JINAV: Journal of Information and Visualization *ISSN: 2746-1440 (online)*
Vol. 4 No. 2 (2023)
https://doi.org/10.35877/454RI.jinav2356

Short-term Forecasting of Covid-19 Cases in East Java of Indonesia using NARX-NN Model

Hermansah[a,*], Dedi Rosadi[b], & Pepi Novianti[c]

[a]*Department of Mathematics Education, University of Riau Kepulauan, Indonesia*
[b]*Department of Mathematics, University of Gadjah Mada, Indonesia*
[c]*Department of Mathematics, University of Bengkulu, Indonesia*

Abstract

The goal of this study is to forecast short-term verified Covid-19 infections in East Java of Indonesia using the NARX-NN model. Here, the external variable used was the weather of East Java. The confirmed data for Covid-19 were obtained from the BNPB, and the weather data of East Java were obtained from the BMKG. Data from July 21st, 2020 to June 20th, 2021, were used for model formation (training data), and data from June 21st to 27th, 2021 were used for validation data. Based on the formatting model results, we can conduct a short-term forecast for three future periods (June 28th to 30th, 2021). This research evaluated the NARX-NN model using the forecasting accuracy of MAPE. The NARX-NN approach is more suitable than the NAR-NN method for predicting daily confirmed Covid-19 cases in East Java, based on the forecasting results of the NAR-NN and NARX-NN methods. The MAPE value was 0.03060 (0.03248 smaller than the MAPE value of the NAR-NN). At the conclusion of the study, the NARX-NN approach was utilized for daily forecasting of Covid-19 instances in East Java from June 28th to June 30th, 2021, namely 1039, 1072, and 1185.

Keywords: Forecasting, covid-19, weather, NARX-NN.

Received: 17 October 2023 Revised: 14 December 2023 Accepted: 23 December 2023

1. Introduction

Corona Virus Diseases (Covid-19), designated as a global pandemic since March 11th, 2020, by the World Health Organization (WHO), is one of the health problems that almost all countries globally face. Indonesia is one of the countries with a relatively high rate of positive confirmed cases of Covid-19. After experiencing the first wave at the end of 2020, Indonesia experienced the second wave in mid-2021. On the Worldometer page, as of August 1st, 2021, Indonesia was in the first position in Asia with 545,447 active cases. After successfully passing the second wave, Indonesia also needs to prepare for the possibility of an increase in the third wave of Covid-19 cases. For that, we need to predict the daily addition of Covid-19 cases.

The contributors to Covid-19 cases in Indonesia are still dominated by provinces on the island of Java, one of which is East Java Province. This province now has the highest total number of Covid-19 cases (22%), around 16,600 cases as of July 13th, ahead of DKI Jakarta (approximately 14,500 cases). Since the end of June, East Java has contributed the newest Covid-19 cases in Indonesia. In an article written by Purwanto et. al. (2021), in general, the transmission of Covid-19 in the community in East Java cannot be controlled. There are two main reasons why East Java is currently a province that surpasses DKI Jakarta in the number of Covid-19 cases and the increase in new cases. First, community compliance with the implementation of health protocols in East Java is still relatively low. Second, weak health policies related to handling this outbreak in East Java (Purwanto et. al., 2021; Toharudin et. al., 2020). In making policies, the government can refer to scientific studies, one of which is statistical predictions.

* Corresponding author.
E-mail address: hermansah@fkip.unrika.ac.id

JINAV: Journal of Information and Visualization is licensed under an
Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)

Forecasting or predictions related to Covid-19 have been studied by various researchers, Fanelli & Piazza (2020) studied the forecasting of the spread of Covid-19 in China, Italy, and France using the SIRD model, Roosa et. al. (2020) studied about Covid-19 real-time forecast in China with generalized logistic growth model (GLM), Benvenuto et. al. (2020) examined the forecast of Covid-19 using ARIMA, and Koczkodaj et. al. (2020) predicted Covid-19 outside of China by using a simple heuristic (exponential curve). According to Hermansah et. al. (2021a) and Hermansah et. al. (2021b), the Nonlinear Auto-Regressive Neural Network with eXogenous inputs (NARX-NN) model outperforms ARIMA and exponential smoothing methods. So in this study, East Java Province Covid-19 data will be applied to the NARX-NN method to predict daily cases of Covid-19 in East Java. The NARX-NN model is a Nonlinear Auto-Regressive Neural Network (NAR-NN) model with exogenous/external variables (Hermansah et. al., 2021a; 2021b; 2021c). Exogenous variables are considered to influence the predicted variables but are not influenced by the predicted variables in the model. In this study, the exogenous variable was used weather in East Java.

Several studies have shown that weather factors can affect Covid-19 cases. According to Wang et. al. (2021), temperature and humidity negatively correlate to Covid-19 transmission both in China and the United States. Sobral et. al. (2020) looked into the link between SARS-CoV-2 transmission and death and environmental factors. The authors found a favorable relationship between rainfall and SARS-CoV-2 transmission but no significant relationship between Covid-19 mortality and temperature. The impact of the environment, socioeconomics, topography, and demography on the Covid-19 cases in the United States of America was investigated using the geographically weighted regression (GWR) approach by Mollalo et. al. (2020). This study aims to determine the characteristics of the Covid-19 data in East Java and how the forecasting results using NARX-NN are confirmed for Covid-19 forecasting in East Java with the exogenous variable of East Java weather. This research hopes to provide good predictive results regarding Covid-19 in East Java in the future period so that policymakers can plan and make the right decisions.

2. Material and Method

2.1. NARX Model

The Artificial Neural Network (ANN) model is capable of obtaining nonlinear relationships in the data by using the following equation:

$$\boldsymbol{y}(\boldsymbol{t}+\mathbf{1}) = \boldsymbol{f}(\boldsymbol{y}(\boldsymbol{t}), \boldsymbol{y}(\boldsymbol{t}-\mathbf{1}), \dots, \boldsymbol{y}(\boldsymbol{t}-\boldsymbol{n_y})) \quad (1)$$

where f is a nonlinear function that maps the previous observation nonlinearly to the following output, the ANN configuration model that has been successfully applied to various time series forecasting is a model focusing on time-lagged feedforward networks and dynamically-driven recurrent networks. Both are feedforward models using time-lagged, but the second model uses recurrent (feedback) in the model. One form of the time-lagged feedforward network model is the Multi-Layer Perceptron (MLP) model, which is also known as the Feed-Forward Neural Network (FFNN) or the Nonlinear Auto-Regressive Neural Network (NARNN). MLP is a form of architecture that is generally the most widely used in applications in engineering. Typically, applications for time series data modeling are based on the MLP architecture.

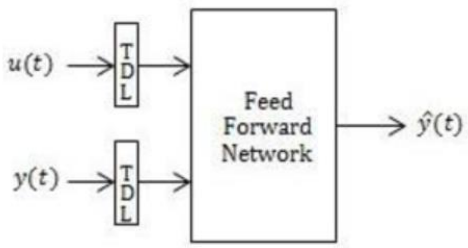

Figure 1. Series-parallel architecture of NARX network.

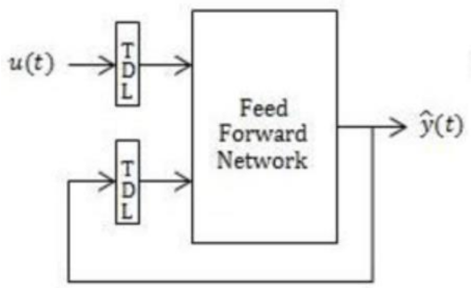

Figure 2. Parallel architecture of NARX network.

One form of the dynamically-driven recurrent network model is the Nonlinear Auto-Regressive with eXogenous input (NARX) model. The NARX model is a discrete nonlinear system defined in the following equation (Siegelmann et. al., 1997):

$$\boldsymbol{y}(\boldsymbol{t}+\mathbf{1}) = \boldsymbol{f}(\boldsymbol{y}(\boldsymbol{t}), \boldsymbol{y}(\boldsymbol{t}-\mathbf{1}), \dots, \boldsymbol{y}(\boldsymbol{t}-\boldsymbol{n_y}), \boldsymbol{u}(\boldsymbol{t}+\mathbf{1}), \boldsymbol{u}(\boldsymbol{t}), \dots, \boldsymbol{u}(\boldsymbol{t}-\boldsymbol{n_u})) \quad (2)$$

where $u(t)$ and $y(t)$ are the input and output of the network at time t, n_u and n_y are the order of input and output, and the function f is a nonlinear function approximated by ANN so that it becomes a NARX network. The NARX model can be implemented in two architectural forms, namely series-parallel and parallel architecture. Illustration of series-parallel and parallel architecture can be seen in Figures 1 and 2, respectively. Series parallel architecture is if the system regressor output is directly used without being fed back into the system (forecasting results are not used as feed or input). The series-parallel architecture is a pure feedforward architecture. The equation of the series-parallel architecture is found in the following equation (Jiang & Song, 2011):

$$\hat{\boldsymbol{y}}(\boldsymbol{t}+\mathbf{1}) = \hat{\boldsymbol{f}}(\boldsymbol{y}(\boldsymbol{t}), \boldsymbol{y}(\boldsymbol{t}-\mathbf{1}), \dots, \boldsymbol{y}(\boldsymbol{t}-\boldsymbol{n_y}), \boldsymbol{u}(\boldsymbol{t}+\mathbf{1}), \boldsymbol{u}(\boldsymbol{t}), \dots, \boldsymbol{u}(\boldsymbol{t}-\boldsymbol{n_u})) \quad (3)$$

Parallel architecture is if the regressor output from the system inputs the predicted value (the forecasting results are used as feedback or input to the system). Parallel architecture is often used in long-term forecasting, where previous forecasts are used as input for longer-term forecasts. The equation of parallel architecture is found in the following equation:

$$\hat{\boldsymbol{y}}(\boldsymbol{t}+\mathbf{1}) = \hat{\boldsymbol{f}}(\hat{\boldsymbol{y}}(\boldsymbol{t}), \hat{\boldsymbol{y}}(\boldsymbol{t}-\mathbf{1}), \dots, \hat{\boldsymbol{y}}(\boldsymbol{t}-\boldsymbol{n_y}), \boldsymbol{u}(\boldsymbol{t}+\mathbf{1}), \boldsymbol{u}(\boldsymbol{t}), \dots, \boldsymbol{u}(\boldsymbol{t}-\boldsymbol{n_u})) \quad (4)$$

The NARX model, in general, has fast convergence and can make good generalizations, and has good performance on problems involving long-term dependencies, see for example in (Bengio et. al., 1994; Lin et. al., 1996). The NARX model has also been widely used in various time-delay neural network applications such as chaotic laser pulsation data and video traffic time series, see in Diaconescu (2008); Ingrassia & Morlini (2007). Meanwhile, this research focused on the NARX model with series-parallel architecture for short-term forecasting of time series data.

2.2. Forecast Measure

Two forecast error measurements, namely Mean Squared Error (MSE) and Mean Absolute Percent Error (MAPE), were used to evaluate the forecast accuracy. MSE is defined as follows:

$$\boldsymbol{MSE} = \sum_{t=1}^{N} \frac{(A_t - F_t)^2}{N} = \sum_{t=1}^{N} \frac{e_t^{\ 2}}{N} \quad (5)$$

where e is error and N is the number of data.

MAPE is defined as follows:

$$\boldsymbol{MAPE} = \frac{1}{N} \sum_{t=1}^{N} \left| \frac{A_t - F_t}{A_t} \right| \quad (6)$$

where A_t is actual values at data time t and F_t is forecast values at data time t.

2.3. Algorithm

This section describes the procedure for the NARX modeling algorithm, assuming that $y(t)$ and $u(t)$ are time-series data to be predicted and exogenous/external data, respectively. The algorithms that we use following closely to Hermansah et. al. (2021b), with some improvements. Furthermore, several steps contained in the algorithm are described as follows:

Step 1: Initial processing begins by identifying the data $y(t)$ has a trend component. If the data $y(t)$ has a trend component, then the first differencing is taken. Furthermore, de-trend data is used to identify the seasonal component, and if the seasonal component is identified, a deterministic seasonal dummy is created as an additional external input. Next, the data is linearly scaled to [-0.8, 0.8]. In addition, $u(t)$ data is given the same treatment as $y(t)$, without identifying the trend and seasonal components.

Step 2: The Partial Auto-Correlation Function (PACF) is used to determine lag (input variable) based on the lag $y(t)$ with significant autocorrelation. Even though the PACF only looks for a linear relationship, experience has found that this is a good technique to choose input variables (Martínez et. al., 2019). It is assumed that the lag $u(t)$ is equal to $y(t)$. Seasonal dummy variables can likewise be used as input variables.

Step 3: The number of neurons in the hidden layer is determined by trial and error from one neuron to the total input variables in Step 2. Data sharing uses a composition of 80% training data and 20% validation data from data processing. Furthermore, the number of neurons in the hidden layer is defined as the value that gives the minimum Mean Squared Error (MSE) of training data. The number of neurons in the hidden layer is also evaluated for the value that minimizes the MSE on the selected validation data.

Step 4: The architecture of the NARX model obtained in Step 3 is combined in three parts, namely the tangent hyperbolic activation function, ensemble mode operator, and resilient backpropagation learning algorithm, to obtain the best single model. Next, the NARX model was fitted 50 times using different initial random weights. The final forecast is created by combining the forecasts acquired using the ensemble mode operator.

Step 5: Recursive or iterative strategies are used for forecasting several steps forward.

Further details showing the application of an algorithm similar to the above algorithm for the NARX model with and without exogenous input are provided in Hermansah et. al. (2021a, 2021b, 2021c) and Crone & Kourentzes (2010); Hermansah et. al. (2020); Kourentzes et. al. (2014), respectively.

3. Results and Discussion

This research was conducted on real cases, namely Covid-19 data confirmed in East Java, with the exogenous/external variable being the East Java weather. Confirmed data for Covid-19 was obtained from BNPB, and East Java weather was obtained from BMKG (Malang climatology station). The data started from 21st July 2020 to 27th June 2021. Data from 21st July 2020 to 20th June 2021 was used for model formation (training data) and validation data from 21st June 2021 to 27th June 2021. Based on the results of the model formation, we can make short-term forecasts for the following three periods (28th June 2021 to 30th June 2021 for Covid-19 in East Java). The following are statistics summary of Covid-19 cases and weather data in East Java (Table 1).

Table 1. Statistics summary of Covid-19 cases and weather data in East Java.

Statistics	**Positive**	**MaxTemp**	**AvgTemp**	**Humidity**
Sample Size	345	345	345	345
Minimum	141.0	24.80	21.10	44.00
Maximum	1203.0	35.000	30.000	94.000
Mean	447.79	29.090	24.360	77.910
Std. Deviation	245.0	1.530	1.390	8.440
DF-test	-0.1	-4.0	-4.0	-4.0
DF-test.p	1.00	0.02	0.01	0.01

Every single variable has a 345 days sample size. In the addition of positive cases, the lowest value of 141 cases occurred on May 14th, 2021, and the highest of 1203 cases on June 30th, 2021. By adding an average of 447.79 cases and a standard deviation of 245, the p-value for the Dickey-Fuller test in positive cases indicates that the data is not stationary. The average daily maximum temperature is 29.09, with the lowest maximum temperature being 24.8 and the highest being 35. The standard deviation of the maximum temperature data is 1.53, and the daily average temperature is 1.39. The average daily temperature value is 24.36, with the lowest average temperature of 21.1 and the highest of 30. For the humidity variable, the lowest value was 44, which occurred on August 21st, 2020, and the highest was 94 on December 30th, 2020. Based on the p-value in the DF test, it was concluded that the variables of maximum temperature, average temperature, and humidity were stationary. The following is a time series plot of the addition of positive cases and the weather variable (Figure 3).

The following results of the correlation analysis of the number of positive cases of Covid-19 and the variables of the three factors used (Figure 4). Figure 4 shows the Pearson correlation of a pair of variables. The three weather variables have a significant correlation with the addition of cases, although they show a relatively low correlation. The correlation of the additional case variables with the maximum and average temperature variables showed a value of -0.36 with a p-value of 0.16 and -0.24 with a p-value of 0.20. With a negative correlation value, it can be said that at high temperatures, the addition of positive cases is relatively low. This is in line with the positive correlation of case addition and humidity, which indicates that when humidity increases, there is a tendency for cases to increase as well. It can also be seen with the correlation value of 0.38 and the p-value of 0.26. Therefore, based on the significant value, the cases and variables have a significant effect of 70%. Based on the correlation value table (Schober et. al., 2018), all variables are categorized as low or weak correlation because each coefficient value is between 0.2-0.4.

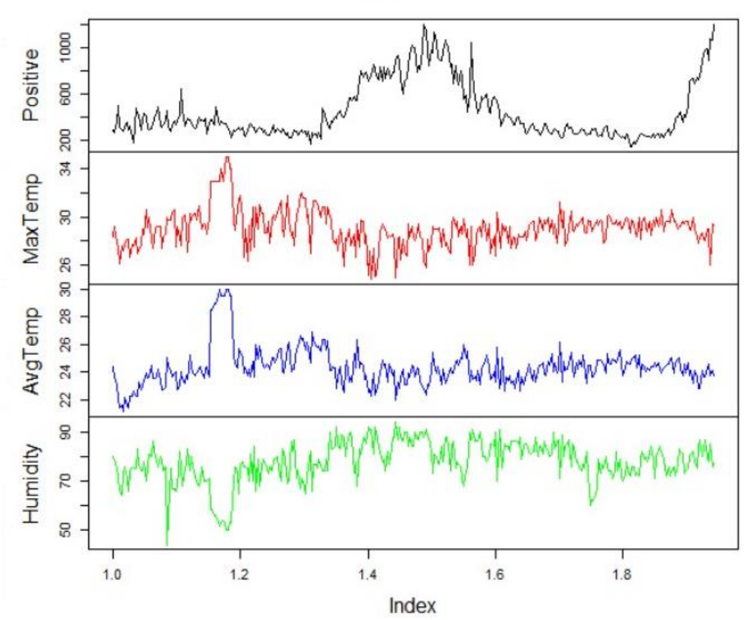

Figure 3. Time series plot of positive cases and weather data in East Java.

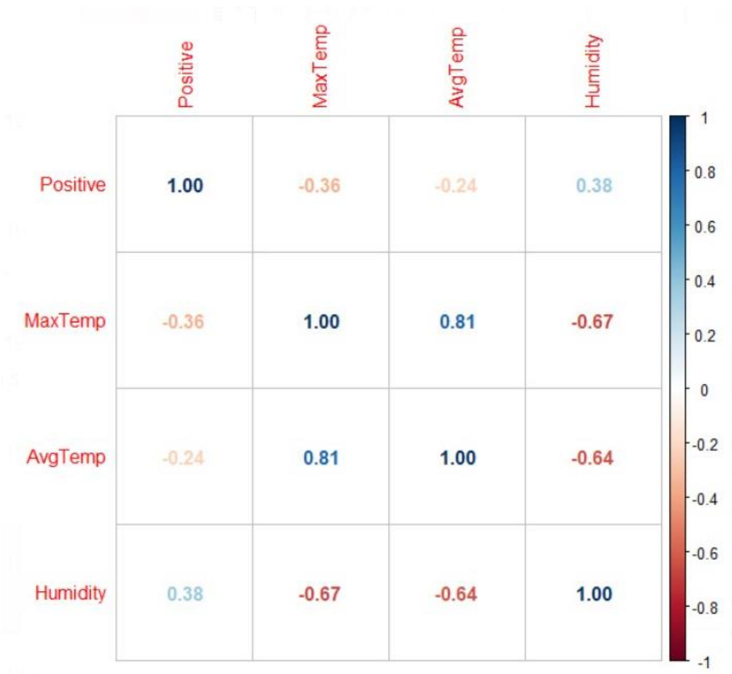

Figure 4. Correlation analysis of positive cases and weather data in East Java.

Time series forecasting utilizes previous experience data as predictive data and becomes study data, so that it might have a weakness in the short prediction period. However, short-term forecasting is still needed to make strategic decisions for the future. In this study, a reasonable prediction for Covid-19 cases is short-term forecasting for the next 3-5 periods. Daily Covid-19 case predictions can provide information to decision-makers and need to be considered in determining how to stop Covid-19 from spreading.

Confirmed data on Covid-19 and weather in East Java, Indonesia, have been processed and forecasted using the NAR-NN and NARX-NN methods. Based on the outline of the algorithm in the previous subsection, the following are the results of calculations for Covid-19 cases in East Java, Indonesia, which are presented in Table 2.

Table 2. Comparison of actual data and daily case prediction results of Covid-19 in East Java of Indonesia.

Date	Actual	NAR-NN	APE	NARX-NN	APE
21/06/2021	719	792	0.10096	758	0.05420
22/06/2021	746	782	0.04798	774	0.03790
23/06/2021	873	797	0.08717	834	0.04454
24/06/2021	945	878	0.07121	928	0.01789
25/06/2021	975	952	0.02404	974	0.00061
26/06/2021	989	960	0.02913	982	0.00714
27/06/2021	889	961	0.08108	935	0.05189
		MAPE	0.06308	**MAPE**	0.03060

Table 2 reveals that the NARX-NN method is better than the NAR-NN method for forecasting positive confirmed cases of Covid-19 in East Java, Indonesia. It can be supported by the MAPE value of 0.03060, which is 0.03248 smaller than the MAPE NAR-NN value. Thus, the NARX-NN algorithm will be applied to predict confirmed cases of Covid-19 from June 28th to 30th, 2021 (Table 3).

Table 3. Prediction results of Covid-19 cases in East Java of Indonesia June 28th to 30th, 2021.

Date	Forecast	Lower 95%	Higher 95%
28/06/2021	1039	836	1131
29/06/2021	1072	832	1213
30/06/2021	1185	848	1277

Based on the prior discussion, The NARX-NN approach is regarded as the best way for forecasting Covid-19 in East Java, Indonesia. The significance of forecasting accuracy measures (MAPE) demonstrates this. Table 3 gives the daily case prediction of Covid-19 in East Java from June 28th to 30th, 2021, based on NARX-NN. The case forecasts for the next three days are 1039, 1072, and 1185 for June 28th, 29th, and 30th, 2021, respectively. Besides, the table displays the lower and higher prediction as to the interval value for each day. On June 30th, 2021, there were 1185 verified Covid-19 cases in East Java, Indonesia, with an interval value of 848 to 1277 instances.

4. Conclusion

The forecasting of Covid-19 in East Java, Indonesia, can help policymakers make decisions in the future. The NARX-NN approach is better than the NAR-NN method for fitting Covid-19 data in East Java, Indonesia, from June 21st to 27th, 2021. The case forecasts for the next three days, using the NARX-NN algorithm, are 1039, 1072, and 1185 for June 28th, 29th, and 30th, 2021, respectively. The policymaker can use the forecast to create a policy for future studies based on it. This method can be evaluated to other algorithms or approaches, such as NARX-NN with parallel architecture or other forecasting methods, for some further studies.

References

Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, *5*(2), 157–166.

Benvenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S., & Ciccozzi, M. (2020). Application of the ARIMA model on the COVID-2019 epidemic dataset. *Data in Brief*, *29*, 105340.

Crone, S. F., & Kourentzes, N. (2010). Feature selection for time series prediction–A combined filter and wrapper approach for neural networks. *Neurocomputing*, *73*(10–12), 1923–1936.

Diaconescu, E. (2008). The use of NARX neural networks to predict chaotic time series. *WSEAS Transactions on Computer Research*, *3*(3), 182–191.

Fanelli, D., & Piazza, F. (2020). Analysis and forecast of COVID-19 spreading in China, Italy and France. *Chaos, Solitons & Fractals*, *134*, 109761.

Hermansah, Rosadi, D., Abdurakhman, & Utami, H. (2020). Selection of input variables of nonlinear autoregressive neural network model for time series data forecasting. *Media Statistika*, *13*(2), 116–124.

Hermansah, Rosadi, D., Abdurakhman, & Utami, H. (2021a). A comparison of learning algorithms for seasonal time series forecasting using NARX model. *J. Math. Comput. Sci.*, *11*(6), 6638–6656.

Hermansah, Rosadi, D., Abdurakhman, & Utami, H. (2021b). Automatic time series forecasting using nonlinear autoregressive neural network model with exogenous input. *Bulletin of Electrical Engineering and Informatics*, *10*(5), 2836–2844.

Hermansah, Rosadi, D., Abdurakhman, & Utami, H. (2021c). Time series forecasting with trend and seasonal patterns using NARX network ensembles. *Math. Stat.*, *9*(4), 511–520.

Ingrassia, S., & Morlini, I. (2007). Equivalent number of degrees of freedom for neural networks. *Advances in Data Analysis: Proceedings of the 30th Annual Conference of the Gesellschaft Für Klassifikation EV, Freie Universität Berlin*, 229–236.

Jiang, C., & Song, F. (2011). Sunspot Forecasting by Using Chaotic Time-series Analysis and NARX Network. *J. Comput.*, *6*(7), 1424–1429.

Koczkodaj, W. W., Mansournia, M. A., Pedrycz, E. W., Wolny-Dominiak, A., Zabrodskii, P. F., Strzałka, D., ... & Mazurek, J. (2020). 1,000,000 cases of COVID-19 outside of China: The date predicted by a simple heuristic. *Global Epidemiology*, *2*, 100023.

Kourentzes, N., Barrow, D. K., & Crone, S. F. (2014). Neural network ensemble operators for time series forecasting. *Expert Systems with Applications*, *41*(9), 4235–4244.

Lin, T., Horne, B. G., Tino, P., & Giles, C. L. (1996). Learning long-term dependencies in NARX recurrent neural networks. *IEEE Transactions on Neural Networks*, *7*(6), 1329–1338.

Martínez, F., Frías, M. P., Pérez, M. D., & Rivera, A. J. (2019). A methodology for applying k-nearest neighbor to time series forecasting. *Artificial Intelligence Review*, *52*(3), 2019–2037.

Mollalo, A., Vahedi, B., & Rivera, K. M. (2020). GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. *Science of The Total Environment*, *728*, 138884.

Purwanto, P., Utaya, S., Handoyo, B., Bachri, S., Astuti, I. S., Utomo, K. S. B., & Aldianto, Y. E. (2021). Spatiotemporal analysis of COVID-19 spread with emerging hotspot analysis and space–time cube models in East Java, Indonesia. *ISPRS International Journal of Geo-Information*, *10*(3), 133.

Roosa, K., Lee, Y., Luo, R., Kirpich, A., Rothenberg, R., Hyman, J. M., ... & Chowell, G. (2020). Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. *Infectious Disease Modelling*, *5*, 256–263.

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: appropriate use and interpretation. *Anesthesia & Analgesia*, *126*(5), 1763–1768.

Siegelmann, H. T., Horne, B. G., & Giles, C. L. (1997). Computational capabilities of recurrent NARX neural networks. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *27*(2), 208–215.

Sobral, M. F. F., Duarte, G. B., da Penha Sobral, A. I. G., Marinho, M. L. M., & de Souza Melo, A. (2020). Association between climate variables and global transmission oF SARS-CoV-2. *Science of The Total Environment*, *729*, 138997.

Toharudin, T., Pontoh, R. S., Zahroh, S., Akbar, A., & Sunengsih, N. (2020). Impact of large scale social restriction on the COVID-19 cases in East Java. *Commun. Math. Biol. Neurosci.*, *2020*, Article-ID.

Wang, J., Tang, K., Feng, K., Lin, X., Lv, W., Chen, K., & Wang, F. (2021). Impact of temperature and relative humidity on the transmission of COVID-19: a modelling study in China and the United States. *BMJ Open*, *11*(2), e043863.