

The PELT-ensemble NARX Hybrid Method for Short-Term Forecasting Time Series Data with Change Points and Its Application

Hermansah¹, Dedi Rosadi^{2,*}, Faihatuz Zuhairah³, Shelton Peiris⁴, Pepi Novianti⁵,
Widyastuti Andriyani⁶, Muhammad Irsyam⁷

¹Department of Mathematics Education, Universitas Riau Kepulauan, Kota Batam, Kepulauan Riau, 29425 Indonesia

²Department of Mathematics, Xiamen University Malaysia, Sepang, 43900 Malaysia

³Statistics RnD, Yogyakarta, 55292 Indonesia

⁴Department of Mathematics Education, STKIP YPUP Makassar, Makassar, 90223 Indonesia

⁵School of Mathematics and Statistics, University of Sydney, Sydney, NSW 2006 Australia

⁶Department of Statistics, Universitas Bengkulu, Bengkulu, 38371 Indonesia

⁷Master of Information Technology, Universitas Teknologi Digital Indonesia, Yogyakarta, 55198 Indonesia

⁷Department of Electrical Engineering, Universitas Riau Kepulauan, Kota Batam, Kepulauan Riau, 29425 Indonesia

(Received: April 30, 2026; Revised: June 30, 2026; Accepted: September 25, 2026; Available online: October 14, 2026)

Abstract

This study proposes a new approach using the PELT-ensemble NARX hybrid method for short-term forecasting of time series data containing change points (CPs). The Pruned Exact Linear Time (PELT) method was employed to detect CPs and to determine cut-off points for the observation data, while the ensemble Nonlinear Autoregressive Neural Network with External Inputs (NARX) was used for forecasting. The method was applied to health data, specifically confirmed Covid-19 cases in the provinces of Java and Bali, Indonesia, from March 2020 to October 2021. Several Covid-19-related variables such as the number of recovered cases and deaths, and the weather variables, i.e., the maximum temperature, the average temperature, and humidity, are used as external inputs for the NARX model. The results indicate that the PELT method effectively identified the first day in the last segmentation for each province, and subsequent forecasting using the ensemble NARX method achieved high accuracy. Forecasting performance was evaluated using the Mean Absolute Percentage Error (MAPE) and several other metrics, showing that the PELT-ensemble NARX hybrid method significantly outperformed the single NARX approach and other forecasting methods, i.e. Exponential Smoothing, Autoregressive Integrated Moving Average (ARIMA) and Long Short-Term Memory (LSTM) model. These findings highlight the effectiveness of integrating change point detection with ensemble neural network modelling to substantially improve short-term forecasting accuracy for time series subject to structural change.

Keywords: Change Point Detection, PELT, Neural Network, NARX, Hybrid Model

1. Introduction

Time series forecasting is an important analytical task for predicting future values based on historical observations. It has been widely applied in various domains, including economics, business, finance, public health, and epidemiology, e.g., in [1], [2], [3], [4]. Forecasting methods can generally be grouped into statistical approaches and machine learning-based approaches. Statistical methods commonly rely on assumptions regarding the structure and distribution of the data [1], [2], whereas machine learning approaches are more flexible in capturing complex and nonlinear patterns [3]. Among machine learning models, Artificial Neural Networks (ANNs) have been widely used for time series forecasting because of their ability to approximate nonlinear relationships in sequential data. One of the ANN-based models frequently used for time series forecasting is the Nonlinear Autoregressive Neural Network with External Inputs

*Corresponding author: Dedi Rosadi (dedi.rosadi@xmu.edu.my)

 DOI: <https://doi.org/10.47738/jads.v7i4.1085>

This is an open access article under the CC-BY license (<https://creativecommons.org/licenses/by/4.0/>).

© Authors retain all copyrights

(NARX) [5], [6], [7], [8], [9]. The NARX model extends the autoregressive neural network framework by incorporating exogenous variables into the learning process. This characteristic allows the model to capture both temporal dependence from past observations and additional information from external factors. Therefore, NARX is considered suitable for forecasting dynamic processes influenced by multiple explanatory variables. In practice, NARX can be implemented using either parallel or series-parallel architecture. This study focuses on the series-parallel architecture because it uses actual observed values as regressors during the training process and is commonly applied in supervised forecasting settings [8], [9].

However, many real-world time series data contain structural changes or change points that may alter their statistical properties over time. A change point refers to a time location at which the underlying characteristics of a sequence of observations change, such as changes in mean, variance, or distributional pattern [10], [11], [12], [13]. When change points occur, a forecasting model trained on the entire historical series may incorporate outdated patterns that are no longer relevant to the most recent data structure. Consequently, ignoring change points may reduce forecasting accuracy and produce biased short-term predictions. Several change point detection methods have been introduced in the literature, including Binary Segmentation [10], Neighborhood Segmentation [11], Optimal Partitioning [12], and Pruned Exact Linear Time (PELT) [13]. In this study, the PELT algorithm will be used because it enables multiple change point detection through an exact optimization approach while maintaining computational efficiency using a pruning mechanism. Compared with the Binary Segmentation method, which performs sequential partitioning or other approaches, PELT evaluates the overall data structure to obtain a segmentation solution that is optimal with respect to the selected cost function and penalty criterion. Therefore, it is considered more suitable for repeated segmentation processes within the modeling framework employed in this study.

The Covid-19 pandemic provides an important example of a time series process with changing data patterns. Confirmed Covid-19 cases have shown multiple waves, shifts, and regional variations across different periods and locations [7], [14], [15], [16], [17], [18], [19]. In Indonesia, the provinces of Java and Bali recorded a large proportion of confirmed cases, making them relevant empirical settings for evaluating short-term forecasting models [7], [14], [15], [16], [17]. The dynamics of Covid-19 cases may be influenced by several external factors, including environmental conditions. Previous studies have reported associations between weather variables, such as temperature and humidity, and Covid-19 transmission patterns [20], [21], [22], [23].

Although NARX has been applied in previous forecasting studies (e.g., [7], [8], [9]), its performance may be limited when the time series contains change points. Conversely, change point detection alone does not directly provide future predictions. This study addresses this methodological gap by integrating PELT and ensemble NARX into a hybrid forecasting framework. In the proposed approach, PELT is used as an initial segmentation procedure to detect change points and identify the most recent statistically homogeneous segment of the data. Subsequently, ensemble NARX is applied to the last segment to produce short-term forecasts. This integration is expected to reduce the influence of obsolete historical patterns and improve forecasting accuracy for time series data containing change points.

Therefore, the main contribution of this study is the development of a PELT–ensemble NARX hybrid method for short-term forecasting of time series data with change points. The proposed method is empirically evaluated using confirmed Covid-19 case data from seven provinces in Java and Bali, Indonesia, with weather variables (maximum temperature, average temperature, and humidity) and Covid-19-related variables (i.e., the number of recovered cases and deaths) included as exogenous inputs. The forecasting performance of the proposed method is compared with the standard NARX approach to examine whether incorporating change point-based segmentation can improve prediction accuracies. The remainder of this paper is organized as follows. Section 2 reviews the theoretical background of the PELT method and the NARX model, followed by the proposed hybrid forecasting procedure. Section 3 describes the empirical results and analysis. Finally, Section 4 provides the conclusions of the study.

2. Literature Review

2.1. Pruned Exact Linear Time (PELT) Method

In statistical analysis, change point detection attempts to identify the time when the probability distribution of a stochastic process or time series changes. Additionally, change point detection involves detecting the number and

location of change points, which are locations in the data where some statistical property, such as the mean or variance, changes. Change point analysis can be applied in various fields, such as finance, oceanography, epidemiology, quality control, genetics, and others (see, e.g., [10], [11], [12], [13], [24], [25]). One application of change point detection using the PELT method was presented in [24], which analyzed time series data in oceanography to automatically identify the beginning and end of the hurricane season.

Suppose that the observed series is denoted by $y_{1:n} = (y_1, y_2, \dots, y_n)$, where n denotes the total number of observations. Let there be m distinct change points in $y_{1:n}$ such that $\tau = \tau_{1:m} = (\tau_1, \tau_2, \dots, \tau_m)$, with $0 = \tau_0 < \tau_1 < \dots < \tau_m < \tau_{m+1} = n$. Thus, the observations are partitioned into $m + 1$ segments, where the i -th segment consists of $(y_{\tau_{i-1}+1}, \dots, y_{\tau_i})$. The PELT method proposed in [13] determines the optimal segmentation by minimizing the following penalized cost function:

$$\min_{\tau \in T_n} \{ \sum_{i=1}^{m+1} [\mathcal{C}(y_{\tau_{i-1}+1:\tau_i}) + \beta] \}. \quad (1)$$

$\mathcal{C}(\cdot)$ denotes the segment cost function, β is the penalty parameter used to avoid overfitting, and T_n represents the set of all possible segmentations. The cost function can be formulated to detect changes in the mean, variance, both the mean and variance, or more general distributional characteristics. In this study, the segmentation procedure is based on a variance-sensitive cost function. To avoid over-segmentation, a penalty term is incorporated, with commonly used criteria including the Akaike Information Criterion (AIC), with $\beta = 2p$, and the Bayesian Information Criterion (BIC), with $\beta = p \log n$, where p denotes the number of additional parameters introduced by adding one change point. In this study, the penalty parameter was determined using the Modified BIC (MBIC). The MBIC penalty is an extension of the BIC that includes an extra term to account for the spacing of the change points. It can be approximated in practice by using the value $\beta = 3 \log n$ as the penalty. The PELT method is developed from the Optimal Partitioning (OP) framework. Define $F(n)$ as the minimum segmentation cost obtained for observations up to time n , with initialization

$$F(0) = -\beta. \quad (2)$$

The recursive representation of the OP method is expressed as

$$F(n) = \min_{\tau < n} \{ F(\tau) + \mathcal{C}(y_{\tau+1:n}) + \beta \}. \quad (3)$$

Equation (2) represents the recursive formulation of the Optimal Partitioning method and serves as the basis for the pruning mechanism in PELT. To improve computational efficiency, PELT applies a pruning strategy that removes candidate change points that cannot become part of the global optimal solution. Based on the theorem in [13], assume there exists a constant K such that, for every $\tau < s < n$, where s denotes an intermediate candidate change point, the following condition holds:

$$\mathcal{C}(y_{\tau+1:s}) + \mathcal{C}(y_{s+1:n}) + K \leq \mathcal{C}(y_{\tau+1:n}). \quad (4)$$

If

$$F(\tau) + \mathcal{C}(y_{\tau+1:s}) + K \geq F(s), \quad (5)$$

then, for any future observation $n > s$, the point τ will never become the most recent optimal change point before n . Therefore, τ can be removed from further optimization steps without affecting the global optimal solution. The PELT algorithm applied in this study consists of the following steps:

Algorithm 1. Pruned Exact Linear Time (PELT) Algorithm for Multiple Change Point Detection

Prepare the observation data $y_1, y_2, \dots, y_n$.

Specify the segment cost function according to the characteristics of the data.

Select the penalty parameter β .

Determine the constant K satisfying Equation (3).

Initialize $F(0) = -\beta$, $cp(0) = \emptyset$, and $R_1 = \{0\}$.

For each candidate $\tau^* = 1, 2, \dots, n$:

(a) Compute $F(\tau^*) = \min_{\tau \in R_{\tau^*}} [F(\tau) + \mathcal{C}(y_{\tau+1:\tau^*}) + \beta]$.

(b) Determine $\tau^{(1)} = \arg \min_{\tau \in R_{\tau^*}} [F(\tau) + \mathcal{C}(y_{\tau+1:\tau^*}) + \beta]$.

(c) Update $cp(\tau^*) = [cp(\tau^{(1)}), \tau^{(1)}]$.

(d) Update the candidate set $R_{\tau^*+1} = \{\tau \in R_{\tau^*} \cup \{\tau^*\} : F(\tau) + \mathcal{C}(y_{\tau+1:\tau^*}) + K \leq F(\tau^*)\}$.

7. Obtain the final change points from $cp(n)$.

2.2. NARX Model

The nonlinear autoregressive neural network with external inputs (NARX) model is a discrete-time nonlinear system defined as [26]:

$$y(t) = f(y_{t-1}, y_{t-2}, \dots, y_{t-n_y}; u_t, u_{t-1}, \dots, u_{t-n_u}; w) + \varepsilon(t), \quad (6)$$

y_t and u_t denote the model's output and input at time t , respectively, $n_u \geq 1$ and $n_y \geq 1$ are the maximum memory orders of the input and output, w denotes the weight matrix, and f denotes the nonlinear function [27]. The NARX network can be trained using the following models [28]. The first is the model with parallel architecture (or with feedback). In this model, the output becomes the feedback for the input of the feed-forward neural network and can be modeled as

$$\hat{y}(t) = \hat{f}(\hat{y}_{t-1}, \hat{y}_{t-2}, \dots, \hat{y}_{t-n_y}; u_t, u_{t-1}, \dots, u_{t-n_u}; w) + \varepsilon(t). \quad (7)$$

The second is the model with series-parallel architecture (or parallel architecture without feedback), where the output regressors enter only through the use of actual output values:

$$\hat{y}(t) = \hat{f}(y_{t-1}, y_{t-2}, \dots, y_{t-n_y}; u_t, u_{t-1}, \dots, u_{t-n_u}; w) + \varepsilon(t). \quad (8)$$

An illustration of the parallel and series-parallel architectures can be seen in figure 1.

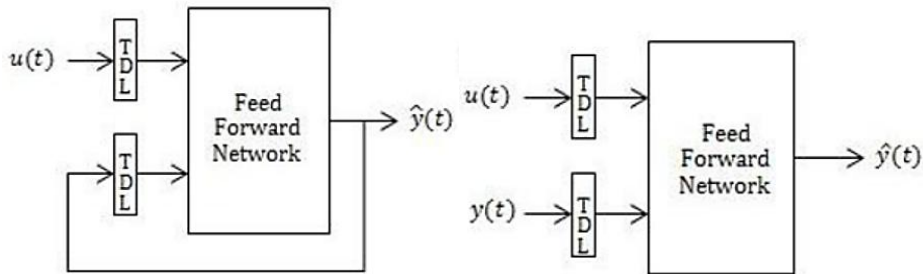


Figure 1. Parallel and series-parallel architectures of the NARX network.

This study considers only the NARX model with series-parallel architecture using three layers, namely the input layer, the hidden layer, and the output layer. The following are the steps of NARX modeling used, where y_t denotes the predicted data and u_t denotes the external data.

Data modeling begins by identifying the frequency of the data y_t . Frequency indicates the number of observations within a certain period. If the frequency of the obtained data y_t is less than 3, successive autoregressive lags of 1 to 4 are considered for the data y_t and u_t . Meanwhile, if the frequency of the obtained data y_t is greater than 3, the autoregressive lags of the data y_t and u_t are considered from 1 to m , where m represents the frequency of the data y_t .

Differencing is first performed on the data y_t and u_t . Furthermore, the first-differenced data are scaled linearly to $[-0.8, 0.8]$ to facilitate the training process. Following [9], this approach provides the best effectiveness in terms of accuracy and convergence speed. The first differencing is calculated using the following equation:

$$\Delta y(t) = y_t - y_{t-1}, \quad (9)$$

$\Delta y(t)$ is the first-differenced value in period t , and y_t and y_{t-1} are the data in periods t and $t - 1$, respectively. The data are scaled using the following equation:

$$y_{sc} = \frac{(mx - mn)(y - mn_{orig})}{(mx_{orig} - mn_{orig})} + mn, \quad (10)$$

y is the input data, mn and mx are the minimum and maximum target values of the input data, and mn_{orig} and mx_{orig} are the minimum and maximum values of the original input data.

The input variables in the input layer are determined based on the autoregressive lags of y_t and u_t identified in Step 1. Furthermore, the optimal input variables in the input layer are selected using a stepwise method. The stepwise method is used to obtain input variables based on the principle of parsimony (see, e.g., [9]).

The optimal parameterization of the hidden layer, particularly the number of hidden neurons, is determined using a trial-and-error procedure. Several alternative NARX architectures are evaluated by varying the number of hidden neurons, and the architecture with the smallest validation Mean Squared Error (MSE) is selected as the final base model. Here, MSE is defined as

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (11)$$

The data from the selected last segment are divided chronologically into 80% training data and 20% validation data. After the optimal NARX architecture is obtained, the same architecture is trained 50 times independently using different random initial weights and biases. Thus, the ensemble does not consist of 50 models selected through an additional post-training ranking process. Instead, it consists of 50 independently trained NARX models generated from the same optimal architecture. The forecasting outputs from these 50 models are then transformed back to the original scale through unscaling and undifferencing procedures and combined using the mode ensemble operator to produce the final forecast. This ensemble procedure is used to reduce prediction instability that may arise from a single neural network model due to random initialization. According to [9], the mode ensemble operator can improve forecasting accuracy compared with a single network. The ensemble mode operator is calculated using the kernel density estimator approach as follows:

$$\hat{f}_{th}(x) = \frac{1}{Mh} \sum_{m=1}^M K\left(\frac{x - \hat{y}_{mt}}{h}\right), \quad (12)$$

M is the number of ensemble models, $h > 0$ is the estimated bandwidth, approximated by

$$h = \left(\frac{4\hat{\sigma}^5}{3M}\right)^{1/5}, \quad (13)$$

following [29], $K(\cdot)$ is a kernel function approximated by the Gaussian kernel,

$$K_h(x) = \frac{1}{\sqrt{2\pi}h} e^{-\frac{x^2}{2h^2}}, \quad (14)$$

and $\hat{y}_{mt}$ is the forecast of the m -th model, where $m = 1, \dots, M$, for period t . The mode ensemble operator is determined by the value of x at which $\hat{f}_{th}(x)$ reaches its maximum. The previous steps are applied recursively to obtain multi-step forecasts.

3. Methodology

The proposed hybrid forecasting framework integrates the Pruned Exact Linear Time (PELT) algorithm and the ensemble Nonlinear Autoregressive Neural Network with External Inputs (NARX) model. The main rationale of this integration is that time series data containing change points may consist of several segments with different statistical properties. If a forecasting model is trained using the entire historical series, older segments with different data structures may influence the learning process and reduce short-term forecasting accuracy. Therefore, PELT is used as a preliminary segmentation procedure to identify the most recent statistically homogeneous segment, while ensemble NARX is applied to this last segment to generate short-term forecasts.

The hybrid procedure consists of six main stages: data preparation, change point detection, last-segment selection, NARX model construction, ensemble forecasting, and model evaluation. Let (y_t) denote the target time series and (x_t) denote the vector of exogenous variables. The complete workflow of the procedure is described as follows.

First, before model training, the daily time series data $(y_t, \mathbf{x}_t)$ were transformed through first-order differencing to reduce non-stationary patterns in the data. The differenced data were then scaled linearly into the interval $([-0.8, 0.8])$ to improve numerical stability and facilitate the neural network training process.

Second, the PELT algorithm was applied to (y_t) , to detect the number and locations of change points. The cost function was defined based on variance changes across segments, as described in Section 2.1. The purpose of this step was to identify structural changes in the time series and divide the observations into several segments with relatively homogeneous statistical characteristics. The detected change points were then used to determine the boundary of the most recent segment.

Third, the last segment obtained from the PELT procedure was selected as the training basis for the ensemble NARX model. This step is central to the proposed hybrid framework. Instead of training NARX on the entire historical data, the model was trained only on the last segment identified by PELT. The rationale is that the last segment represents the most recent data regime and is expected to contain patterns that are more relevant for short-term forecasting. This procedure reduces the influence of previous segments that may reflect obsolete case dynamics.

Fourth, the NARX model was constructed using a series-parallel architecture consisting of an input layer, a hidden layer, and an output layer. The input layer was formed using lagged values of the target variable (y_t) and lagged values of the exogenous variables $(\mathbf{x}_t)$. The candidate input variables were determined based on the autoregressive lag structure, and the optimal input combination was selected using a stepwise procedure based on the parsimony principle. The number of hidden neurons was determined through a trial-and-error procedure by comparing the validation mean squared error (MSE) across alternative network configurations. The data in the selected segment were divided into 80% training data and 20% validation data.

Fifth, after the optimal NARX architecture was obtained, the ensemble forecasting procedure was implemented. The selected (i.e. optimal) NARX architecture was trained 50 times using different random initial weights and biases. Each trained network produced a forecast for the target variable. The forecasts from the 50 independently trained NARX models were then combined using the mode ensemble operator based on the kernel density estimator approach. This ensemble procedure was used to reduce the instability that may arise from a single neural network model due to random initialization and local convergence.

Sixth, the final forecasts were transformed back into the original scale through unscaling and undifferencing procedures. The forecasting performance was then evaluated by comparing the predicted values with the observed values in the testing period. The evaluation metric used in the analysis was the Mean Absolute Percentage Error (MAPE) (and possibly, other metrics). A lower MAPE value indicates a smaller forecasting error and better predictive performance. For comparison, the forecasting results of the proposed PELT–ensemble NARX method were compared with those of the standard NARX model trained without PELT-based segmentation, and/or other forecasting methods. Here, MAPE is defined as

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \quad (15)$$

y_i denotes the actual value, $\hat{y}_i$ denotes the predicted value using the chosen model and n denotes total number of data points. In addition to MAPE, three additional evaluation metrics were calculated to provide a more comprehensive assessment of forecasting performance, namely MAE (Mean Absolute Error), RMSE (Root Mean Squared Error), and sMAPE (Symmetric Mean Absolute Percentage Error). The metric MAE, RMSE, and sMAPE are defined as

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (16)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (17)$$

$$sMAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{(|y_i| + |\hat{y}_i|)/2}. \quad (18)$$

The proposed hybrid framework can be summarized in the following algorithm.

Algorithm 2. PELT–Ensemble NARX Hybrid Forecasting Procedure

1. Prepare the target variable (y_t) and the exogenous variables (x_t).
2. Apply first-order differencing to the target and exogenous variables.
3. Scale the differenced data into the interval $([-0.8, 0.8])$.
4. Apply the PELT algorithm to the target variable (y_t) to detect change points.
5. Identify the last segment based on the final detected change point.
6. Use only the last segment as the modeling data for short-term forecasting.
7. Determine candidate lagged inputs from the target and exogenous variables.
8. Select the optimal input variables using a stepwise selection procedure.
9. Determine the optimal number of hidden neurons based on validation MSE.
10. Train the selected NARX architecture 50 times using different random initial weights and biases.
11. Combine the 50 forecasting outputs using the mode ensemble operator.
12. Transform the forecasts back to the original data scale.
13. Evaluate forecasting performance using MAPE and/or other metrics and compare the results with the other forecasting methods.

This procedure ensures that the forecasting model is trained on the most recent statistically homogeneous segment of the time series. Thus, the proposed method is expected to be more adaptive for short-term forecasting when the data contain change points.

4. Results and Discussion

The empirical study was applied to Covid-19 data from Java and Bali provinces in Indonesia, which represented regions with high cumulative Covid-19 cases. The proposed PELT–ensemble NARX hybrid method was used to forecast confirmed Covid-19 cases by integrating change point detection and neural network-based forecasting. The PELT method served as an initial segmentation procedure to identify structural changes in the time series, while the ensemble NARX model was applied to the most recent segment to produce short-term forecasts.

In the empirical study, the daily confirmed Covid-19 cases were used as the response variable, while the daily data of weather variables (i.e. maximum temperature, average temperature, and humidity) and the daily data of Covid-19-related variables (i.e. the number of recovered cases and deaths) are considered as the exogeneous variables. These variables are used as external inputs in the NARX model to provide additional explanatory information beyond the autoregressive structure of the confirmed case series and are expected to improve short-term forecasting accuracy. The weather data for each province were obtained from Dataonline-BMKG (www.dataonline.bmkg.go.id) and Covid-19 data were obtained from BNPB (www.covid19.bnpb.go.id).

It should be noted that other exogeneous variables such as epidemiological and social factors or other potentially influential factors were not incorporated into the current studies. Some important variables, such as government policy interventions, population mobility patterns, vaccination rates and coverage, testing intensity, public compliance with health protocols, public health interventions, policy restrictions, etc., may have direct impact to the confirmed Covid-19 cases. However, these variables were not included in the present study because consistent daily province-level data were not (fully) available for all provinces in Java and Bali during the observation period. The complete information of time interval of data for each province is presented in [table 1](#).

Table 1. Time interval of Data used in Seven Provinces on Java and Bali Islands

Province	Time Interval
Bali	March 10 th , 2020 - October 18 th , 2021
Banten	March 9 th , 2020 - October 18 th , 2021
DI Yogyakarta	March 14 th , 2020- October 18 th , 2021
DKI Jakarta	March 2 nd , 2020 - October 18 th , 2021
West Java	March 10 th , 2020 - October 18 th , 2021
Central Java	March 13 th , 2020 - October 18 th , 2021
East Java	March 17 th , 2020 - October 18 th , 2021

The first step in the empirical analysis here is to plot the number of confirmed cases of Covid-19 in each province together with the segmentation of the data generated using the PELT method steps. The results for each province can be seen in [figure 2 \(a\)-\(g\)](#). The blue curve represents the daily data for confirmed cases of Covid-19, while the red line is the cut-off point for each data segment. For example, in the case of DKI Jakarta, the first change point occurred on the 179th day, it means that the first to the 178th data had the same data properties as the cost function determined based on the data variance value. Likewise for the second segment, starting from the 179th to 310th and so on. A similar interpretation can be applied to data of other provinces.

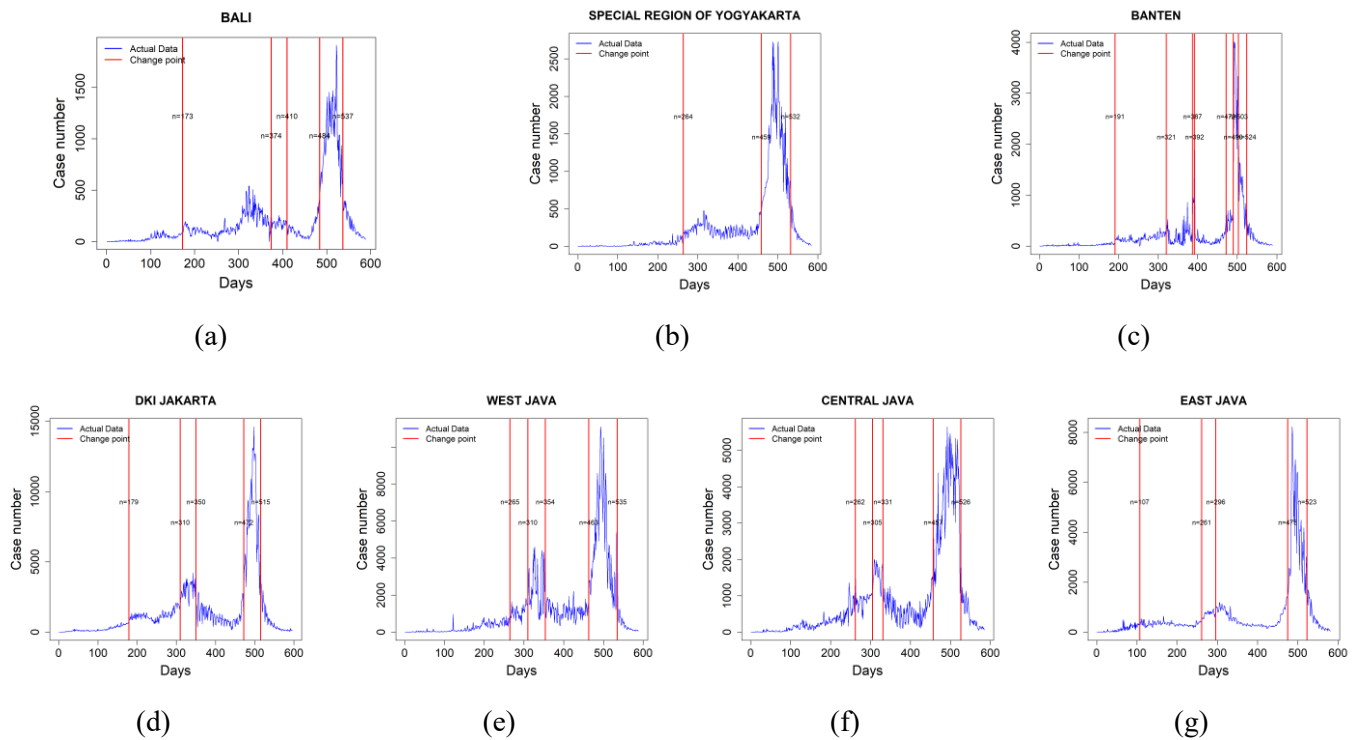


Figure 2. Plot of the number of confirmed Covid-19 cases in each province together with the segmentation using PELT

In addition to identifying structural changes in the data and showing the number of segments formed from the data set, the segmentation results obtained using the PELT method provide the cut-off point of the final segment used for forecasting. Therefore, forecasting performance may depend on the detected change point locations since different change point positions alter the number of observations and data patterns used in the forecasting process. In this study, the final segments obtained for each province consistently preserved the declining trend of confirmed Covid-19 cases, providing relatively stable conditions for short-term forecasting. The segments of data in each province are presented in [table 2](#).

Table 2. Change points output from PELT

#	Province	Formed segmentation								The beginning of last segment data
		1	2	3	4	5	6	7	8	
1	Bali	173	374	410	484	537	-	-	-	August 28, 2021
2	Banten	191	321	387	392	472	490	503	524	August 14, 2021
3	DI Yogyakarta	264	459	532	-	-	-	-	-	August 27, 2021
4	DKI Jakarta	179	310	350	472	515	-	-	-	July 29, 2021
5	West Java	265	310	354	463	535	-	-	-	August 26, 2021
6	Central Java	262	305	331	457	526	-	-	-	August 20, 2021
7	East Java	107	261	296	475	523	-	-	-	August 21, 2021

After the last segment for each province was identified using the PELT method, the ensemble NARX model was trained using only the data from the corresponding last segment. To improve the reproducibility of the proposed hybrid framework, the final NARX architecture for each province is reported in [table 3](#). The table includes the target lag

structure, exogenous input lag structure, number of hidden neurons, activation functions, training algorithm, training–validation split, and the number of ensemble models used in the forecasting process.

Table 3. Final NARX Architecture for Each Province

Province	Beginning of Last Segment	Target Lag Structure	Exogenous Input Lag Structure	Number of Hidden Neurons	Hidden Layer Activation Function	Output Layer Activation Function	Training Algorithm	Training–Validation Split	Number of Ensemble Models
Bali	August 28, 2021	1–4	Lag 1–4 (all exogenous variables Xreg1–Xreg5); seasonal dummies D1.1–D1.3	Rounded number of input features	tanh	linear	resilient backpropagation	80%–20%	50
Banten	August 14, 2021	1–4	Lag 1–4 (all exogenous variables Xreg1–Xreg5); seasonal dummies D1.1–D1.3	Rounded number of input features	tanh	linear	resilient backpropagation	80%–20%	50
DI Yogyakarta	August 27, 2021	1–4	Lag 1–4 (all exogenous variables Xreg1–Xreg5); seasonal dummies D1.1–D1.3	Rounded number of input features	tanh	linear	resilient backpropagation	80%–20%	50
DKI Jakarta	July 29, 2021	1–4	Lag 1–4 (all exogenous variables Xreg1–Xreg5); seasonal dummies D1.1–D1.3	Rounded number of input features	tanh	linear	resilient backpropagation	80%–20%	50
West Java	August 26, 2021	1–4	Lag 1–4 (all exogenous variables Xreg1–Xreg5); seasonal dummies D1.1–D1.3	Rounded number of input features	tanh	linear	resilient backpropagation	80%–20%	50
Central Java	August 20, 2021	1–4	Lag 1–4 (all exogenous variables Xreg1–Xreg5); seasonal dummies D1.1–D1.3	Rounded number of input features	tanh	linear	resilient backpropagation	80%–20%	50
East Java	August 21, 2021	1–4	Lag 1–4 (all exogenous variables Xreg1–Xreg5); seasonal dummies D1.1–D1.3	Rounded number of input features	tanh	linear	resilient backpropagation	80%–20%	50

Note. The target lag structure refers to the lagged values of confirmed Covid-19 cases used as autoregressive inputs. The exogenous input lag structure refers to the lagged values of weather variables, recovered cases, and death cases used as external inputs. The optimal number of hidden neurons was selected based on the minimum validation Mean Squared Error (MSE), i.e. $MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$. The ensemble model was generated by training the selected NARX architecture 50 times using different random initial weights and biases.

The next step is to forecast using the NARX method. The forecasting process was carried out based on the last segment of PELT method which is summarized in table 2. The graphic which compare real data and forecasts of confirmed Covid-19 cases for each province using the PELT-NARX hybrid method is shown in [figure 3 \(a\)-\(g\)](#). The summaries of forecasting performances are given in Table 4. For each province, Mean Absolute Percentage Error (MAPE) of the forecast were computed, based on the last 7 available data from October 12, 2021 until October 18, 2021. Table 4 shows that the PELT-NARX hybrid method is relatively accurate (significantly smaller) than the alternative approach to predict confirmed cases of Covid-19 in the provinces of Java and Bali, Indonesia. For example, MAPE values of the PELT-NARX hybrid method in the province of Bali was obtained at 0.001485 (much less than MAPE values of the NARX method). The same results were also obtained for provinces in Java. Here we may conclude, the PELT–Ensemble NARX method consistently yielded lower MAPE values across all provinces, indicating improved forecasting accuracy.

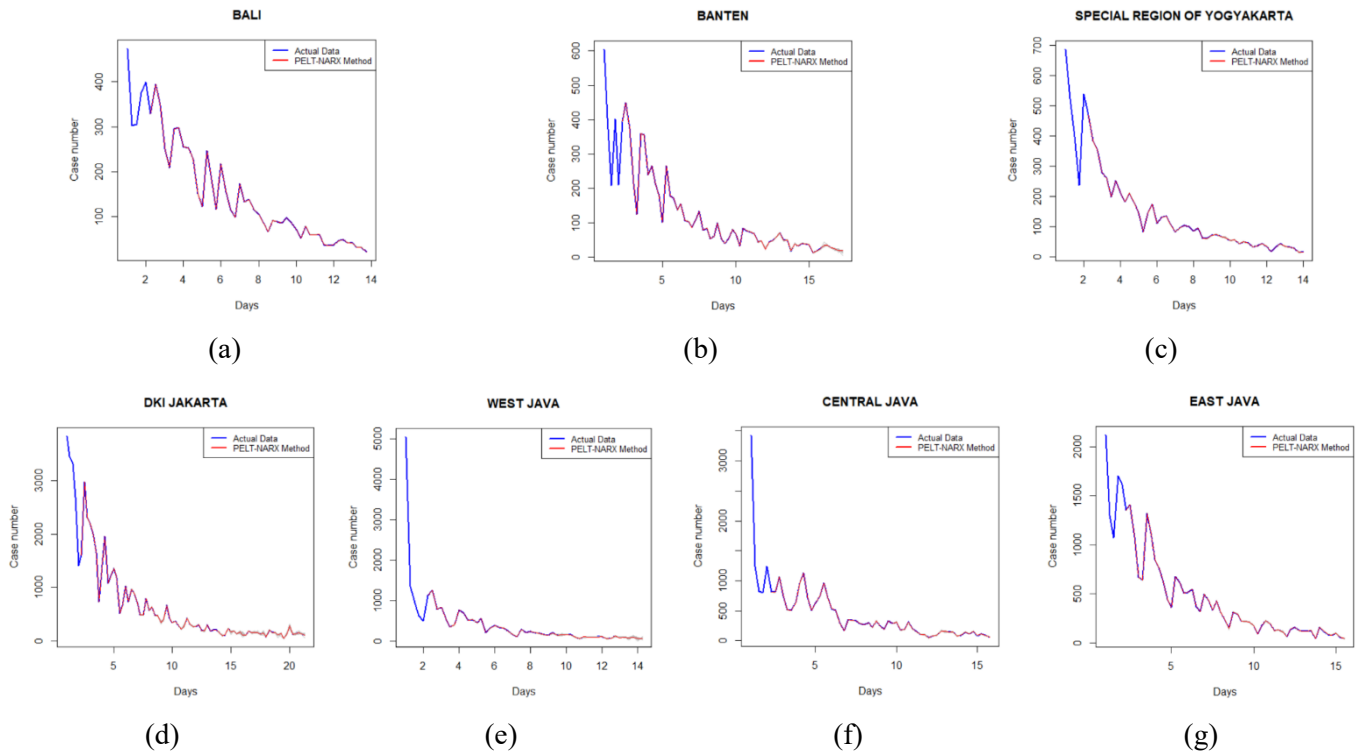


Figure 3. Plot of real data and fitted values using PELT-NARX method for each province

Table 4 also shows that the PELT–Ensemble NARX method consistently produced lower MAE, RMSE, MAPE, and sMAPE values than the standard NARX model across all provinces. This indicates that the improvement of the proposed hybrid method is not only reflected in MAPE but is also supported by other error measures in the original scale and by a more symmetric percentage-based metric. For example, in DKI Jakarta, the RMSE decreased from 49.145361 using the standard NARX model to 5.040599 using the PELT–Ensemble NARX model. Similarly, in Central Java, the MAE decreased from 16.007640 to 0.454567, showing a substantial reduction in average forecasting error. The lower sMAPE values further confirm that the proposed hybrid method remains superior when evaluated using a symmetric percentage-based metric.

Table 4. Forecasting Performance Based on MAE, RMSE, MAPE, and sMAPE

Province	Model	MAE	RMSE	MAPE	sMAPE
Bali	NARX	6.188331	6.871675	0.165450	0.166762
	PELT-NARX	0.049046	0.064099	0.001485	0.001484
Banten	NARX	6.746530	9.218503	0.250864	0.266643
	PELT-NARX	0.620100	0.832862	0.030615	0.030070
DI Yogyakarta	NARX	6.448189	7.350801	0.262961	0.246876
	PELT-NARX	0.107026	0.136382	0.004236	0.004226
DKI Jakarta	NARX	39.22246	49.14536	0.276914	0.272470
	PELT-NARX	3.341986	5.040599	0.021853	0.022394
West Java	NARX	23.16024	29.18351	0.251479	0.316617
	PELT-NARX	2.939820	3.640016	0.034439	0.035447
Central Java	NARX	16.00766	19.00875	0.166847	0.152564
	PELT-NARX	0.454571	0.541128	0.004025	0.004026
East Java	NARX	17.35522	27.57671	0.154919	0.171910
	PELT-NARX	0.562971	0.642368	0.007022	0.007055

This is important because Covid-19 confirmed case data may contain relatively small actual values in some periods, which can make MAPE unstable. Therefore, the additional evaluation using MAE, RMSE, and sMAPE strengthens the conclusion that the PELT–Ensemble NARX hybrid method provides more accurate and stable short-term forecasts than the standard NARX model.

These findings are consistent with prior studies highlighting the effectiveness of hybrid methods combining statistical segmentation and neural network forecasting. The results further validate the use of external variables, such as weather and epidemiological data, in improving NARX model performance. Additionally, the ensemble operator addresses potential variability associated with single network forecasts, reinforcing the robustness of predictions [9].

To statistically validate the improvement in forecasting performance, a paired t-test was conducted for each province comparing the MAPE values of the NARX and PELT–Ensemble NARX methods. Table 5 presents the mean MAPE, mean difference, t-value, degrees of freedom, p-value, and decision of significance for each province. The results show that the PELT–Ensemble NARX method significantly reduces forecast errors compared to the standard NARX model ($p < 0.05$ for all provinces). The paired t-test results demonstrate that integrating change point detection through PELT with an ensemble NARX approach provides a statistically significant improvement in short-term forecasting of Covid-19 confirmed cases. The reduction in MAPE across all seven provinces, confirmed by the t-test, indicates that the hybrid approach consistently enhances predictive accuracy.

Table 5. Paired t-test Results Comparing Forecast Errors (APE) Between NARX and PELT–Ensemble NARX

Province	MAPE (NARX)	MAPE (PELT–Ensemble NARX)	Mean Difference	t-value	df	p-value	Significant ($\alpha=0.05$)
Bali	0.165450	0.001484	0.163966	6.2939	6	0.00074	Yes
Banten	0.250863	0.030615	0.220249	2.6322	6	0.03894	Yes
DI Yogyakarta	0.262960	0.004235	0.258725	3.5195	6	0.01253	Yes
DKI Jakarta	0.276914	0.021853	0.255061	2.6267	6	0.03923	Yes
West Java	0.251479	0.034439	0.217040	2.7211	6	0.03459	Yes
Central Java	0.166847	0.004025	0.162822	3.2933	6	0.01655	Yes
East Java	0.154918	0.007022	0.147897	2.8098	6	0.03077	Yes

Last but not least, we further compare the performance of forecasting accuracy of the proposed method with several well-known forecasting methods, ranging from the classical Exponential Smoothing and ARIMA (AutoRegressive Integrated Moving Average) method, to more advanced method such as the LSTM (Long Short-Term Memory) approach. The results are summarized in table 6, showing the proposed method outperforms other considered approaches. Overall, these results confirm that the PELT–Ensemble NARX hybrid framework can not only improves accuracy but also provides a reliable approach for real-time short-term forecasting in the context of dynamic, multi-wave epidemic data.

Table 6. Mean Absolute Percentage Error (MAPE) for Each Province Using Different Forecasting Models

Province	Exponential Smoothing	ARIMA	LSTM	NARX	PELT–Ensemble NARX
Bali	0.211598	0.153264	0.186839	0.165450	0.001484
Banten	0.238974	0.238945	0.235419	0.250863	0.030615
DIY	0.336772	0.475981	0.200147	0.262960	0.004235
DKI Jakarta	0.448326	0.515446	0.117870	0.276914	0.021853
West Java	0.253011	0.251991	0.235278	0.251479	0.034439
Central Java	0.306962	0.281533	0.215771	0.166847	0.004025
East Java	0.391614	0.321882	0.242803	0.154918	0.007022

5. Conclusion

This study proposed a hybrid PELT–ensemble NARX framework for short-term forecasting of time series containing structural change points. The method combines the Pruned Exact Linear Time (PELT) algorithm for change-point detection with an ensemble Nonlinear Autoregressive Neural Network with Exogenous Inputs (NARX) for forecasting. Its performance was evaluated using confirmed COVID-19 cases in the provinces of Java and Bali, Indonesia.

The empirical results demonstrate that incorporating change-point detection prior to model fitting substantially improves short-term forecasting performance. Compared with alternative forecasting methods, the proposed PELT–ensemble NARX approach consistently achieved lower forecasting errors across several commonly used performance measures, highlighting the importance of explicitly accounting for structural changes in time series forecasting.

Despite these encouraging results, the study has several limitations. First, the empirical analysis considered only COVID-19-related variables (recoveries and deaths) together with weather variables (maximum temperature, average

temperature, and humidity) as exogenous inputs to the NARX model. Other important epidemiological and socio-economic factors, including government intervention policies, population mobility, vaccination coverage, testing intensity, and public behavioural responses, were not incorporated. These variables are known to influence COVID-19 transmission dynamics, particularly during periods of lockdowns, vaccination campaigns, and changes in testing capacity. Their exclusion was primarily due to the limited availability and inconsistency of daily province-level data across Java and Bali during the study period.

Second, the inclusion of weather variables should not be interpreted as implying that meteorological conditions are the primary drivers of COVID-19 transmission. Rather, they were selected because they were consistently available throughout the observation period and served as informative exogenous predictors within the proposed forecasting framework. Future research should investigate the integration of broader epidemiological, behavioural, mobility, vaccination, and policy-related variables to assess whether they further enhance the forecasting accuracy, robustness, and generalisability of the proposed hybrid approach.

Finally, this study focused exclusively on short-term forecasting of time series with structural change points, using COVID-19 data as an illustrative application. Future work should examine the performance of the proposed framework for medium- and long-term forecasting, particularly for multi-wave epidemic data and other complex non-stationary time series. A comprehensive discussion of medium- and long-horizon forecasting of COVID-19 data can be found in the recent work of e.g., Kalahasti et al. [30].

6. Declarations

6.1. Author Contributions

Conceptualization: D.R.; Methodology: D.R., H., F.Z. P.N.; Software: H., F.Z., P.N.; Validation: D.R., S.P.; Formal Analysis: P.N., W.A., M.I.; Investigation: P.N., W.A., M.I.; Resources: H., F.Z.; Data Curation: P.N., W.A., M.I.; Writing Original Draft Preparation: D.R., H., F.Z.; Writing Review and Editing: D.R., S.P.; Visualization: H., F.Z.; All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

D. Rosadi was supported by the Program Riset Keilmuan, grant numbers 001/E4.1/AK.04.RA/2021 and KEP-2/LPDP.4/2021 to initiate the study. Hermansah was supported by the Hibah Fundamental DIKTI, grant numbers 0419/C3/DT.05.00/2025 and 003/KP-PFR/LPPM/UNRIKA/VI/2025 to finalize the findings.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] D. C. Montgomery, C. L. Jennings, and M. Kulahci, *Introduction to Time Series Analysis and Forecasting*, 3rd ed. Hoboken, NJ, USA: John Wiley & Sons, 2024.
- [2] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, 5th ed. Hoboken, NJ, USA: John Wiley & Sons, 2015.

-
- [3] J. Kim, H. Kim, H. Kim, D. Lee, and S. Yoon, "A comprehensive survey of deep learning for time series forecasting: Architectural diversity and open challenges," *Artif. Intell. Rev.*, vol. 58, no. 1, pp. 1–216, 2025, doi: 10.1007/s10462-025-11223-9.
- [4] L. Tomov, L. Chervenkov, D. G. Miteva, H. Batselova, and T. Velikova, "Applications of time series analysis in epidemiology: Literature review and our experience during COVID-19 pandemic," *World J. Clin. Cases*, vol. 11, no. 29, pp. 6974–6983, Oct. 2023, doi: 10.12998/wjcc.v11.i29.6974.
- [5] S. M. Guzman, J. O. Paz, and M. L. M. Tagert, "The use of NARX neural networks to forecast daily groundwater levels," *Water Resour. Manage.*, vol. 31, no. 5, pp. 1591–1603, Mar. 2017, doi: 10.1007/s11269-017-1598-5.
- [6] C. Jiang and F. Song, "Sunspot forecasting by using chaotic time-series analysis and NARX network," *J. Comput.*, vol. 6, no. 7, pp. 1424–1429, Jul. 2011, doi: 10.4304/jcp.6.7.1424-1429.
- [7] M. A. Shbool, F. Altarazi, and W. Alalaween, "A dynamic nonlinear autoregressive exogenous model to analyze the impact of mobility during COVID-19 pandemic on the electricity consumption prediction in Jordan," *J. Sci. Ind. Res.*, vol. 83, no. 2, pp. 164–173, Feb. 2024, doi: 10.56042/jsir.v83i2.84.
- [8] X. Wang, R. J. Hyndman, F. Li, and Y. Kang, "Forecast combinations: An over 50-year review," *Int. J. Forecast.*, vol. 39, no. 4, pp. 1518–1547, Oct. 2023, doi: 10.1016/j.ijforecast.2022.11.005.
- [9] N. Kourentzes, B. K. Barrow, and S. F. Crone, "Neural network ensemble operators for time series forecasting," *Expert Syst. Appl.*, vol. 41, no. 5, pp. 4235–4244, Apr. 2014, doi: 10.1016/j.eswa.2013.12.011.
- [10] A. J. Scott and M. Knott, "A cluster analysis method for grouping means in the analysis of variance," *Biometrics*, vol. 30, no. 3, pp. 507–512, Sep. 1974, doi: 10.2307/2529204.
- [11] I. E. Auger and C. E. Lawrence, "Algorithms for the optimal identification of segment neighborhoods," *Bull. Math. Biol.*, vol. 51, no. 1, pp. 39–54, Jan. 1989, doi: 10.1007/BF02458835.
- [12] B. Jackson et al., "An algorithm for optimal partitioning of data on an interval," *IEEE Signal Process. Lett.*, vol. 12, no. 2, pp. 105–108, Feb. 2005, doi: 10.48550/arXiv.math/0309285.
- [13] R. Killick, P. Fearnhead, and I. A. Eckley, "Optimal detection of changepoints with a linear computational cost," *J. Amer. Stat. Assoc.*, vol. 107, no. 500, pp. 1590–1598, Dec. 2012, doi: 10.1080/01621459.2012.737745.
- [14] M. Rayungsari, M. Aufin, and N. Imamah, "Parameters estimation of generalized Richards model for COVID-19 cases in Indonesia using genetic algorithm," *Jambura J. Biomath.*, vol. 1, no. 1, pp. 25–30, Jun. 2020, doi: 10.34312/jjbm.v1i1.6910.
- [15] Purwanto, I. S. Astuti, A. Tanjung, F. Rohman, and K. S. B. Utomo, "Mapping the spatial transmission risk and public spatial awareness in the use of personal protective equipment: COVID-19 pandemic in East Java, Indonesia," *Int. J. Disaster Risk Reduct.*, vol. 97, no. 1, pp. 1–12, Nov. 2023, doi: 10.1016/j.ijdr.2023.104018.
- [16] B. Fatimah, P. Aggarwal, P. Singh, and A. Gupta, "A comparative study for predictive monitoring of COVID-19 pandemic," *Appl. Soft Comput.*, vol. 122, no. 1, pp. 1–12, Jul. 2022, doi: 10.1016/j.asoc.2022.108806.
- [17] B. A. Sunjaya, S. D. Permai, and A. A. S. Gunawan, "Forecasting of COVID-19 positive cases in Indonesia using long short-term memory (LSTM)," *Procedia Comput. Sci.*, vol. 216, no. 1, pp. 177–185, 2023, doi: 10.1016/j.procs.2022.12.125.
- [18] P. Thapa, "Time series analysis and forecast of COVID-19 pandemic," in *Decision Sciences for COVID-19*. Cham, Switzerland: Springer International Publishing, 2022, ch. 6, pp. 97–106.
- [19] W. Yang and X. Chang, "Time series analysis and prediction of the trends of COVID-19 epidemic in Singapore based on machine learning," *Comput. Methods Programs Biomed. Update*, vol. 7, no. 1, pp. 1–10, 2025, doi: 10.1016/j.cmpbup.2025.100190.
- [20] J. Wang et al., "Impact of temperature and relative humidity on the transmission of COVID-19: A modelling study in China and the United States," *BMJ Open*, vol. 11, no. 2, pp. 1–10, Feb. 2021, doi: 10.1136/bmjopen-2020-043863.
- [21] M. F. F. Sobral, G. B. Duarte, A. I. G. da Penha Sobral, M. L. M. Marinho, and A. de Souza Melo, "Association between climate variables and global transmission of SARS-CoV-2," *Sci. Total Environ.*, vol. 729, no. 1, pp. 1–7, Aug. 2020, doi: 10.1016/j.scitotenv.2020.138997.
- [22] A. Mollalo, B. Vahedi, and K. M. Rivera, "GIS-based spatial modeling of COVID-19 incidence rate in the continental United States," *Sci. Total Environ.*, vol. 728, no. 1, pp. 1–8, Aug. 2020, doi: 10.1016/j.scitotenv.2020.138884.
- [23] M. Ujiie, S. Tsuzuki, and N. Ohmagari, "Effect of temperature on the infectivity of COVID-19," *Int. J. Infect. Dis.*, vol. 95, no. 1, pp. 301–303, Jun. 2020, doi: 10.1016/j.ijid.2020.04.068.

- [24] R. Killick, I. A. Eckley, and P. Jonathan, "Efficient detection of multiple changepoints within an oceanographic time series," *Proc. 58th Session Int. Stat. Inst.*, vol. 2011, no. 1, pp. 1–10, Aug. 2011.
- [25] N. Pawar, G. V. Hollweg, A. Hussain, W. Sub, and V. Buib, "Enhancing forecasting accuracy in dynamic environments via PELT-driven drift detection and model adaptation," *Next Energy*, vol. 9, no. 1, pp. 1–12, 2025, doi: 10.1016/j.nxener.2025.100462.
- [26] H. T. Siegelmann, B. G. Horne, and C. L. Giles, "Computational capabilities of recurrent NARX neural networks," *IEEE Trans. Syst., Man, Cybern.*, vol. 27, no. 2, pp. 208–215, Apr. 1997, doi: 10.1109/3477.558801.
- [27] K. S. Narendra and K. Parthasarathy, "Identification and control of dynamical systems using neural networks," *IEEE Trans. Neural Netw.*, vol. 1, no. 1, pp. 4–27, Mar. 1990, doi: 10.1109/72.80202.
- [28] C. Jiang and F. Song, "Sunspot forecasting by using chaotic time-series analysis and NARX network," *J. Comput.*, vol. 6, no. 7, pp. 1424–1429, Jul. 2011, doi: 10.4304/jcp.6.7.1424-1429.
- [29] B. W. Silverman, "Using kernel density estimates to investigate multimodality," *J. R. Stat. Soc. Ser. B Methodol.*, vol. 43, no. 1, pp. 97–99, 1981, doi: 10.1111/j.2517-6161.1981.tb01155.x.
- [30] S. Kalahasti et al., "Foundation models for time series forecasting and policy evaluation in infectious disease epidemics," *Epidemics*, vol. 55, no. 1, pp. 1–12, 2026, doi: 10.1016/j.epidem.2026.100916.