architecture

2. 5. Mean Squared Error

Mean Square Error (MSE) is used to test the accuracy of prediction results [12]. The smaller the MSE value, the more accurate the prediction results will be [13]. The MSE is the average of the squared error values. MSE is formulated in the following Equation 5.

$$MSE = \frac{\sum_{i=1}^{n}(Y_i - \overline{Y}_i)^2}{n} \tag{5}$$

2.6. Architecture Testing

Architectural testing is divided into two parts [14]. The first part is to test the activation function that is most suitable for precipitation prediction cases [15]. The activation functions studied are Sigmoid, Tanh, ReLu, and without activation function. The activation function with the smallest MSE Value will be used in the next step [16].

The second part is to find a hidden layer. The stage will conduct testing for a wide variety of architectures ranging from 1 hidden layer to 3 layers of multiples of 10 neurons with a limit of 110 each neuron. Best results are characterized by the lowest MSE

Jurnal Informatika
Vol. 16, No. 2, May 2022, pp. 63-73

2.7. Parameter Testing

The purpose of this test is to find out the combination of parameters that give the most optimal results. Parameter combination testing is performed by comparing the MSE values of each combination result [17].

3. Results and Discussion

Testing of the activation function by comparing mse results between sigmoid, hyperbolic tangent, and ReLu obtained with the results as in **Table 1**. The input parameter data used are eleven. The number of neurons in the *hidden layer* used as a fixed variable is 50 neurons. The most optimal results obtained were 0.00896 and R-Square of 0.59269, namely using the ReLu activation function on the hidden layer only, but the prediction results still have negative values where rainfall cannot be negative. Thus, the architecture with the second lowest MSE value with a value of 0.00897 and an R-Square of 0.59217 will be used because all the predicted results are positive. This is caused by the ReLu activation function on the output layer where the resulting range is zero to one [0, x].

Table 1. Activation Function Test Results

No.	Hid Layer	Act Func (Hid)	Act Func (Out)	MSE	R-Square
1	50	-	-	0.01052	0.52158
2	50	Sigmoid	-	0.01614	0.26608
3	50	-	Sigmoid	0.01462	0.33540
4	50	Sigmoid	Sigmoid	0.01544	0.29797
5	50	tanh	-	0.01051	0.52230
6	50	-	tanh	0.01064	0.51631
7	50	tanh	tanh	0.01062	0.51072
8	**50**	**ReLu**	**-**	**0.00896**	**0.59269**
9	50	-	ReLu	0.00935	0.57490
10	**50**	**ReLu**	**ReLu**	**0.00897**	**0.59217**
11	50	Sigmoid	tanh	0.01593	0.27563
12	50	Sigmoid	ReLu	0.01627	0.26035
13	50	tanh	Sigmoid	0.01441	0.34492
14	50	tanh	ReLu	0.00935	0.57492
15	50	ReLu	Sigmoid	0.02248	-0.02205
16	50	ReLu	tanh	0.00908	0.58731

The results of thecorrelation of each parameter are listed in **Table 2** in order from those that have the most effect on rainfall. The supporting parameters with the highest influence to the lowest in order are dew point temperature, clouds, wind speed, air pressure, wind gusts, wind direction, and temperature.

Table 2. Sequence of Correlation Coefficients from the Most Influential

	Main Parameters			Supporting Parameters							
	t-1 (mm)	t-2 (mm)	t-3 (mm)	Dew point temperature (°C)	Cloud Cover (%)	Wind Speed (Kmph)	Humidity	Air Pressure	Gusts of Wind (Kmph)	Wind Direction	Temperature (°C)
Correlation Coefficient	0.560697	0.380663	0.324084	0.472924	0.395159	-0.32088	0.318868	-0.27842	-0.21799	0.114121	0.011389

Hidden layer neuron testing has been carried out. The experiment was carried out using 11 input parameters and using the ReLu activation function on the hidden layer and output layer. Neurons in the hidden layer were tested for up to three hidden layers of multiples of 10 to 110. Based on the results, it is known that the architecture of one hidden layer consisting of 60 neurons using the ReLu activation function of 100 epochs obtained the smallest MSE with an MSE value of 0.008806448 and an R-Square of 0.599657059. A summary of the results of this experiment can be seen in **Table 3**.

Table 3. Hidden Layer Test Results

No	Neuron Layer 1	Neuron Layer 2	Neuron Layer 3	time(s)	ActivationFunction	MSE	R-Square
1	60	0	0	18.59481049	relu	0,008806448	0.599657059
2	110	0	0	20.90722799	relu	0,008892794	0.595731735
3	100	0	0	21.15185452	relu	0,008933042	0.593902051
4	90	0	0	20.93208194	relu	0,008938864	0.593637347
5	70	0	0	18.69234419	relu	0,008940457	0.593564987
6	50	0	0	21.26204944	relu	0,008971903	0.59213537
7	40	0	0	18.57602715	relu	0,008993756	0.591141939
8	80	0	0	18.61658835	relu	0,009011503	0.590335131
9	30	0	0	18.49914312	relu	0,009113772	0.585686028
10	20	0	0	19.03775978	relu	0,009306462	0.576926231
11	70	10	0	20.54806709	relu	0,00938798	0.573220432
12	80	10	0	20.63078928	relu	0,009390667	0.573098302
13	60	10	0	20.73282552	relu	0,009391815	0.573046088

By using an ANN Architecture one hidden layer totaling 60, thetraining performance can be seen in **Fig. 6** the y-axis is of the error value, while the x-axis is the epoch priode. Training errors are shown in blue and predictions/testing errors are shown in orange [18]. Pthere is a 100th epoch, the image shows optimal performance, where the training and testing performance intersect each other.

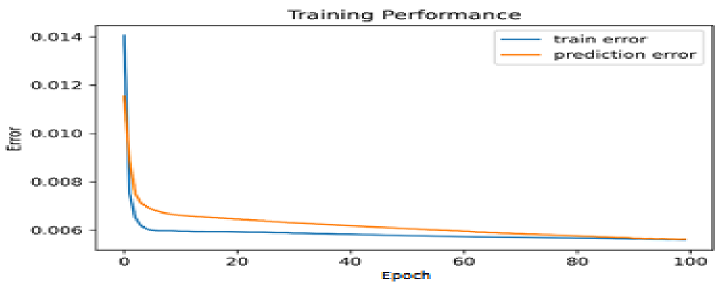

Fig 6. Performance Training

The results of the parameter testing are written in **Table 4**. The training data in this test uses data from 2009 to 2018. TheD ata testing used is 2019. In combination1, the parameter used is a three-day time series only as the main parameter with an MSE result of 0.01654. Gradually, the study added weather input parameters one by one such as dew point temperature, clouds, wind speed, air pressure, wind gusts, wind direction, and temperature. With the addition of weather parameters, the MSE value

Jurnal Informatika
Vol. 16, No. 2, May 2022, pp. 63-73

of each test tends to decrease further. The smallest MSE value is successfully obtained when Combination 9 is used [19]. This combination of 9 uses the main parameters of the time series for three days and eight additional parameters, namely dew point temperature, cloud cover, humidity, wind direction, and berhasi temperature getting an MSE of 0.00884.

Table 4. Parameter Testing

	Main Parameters	Supporting Parameters								MSE
Combination	**Rainfall Time Series**	**Dew (ºC)**	**Cloud (%)**	**Wind Speed (Kmph)**	**Humidity**	**Air Pressure**	**Gusts of Wind (Kmph)**	**Wind Direction**	**Temperature (ºC)**	
1	✓									0.01654
2	✓	✓								0.01604
3	✓	✓	✓							0.01689
4	✓	✓	✓	✓						0.01606
5	✓	✓	✓	✓	✓					0.01479
6	✓	✓	✓	✓	✓	✓				0.01502
7	✓	✓	✓	✓	✓	✓	✓			0.01507
8	✓	✓	✓	✓	✓	✓	✓	✓		0.00896
9	✓	✓	✓	✓	✓	✓	✓	✓	✓	**0.00884**

The results of the predictions can be seen in **Table 5**. The table lists the columns of date, actual rainfall, predicted rainfall, and the absolute difference between actual rainfall and predicted results so that differences in millimeters can be seen [20]. The average absolute difference of all predicted results for one year was obtained at 3.312573 millimeters.

Table 5. Prediction Results (millimeters)

Date	Actual	Predictions	Absolute Difference
01-01-2019	0.3	2.93	2.629
02-01-2019	3.7	5.85	2.149
03-01-2019	2.3	4.70	2.397
04-01-2019	0.4	1.33	0.929
05-01-2019	20.6	12.45	8.146
06-01-2019	16.3	14.95	1.348
07-01-2019	8.8	15.39	6.590
08-01-2019	0.7	7.61	6.908
09-01-2019	2.9	6.83	3.931
10-01-2019	3.3	5.89	2.585
11-01-2019	1.9	5.40	3.496
12-01-2019	1.5	5.92	4.423
13-01-2019	19.2	13.22	5.980
14-01-2019	13.6	14.31	0.710
15-01-2019	0.9	6.97	6.070
16-01-2019	2.6	3.06	0.456
17-01-2019	3.6	6.49	2.885

Vol. 16, No. 2, May 2022, pp. 63-73 ISSN 2086-8138 (print) | 2745-7249 (online)

Linear regression isshown in **Fig. 7** with an R-Square result of 0.59972. The red line is the result of regression, while the blue dot is the distribution of actual data to the predicted result.

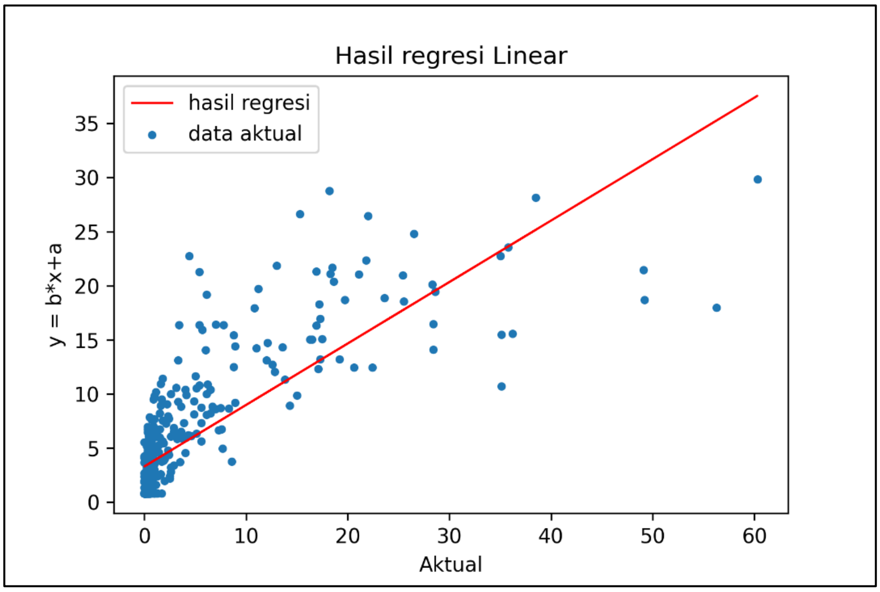

Fig 7. Linear Regression

The result of combination 1 (not using supporting parameters) with combination 9 (using supporting parameters) is depicted in **Fig. 8**. Combination 9 (in purple) tends to be better at following the pattern of actual rainfall data (pink) than combination 1 (colored green). From the results of this study, it can be said that adding weather parameters as supporting parameters can provide a more accurate prediction of rainfall.

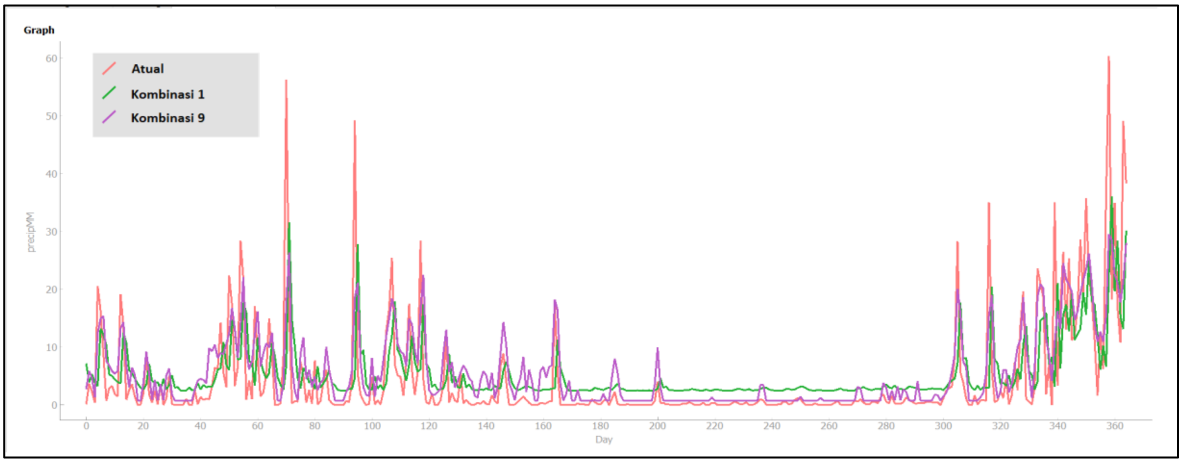

Fig 8. Monthly Prediction Results for 2019

4. Conclusion

Based on the results of the study, rainfall prediction using ANN without weather parameters obtained AN MSE of 0.01652, while predictions with a combination of weather parameters (additional parameters) such as cloud cover, humidity, dew point temperature, wind gusts, rainfall, air pressure, temperature, wind direction, and wind speed obtained MSE of 0.00881. From these results it can be concluded that with weather parameters it is proven to be able to provide a smaller MSE value. The next suggestion for researchers is to conduct further testing of the weight value and initial bias so that the prediction results remain consistent

Declarations

Author contribution. All authors contributed equally to the main contributor to this paper. All authors read and approved the final paper.
Funding statement. None of the authors have received any funding or grants from any institution or funding body for the research.
Conflict of interest. The authors declare no conflict of interest.
Additional information. No additional information is available for this paper.

References

[1] H. Mutmainah, "PERBEDAAN KARAKTERISTIK PARAMETER LINGKUNGAN LAUT PADA DUA PERIODE MUSIM BASAH DI DAERAH EKUATORIAL Periode April dan September 2016 di Selat Pagai, Mentawai," *Semin. Nas. Geomatika*, vol. 2, no. September 2016, p. 205, 2018, doi: 10.24895/sng.2017.2-0.413.

[2] I. N. Hidayati and Suryanto, "Pengaruh Perubahan Iklim Terhadap Produksi," *J. Ekon. dan Stud. Pembang.*, vol. 16, no. 1, pp. 42–52, 2015.

[3] N. S. Pradipta, P. Sembiring, and P. Bangun, "Suhu Udara," *Cuaca Iklim*, vol. 1, no. 5, pp. 459–468, 2013.

[4] D. Susilokarti, S. S. Arif, S. Susanto, and L. Sutiarso, "Studi Komparasi Prediksi Curah Hujan Metode Fast Fourier Transformation (Fft), Autoregressive Integrated Moving Average (Arima) Dan Artificial Neural Network (Ann)," *J. Agritech*, vol. 35, no. 02, p. 241, 2015, doi: 10.22146/agritech.9412.

[5] A. Rachmawati, "Pada Skala Bulanan , Dasarian Dan Harian," vol. V, no. 2, pp. 50–57, 2015.

[6] Mislan, Haviluddin, S. Hardwinarto, Sumaryono, and M. Aipassa, "Rainfall Monthly Prediction Based on Artificial Neural Network: A Case Study in Tenggarong Station, East Kalimantan - Indonesia," *Procedia Comput. Sci.*, vol. 59, no. Iccsci, pp. 142–151, 2015, doi: 10.1016/j.procs.2015.07.528.

[7] L. C. P. Velasco, R. P. Serquiña, M. S. A. Abdul Zamad, B. F. Juanico, and J. C. Lomocso, "Week-ahead rainfall forecasting using multilayer perceptron neural network," *Procedia Comput. Sci.*, vol. 161, pp. 386–397, 2019, doi: 10.1016/j.procs.2019.11.137.

[8] Y. A. Lesnussa, L. J. Sinay, and M. R. Idah, "Aplikasi Jaringan Saraf Tiruan Backpropagation untuk Penyebaran Penyakit Demam Berdarah Dengue (DBD) di Kota Ambon," *J. Mat. Integr.*, vol. 13, no. 2, p. 63, 2017, doi: 10.24198/jmi.v13.n2.11811.63-72.

[9] M. G. M. Abdolrasol *et al.*, "Artificial Neural Networks Based Optimization Techniques: A Review," *Electron. 2021, Vol. 10, Page 2689*, vol. 10, no. 21, p. 2689, Nov. 2021, doi: 10.3390/ELECTRONICS10212689.

[10] S. Walczak and N. Cerpa, "Artificial Neural Networks," *Encycl. Phys. Sci. Technol.*, pp. 631–645, Jan. 2003, doi: 10.1016/B0-12-227410-5/00837-1.

[11] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, "Activation Functions: Comparison of trends in Practice and Research for Deep Learning," Nov. 2018, doi: 10.48550/arxiv.1811.03378.

[12] S. Hu, L. Jin, H. Wang, Y. Zhang, S. Kwong, and C. C. J. Kuo, "Objective Video Quality Assessment Based on Perceptually Weighted Mean Squared Error," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 9, pp. 1844–1855, Sep. 2017, doi: 10.1109/TCSVT.2016.2556499.

[13] Z. Tian, L. Wong, and N. Safaei, "A neural network approach for remaining useful life prediction utilizing both failure and suspension histories," *Mech. Syst. Signal Process.*, vol. 24, no. 5, pp. 1542–1555, Jul. 2010, doi: 10.1016/J.YMSSP.2009.11.005.

[14] S. Hatami, M. Alisafaee, E. Atoofian, Z. Navabi, and A. Afzali-Kusha, "A Low-Power Scan-Path architecture," *Proc. - IEEE Int. Symp. Circuits Syst.*, pp. 5278–5281, 2005, doi: 10.1109/ISCAS.2005.1465826.

[15] M. C. Valverde Ramírez, H. F. De Campos Velho, and N. J. Ferreira, "Artificial neural network technique for rainfall forecasting applied to the São Paulo region," *J. Hydrol.*, vol. 301, no. 1–4, pp. 146–162, Jan. 2005, doi: 10.1016/J.JHYDROL.2004.06.028.

[16] M. A. Mercioni and S. Holban, "The Most Used Activation Functions: Classic Versus Current," *2020 15th Int. Conf. Dev. Appl. Syst. DAS 2020 - Proc.*, pp. 141–145, May 2020, doi: 10.1109/DAS49615.2020.9108942.

[17] E. P. Martins, J. A. F. Diniz-Filho, and E. A. Housworth, "ADAPTIVE CONSTRAINTS AND THE PHYLOGENETIC COMPARATIVE METHOD: A COMPUTER SIMULATION TEST," *Evolution (N. Y).*, vol. 56, no. 1, pp. 1–13, Jan.

2002, doi: 10.1111/J.0014-3820.2002.TB00844.X.

[18] L. Hou, Y. Cheng, X. Wang, J. Ren, and X. Geng, “Effect of slickwater-alternate-slurry injection on proppant transport at field scales: A hybrid approach combining experiments and deep learning,” *Energy*, vol. 242, p. 122987, Mar. 2022, doi: 10.1016/J.ENERGY.2021.122987.

[19] A. Timmermann, “Chapter 4 Forecast Combinations,” *Handb. Econ. Forecast.*, vol. 1, pp. 135–196, Jan. 2006, doi: 10.1016/S1574-0706(05)01004-9.

[20] J. D. Brown, J. Demargne, D. J. Seo, and Y. Liu, “The Ensemble Verification System (EVS): A software tool for verifying ensemble forecasts of hydrometeorological and hydrologic variables at discrete locations,” *Environ. Model. Softw.*, vol. 25, no. 7, pp. 854–872, Jul. 2010, doi: 10.1016/J.ENVSOFT.2010.01.009.