Journal of
Information Systems Engineering
and Business Intelligence

Vol.9, No.1, April 2023
Available online at: http://e-journal.unair.ac.id/index.php/JISEBI

Prediction of COVID-19 Using the Artificial Neural Network (ANN) with K-Fold Cross-Validation

Nur Alifiah [1], Dian Kurniasari [2] *, Amanto [3], Warsono [4]

[1)2)3)4)] *Mathematics, Faculty of Mathematics and Natural Sciences, University of Lampung, Indonesia*
Jl. Prof. Dr. Sumantri Brojonegoro No. 1, Bandar Lampung
[1)]alifiah.alif.14.aa@gmail.com, [2)]dian.kurniasari@fmipa.unila.ac.id, [3)]amanto.1973@fmipa.unila.ac.id, [4)]warsono.1963@fmipa.unila.ac.id

Abstract

Background: COVID-19 is a disease that attacks the respiratory system and is highly contagious, so cases of the spread of COVID-19 are increasing every day. The increase in COVID-19 cases cannot be predicted accurately, resulting in a shortage of services, facilities and medical personnel. This number will always increase if the community is not vigilant and actively reduces the rate of adding confirmed cases. Therefore, public awareness and vigilance need to be increased by presenting information on predictions of confirmed cases, recovered cases, and cases of death of COVID-19 so that it can be used as a reference for the government in taking and establishing a policy to overcome the spread of COVID-19.
Objective: This research predicts COVID-19 in confirmed cases, recovered cases, and death cases in Lampung Province
Method: This study uses the ANN method to determine the best network architecture for predicting confirmed cases, recovered cases, and deaths from COVID-19 using the k-fold cross-validation method to measure predictive model performance.
Results: The method used has a good predictive ability with an accuracy value of 98.22% for confirmed cases, 98.08% for cured cases, and 99.05% for death cases.
Conclusion: The ANN method with k-fold cross-validation to predict confirmed cases, recovered cases, and COVID-19 deaths in Lampung Province decreased from October 27, 2021, to January 24, 2022.

Keywords: Artificial Intelligence, Artificial Neural Network (ANN) K-Fold Cross Validation, COVID-19 Cases, Data Mining, Prediction.

Article history: Received 22 September 2022, First decision 12 March 2023, Accepted 10 April 2023, Available online 28 April 2023

I. INTRODUCTION

COVID-19 is a disease that attacks the respiratory system caused by SARS-CoV-2. SARS-CoV2 is a coronavirus with a crown-like structure that is highly contagious [1]. The World Health Organization (WHO) officially declared that COVID-19, which had spread to various countries worldwide, was declared a pandemic on March 11, 2020 [2]. The first case of COVID-19 in Indonesia was reported on March 2, 2020, with one positive confirmed case. Cases of the spread of COVID-19 are increasing every day. The increase in COVID-19 cases cannot be predicted accurately, resulting in a shortage of services, facilities and medical personnel. This number will always increase if the community is not vigilant and plays an active role in reducing the rate of addition of confirmed cases [3]. Therefore, public awareness and vigilance need to be increased by presenting information on predictions of confirmed cases, recovered cases, and death cases in COVID-19 so that it can be used as a reference for the government in taking and establishing a policy to overcome the spread of COVID-19. Information, studies, and insights based on influencing factors can be obtained by making predictions. In addition, time series models are prone to overfitting, and if the outliers are not handled properly, it can result in inaccurate prediction results [4].

Along with developments in Artificial Intelligence (AI) and Machine Learning (ML), better techniques and algorithms have been developed. The ML model has an excellent track record as a predictive model [5]. Artificial intelligence is a technology-based machine with intelligence like human thinking. Artificial intelligence is very effective to use because it can minimize human error. Sub-fields in artificial intelligence include machine learning. Machine learning is an innovation from machines developed to learn to produce a model from a data set. The learning process in machine learning uses a special algorithm, more commonly called a machine learning algorithm [6].

The most frequently used machine learning algorithms are artificial neural networks (ANN). ANN can solve complex problems in various applications such as optimization, prediction, simulation, modelling, clustering, pattern

* Corresponding author

ISSN 2443-2555 (online) 2598-6333 (print)
doi: http://dx.doi.org/10.20473/jisebi.9.1.16-27

recognition, and classification. We implemented ANN in this study to predict COVID-19 cases using the k-fold cross-validation method to measure predictive model performance. K-Fold cross-validation is a popular method by folding data as much as k parts, where data testing uses one part and k-1 parts as training data, then each k will look for the accuracy levels to obtain an average accuracy [7]. Therefore, this allows each piece of data to become training and testing data. It is different using a data partition scheme of 80%: 20%, 70%: 30%, or 60: 40%. While partitioning the dataset into training data and testing data with this scheme, it is possible to lose some crucial data points for research purposes. Because there is data that are not included in the training data, the model cannot detect some patterns. As is well known, the accuracy of analysis and modelling is greatly influenced by data, so it is hoped that the prediction model built into this study will obtain more accurate results.

This study aims to predict and forecast COVID-19 cases using the Artificial Neural Network (ANN) method, especially for confirmed, recovered, and COVID-19 death cases in Lampung Province. This research is expected to get a good level of accuracy by applying K-Fold cross-validation as a testing method.

II. RELATED WORKS

A. Spread of COVID-19 Cases in Indonesia

World Health Organization China Country Office reported a case of pneumonia of unknown etiology in Wuhan City, Hubei Province, China, on December 31, 2019. China identified pneumonia of unknown cause as a new type of coronavirus on January 7, 2020. In March 2020, WHO declared COVID-19 a global pandemic.

Several countries, such as France [8], Japan [9], and India [10], have conducted research related to the development of COVID-19 in these countries. The spread of the COVID-19 virus also occurred in Indonesia. About 50% of cases were from cities at the outbreak's start. The virus quickly spread to all 34 provinces, with more than 450,000 confirmed cases and 15,000 deaths as of November 12, 2020 [11]. As the virus that causes COVID-19 begins to spread, scientists need to track the disease and try to slow its spread. Therefore, a standard definition of COVID-19 cases must ensure that case counts are carried out similarly. Several definitions form the basis for determining the success of handling the spread of COVID-19, namely confirmed cases, recovered cases, and death cases.

Based on the WHO definition, a confirmed case is a condition where a person meets the clinical and epidemiological criteria using positive professional use or self-testing for the SARS-CoV-2 Antigen-RDT. Whereas cured cases were defined as cases where 14 days had passed between confirmed cases and no hospitalization or additional treatment was required, 10 days since hospital discharge, or two negative tests for at least 24 consecutive hours.

Finally, a COVID-19 case death by the Pan American Health Organization is defined as death from clinically appropriate disease in a probable or confirmed case of COVID-19 unless there is a clear alternative cause of death that cannot be linked to COVID-19.

B. Artificial Neural Networks for Prediction

Studies on applying the ANN algorithm to predict disease, especially COVID-19, have been carried out before. Ardabili [12] predicts COVID-19 with the Hybrid Artificial Intelligence Method from ANN, which is trained with the Gray Wolf Optimizer to get the highest performance. Based on the training, they project the outbreak to the end of May. Then, the training results were evaluated using the mean absolute percentage error (MAPE) and the correlation coefficient value.

Hasan [13] proposed an efficient and cost-effective time series modelling for the prediction of the daily trend of COVID-19 using an empirical mode decomposition (EEMD) ensemble based on an artificial neural network (ANN). Based on the output obtained, they concluded that the EEMD-ANN model provides promising indications for predicting COVID-19. Dharani [14] conducted an ANN-based prediction of COVID-19 and Symptom Relevance Survey and Analysis. The main focus is to see how someone can easily predict whether he is in good health or has been infected with COVID-19.

Niazkar [15] applied ANN to predict COVID-19 outbreaks by estimating the incubation period of COVID-19. The results show that applying the COVID-19 incubation period to ANN-based prediction models produces more accurate estimates. Furthermore, Prakash [16] built a prediction model to analyze the global spread of the COVID-19 pandemic in India and the world. They started with simple models like regression and eventually used the ANN method. In addition, they built a growth ratio metric to predict the peak and end of the pandemic.

To find the most efficient method and create a system that can assist in the early detection of COVID-19 conditions, Samsani [17] used various machine learning approaches such as Support Vector Machine, Naïve Bayes, AdaBoost, and ANN. The analysis sample was selected from the Coswara database, which is freely available. The results show that the ANN-based approach predicts COVID-19 with the highest accuracy.

III. METHODS

This section explains the proposed dataset and method, namely preprocessing, training, prediction, denormalization and forecasting, as shown in Fig. 1.

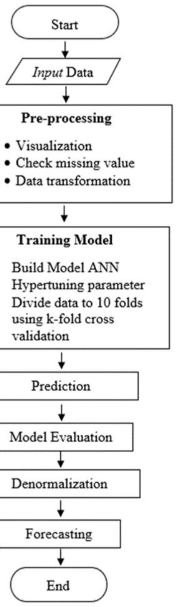


Fig. 1 ANN method process

A. Datasets

To predict the development of confirmed, recovered, and COVID-19 death cases in Lampung Province, we use daily data on COVID-19 cases obtained from the Indonesian Ministry of Health's website https://infectinemerging.kemkes.go.id/category/situasi- up to date, as of March 19, 2020, to January 24, 2022. The data obtained contains six variables: confirmed cases, recovered cases, death cases, number of confirmed cases, number of recovered cases, and number of death cases. However, the data variables for COVID-19 cases used in this study were only three: confirmed cases, recovered cases, and death cases, from March 19, 2020, to October 26, 2021. Meanwhile, data on confirmed, recovered, and death cases from October 27 to January 24, 2022, will be compared with the forecast data obtained. Data is shown in Table 1.

TABLE 1
DAILY DATA OF COVID-19 CASES IN LAMPUNG PROVINCE

Date	Confirmed Case	Recovered Case	Death Case	Number of Confirmed Case	Number of Recovered Case	Number of Death Case
19/03/2020	1	0	0	1	0	0
20/03/2020	0	0	0	1	0	0
21/03/2020	0	0	0	1	0	0
22/03/2020	0	0	0	1	0	0
23/03/2020	0	0	0	1	0	0
…	…	…	…	…	…	…
20/01/2022	1	2	0	49654	45149	3821
21/01/2022	3	1	0	49657	45150	3821
22/01/2022	6	0	1	49663	45150	3822
23/01/2022	2	1	0	49665	45151	3822
24/01/2022	3	1	0	49668	45152	3822

B. Method

1) Preprocessing

Before entering into the process of building the ANN model, the data need to be preprocessed first. First, visualization: the data that have been input will be visualized to see whether confirmed cases, recovered cases, and COVID-19 death cases have increased or decreased every day. Second, check missing value: this process is carried out to determine whether or not data is missing and then to repair the data. Third, data transformation: Transformation is changing the scale of data measurement from the original form into another form for analysis and certain assumptions. This study transforms the data using a standard scaler. Standard Scaler is rescaling distribution values so that the average observed value is 0 and the standard deviation is 1. Equation (1) formulates normalization calculations with a standard scaler [18].

$$x^* = \frac{x - \mu_x}{\sigma_x} \quad (1)$$

Where:
x^* = value x new
μ_x = average value x
σ_x = standard deviation of x

2) Training Model

First, we build the models. According to [19], an artificial neural network is an information processing system with biological neural network performance characteristics. Artificial Neural Network (ANN) is a description of the functions of the human brain in the form of mathematical functions which can be used to solve complex tasks, such as prediction, modelling, forecasting and classification.

One of the critical components in building the ANN model is determining the number of hidden layers, the number of nodes in each hidden layer, and the activation function. Determination of the number of layers should be adjusted to the complexity of the problem. However, it should be noted that the more hidden layers, the longer the computation time required. Modification of the number of hidden layers can be used to overcome the problem of low model performance.

Based on Heaton's rules [20], the number of nodes in each hidden layer are determined as follows: The number of nodes in the hidden layer must be less than double the number of input layers, The number of nodes in the hidden layer must be 2/3 of the number of input layers plus the output layers (2/3 (input+output), The number of nodes in the hidden layer must be between the size of the input and output layers, and because two input nodes are used, so the number of nodes in each hidden layer meets the requirements, namely 3,2, and 1.

The activation function activates neurons in the input layer connected to neurons in the hidden layer and neurons in the hidden layer connected to neurons in the output layer. The Rectified Linear Unit (ReLU) is the activation function in this study. Fig. 2 showed the ReLU activation function. ReLU is an activation function used to normalize the value generated by the layer. ReLU only creates a limit on zero, meaning when $x \leq 0$, the value $x = 0$, and when $x > 0$, then $x = x$. Equation (2) is the ReLU activation function [21].

$$f(x) = \max(0, x) \quad (2)$$

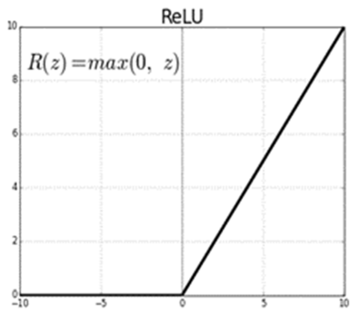


Fig. 2 Rectified Linear Unit (ReLU) activation function

Second step is hyperparameter tuning. Hyperparameters are parameters contained in a model and have been determined before the training process is carried out [22]. The optimal combination of parameters will produce a good model. The optimal combination of parameters can be searched using hyper-tuning. The hyperparameters that we use in this study include dropouts, epochs, batch size, and learning rate.

Dropout is a random abandonment of specific nodes in the layer during the training process [23]. Dropout can be done to prevent overfitting that occurs during the training process. Epochs are the number of iterations or iterations during the training process that provides input from the network and update the network weights. The training process takes place on the neural network to the beginning again [24]. According to [25], the batch size is one of the most critical hyperparameters used in machine learning and refers to the number of training examples used in one iteration. The learning rate is a significant parameter in the performance of a network concerning the time needed to achieve an optimal target. The learning rate parameter influences learning speed. If the learning rate is too low, it takes a long time to approach the minimum error, but if the learning rate is too large, the update weight will exceed the minimum error, and the weight will oscillate. Learning rate values are commonly 0.1, 0.01, and 0.001 [26].

Third step is K-Fold Cross-Validation. The following process is the division of the dataset using the k-fold cross-validation method. The dataset is divided into 'k' subsets with the same data. This study will use 10 k-fold. The data are divided into 10 k-folds roughly the same size for each fold, so they have 10 data subsets. For each of the 10 data subsets, the cross-validation test will use 9 k-folds for training and 1 k-fold for testing, as illustrated in Fig. 3. That is repeated k times to obtain a value in each part; the final result obtained is the average of all repetitions [27].

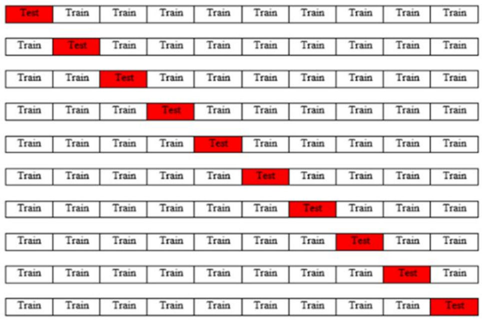


Fig. 3 Schema of K-fold cross-validation

3) Predictions

After getting the best parameters from hyper-tuning, ANN model training was carried out with K-fold cross-validation. Then, the model that has been trained can be used to predict the data.

4) Model Evaluation

The model evaluation process is vital in developing a good machine learning model. However, forecasting techniques that use quantitative data with specific time series data have errors caused by these techniques [28]. Thus, a method is needed to measure how much error is obtained after training the model. The methods used to evaluate errors in this study include accuracy, root mean square error (RMSE), mean square percentage error (MAPE).

Accuracy can be obtained by comparing the correct predictions to those made and expressed as a percentage.

$$a = \frac{t}{n} \times 100\% \tag{3}$$

with:
a is the accuracy in percent,
t is the number of correct prediction attempts, and
n is the number of trials.

However, accuracy is not the only value considered in evaluation because high accuracy values can be deceptive due to dataset imbalances. Therefore, other evaluation metrics are needed, such as RMSE and MAPE.

Root Mean Square Error can be interpreted as the average distance MSE [29], and can be calculated by squaring the error (prediction) divided by the amount of data, then taking the root. Root Mean Square Error is the magnitude of the error rate of the prediction results, where the smaller (closer to 0) the RMSE value, the more accurate the prediction results will be.

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(\hat{y}_t - y_t)^2} \tag{4}$$

with:
y^_t: predicted value in period t
y_t: actual value in period t

The MAPE calculation aims to measure statistics about prediction accuracy using the absolute error in each period divided by the actual value and then averaging the percentage error. The MAPE calculation formula is as follows:

$$MAPE = \frac{\sum_{t=1}^{n} \frac{|y_i - t_i|}{y_i}}{n} \times 100\% \quad (5)$$

Forecasting is successful when it produces a low MAPE [30].

5) Denormalization

In order to compare the predicted data with actual data, it is necessary to process the data denormalization first so that the previously normalized data return to the actual value. The denormalized data on the predicted data obtained from building the model using the best parameters will be compared with the actual data. If the predicted data are close to the actual data, then this model can be used to forecast COVID-19 cases in the future.

6) Forecasting

After the best model is obtained, then the model is used for forecasting. Forecasting is part of statistical modelling, widely used in various fields because of its benefits in decision-making. Forecasting aims to predict the future value of a specific variable that ranges over time using the previous value [31].

IV. RESULTS

A. Preprocessing

1) Visualization

The COVID-19 case data variables used in this study are only three: confirmed cases, recovered cases, and death cases, which began from March 19, 2020 to October 26, 2021. Then, data on confirmed cases, recovered cases and death cases from October 27 to January 24, 2022, will be compared with the forecast data obtained. The data that have been input will be visualized by making plots in Fig. 4 to see data patterns of confirmed cases, recovered cases, and death cases.

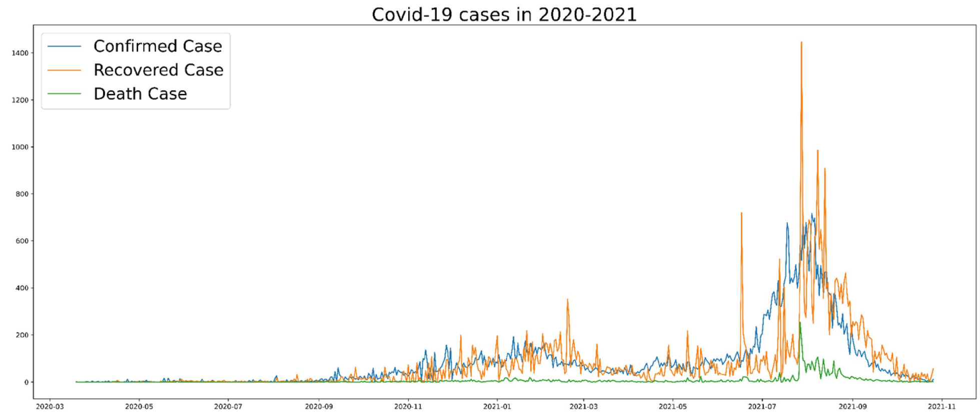


Fig. 4 Plot of COVID-19 data in Lampung Province

2) Check Missing Value

Before building an ANN, the data must be processed by looking at the presence or absence of a missing value. This process must be done to view the missing data and correct errors.

3) Data Transformation

Next, data transformation will be carried out using a standard scaler as shown in Table 2.

TABLE 2
RESULT OF DATA SCALING

	Confirmed Case	Recovered Case	Death Case
0	-0.6565246	-0.55131978	-0.3490311
1	-0.6644127	-0.55131978	-0.3490311
2	-0.6644127	-0.55131978	-0.3490311
3	-0.6644127	-0.55131978	-0.3490311
4	-0.6644127	-0.55131978	-0.3490311
…	…	…	…
582	-0.6249723	-0.29812062	-0.3490311
583	-0.5855319	-0.53685123	-0.2952141
584	-0.6565246	-0.49344567	-0.2952141
585	-0.6486365	-0.34152618	-0.2952141
586	-0.5776438	-0.14620110	-0.2413971

B. Training Model

This research will build an artificial neural network using K-fold validation, one input layer, one output layer, and the ReLU activation function, as well as several other parameters such as a hidden layer, batch size, epoch, learning rate, and dropout. To get the best combination of parameters, we apply hyper-tuning. In addition, this hyper-tuning process will use early stopping so that the model's learning process stops when the conditions are met. The following parameters will be used in the hyper-tuning process, and Table 3 shows the optimal parameters obtained.

1. Number of batch sizes: 16 and 32
2. Number of epochs: 50 and 100
3. Number of dropouts: 0.2 and 0.3
4. Value of learning rate: 0.001 and 0.01

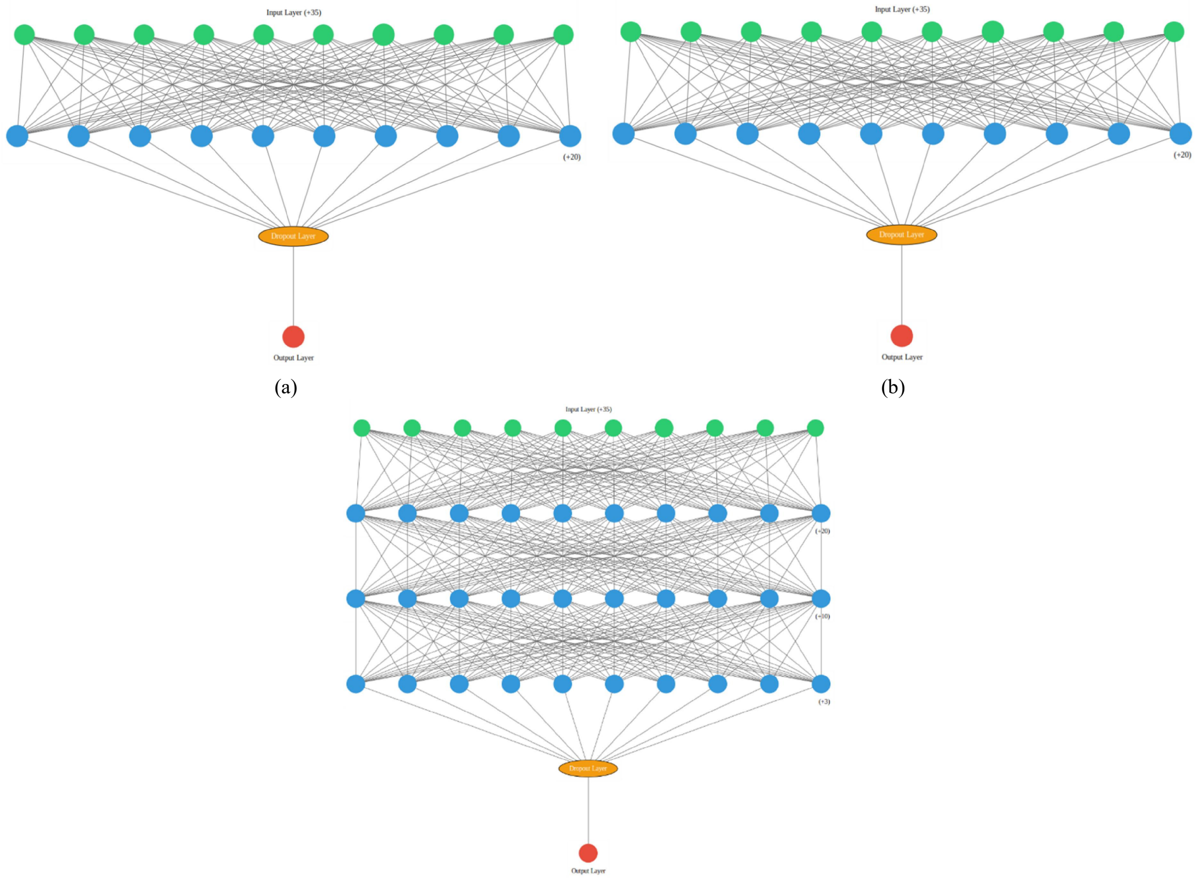
TABLE 3
OPTIMAL PARAMETER OF HYPERTUNING PROCESS

Parameter	Confirmed Case	Recovered Case	Death Case
k-fold	*k* (10 folds)	*k* (10 folds)	*k* (10 folds)
Batch size	16	32	32
Epochs	100	50	100
Learning rate	0.01	0.01	0.01
dropout	0.2	0.2	0.2
layer	*Input* : 1 *layer* (45 *units*) *Hidden* : 1 *layer* (30 *units*) *Output* : 1 *layer* (1 *units*)	*Input* : 1 *layer* (45 *units*) *Hidden* : 1 *layer* (30 *units*) *Output* : 1 *layer* (1 *units*)	*Input* : 1 *layer* (45 *units*) *Hidden* : 3 *layer* (30, 20, and 13 *units*) *Output* : 1 *layer* (1 *units*)

Fig. 5 shows an artificial neural network for confirmed cases, cases recovering, and cases of death in COVID-19. After getting the best parameters from hyper-tuning, testing the artificial neural network with k-fold validation will see each fold's MAPE value and accuracy, and the final result obtained is the average of all folds as shown in Table 4.

TABLE 4
THE ACCURACY VALUE OF THE ARTIFICIAL NEURAL NETWORK MODEL TEST RESULT WITH K-FOLD VALIDATION

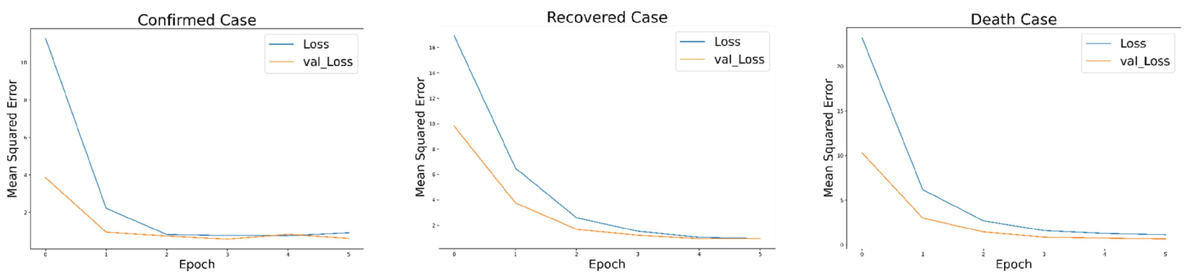
	Confirmed Case	Recovered Case	Death Case
Fold 1	98.79%	98.79%	98.88%
Fold 2	99.45%	98.61%	98.76%
Fold 3	99.51%	99.12%	98.92%
Fold 4	98.38%	96.49%	99.03%
Fold 5	97.18%	99.44%	98.68%
Fold 6	99.46%	99.21%	98.47%
Fold 7	99.47%	99.17%	99.18%
Fold 8	98.56%	98.90%	99.12%
Fold 9	99.32%	97.80%	99.07%
Fold 10	98.07%	97.64%	98.77%
Average	98.82%	98.52%	98.89%



(a) (b) (c)

Fig. 5 Structure of Artificial Neural Network (ANN) in (a) confirmed cases, (b) recovered cases, (c) death cases.

Then, how well the model has been built is tested by looking at the loss value and the validation loss, where the mean squared error function loss when the model is built. The visualization result can be seen in Fig. 6.

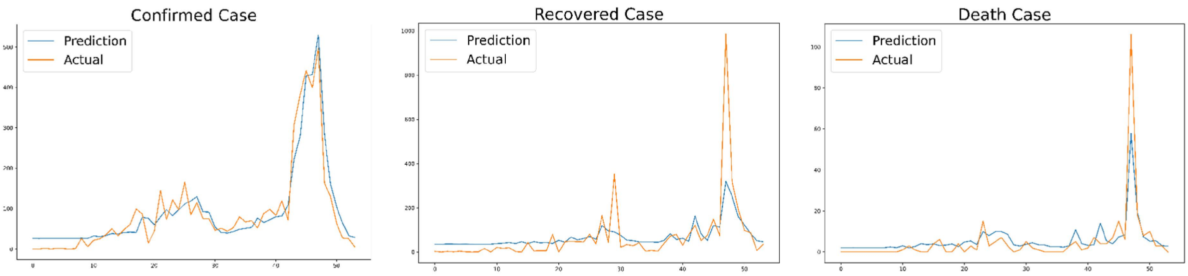


(a) (b) (c)

Fig. 6 Loss graphic for (a) confirmed cases, (b) recovered cases, and (c) death cases

C. Results of Predicted COVID-19 Cases in Lampung Province

Predicted data must first be denormalized so the previous data are normalized to the actual value. Denormalized data on predictive data obtained from model development using the best parameters will be compared with actual data. If the predicted data are close to the actual data, then this model can predict COVID-19 cases in the future. The result of comparison plot between actual data and predictive data can be seen in Fig. 7.



(a) (b) (c)

Fig. 7 Comparison plot of actual data with predictive data for (a) confirmed cases, (b) recovered cases, and (c) death cases

D. Evaluation of Artificial Neural Network with K-Fold Validation

The evaluation result is shown in Table 5.

TABLE 5
EVALUATION OF COVID-19 CASES IN LAMPUNG PROVINCE

Model Evaluation	Confirmed Case	Recovered Case	Death Case
RMSE	0.322	0.698	0.518
MAPE	1.18%	1.48%	1.11%
Accuracy	98.82%	98.52%	98.89%

E. Forecasting COVID-19 Cases in Lampung Province

The accuracy of the model built shows promising results. Then forecasting is carried out for the next 90 days. The following are the results of forecasting cases of COVID-19 in Lampung Province. The Fig. 8 is a graph of the overall actual data and forecasting data on confirmed cases, recovered cases, and cases of death based on Table 6.

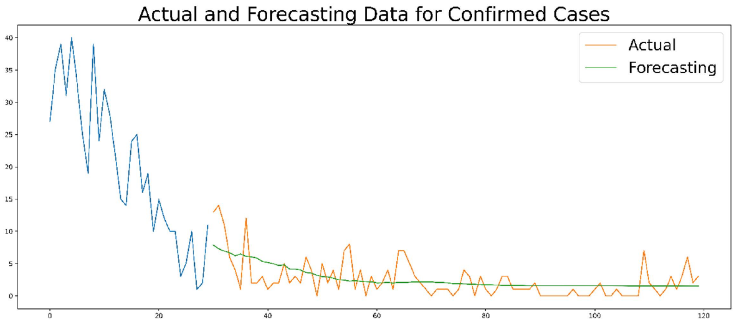
TABLE 6
FORECASTING COVID-19 CASES IN LAMPUNG PROVINCE

Date	Confirmed Case	Recovered Case	Death Case
27/10/2021	10.69	20.97	1.14
28/10/2021	10.45	10.40	1.10
29/10/2021	9.79	12.75	1.36
30/10/2021	9.84	16.12	1.25
31/10/2021	9.28	15.14	1.44
1/11/2021	9.02	13.47	1.10
2/11/2021	8.82	9.90	1.32
…	…	…	…
18/1/2022	0.94	0.65	0.18
19/1/2022	0.92	0.63	0.18
20/1/2022	0.91	0.64	0.17
21/1/2022	0.89	0.65	0.17
22/1/2022	0.88	0.65	0.16
23/1/2022	0.86	0.64	0.16
24/1/2022	0.85	0.62	0.15

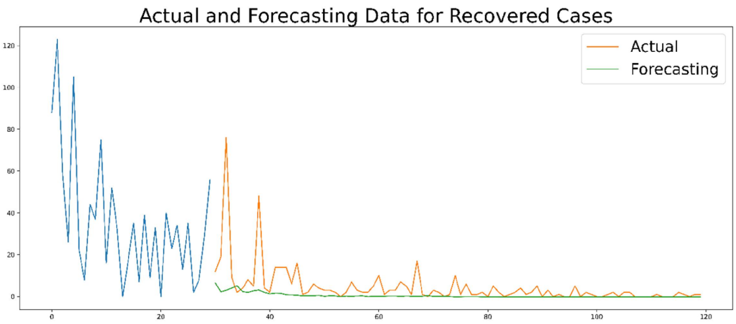
Based on Fig. 8, the forecast for confirmed, recovered, and death cases for October 27, 2021, to January 24, 2022, has decreased.

V. DISCUSSION

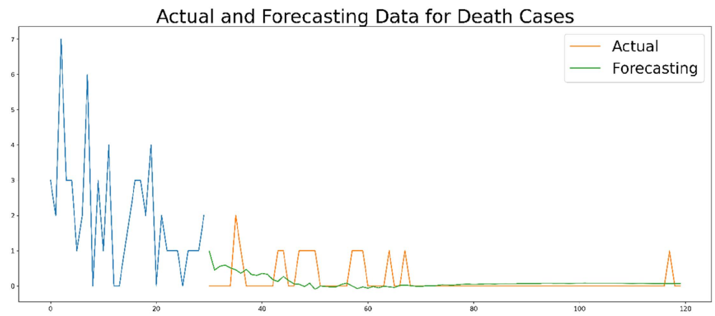
We compared the test results in this study with other research methods and results, as shown in Table 7. For example, Ardabili [12] used the ANN Hybrid Artificial Intelligence Method of ANN Trained with Gray Wolf Optimizer model for COVID-19 Global Prediction with a dataset based on time series with a total of 300,000 datasets. The results show that ANN-GWO obtains a MAPE value of 13.15%. Hasan [13] used the EEMD-ANN Hybrid Model on the cumulative global data of daily level information by country and obtained an RMSE value of 0.00559. Furthermore, a study by Dharani [14] using ANN based on Symptoms Relevance Survey and Analysis with 300,000 datasets obtained an accuracy of 66%. Samsani [17], in his research, tried to compare several machine learning approaches such as SVM, Adaboost, Naïve Bayes, and ANN using the Coswara Dataset developed by the Indian Institute of Science in Bangalore. As a result, the highest accuracy is obtained using the ANN method of 93%. Meanwhile, the prediction results from the proposed method obtained a MAPE value of 1.78% for confirmed cases, 1.92% for recovered cases and 0.95% for death cases, so it can be interpreted that these results have a good predictive ability with an accuracy value of 98.22% for cases confirmed cases, 98.08% for cured cases, and 99.05% for death cases. However, this study has a limited dataset, where the number of datasets is one of the determining factors.



(a)



(b)



(c)

Fig. 8 Plot of forecasting results with actual data on (a) confirmed cases, (b) recovered cases, and (c) death cases.

TABLE 7
PREDICTIVE MODEL PERFORMANCE

Reference Study	Model	MAPE	RMSE	Accuracy
Ardabili	ANN + Gray Wolf Optimizer	13.15%	NA	NA
Dharani	ANN + Symptoms Analysis	NA	NA	66 %
Hasan	EEMD + ANN	NA	0.00559	NA
Samsani	ANN, SVM, Adaboost, Naïve Bayes	NA	NA	93%
Proposed Method	ANN + K-Fold Cross--Validation	Confirmed case = 1.18% Recovered case = 1.42% Death case = 1.11%	Confirmed case = 0.322 Recovered case = 0.698 Death case = 0.518	Confirmed case = 98.82 % Recovered case = 98.52 % Death case = 98.89 %

VI. CONCLUSION

Applying the artificial neural network method with k-fold cross-validation to predict confirmed cases, recovered cases, and deaths in COVID-19 in Lampung Province resulted in MAPE values of 1.18%, 1.42%, and 1.11%, respectively. RMSE obtained a value of 0.322 for confirmed cases, 0.698 for cured cases, and 0.518 for death cases, so it can be interpreted that these results have better predictive ability than existing methods. The prediction accuracy for confirmed cases is 98.82%, 98.52% for recovered cases, and 98.89% for death cases. Experiments of the proposed method show that this method can also be used for forecasting. As a result, forecasts for confirmed, recovered, and death cases from October 27, 2021, to January 24, 2022, have decreased according to actual data.

Author Contributions: *Alifiah*: Methodology, Programming, Writing - Original Draft, Writing - Review & Editing, Supervision. *Kurniasari*: Conceptualization, Methodology, Writing - Reviewing & Editing. *Amanto*: Writing - Review & Editing. *Warsono*: Writing – Review & Editing.

Funding: This research received no specific grant from any funding agency.

Conflict of Interest: The authors declare no conflict of interest.

REFERENCES

[1] Y. Chen, Q. Liu and D. Guo, "Emerging Coronaviruses: Genome Structure, Replication, and Pathogenesis.," *Journal of Medical Virology,* vol. 92, no. 4, pp. 418-423, 2020.

[2] M. Z. Virtyani, S. Muljaningsih and K. Asmara, "Studi Peristiwa Penetapan COVID-19 Sebagai Pandemi Oleh World Health Organization Terhadap Saham Sektor Healthcare di Bursa Efek Indonesia," *Jurnal Sekuritas,* vol. 4, no. 3, 2021.

[3] Kementerian Kesehatan, "*Positif COVID-19 Usai Divaksinasi? Begini Penjelasan Komnas KIPI dan Kemenkes*," 2021.

[4] A. S. Bharatpur, *"A Literature Review on Time Series Forecasting Methods*," United Kingdom, 2022.

[5] S. Vartak and A. Sapre, "An Overview of Predictive Analysis: Techniques and Applications," *International Journal for Research in Applied Science & Engineering Technology (IJRASET),* vol. 8, no. 11, pp. 652-662, 2020.

[6] M. M. Santoni, N. Chamidah and N. Matondang, "Prediksi Hipertensi menggunakan Decision Tree, Naïve Bayes dan Artificial Neural Network pada software KNIME," *IEEE Transactions on Computers,* vol. 19, pp. 352-363, 2020.

[7] G. Romadhona, B. D. Setiawan and F. A. Bachtiar, "Prediksi Produktivitas Padi Menggunakan Jaringan Syaraf Tiruan Backpropagation," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer,* vol. 2, no. 12, pp. 6048-6057, 2018.

[8] S. B. Stoecklin, P. Rolland, Y. Silue, A. Mailles, C. Campese, A. Simondon, M. Mechain, L. Meurice, M. Nguyen, C. Bassi, E. Yamani, S. Behillil, S. Ismael, D. Nguyen, D. Malvy, F. X. Lescure, S. Georges, C. Lazarus, A. Tabai, M. Stempfelet, V. Enouf, B. Coignard and D. Levy-Bruhl, "First cases of coronavirus disease 2019 (COVID-19) in France: surveillance, investigations and control measures, January 2020," *Eurosurveillance,* vol. 25, no. 6, 2020.

[9] K. Kakimoto, H. Kamiya, T. Yamagishi, T. Matsui, M. Suzuki and T. Wakita, "Initial Investigation of Transmission of COVID-19 Among Crew Members During Quarantine of a Cruise Ship — Yokohama, Japan, February 2020," *Morbidity and Mortality Weekly Report,* vol. 69, no. 11, pp. 312-313, 2020.

[10] A. Rajesh, H. Pai, V. Roy, S. Samanta and S. Ghosh, *"CoVID-19 Prediction for India from The Existing Data and SIR(D) Model Study*," medRxiv. Cold Spring Harbor Laboratory Press, 2020.

[11] A. Hasan, H. Susanto, M. F. Kasim, N. Nuraini, B. Lestari, D. Triany and W. Widyastuti, "Superspreading in early transmissions of COVID-19 in Indonesia," *Scientific Reports,* 2020.

[12] S. Ardabili, A. Mosavi, S. S. Band and A. R. Varkonyi-Koczy, "Coronavirus Disease (COVID-19) Global Prediction using Hybrid Artificial Intelligence Method of ANN Trained with Grey Wolf Optimizer," in *2020 IEEE 3rd International Conference and Workshop in Óbuda on Electrical and Power Engineering (CANDO-EPE)*, Budapest, Hungary, 2020.

[13] N. Hasan, "A Methodological Approach for Predicting COVID-19 Epidemic Using EEMD-ANN Hybrid Model," *Internet of Things,* vol. 11, 2020.

[14] D. N and K. K. G, "ANN-based COVID -19 Prediction and Symptoms Relevance Survey and Analysis," in *Proceedings of the Fifth International Conference on Computing Methodologies and Communication (ICCMC 2021)*, Erode, India, 2021.

[15] H. R. Niazkar and M. Niazkar, "Application of Artificial Neural Networks to Predict the COVID-19 Outbreak," *Global Health Research and Policy,* vol. 5, no. 50, 2020.

[16] A. Prakash, P. Sharma, I. K. Sinha and U. P. Singh, "Spread & Peak Prediction of Covid-19 using ANN and Regression," in *2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM)*, New Delhi, India, 2020.

[17] S. Samsani and S. S. Ankadi, "Covid-19 Prediction Using Artificial Intelligence Techniques," *International Journal for Research in Applied Science & Engineering Technology (IJRASET),* vol. 10, no. 6, pp. 3543-3546, 2022.

[18] K. N. A. Halim, A. S. M. Mohd Jaya and A. F. A. Fadzil, "Data Preprocessing Algorithm for Neural Network Binary Classification Model in Bank Tele-Marketing," *International Journal of Innovative Technology and Exploring Engineering,* vol. 9, no. 3, pp. 272-277, 2020.

[19] J. F. N. Santoso, A. Setiawan and S. Rostianingsih, "Perbandingan dan Analisis Metode Artificial Neural Network dan SIRD pada Kasus Covid-19 di Surabaya," *Jurnal Infra,* vol. 9, no. 1, 2021.

[20] J. T. Heaton, I*ntroduction to Neural Networks for Java*, Florida: Heaton Research Inc, 2008.

[21] A. C. Sitepu and M. Sigiro, "Analisis Fungsi Aktivasi Relu dan Sigmoid menggunakan Optimizer SGD dengan Representasi MSE pada Model Backpropagation," *JUTISAL (Jurnal Teknik Informatika Komputer Universal),* vol. 1, no. 1, pp. 12-25, 2021.

[22] U. Michelucci, *Applied Deep Learning: A Case-Based Approach to Understanding Deep Neural Networks*, California: Apress Berkeley, 2018.

[23] N. Srivastava, G. Hinton, A. Krizhevsky and I. Sutskever, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *The Journal of Machine Learning Research,* vol. 15, no. 1, pp. 1929-1958, 2014.

[24] V. Vijayalakshmi and K. Venkatachalapathy, "Deep Neural Network for Multi-Class Prediction of Student Performance in Educational Data," *International Journal of Recent Technology and Engineering (IJRTE),* vol. 8, no. 2, pp. 2277-3878, 2019.

[25] N. Rochmawati, H. B. Hidayati, Y. Yamasari, H. P. A. Tjahyaningtijas, W. Yustanti and A. Prihanto, "Analisa Learning Rate dan Batch Size pada Klasifikasi Covid menggunakan Deep Learning dengan Optimizer Adam," *JIEET (Journal Information Engineering and Educational Technology),* vol. 5, no. 2, pp. 44-48, 2021.

[26] C. D. Suhendra and A. C. Saputra, "Penentuan Parameter Learning Rate selama Pembelajaran Jaringan Syaraf Tiruan Backpropagation menggunakan Algoritma Genetika," *Jurnal Teknologi Informasi : Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika,* vol. 14, no. 2, pp. 202-212, 2020.

[27] J. W. G. Putra, *Pengenalan Konsep Pembelajaran Mesin dan Deep Learning*, Edisi 1.4, Tokyo, 2020.

[28] J. E. Hanke and D. W. Wichern, *Business Forecasting*, New Jersey: Prentice Hall, 2014.

[29] N. Lestari and N. Wahyuningsih, "Peramalan Kunjungan Wisata dengan Pendekatan Model SARIMA (Studi kasus : Kusuma Agrowisata)," *Jurnal Sains dan Seni ITS,* vol. 1, no. 1, pp. 29-33, 2012.

[30] P. C. Chang, C. H. Liu and Y. Y. Wang, "The Development of a Weighted Evolving Fuzzy Neural Network for PCB Sales Forecasting," *Expert Systems with Applications,* vol. 32, pp. 86-96, 2007.

[31] J. Patil and N. Shinde, "Time Series Weather Forecasting Techniques: Literature Survey," *International Research Journal of Engineering and Technology (IRJET),* vol. 9, no. 6, pp. 3097-3100, 2022.