International Journal of Advances in Intelligent Informatics
Vol. 8, No. 1, March 2022, pp. 69-83 ISSN 2442-6571

The mortality modeling of covid-19 patients using a combined time series model and evolutionary algorithm

Imam Tahyudin [a,1,*], Rizki Wahyudi [b,2], Wiga Maulana [c,3], Hidetaka Nambo [d,4]

[a] Department of Information System, Universitas Amikom Purwokerto, Purwokerto 53127, Indonesia
[b] Department of Informatics, Universitas Amikom Purwokerto, Purwokerto 53127, Indonesia
[c] Department of Information Technology, Universitas Amikom Purwokerto, Purwokerto 53127, Indonesia
[d] Artificial Intelligence laboratory, Kanazawa University, Japan
[1] imam.tahyudin@amikompurwokerto.ac.id; [2] rizkiw@amikompurwokerto.ac.id; [3] wiga@amikompurwokerto.ac.id;
[4] nambo@blitz.ec.t.kanazawa-u.ac.jp
* corresponding author

ARTICLE INFO

Selected paper from The 2020 3rd International Symposium on Advanced Intelligent Informatics (SAIN'20), Nanjing-China (Virtually), 25-26 November 2020, http://sain.ijain.org/2020/. Peer-reviewed by SAIN'20 Scientific Committee and Editorial Team of IJAIN journal.

Article history
Received April 29, 2021
Revised June 7, 2021
Accepted November 21, 2021
Available online March 31, 2022

Keywords
Time Series Analysis
Evolutionary Algorithm
SARIMA-PARCD
COVID-19 Patients
LSTM

ABSTRACT

COVID-19 pandemics for as long as two years ago since 2019 gives many insights into various aspects, including scientific development. One of them is the fundamental research of computer science. This research aimed to construct the best model of COVID-19 patients' mortality and obtain less prediction errors. We performed the combination methods of time series, SARIMA, and Evolutionary algorithm, PARCD, to predict male patients who died because of COVID-19 in the USA, containing 1.008 data. So, this research proposed that SARIMA-PARCD has a powerful combination for addressing the complex problem in a dataset. The prediction error of SARIMA-PARCD was compared with other methods, i.e., SARIMA, LSTM, and the combination of SARIMA-LSTM. The result showed that the SARIMA-PARCD has the smallest MSE value of 0.0049. Therefore, the proposed method is competitive to implement in other cases with similar characteristics. This combination is robust for solving linear and non-linear problems.

1. Introduction

The time series method is used in many fields, including the economic sector for unemployment and hospital financial turnover datasets [1]–[3]. One of the discussion topics in time series methods is the ARIMA model. ARIMA is generally divided into non-seasonal (ARIMA) and seasonal ARIMA (SARIMA) [4]. ARIMA has been used to complete data modeling of potential bioelectric plants. However, some of these models have not yielded the best value. The mean error rate for the mean square error (MSE) and mean absolute error (MAE) is still high. Meanwhile, the average prediction accuracy is 75% [5]–[7]. Authors have been conducting many studies to improve accuracy or gain less error forecasting. For example, the implementation of ARIMA for solving the COVID-19 case has been used for predicting the Johns Hopkins epidemiological trend of the prevalence and incidence of COVID-19 [8], [9].

Furthermore, the ARIMA model was used to estimate COVID-19 prevalence in Italy, Spain, and France in the same case. This study performed that ARIMA (0,2,1) has the lowest MAPE values by 4.7520 [10]. The hybrid method with other methods has been conducted to improve the accuracy, such as the hybrid SSA-TSR-ARIMA for water demand forecasting. And then, the combination of ARIMA with RBFNN for short-term Forecasting uses tourist arrival to Indonesia. Combined with other methods, ARIMA showed accuracy improvement [11], [12]. Those discussions still have a problem because it does not cover the seasonal problem and when the data is non-linear.

Meanwhile, the implementation of SARIMA has already been conducted to estimate human position modeling using the bioelectric potential of the plant [4]. Then, the SARIMA model was compared with the LSTM model for modeling the data of the bioelectric potential of the plant. The SARIMA model performed better than the LSTM method [7]. To improve the SARIMA model's accuracy, some authors conduct experiments and observations such as using the combination method or optimization method. This method also has a drawback because it was robust for solving a linear problem not for non-linear ones. Therefore, to address this problem some authors tried to combine SARIMA with other methods.

Among them performed the combination of the SARIMA method with the SVM method to predict the production value of the machine industry in Taiwan [13]. The results obtained that the accuracy of the SARIMA - SVM hybrid is better than with each method—another study hybrid ARIMA with ANN for forecasting pollution index in cities in Southeast Asia [14]. Furthermore, the study used the same method to predict tourists arriving at Minangkabau international airport [15]. Moreover, for modeling and evaluating the data from a residential district in Berlin [16]. The results showed that the accuracy of the hybrid method is better. The next research compares the SARIMA method and ANN to predict the power absorption in Turkey's electricity use [17]. The results obtained show the results of observations for 12 weeks that the MAPE value of the ANN method is 1.8% better than SARIMA with a MAPE value of 2.6%. However, the result is the opposite in certain conditions, such as after a holiday. Another study combines SARIMA with SVM and then analyzes using clustering [18]. This research is used to predict passengers at North Iran Station. The result is that a combination of these methods is better than alone. Moreover, the SVR-SARIMA model's combination was conducted to forecast tourists to determine the best model using the PROMETHEE II decision support system method. The result is the same; the proposed model is better than the baselines [19]. The other study compares SARIMA with exponential smoothing (ETS) models. This research explores and predicts accurate hemorrhagic conjunctivitis's demographic and distributive features (AHC). The results showed that RMSE and MAPE the SARIMA (0,1,1) (2,0,0)12 model is lower than ETS (M,N,M) [20]. In addition, the study for improving the accuracy of SARIMA has been conducted using Genetic algorithm optimization. This research was used to forecast the Singapore tourist arrivals to Malaysia. The result presented that the MAPE, MAE, and MSE of GA-SARIMA (0,1,1)(1,0,0) was lowest than SARIMA (0,1,1)(1,0,0) [21]. Those observations already covered the complex problem; however, the error predictions are still more than 1. Therefore, we proposed the combination of SARIMA-PARCD to obtain less error prediction.

The proposed method is compared with other methods, such as the LSTM method. The implementation of LSTM to predict the COVID-19 case has been presented in India. They use LSTM and its variant to predict and analyze COVID-19 positive cases for 32 states and union territories. Based on their research obtained, the accuracy of bi-directional LSTM is better than others which the error is less than 3% [22]. Moreover, the implementation of LSTM was conducted to address dynamic resource scaling and power consumption. The high accuracy was successfully obtained by reducing RMSE up to 3.17x10-3[23]. The other research tried to optimize the LSTM method by using some metaheuristic algorithms such as Harmony Search (HS), Gray Wolf Optimizer (GWO), Sine Cosine (SCA), and Ant Lion Optimization algorithms (ALOA). The study presented that LSTM, which ALOA optimizes, is better than others [24]. The other author tried to combine LSTM with CNN to improve the accuracy. This combination confirmed that the feature fusion LSTM-CNN model performs better than the single models (SC-CNN and ST-LSTM) in predicting stock prices [25].

Furthermore, the robust model of LSTM was proved when it was compared with ARIMA and NARNN methods. The LSTM approach is much higher than both when used to forecast COVID-19 cases in European countries [26]. Additionally, the combination of LSTM with CNN successfully obtained a high accuracy of 0.92 for generating images of Bangla text [27]. The study, a combination of the SARIMA and PARCD methods, has not been done before. The SARIMA model successfully predicts a person's position for the linear data set type, which is better than the deep learning method

(LSTM method) [7], and also the SARIMA model has been tested with a reasonably high accuracy of about 80% [4].

Meanwhile, the PARCD method can also estimate a person's position with a non-linear data set with an accuracy of about 75% [28]; hence, In this paper, we proposed a combined SARIMA time series model and PARCD evolutionary algorithm (SARIMA-PARCD) that aims to solve for linear and non-linear problems. The proposed model accuracy is compared to the other baseline models, SARIMA, LSTM, and SARIMA-LSTM.

2. Method

2.1. SARIMA Model

Seasonal ARIMA capital or SARIMA is a model or pattern that constantly repeats itself at certain intervals. For the stationary Dataset, seasonality can be detected from the ACF plot. If the ACF visualization shows a seasonal pattern, it will be done with different solutions. In general, the seasonal ARIMA is shown in (1).

$$ARIMA(p,d,q)(P,D,Q)^S \tag{1}$$

where (p,d,q) is the non-seasonal ARIMA model index, while (P,D,Q) is the seasonal ARIMA model, and S is a number of periods in a seasonal model. For example, if ARIMA (1,0,0), then the model follows equations 2 and 3 below:

$$(1-\emptyset_1 B)Y_t = c \tag{2}$$

Where $BY_t = Y_{t-1}$. So that

$$Y_t = c + \emptyset_1 Y_{t-1} \tag{3}$$

Several graphical techniques are used to detect seasonal datasets, such as sequential plots, seasonal subseries plots, multiple box plots, and autocorrelation plots. This study will use an autocorrelation plot to detect seasonality. One of the solutions for this autocorrelation plot is using seasonal differential operators [4].

2.2. LSTM

This algorithm is one of the kinds of RNN methods. The RNN method addresses the problem of a data sequence. It can connect the previous information to the current time and predict the next event. RNN has a network with loops, allowing information to persist. There are repeating modules of neural networks. In standard RNN, the repeating modules have a very simple structure, such as a single tanh layer. However, RNN cannot determine the next result in some cases when the gap is very large. The possibility of the next word regarding the recent information is the name location, but if we want to decide that word, what kind of location is complicated. RNN became unable to connect the information because there was some similar information in the previous time. Therefore, LSTM is a solution to handle this drawback. LSTM has a capability for learning long-term dependencies. Remembering information for long periods is the default behavior. LSTM has an internal mechanism which is cell states and gates. This mechanism can organize the memory in each input. The first process in LSTM is determining the unused value (forget gate). Decided the delete information from the cell gate by using formula (4). Which is h_{t-1} and x_t will pass the sigmoid gate (5). The forget gate will process h_{t-1} and x_t as the input, and it will obtain the output of 0 and 1 values on the cell gate C_{t-1}. [7], [29].

$$(x) = 1/(1 + exp^{-x}) \tag{4}$$

Where x is input data, and exp is the mathematical Constanta (2.7182818).

$$f_t = \sigma(W_f . [h_{t-1,} x_t] + b_f) \quad (5)$$

where σ is sigmoid function, W_f is weight value for forget gate, h_{t-1} is output value before order t, x_t is input value on the order t, and b_f is bias value on forget gate. The value of weight is explained in formula (6).

$$W = (-1/d^{1/2}), 1/d^{1/2} \quad (6)$$

After that, decide on the new information saved in the cell state formula. Two components are input gates for determining the renewal value. Then, the tanh makes the new vector value C_t for adding to cell state (8). The next step is that both of them are combined to renew the cell state. The formula (7) is the input gate.

$$i_t = \sigma(W_i . [h_{t-1,} x_t] + b_i) \quad (7)$$

where it is input gate, σ is the sigmoid function, W_i is weight value for the input gate, $ht - 1$ is the output value before order t , xt input value on order t, and bi is the bias value on the input gate.

$$C_t = \tanh\ (W_{c.}[h_{t-1,} x_t] + b_c) \quad (8)$$

where C_t is the new value that adds to cell state, tanh is tanh function, $W_{c.}$ Is weight value on cell state, $h_{t-1,}$ is output value before order t, $x_t$is input value on order t, and b_c is bias value on cell state. The formula (9) is the cell state.

$$C_t = f_t * C_{t-1} + i_t * C_t \quad (9)$$

where cell state is denoted by the C_t, f_t for forget gate, C_{t-1} as a cell state before order t, i_t is input gate, and C_t is the new value for adding cell state. The next step is determining the output value (output gate).

$$O_t = \sigma(W_{o.}[h_{t-1,} x_t] + b_o) \quad (10)$$

$$h_t = O_t * \tanh\ (C_t) \quad (11)$$

where O_t is output gate, σ is sigmoid function, W_o as a weight value for the output gate, $h_{t-1,}$ is output value before order t, x_t is input value on order t, b_o is bias value on output gate, h_t is output value on order $t, tanh$ is tanh function, and the new value for adding output gate is denoted by the C_t.

2.3. PARCD

This method extends the PSO method by nominating it with the Cauchy distribution to solve the numerical data association analysis problem. The aim is to prevent the premature search for optimal values because they are trapped in optimal local values. This method uses the concept of PSO, but a modification process is carried out on the velocity equation by including the Cauchy distribution [28], [30]. The velocity function can be seen from function (12) as follows.

$$V_i(t+1) = \omega(t)V_i(t) + C1_{rand}()(pBest - X_{i(t)} + C_{2rand}()(gBest - X_{i(t)}) \quad (12)$$

The next step is the normalization process using the resulting value of $V_i(t+1)$ as seen in the following equation.

$$U_i(t+1) = \frac{V_i(t+1)}{\sqrt{V_i1(t+1)^2 + V_i2(t+1)^2 \cdots + V_iK(t+1)^2}} \quad (13)$$

After that, the normalization process results are multiplied by the value of Cauchy's random variable.

$$S_i(t+1) = U_i(t+1) \cdot tan\left(\frac{\pi}{2} \cdot rand[0,1]\right) \quad (14)$$

Then the velocity value combined with the Cauchy distribution is used to determine the particle's current position as follows.

$$X_i(t+1) = X_i(t) + S_i(t+1) \quad (15)$$

2.4. Combination of SARIMA - PARCD

The steps for this combination are described in Fig. 1. The Dataset used is data on the mortality of COVID-19 patients in the USA. This step is pre-processing Dataset for the cleaning process. Furthermore, this process is checked using the Box-Jenkins process. The analysis step of the fit Dataset uses the SARIMA model. The process carried out includes the identification process. Namely, the process for determining stationary data. Then, check the seasonal Dataset by looking at the ACF and PACF values. If the Dataset is seasonal, the next step is to determine the best SARIMA model.

Furthermore, the fit SARIMA model is continued to use the forecasting process for linear components. While the residual data is used for the forecasting process using the PARCD method for non-linear components. The final step is the combined result of forecasting the two methods. Finally, the MSE values of SARIMA PARCD are compared with other methods.

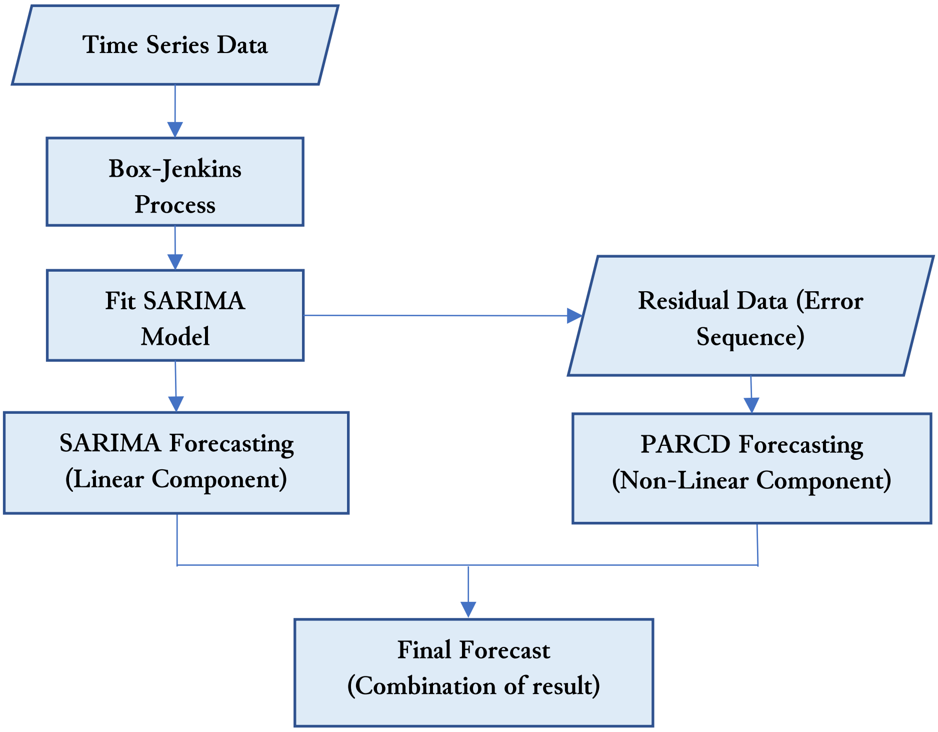

Fig. 1. The process of combining SARIMA-PARCD

3. Results and Discussion

3.1. Implementation of the SARIMA model

In implementing the SARIMA method, data on Covid-19 patients who died based on sex in the United States were used. The data is obtained from this link: https://www.genderscilab.org/gender-and-sex-in-covid19 (accessed on Sept 10th, 2020). The data consists of confirmed cases and mortality in the count, percentage, and crude rate for males and females. The data come from 53 country states in the USA from Apr 13th, 2020 to 24th Augustus 2020. To implement the SARIMA-PARCD method, we used only the Male Dataset. The amount of data used is 1008 datasets; then, the blank data is deleted because there are many missing values. The clean data are 711 datasets. The analysis process uses R software and Minitab.

3.1.1. Datasets

The visualization of male covid-19 patients mortality in the USA is shown in Fig. 2. The data which used for four months, from April to August 2020. Generally in every month, two times reached a peak number at the half and the end of the month. The average number is around 14.000 people.

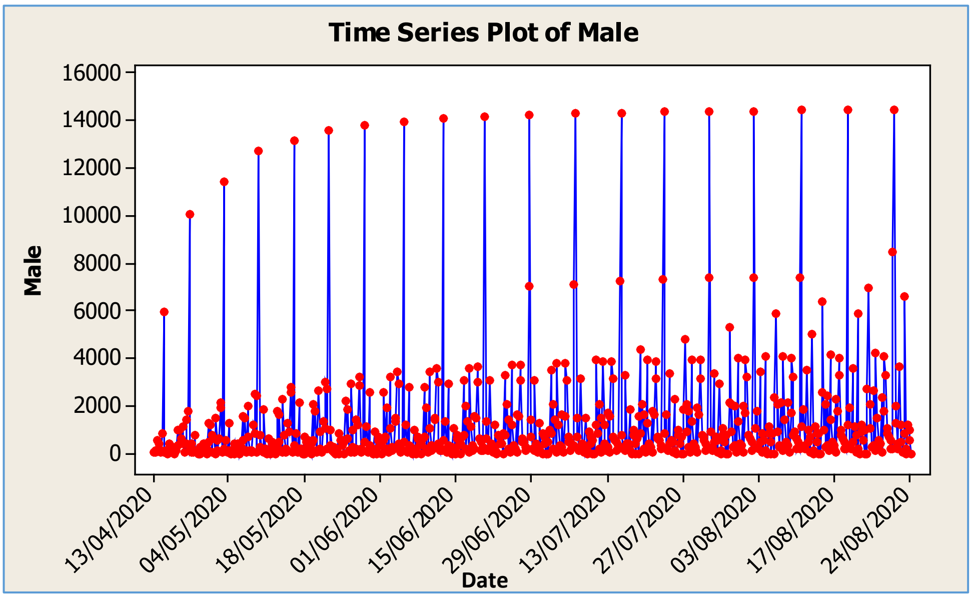

Fig. 2. Visual Dataset of male covid-19 patients who died in the USA

Furthermore, the Stationary Check of the male Covid-19 patient dataset. Check stationary against variety with the Box cox test; the results obtained are as Fig. 3. This process ensures that the data is fixed on the mean or the variance with 95% confidence and obtained lower CL is 0.06, and upper CL is 0.15.

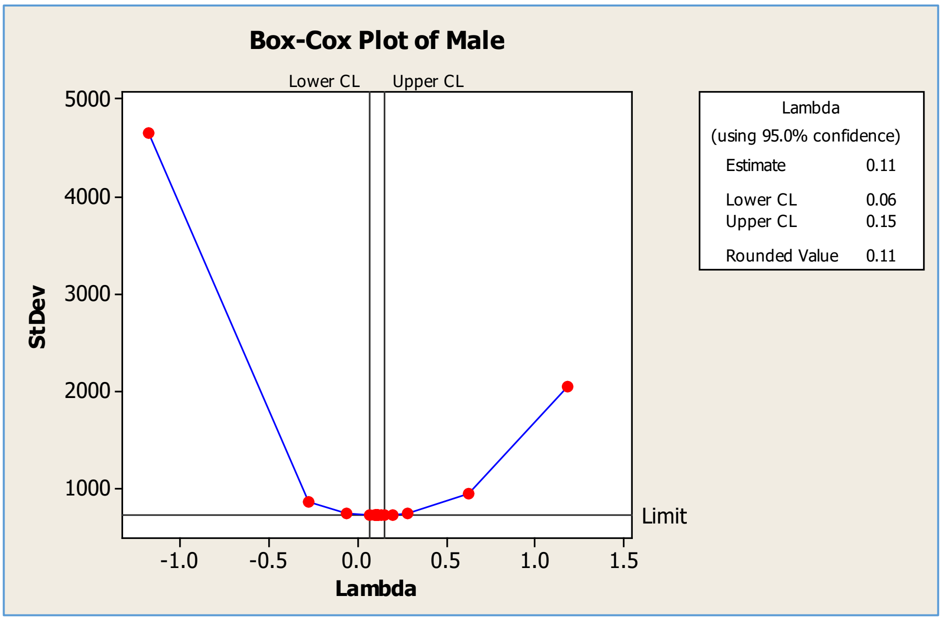

Fig. 3. Visual Dataset of male covid-19 patients who died in the USA

Because the value is still 0.11 (less than 1), the transformation process is carried out as in Fig. 4. After this process, performed the good result which the lower CL was 0.55 and the upper CL was 1.39. Finally, the value reached 1.

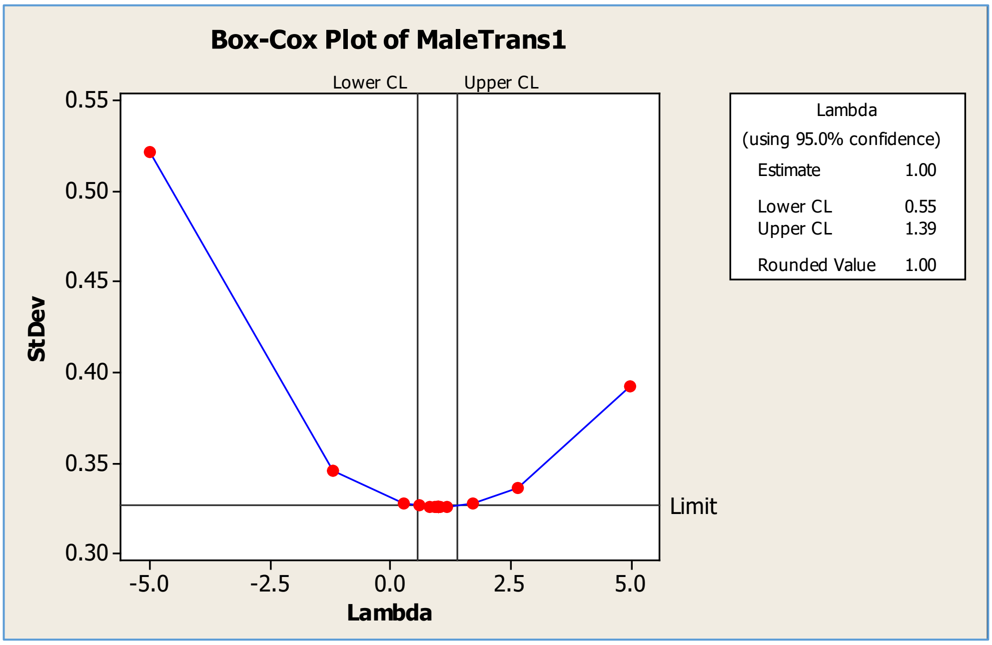

Fig. 4. Box cox transformation

Because the value of lambda is equal to 1, it is stationary. Then it is checked against the average stationery by looking at the ACF and PACF values. Check stationary against the mean by checking the ACF and PACF values. Based on the ACF (Fig. 5) and PACF (Fig. 6), lag 1-3 is still within the significance interval. Then it has been declared stationary to the average.

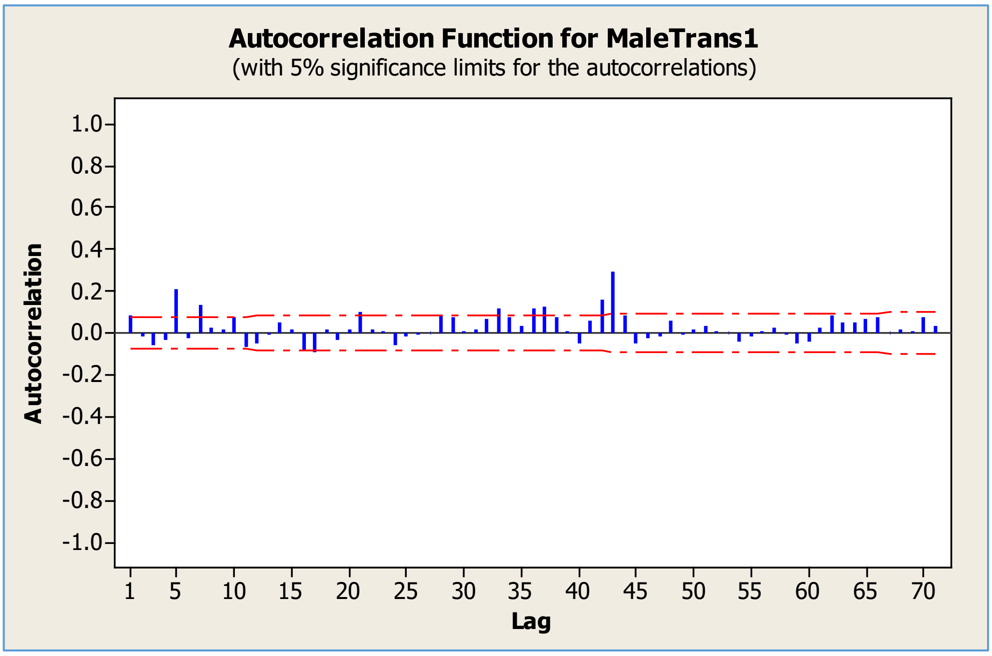

Fig. 5. ACF

Based on Fig. 6, it can be seen that there is a trend. Namely, an increase in the number of Covid-19 patients who died per 10 datasets. Therefore, a further identification process is carried out utilizing differencing.

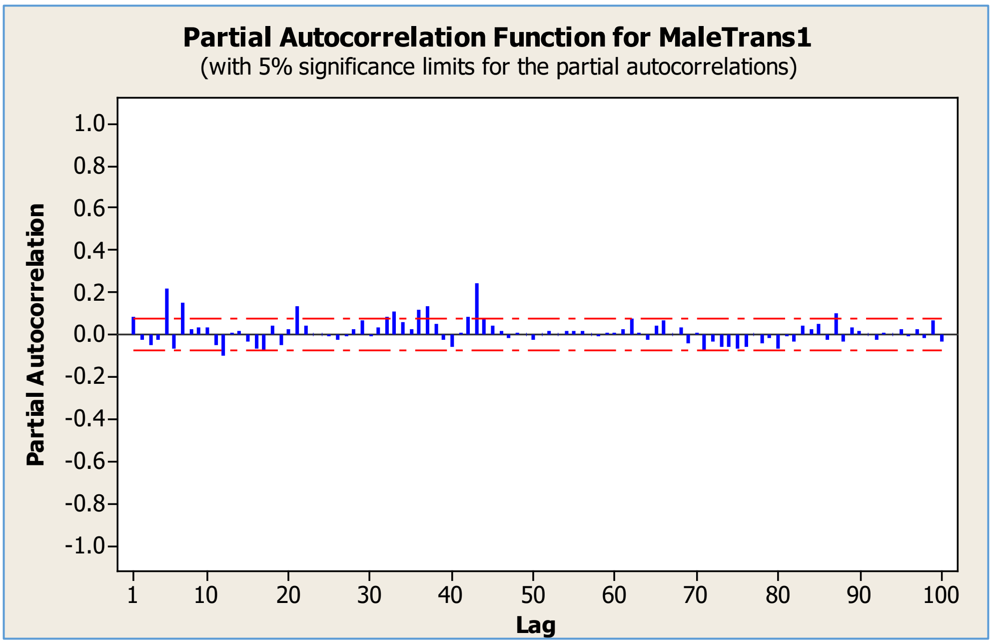

Fig. 6. PACF

3.1.2. Estimation

Fig. 7 shows the time series plot of the male and female Dataset. However, we only use male of covid-19 patients in the USA for performing the proposed method.

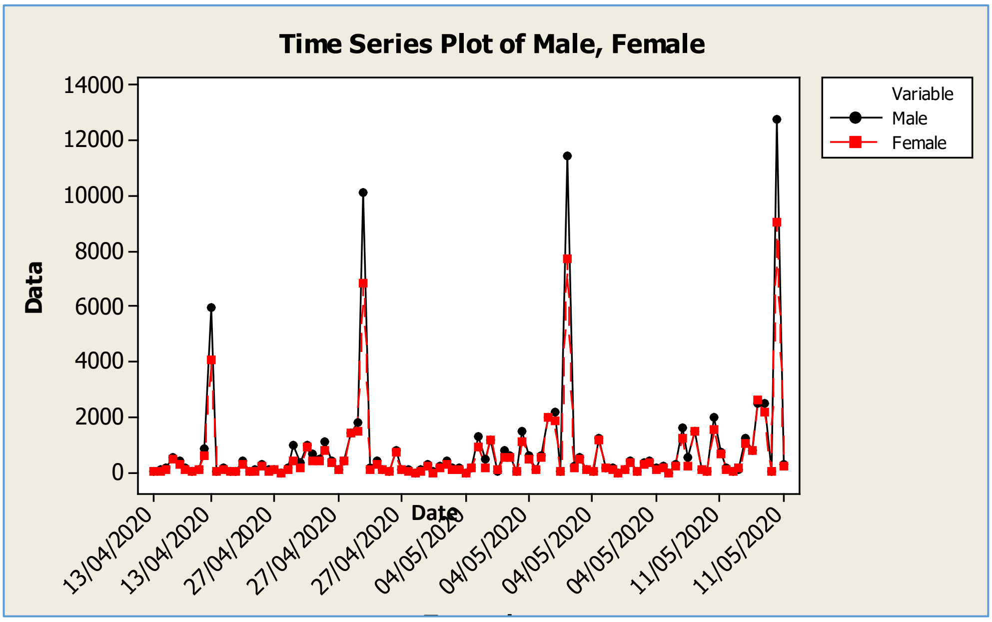

Fig. 7. Dataset of covid-19 male patients in the USA

After knowing the trend of the first differencing process, lag = 1, the formation of the model was made with a non-seasonal model first by checking the autocorrelation function (ACF) and the partial autocorrelation function (PACF) graph. We Obtained ACF and PACF graphs as in Fig. 8 and Fig. 9.

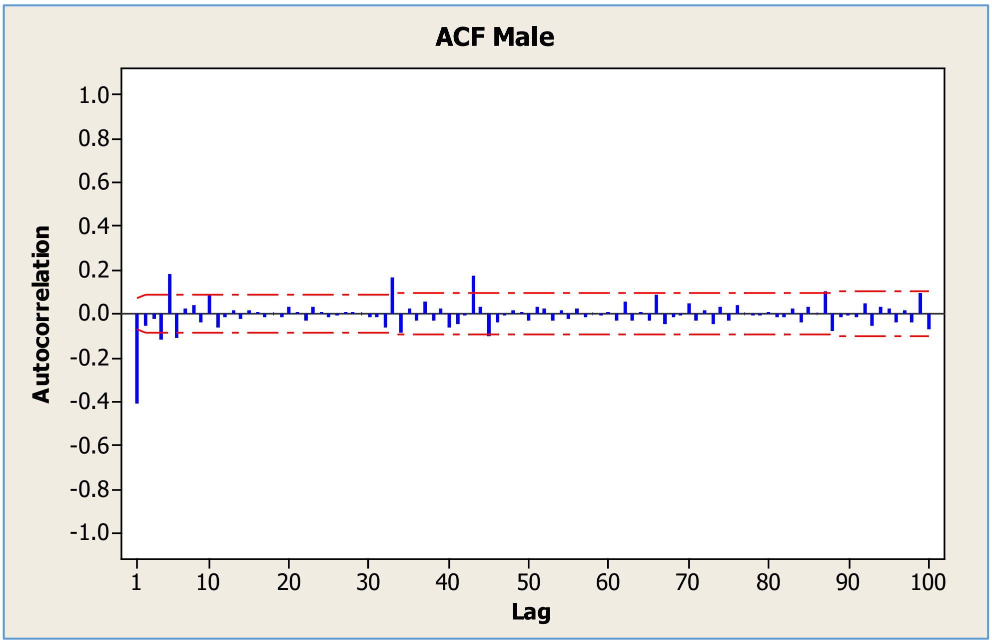

Fig. 8. ACF graph of male covid-19 patients in the United States, non-seasonal ARIMA model

The ACF graph in Fig. 8 shows a dying down of 5 lags, while the PACF (Fig. 9) chart shows a cut-off pattern. Thus, the non-seasonal ARIMA model formed is ARIMA (5,1,0). Then check the non-seasonal ARIMA model. The steps to determine the seasonal ARIMA model are the same as those used in the search process for the best model in non-seasonal ARIMA, namely by determining the ACF and PACF charts. The seasonal value determines the difference. In this case, it is ten because in every 10 datasets, there is a significant increase Covid-19 patients mortality (Fig. 10).

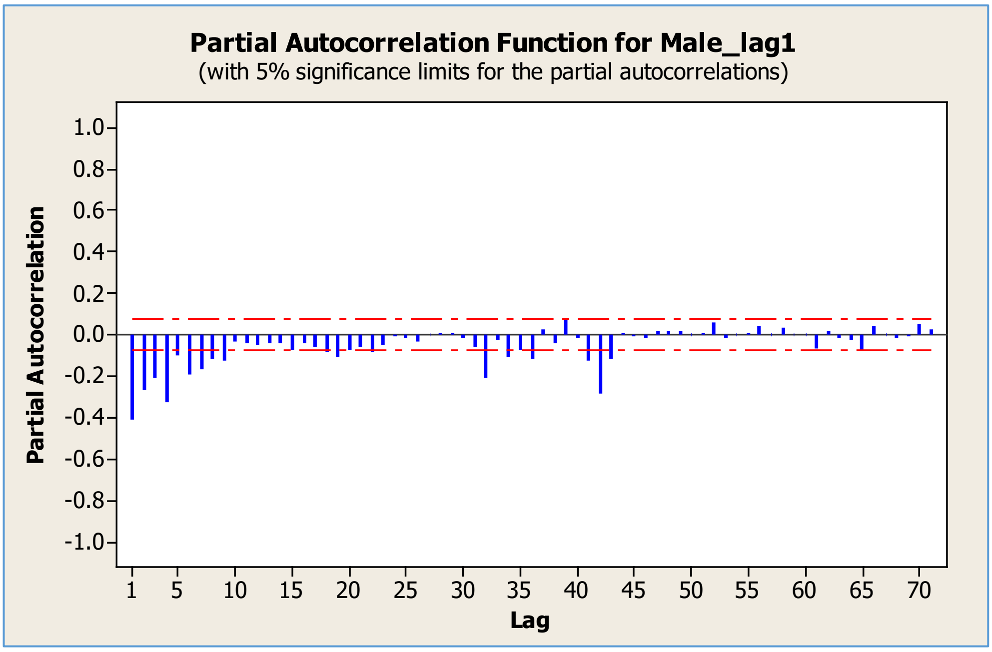

Fig. 9. PACF graph of male covid-19 patients in the United States, non-seasonal ARIMA model

Based on Fig. 10 and Fig. 11, information on the ACF graph shows that a cut-off occurs after lag 1, which occurs in the PACF graph. The best model we get is ARIMA (5,1,0) $(1,1,0)^{10}$.

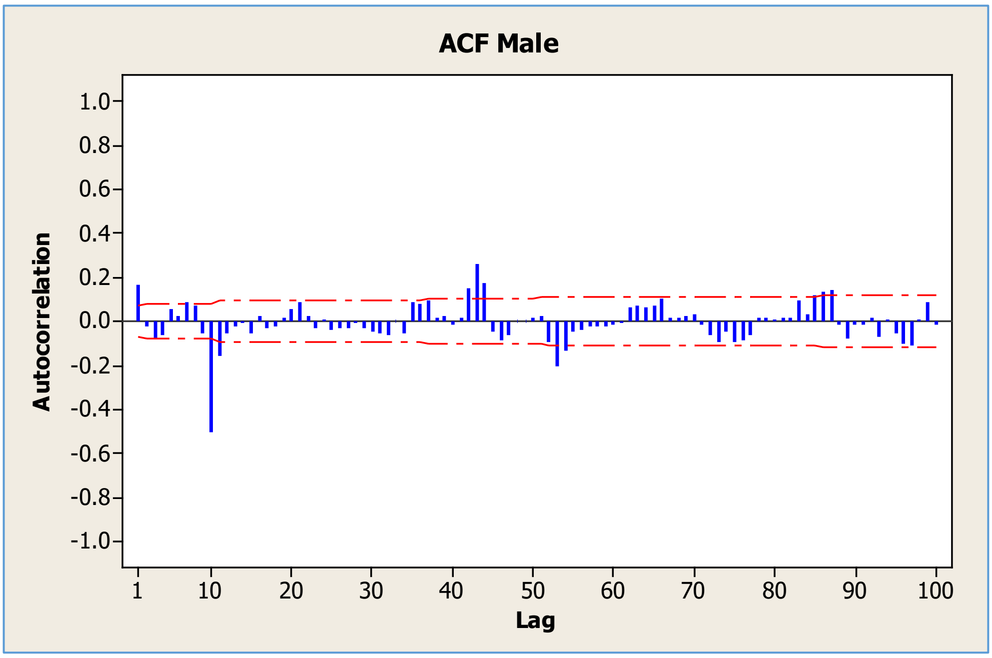

Fig. 10. ACF graph of male covid-19 patients with seasonal ARIMA model

3.1.3. Model Evaluation

At this stage, the error value and other values are checked. Based on the results of the software output, the following evaluation values are obtained. Final Estimates of Parameters.

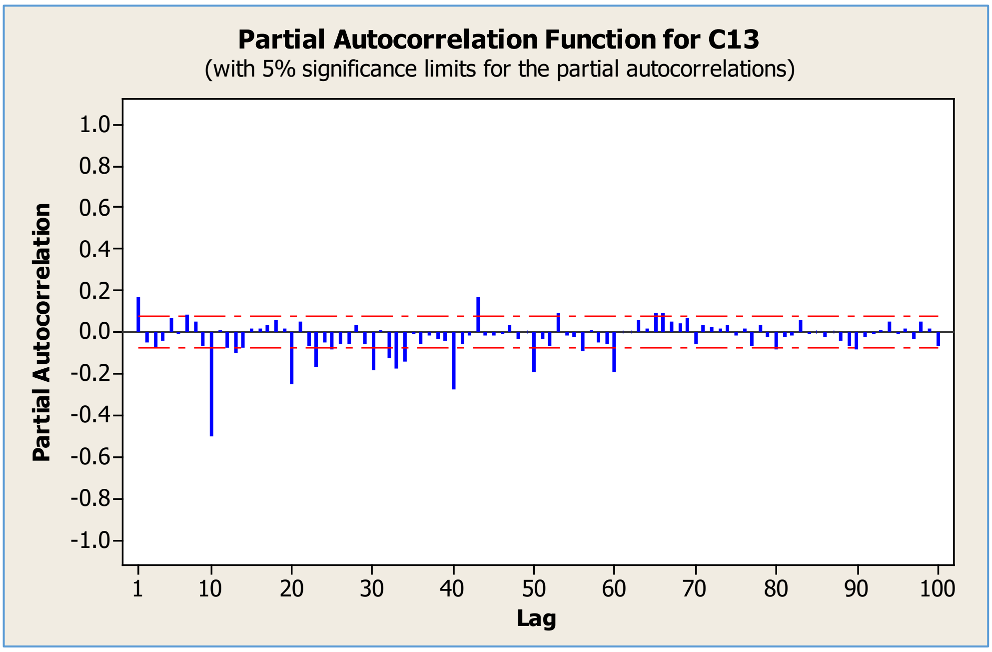

Fig. 11. PACF graph of male covid-19 patients with seasonal ARIMA model

The output of the best model of SARIMA is shown in Fig. 12. Based on the output, the best model of SARIMA is (5,1,0) $(1,1,0)^{10}$. All p-values of autoregressive (AR 1 to AR 5) and seasonal autoregressive (SAR) are less than 0.05. Furthermore, the differencing value is 1, and the seasonal order is 10. The value of AIC is 13179.28. Therefore, the model is accepted as the best model.

AR 1 -0.6979 0.0377 -18.53 p-value = 0.000

AR 2 -0.5601 0.0433 -12.93 p-value = 0.000

AR 3 -0.4816 0.0447 -10.77 p-value = 0.000

AR 4 -0.4182 0.0433 -9.67 p-value = 0.000

AR 5 -0.1343 0.0381 -3.53 p-value = 0.000

SAR 10 -0.5159 0.0333 -15.48 p-value = 0.000

Differencing: 1 regular, 1 seasonal of order 10

AIC = 13179.28 AICc = 13179.4 BIC = 13206.68

Fig. 12. The best model of SARIMA

3.1.4. Forecasting

Fig. 13 shows that the number of Covid-19 male patients who died in the USA generally decreased for the following 100 datasets. The average difference is around 2500 people.

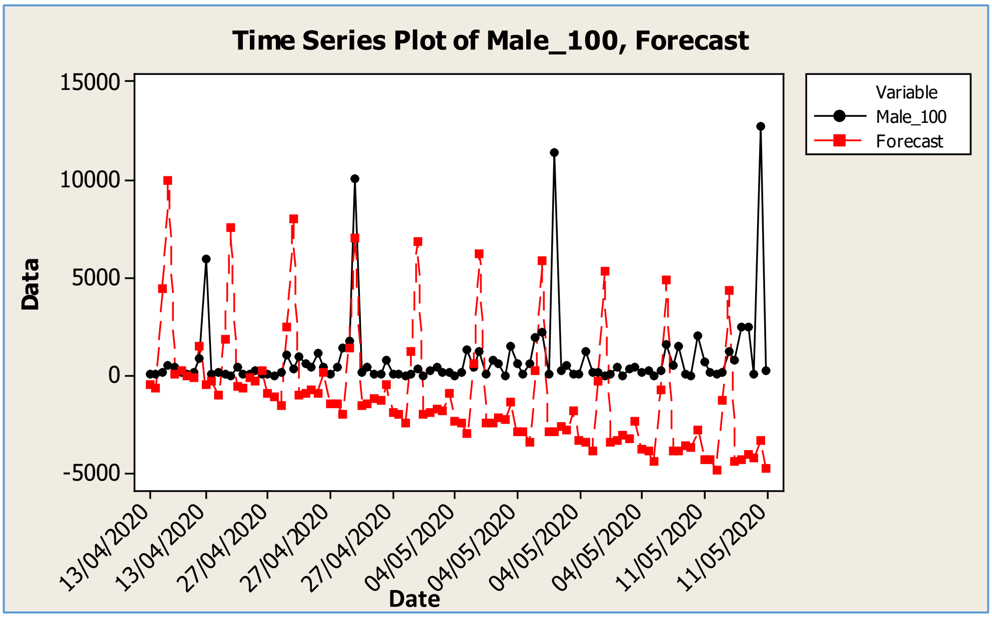

Fig. 13. Graph of male Covid-19 patient forecast

3.2. Implementation of LSTM, SARIMA-LSTM method, and the SARIMA-PARCD method

With the same data, namely a male covid-19 patient in the USA analyzed using LSTM, the MSE value was 8898744.43. Based on the MSE results of LSTM, the value is smaller than SARIMA (MSE=9344238), so it can be said that LSTM is better than SARIMA. Forecasting results using LSTM (Fig. 14) shows unfavorable results. This can be seen from the forecasting data that tends to be monotonous, close to zero, and the average number of Covid-19 patients mortality decreases.

For the next test, a combination of the SARIMA method with LSTM will be carried out to determine how much accuracy the MSE value is obtained and the prediction results. Based on the calculation results, the MSE value is 0.14. The results are shown in Fig. 15.

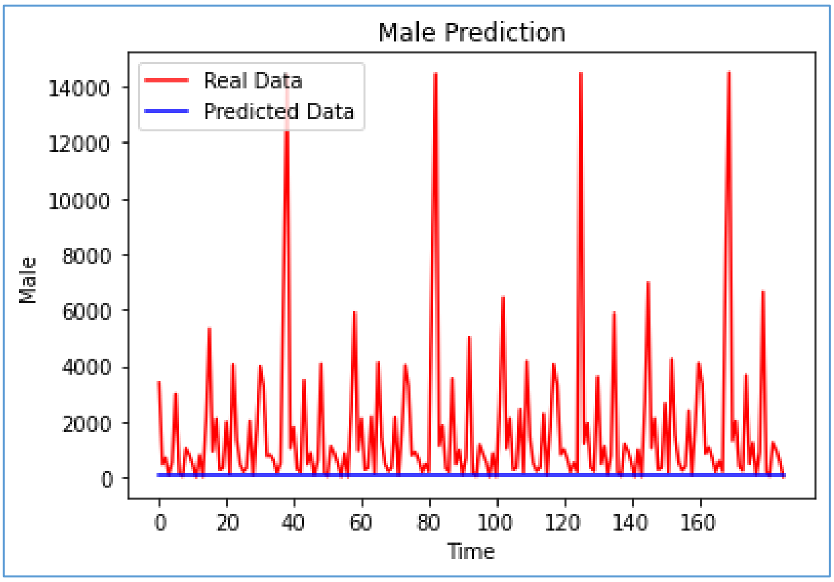

Fig. 14. Graph of male Covid-19 patient forecasting using the LSTM method

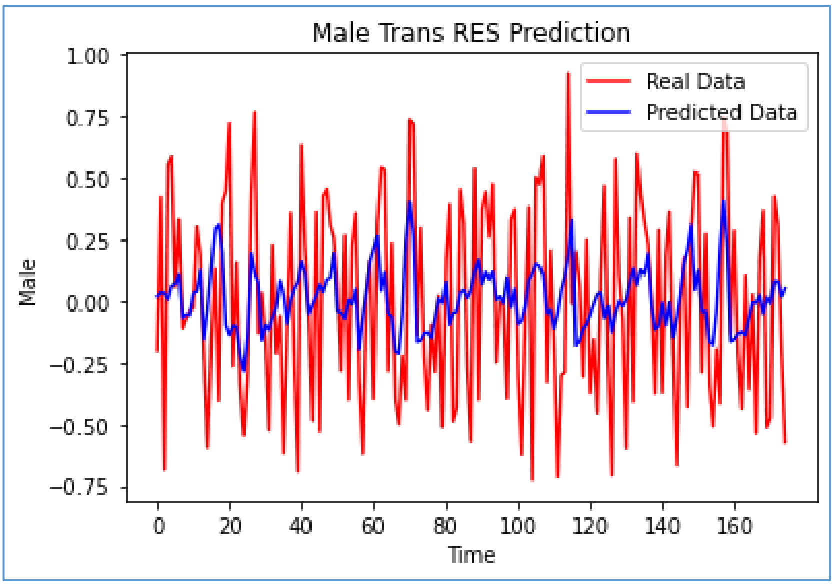

Fig. 15. Forecasting male covid-19 patients using the SARIMA-LSTM combination method

In this study, the combination of SARIMA and PARCD was also tested. The residual Dataset of covid-19 male patients who died was run using the PARCD method for non-linear data solutions. The experiment was performed using a computer with an Intel Core i5 processor with 8 GB of main memory running Windows 7. The algorithms were implemented using MATLAB. For the proposed algorithm, we set the population size, external repository size, number of iterations, C1 and C2, ω, velocity limit, and xRank parameters to 40, 100, 2000, 2, 0.63, 3.83, and 13.33, respectively. Based on the calculations, the MSE value of SARIMA-PARCD is 0.0049. Therefore, the proposed SARIMA-PARCD model is better than other studies (Table 1).

Table 1. Comparison MSE result

Method	MSE
SARIMA	0.048643
LSTM	0.206617
SARIMA-LSTM	0.14
The proposed SARIMA-PARCD	**0.0049**

4. Conclusion

Based on the results of these studies, it is found that the LSTM method is better than the SARIMA method. The SARIMA - LSTM combination method is better than the two methods separately. Furthermore, the proposed SARIMA - PARCD method is better than the other models in this study, with an MSE value of 0.0049. Based on the results of general predictions using these methods, it is found that there is a decrease in the number of male covid-19 patients who die in the USA on average. In the next research, the SARIMA-PARCD combination method will be used to the predicted temperature by using natural sensors, i.e., the bioelectric potential of the plant. This innovation would like to observe the utility of the bioelectric potential of plants for differentiating the temperature as one of the identification factors of COVID-19 cases.

Acknowledgment

The authors thank Directorate Research and Community Services, Deputy for strengthening research and development, Kemenristek-BRIN, and LPPM Universitas Amikom Purwokerto for supporting this research.

Declarations

Author contribution. This research has some authors, Conceptualization, Imam Tahyudin; methodology, Imam Tahyudin and Hidetaka Nambo.; software, Imam Tahyudin, Wiga Maulana Baihaqi; validation, Imam Tahyudin, and Rizki Wahyudi; formal analysis, Imam Tahyudin, and Wiga Maulana Baihaqi; writing—original draft preparation, Imam Tahyudin and Rizki Wahyudi; writing—review and editing, Imam Tahyudin, and Hidetaka Nambo.; visualization, Wiga Maulana Baihaqi.
Funding statement. This research is granted by Directorate Research and Community Services, Deputy for strengthening research and development, Kemenristek-BRIN and LPPM Universitas Amikom Purwokerto no. 021/LL6/PG/PG/SP2H/PL.I/2020; 030/LL6/PG/PG/SP2H/PL.I/2020; and 318/AMIKOMPWT/LPPM/06/VII/2020.
Conflict of interest. The authors declare no conflict of interest.
Additional information. No additional information is available for this paper.

References

[1] R. A. Davis, "Introduction to Statistical Analysis of Time Series." Department of Statistics Columbia University, pp. 1–24, 2014. Available at: Google Scholar.

[2] C. B. Borkowf, "Time-Series Forecasting," *Technometrics*, vol. 44, no. 2, pp. 194–195, 2002, doi: 10.1198/tech.2002.s718.

[3] T. Schluter, "Knowledge discovery from time series," Universit¨at D¨usseldorf, 2012. Available at: Google Scholar.

[4] I. Tahyudin, Berlilana, and H. Nambo, "Sarima model of bioelectic potential dataset," *Commun. Comput. Inf. Sci.*, vol. 872, pp. 367–378, 2018, doi: 10.1007/978-3-319-96292-4_29.

[5] X. Li, Y. Liu, L. Fan, S. Shi, T. Zhang, and M. Qi, "Research on the prediction of dangerous goods accidents during highway transportation based on the ARMA model," *J. Loss Prev. Process Ind.*, vol. 72, p. 104583, 2021, doi: 10.1016/j.jlp.2021.104583.

[6] D. Chakraborty and S. K. Sanyal, "Time-series data optimized AR/ARMA model for frugal spectrum estimation in Cognitive Radio," *Phys. Commun.*, vol. 44, p. 101252, 2021, doi: 10.1016/j.phycom.2020.101252.

[7] I. Tahyudin and H. Nambo, "Comparison study of deep learning and time series for bioelectric potential analysis," *Proc. - 2018 3rd Int. Conf. Inf. Technol. Inf. Syst. Electr. Eng. ICITISEE 2018*, pp. 79–83, 2019, doi: 10.1109/ICITISEE.2018.8720998.

[8] D. Benvenuto, M. Giovanetti, L. Vassallo, S. Angeletti, and M. Ciccozzi, "Application of the ARIMA model on the COVID-2019 epidemic dataset," *Data Br.*, vol. 29, p. 105340, 2020, doi: 10.1016/j.dib.2020.105340.

[9] Z. Malki *et al.*, "ARIMA models for predicting the end of COVID-19 pandemic and the risk of second rebound," *Neural Comput. Appl.*, vol. 0, 2020, doi: 10.1007/s00521-020-05434-0.

[10] Z. Ceylan, "Estimation of COVID-19 prevalence in Italy , Spain, and France," *Sci. Total Environ.*, no. January, 2020. doi: 10.1016/j.scitotenv.2020.138817

[11] Haviluddin and A. Jawahir, "Comparing of ARIMA and RBFNN for short-term forecasting," *Int. J. Adv. Intell. Informatics*, vol. 1, no. 1, pp. 15–22, 2015, doi: 10.26555/ijain.v1i1.10.

[12] Suhartono, S. Isnawati, N. A. Salehah, D. D. Prastyo, H. Kuswanto, and M. H. Lee, "Hybrid SSA-TSR-ARIMA for water demand forecasting," *Int. J. Adv. Intell. Informatics*, vol. 4, no. 3, pp. 238–250, 2018, doi: 10.26555/ijain.v4i3.275.

[13] K. Y. Chen and C. H. Wang, "A hybrid SARIMA and support vector machines in forecasting the production values of the machinery industry in Taiwan," *Expert Syst. Appl.*, vol. 32, no. 1, pp. 254–264, 2007, doi: 10.1016/j.eswa.2005.11.027.

[14] N. H. A. Rahman, M. H. Lee, Suhartono, and M. T. Latif, "Hybrid seasonal ARIMA and artificial neural network in forecasting southeast Asia City Air Pollutant Index," *ASM Sci. J.*, vol. 12, no. Special Issue 1, pp. 215–226, 2019. Available at: Google Scholar.

[15] M. Yollanda and D. Devianto, "Hybrid Model of Seasonal ARIMA-ANN to Forecast Tourist Arrivals through Minangkabau International Airport," 2020, doi: 10.4108/eai.2-8-2019.2290473.

[16] M. Braun, T. Bernard, O. Piller, and F. Sedehizade, "24-hours demand forecasting based on SARIMA and support vector machines," *Procedia Eng.*, vol. 89, pp. 926–933, 2014, doi: 10.1016/j.proeng.2014.11.526.

[17] A. Ozozen, G. Kayakutlu, M. Ketterer, and O. Kayalica, "A combined seasonal ARIMA and ANN model for improved results in electricity spot price forecasting: Case study in Turkey," *PICMET 2016 - Portl. Int. Conf. Manag. Eng. Technol. Technol. Manag. Soc. Innov. Proc.*, pp. 2681–2690, 2017, doi: 10.1109/PICMET.2016.7806831.

[18] L. Parviz, "Comparative evaluation of hybrid sarima and machine learning techniques based on time varying and decomposition of precipitation time series," *J. Agric. Sci. Technol.*, vol. 22, no. 2, pp. 563–578, 2020. doi: 10.3390/e22050578

[19] D. P. M. Abellana, D. M. C. Rivero, M. E. Aparente, and A. Rivero, "Hybrid SVR-SARIMA model for tourism forecasting using PROMETHEE II as a selection methodology: a Philippine scenario," *J. Tour. Futur.*, 2020, doi: 10.1108/JTF-07-2019-0070.

[20] H. Liu *et al.*, "Forecast of the trend in incidence of acute hemorrhagic conjunctivitis in China from 2011–2019 using the Seasonal Autoregressive Integrated Moving Average (SARIMA) and Exponential Smoothing (ETS) models," *J. Infect. Public Health*, vol. 13, no. 2, pp. 287–294, 2020, doi: 10.1016/j.jiph.2019.12.008.

[21] M. Z. Md Maarof, Z. Ismail, and M. Fadzli, "Optimization of SARIMA model using genetic algorithm method in forecasting Singapore tourist arrivals to Malaysia," *Appl. Math. Sci.*, vol. 8, no. 169–172, pp. 8481–8491, 2014, doi: 10.12988/ams.2014.410847.

[22] P. Arora, H. Kumar, and B. K. Panigrahi, "Prediction and analysis of COVID-19 positive cases using deep learning models: A descriptive case study of India," *Chaos, Solitons and Fractals*, vol. 139, p. 110017, 2020, doi: 10.1016/j.chaos.2020.110017.

[23] J. Kumar, R. Goomer, and A. K. Singh, "Long Short Term Memory Recurrent Neural Network (LSTM-RNN) Based Workload Forecasting Model for Cloud Datacenters," *Procedia Comput. Sci.*, vol. 125, pp. 676–682, 2018, doi: 10.1016/j.procs.2017.12.087.

[24] T. A. Rashid, P. Fattah, and D. K. Awla, "Using accuracy measure for improving the training of LSTM with metaheuristic algorithms," *Procedia Comput. Sci.*, vol. 140, pp. 324–333, 2018, doi: 10.1016/j.procs.2018.10.307.

[25] T. Kim and H. Y. Kim, "Forecasting stock prices with a feature fusion LSTM-CNN model using different representations of the same data," *PLoS One*, vol. 14, no. 2, pp. 1–23, 2019, doi: 10.1371/journal.pone.0212320.

[26] I. Kirbas, A. Sozen, A. D. Tuncer, and F. S. Kazancioglu, "Comparative analysis and forecasting of COVID-19 cases in various European countries with ARIMA , NARNN and LSTM approaches," no. January, 2020. doi: 10.1016/j.chaos.2020.110015

[27] M. A. Jishan, K. R. Mahmud, A. K. Al Azad, M. S. Alam, and A. M. Khan, "Hybrid deep neural network for bangla automated image descriptor," *Int. J. Adv. Intell. Informatics*, vol. 6, no. 2, 2020, doi: 10.26555/ijain.v6i2.499.

[28] I. Tahyudin, Berlilana, and H. Nambo, "Predicting human position using improved numerical association analysis for bioelectric potential data," *Lect. Notes Networks Syst.*, vol. 69, pp. 655–666, 2020, doi: 10.1007/978-3-030-12388-8_46.

[29] J.-Y. Yeh, T.-H. Wu, and C.-W. Tsao, "Using data mining techniques to predict hospitalization of hemodialysis patients," *Decis. Support Syst.*, vol. 50, no. 2, pp. 439–448, Jan. 2011, doi: 10.1016/j.dss.2010.11.001.

[30] Q. Wu and R. Law, "Cauchy mutation based on objective variable of Gaussian particle swarm optimization for parameters selection of SVM," *Expert Syst. Appl.*, vol. 38, no. 6, pp. 6405–6411, 2011, doi: 10.1016/j.eswa.2010.08.069.