CAUCHY – Jurnal Matematika Murni dan Aplikasi
Volume 10 (2) (2025), Pages 790–804
p-ISSN: 2086-0382; e-ISSN: 2477-3344

Rainfall Forecasting using Spatio-Temporal and Neural Network Study Case: Meteorological Data of Madura Island

Ryanta Meylinda Savira[1], Regita Putri Permata*[1], Amalia Nur Alifah[1,2], Yohanes Setiawan[3], and Adzanil Rachmadhi Putra[3]

[1] *Department of Data Science, School of Computing, Telkom University, Surabaya, Indonesia*
[2] *Department of Information Technology, School of Computing, Telkom University, Surabaya, Indonesia*
[3] *Department of Information System, School of Computing, Telkom University, Surabaya, Indonesia*

Abstract

Rainfall forecasting is crucial in meteorological studies due to its significant impact on sectors such as agriculture, which is the main livelihood on Madura Island. This study aims to forecast rainfall on Madura Island using a hybrid approach that combines the Generalized Space-Time Autoregressive-X (GSTARX) model and Neural Network (NN). The data used consist of daily rainfall records from Bangkalan, Sampang, Pamekasan, and Sumenep, covering the period from January 2013 to December 2023. Data from January 2013 to September 2023 were used for training, while data from October to December 2023 were used for testing. The GSTARX model was employed to capture spatio-temporal patterns, while the NN was applied to learn the non-linear relationships in the residuals. The results show that the GSTARX model effectively captures rainfall patterns, though some differences remain compared to the actual data, with RMSE values of Bangkalan (1.514), Sampang (0.256), Pamekasan (0.477), and Sumenep (0.127). Meanwhile, the hybrid GSTARX-FFNN model achieved improved forecasting performance in Sampang (0.392), Pamekasan (0.679), and Sumenep (0.412), although Bangkalan recorded a higher RMSE (1.359). Overall, the GSTARX model proved more effective in forecasting rainfall on Madura Island, delivering smaller and more consistent prediction errors.

Keywords: GSTARX; Hybrid Model; Neural Network (NN); Rainfall; Spasio-Temporal.

Copyright © 2025 by Authors, Published by CAUCHY Group. This is an open access article under the CC BY-SA License (https://creativecommons.org/licenses/by-sa/4.0)

1. Introduction

Rainfall is one of the key weather parameters that plays a crucial role in natural resource management, agriculture, and disaster mitigation. The uncertainty in rainfall patterns due to global climate change demands the development of forecasting methods that are adaptive to both spatial [1, 2]. In regions such as Madura Island, where the majority of the population relies on the agricultural sector, accurate rainfall prediction is essential to support productivity and food security. The complexity of weather dynamics in this area including geographical influences and seasonal wind patterns—makes rainfall forecasting a challenge that requires more advanced and flexible computational approaches [3].

Spatio-temporal approaches have been extensively applied in meteorological forecasting, owing to their capability to integrate spatial and temporal dimensions concurrently [4]. The

*Corresponding author. E-mail: regitapermata@telkomuniversity.ac.id

Received: May 3, 2025 Revised: June 7, 2025 Accepted: July 30, 2025
DOI: https://doi.org/10.18860/cauchy.v10i2.35091

Generalized Spatio-Temporal Autoregressive with Exogenous variables (GSTARX) model serves as an effective method for capturing spatial interdependencies among observation sites and modeling the linear effects of environmental variables [5]. However, the limitation of this model lies in its inability to capture non-linear patterns that are commonly found in weather data. Therefore, integrating it with artificial intelligence methods such as the Feed Forward Neural Network (FFNN) is considered to enhance the model's performance in predicting complex and non-linear patterns [6, 7].

The hybrid GSTARX-FFNN model integrates the advantages of both methods—namely, the ability of GSTARX to capture spatial and temporal dynamics, and the strength of FFNN in modeling non-linear patterns [8]. Numerous prior studies have demonstrated that this hybrid approach can significantly reduce forecasting errors compared to conventional models, especially in the context of rainfall prediction [9, 10]. The inclusion of environmental variables such as temperature, humidity, and wind speed as additional inputs has also been proven to enhance model accuracy, particularly in tropical regions characterized by diverse microclimates [11, 12].

This study also adopts the transfer function approach, in which the transfer function is a forecasting method that combines ARIMA and multiple regression, predicting the output y_t based on its past values and the influence of input variable x_t [13, 14].The transfer function is a key component of dynamic regression modeling, where time-lagged input variables are explicitly incorporated to capture the temporal influence of external factors on the target variable [15, 16]. This formulation enables a more accurate representation of the dynamic relationship between inputs and outputs, which is critical before further processing within a hybrid modeling structure. The integration of hybrid machine learning frameworks has recently proven effective in boosting the accuracy of rainfall forecasting, primarily by utilizing robust data pre-processing, parameter optimization techniques, and ensemble modeling architectures [17, 18].

This research aims to develop a hybrid GSTARX-FFNN model for daily rainfall forecasting on Madura Island by utilizing meteorological data collected between 2013 and 2023. Observations were taken from one station in each regency, with the study limited to the local setting of Madura Island, excluding broader regional generalizations. The novelty of this study lies in its localized application of the hybrid GSTARX-FFNN approach, which integrates spatio-temporal linear modeling with nonlinear neural networks specifically tailored to regional-scale rainfall dynamics. Unlike previous studies that often focus on national or global settings, this research highlights the value of hybrid models in capturing complex weather behavior within a limited geographical context. The main contribution of this work is the implementation and evaluation of a hybrid spatio-temporal and neural network framework that improves rainfall forecast accuracy in data-scarce, agriculture-dependent regions.

The remainder of this paper is organized as follows: Section 2 presents the methodology, including the data description, the modeling framework of GSTARX, and the neural network model based on the Feed Forward Neural Network (FFNN) architecture, followed by the hybrid integration approach. In the proposed hybrid structure, the GSTARX model first captures linear spatio-temporal patterns using a transfer function, whose residuals are computed and modeled accordingly. Meanwhile, the FFNN-based neural network is employed independently to learn non-linear patterns from the original data. Section 3 discusses the results of model implementation and evaluation, including a comparative analysis of forecasting performance. Finally, Section 4 concludes the study by summarizing the findings and offering directions for future research.

2. Methods

The data used in this study consist of daily rainfall records from four locations (based on administrative regencies) on Madura Island, namely Bangkalan, Sampang, Pamekasan, and Sumenep, as shown in Figure 1 and Table 1. The dataset covers the period from January 2013 to December 2023. In this study, data from January 2013 to September 2023 were used for training,

while data from October 2023 to December 2023 were used for testing. The data were obtained from the open-source NASA POWER (Prediction of Worldwide Energy Resources) platform.



Fig. 1: Map of Four Locations
Source: Personal Processing Data

Table 1: Coordinates of Rainfall Observation Locations

Location	Latitude	Longitude
Bangkalan	-7.0284	112.7431
Sampang	-7.1919	113.2512
Pamekasan	-7.1620	113.4847
Sumenep	-6.9139	113.9173

The stages in the research are as follows:

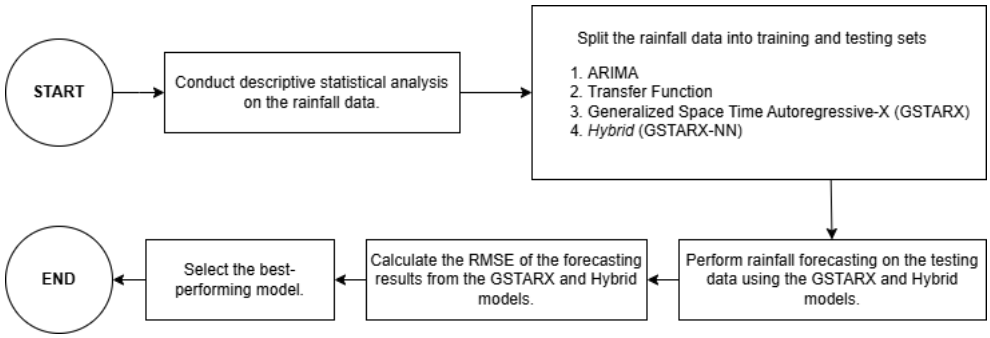


Fig. 2: Research Flowchart

2.1. Transfer Function

The stages carried out in transfer function modeling are:

1. Stationarity testing on the input variables (RH2M, T2M, and WS2M) was conducted using the Augmented Dickey-Fuller (ADF) test to ensure suitability for ARIMA-based pre-whitening. First-order differencing was applied when necessary to achieve stationarity.
2. Perform *pre-whitening* on the input time series (wind speed, temperature, humidity) and the output time series (rainfall).
3. Calculate the cross-correlation between the input and output series to examine the dynamic relationships among variables.
4. Determine the impulse response weights (b, r, s) linking each input series to the rainfall output series, and conduct initial estimation of the error series u_t.
5. Construct a preliminary transfer function model based on the identified orders (b, r, s), and fit an ARIMA model to the error term u_t.
6. Estimate the parameters of the multi-input transfer function model, as shown in Equation

1, which is formulated as follows:

$$y_t = \sum_{h=1}^{H} \frac{\omega_{sh}(B)B^{b_h}}{\delta_{rh}(B)} x_t + \frac{\theta_q(B)}{\phi_p(B)} a_t \tag{1}$$

where:

- y_t = output series at time t,
- x_{ht} = input series for the h-th variable,
- $\omega_{sh}(B)$ = a polynomial operator of order s representing the influence of past values of x_{ht} on y_t,
- $\delta_{rh}(B)$ = a polynomial operator of order r representing the dynamics of the output series in relation to the h-th input,
- $v_h(B) = v_{0h} + v_{1h}B + v_{2h}B^2 + \ldots$ = transfer function weights or impulse response coefficients for the h-th input series,
- u_t = error term assumed to follow a certain ARIMA process.

7. Perform diagnostic checking to obtain a valid model by examining the residuals to ensure that the assumptions of white noise and normality are satisfied.
8. Forecast rainfall using the best-fitting model and calculate the RMSE on the testing data to evaluate the model's accuracy.

2.2. Generalized Space-Time Autoregressive with Exogenous Variables (GSTARX)

In general, this stage involves the identification process of a multi-input transfer function model to obtain residuals from the initial model. The residuals resulting from the transfer function forecasting will then be used in the modeling process to examine the influence of location on rainfall across the four regencies on Madura Island using the GSTARX approach. The steps of the GSTARX model in this study include:

1. The first stage consists of the following steps:
   (a) Identifying the transfer function model for each regency in Madura Island (Bangkalan, Sampang, Pamekasan, and Sumenep) based on input variables including wind speed, temperature, and humidity in relation to rainfall as the output.
   (b) Estimating the parameters of the transfer function model for each regency without modeling the noise series u_t using ARIMA, resulting in an initial model as represented in Equation (2):

$$y_t = \sum_{h=1}^{H} \nu_h(B)x_t + u_t \tag{2}$$

   or

$$y_t = \sum_{h=1}^{H} \frac{\omega_{sh}(B)B^{b_h}}{\delta_{rh}(B)} x_t + u_t \tag{3}$$

   (c) Forecasting y_t based on model (3) to obtain the residual u_t at location i, denoted as $u_{i,t}$.

2. The second stage consists of the following steps:
   (a) Examining the stationarity of the residual series $u_{i,t}$ in terms of its mean using the Multiple Cross-Correlation Function (MCCF) diagram.
   (b) Determining the autoregressive order p of the stationary residual series $u_{i,t}$ using the MPCCF scheme and the minimum AIC criterion.
   (c) Defining the type of spatial weight matrix W to be used, either based on spatial adjacency or distance.

(d) Calculating the spatial weight matrix $W^{(1)}$. The spatial weights used in this study are based on the inverse of the distance between locations. These weights quantify spatial influence, where closer locations exert stronger effects. The weighting scheme is defined in Equation (4):

$$w_{ij} = \frac{1}{d_{ij}} \tag{4}$$

where d_{ij} is the distance between location i and j.
To standardize influence across locations, the weights are row-normalized, ensuring that the sum of each row equals 1. The normalized form is given in Equation (5):

$$w_{ij} = \frac{\frac{1}{d_{ij}}}{\sum_{j=1}^{n} \frac{1}{d_{ij}}}, \quad i \neq j \tag{5}$$

To avoid self-influence, diagonal elements are set to zero ($w_{ij} = 0$). Hence, the resulting weight matrix is row-normalized and asymmetric, as shown in Table 2:

Table 2: Inverse Distance Weights Between Four Locations on Madura Island

Location	**Bangkalan**	**Sampang**	**Pamekasan**	**Sumenep**
Bangkalan	0	$d_{12} = 59.45$	$d_{13} = 83.93$	$d_{14} = 124.13$
Sampang	$d_{21} = 59.45$	0	$d_{23} = 26.15$	$d_{24} = 70.58$
Pamekasan	$d_{31} = 83.93$	$d_{32} = 26.15$	0	$d_{34} = 44.96$
Sumenep	$d_{41} = 124.13$	$d_{42} = 70.58$	$d_{43} = 44.96$	0

The spatial weight between the four locations is calculated using equation (5) as follows:

$$w_{12} = \frac{\frac{1}{d_{12}}}{\frac{1}{d_{12}} + \frac{1}{d_{13}} + \frac{1}{d_{14}}} = \frac{\frac{1}{59.45}}{\frac{1}{59.45} + \frac{1}{83.93} + \frac{1}{124.13}} = 0.4571$$

$$w_{13} = \frac{\frac{1}{d_{13}}}{\frac{1}{d_{12}} + \frac{1}{d_{13}} + \frac{1}{d_{14}}} = \frac{\frac{1}{83.93}}{\frac{1}{59.45} + \frac{1}{83.93} + \frac{1}{124.13}} = 0.3238$$

$$w_{14} = \frac{\frac{1}{d_{14}}}{\frac{1}{d_{12}} + \frac{1}{d_{13}} + \frac{1}{d_{14}}} = \frac{\frac{1}{124.13}}{\frac{1}{59.45} + \frac{1}{83.93} + \frac{1}{124.13}} = 0.2189 \quad \text{and so on.}$$

Thus, the resulting spatial weight matrix W for the four observation locations is as follows:

$$W = \begin{bmatrix} 0 & 0.4571 & 0.3238 & 0.2189 \\ 0.2429 & 0 & 0.5523 & 0.2046 \\ 0.1645 & 0.5282 & 0 & 0.3072 \\ 0.1811 & 0.3186 & 0.5001 & 0 \end{bmatrix}$$

(e) Estimating the parameters of the model using the selected order p from step (b) in the GSTAR-GLS model, as defined in Equation (6), involves capturing both temporal and spatial dependencies in the residual series. The model is formulated as follows:

$$\mathbf{u}(t) = \sum_{k=1}^{p} \left[\Phi_{k0} \mathbf{u}(t-k) + \Phi_{k1} W^{(1)} \mathbf{u}(t-k) \right] + \mathbf{e}(t) \tag{6}$$

(f) Testing the significance of the parameters in the GSTARX-GLS model. If any parameters are found to be insignificant, a restricted model is applied using only the significant parameters.

(g) The final model was selected based on the lowest RMSE, and the GSTARX model with parameters estimated using GLS was then employed to generate forecasts of the residual series $\hat{u}_{i,t}$.

(h) Forecasting rainfall data at the four locations using the GSTARX model, as represented as represented in Equation (7) following:

$$\hat{Y}_{i,t} = \hat{y}_{i,t} + \hat{u}_{i,t} \tag{7}$$

where:
$\hat{Y}_{i,t}$ = forecasted value at time t and location i from the GSTARX model
$\hat{y}_{i,t}$ = forecasted value at time t and location i from Stage I
$\hat{u}_{i,t}$ = forecasted residual at time t and location i

2.3. Hybrid GSTARX-FFNN

The stages of the GSTARX-FFNN hybrid modeling in this study include:

1. Determining the input variables for the hybrid GSTARX-FFNN model based on the residuals from the GSTARX model.
2. Hybrid modeling using the GSTARX-FFNN approach is performed with a neural network consisting of one hidden layer with 2-10 neurons and using the ReLU activation function defined in (8).

$$f(x) = \max(0, x) \tag{8}$$

A linear activation function was used in the output layer, as shown in (9).

$$f(x) = x \tag{9}$$

The model was trained using the Adam optimizer with Mean Squared Error (MSE) as the loss function, formulated in (10).

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{10}$$

Here, n is the number of observations, y_i is the actual value, and $\hat{y}_i$ is the predicted value. Forward propagation in the neural network is represented by (11):

$$a^{(l)} = f(W^{(l)}a^{(l-1)} + b^{(l)}) \tag{11}$$

where $a^{(l)}$ is the output of layer l, $W^{(l)}$ is the weight matrix, $b^{(l)}$ is the bias vector, and f is the activation function. The model was trained for 50 epochs to ensure convergence. As illustrated in Figure 3.

3. Selecting the best hybrid (GSTARX-FFNN) model based on the RMSE value on the testing data.

The modeling flow illustrated in the GSTARX-FFNN hybrid architecture is presented in Figure 3 below:

3. Results and Discussion

To systematically evaluate the performance of each modeling stage, this section is divided into several parts. It begins with the results of the multi-input transfer function model that captures linear input-output relationships. Then, it proceeds to the GSTARX model to incorporate spatial and temporal dependencies using the residuals. Finally, the hybrid GSTARX-FFNN model is implemented to address non-linear patterns and improve forecasting accuracy. Each stage is evaluated using Root Mean Square Error (RMSE) as the primary performance metric.

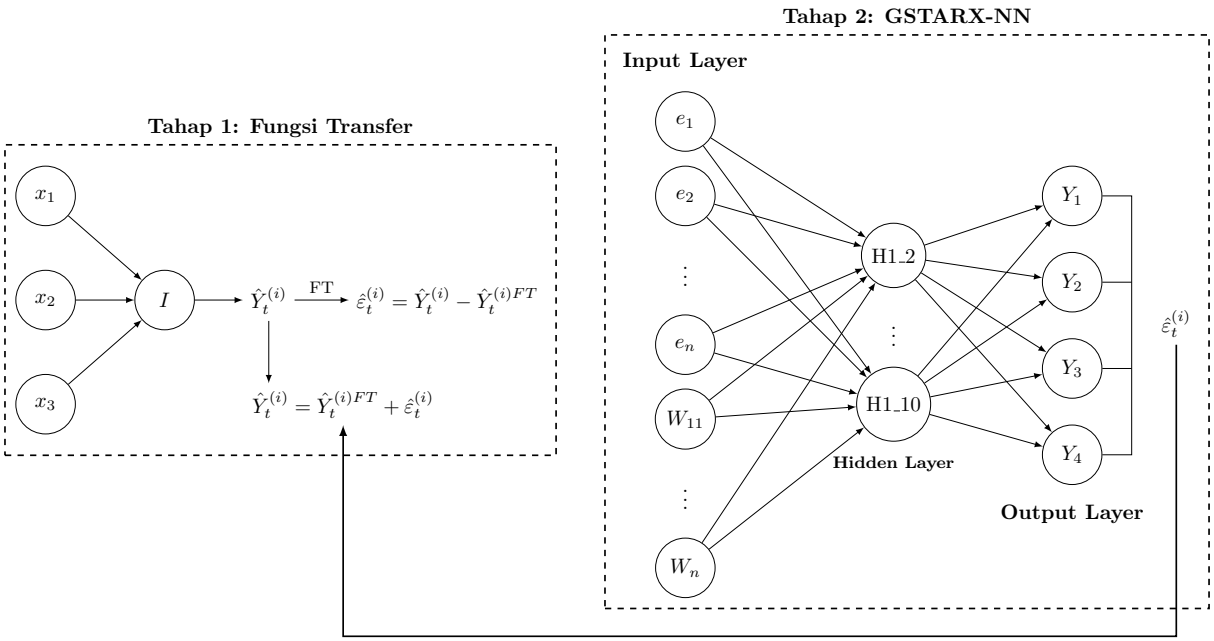


Fig. 3: Hybrid GSTARX-FFNN Modeling Architecture

3.1. Multi-Input Transfer Function Modeling

The initial stage of modeling involves constructing a multi-input transfer function model to capture the linear relationship between rainfall (output) and exogenous variables: relative humidity (RH2M), air temperature (T2M), and wind speed (WS2M). Prior to model construction, a stationarity test was conducted on each exogenous variable using the Augmented Dickey-Fuller (ADF) test, a widely accepted method to assess the presence of unit roots in time series data. The test indicated that the original input variables were non-stationary at level, but became stationary after first-order differencing, as shown by p-values below 0.05 and ADF test statistics exceeding the critical threshold at the 5% significance level.

This stationarity step is critical to avoid spurious regression results and to ensure that the linear modeling assumptions are met. Subsequently, a pre-whitening process was applied to both the input and output series. This step involves fitting ARIMA models to remove autocorrelation and seasonal trends, thereby enabling valid cross-correlation analysis. This process removes autocorrelation and seasonal effects, allowing for more accurate identification of genuine lag relationships in the cross-correlation function (CCF) [5].

After differencing and pre-whitening, cross-correlation analysis (CCF) was performed to determine the optimal lag for each input variable at each location. Table 3 presents the ARIMA model and the temporary estimated parameters of the transfer function.

The parameter estimation results at each location, with a significance level of $\alpha = 0.05$, show that all input variables are statistically significant since p-values were less than 0.0001. The resulting forecast from the transfer function model on the testing set for all input variables across all locations is presented in Figure 4 with RMSE values in 4.

Table 3: ARIMA Model Orders and Transfer Function Parameters for Each Input Variable

Location	Input Variable	ARIMA Model	b	s	r
Bangkalan	RH2M	(2,1,0)	1	0	0
	T2M	(1,1,0)	1	0	0
	WS2M	(3,1,0)	1	0	0
Sampang	RH2M	(2,1,0)	2	0	0
	T2M	(1,1,1)	0	2	0
	WS2M	(2,1,0)	1	0	0
Pamekasan	RH2M	(0,1,1)	1	0	0
	T2M	(1,1,1)	0	2	0
	WS2M	(2,1,0)	2	0	0
Sumenep	RH2M	(0,1,1)	0	9	0
	T2M	(3,1,0)	0	2	0
	WS2M	(1,1,0)	1	0	0

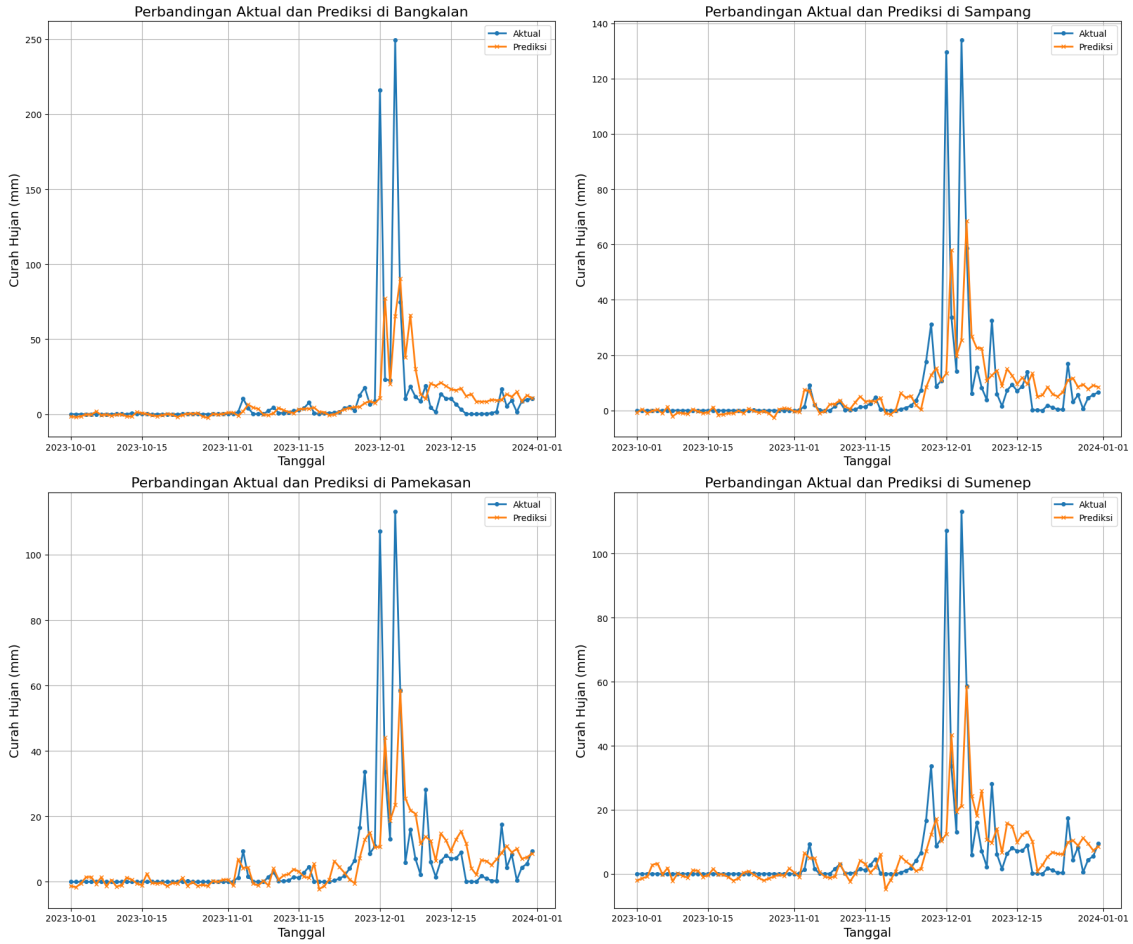


Fig. 4: Forecasting Using Multi-Input Transfer Function

Table 4: RMSE Evaluation of FTMI Model at Four Locations on Madura Island

Location	RMSE
Bangkalan	30.413
Sampang	17.649
Pamekasan	14.741
Sumenep	14.871

The transfer function model effectively captured most of the linear relationships between input and output. However, the results also indicated the presence of significant non-linear

patterns and spatial dependencies. Therefore, the residuals from the transfer function model were used as inputs in the GSTARX model.

3.2. GSTARX Modeling

The initial step in the GSTARX modeling process involved checking the stationarity of the rainfall data using residual analysis. Once the data were confirmed to be stationary, model order selection was conducted by setting the spatial order $q = 1$ for interpretability, and the temporal order $p = 1$ was selected based on the MPCCF method and minimum AIC criterion. Hence, the model used was GSTARX$(1, 1)$.

Based on Table 5, the smallest AIC value was obtained for the AR(6) and MA(6) model, indicating that a longer lag structure provides a better fit to the data. However, the MPCCF analysis in Figure 5 revealed that the strongest temporal dependence occurs at lag 1. This suggests that the relationship between consecutive days is more significant, leading to the selection of a temporal order of $p = 1$. Together with the previously identified spatial order of 1, the resulting GSTARX model is GSTARX$(1, 1)$.

Schematic Representation of Partial Cross Correlations

Variable/Lag	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Y1	-.++	.-..	-+++	....	+-..	+.-+	..+-	.-.-	....	....	.-.-	....	.+-.	....
Y2	.-+.	--+.	-...	.-..	-...	.+-+	-+.-	.+..	-+-.	-+..	.+..	....	.+..	....
Y3	++--	+...	.+..	..-.	..+.	+.-+	.++-	....	..-.	.+..	..+.	....	....	....
Y4	.--+	...+	....	..--	-..+	..-+	...-	....	..+-	.+..	..-.	....	....	+...

\+ is > 2*std error, - is < -2*std error, . is between

Fig. 5: MPCCF residual model plot of rainfall data

Table 5: Minimum Information Criterion Based on AICC

Lag	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5	MA 6
AR 0	8.4686	8.3223	8.3140	8.2859	8.2820	8.2651	8.2065
AR 1	8.3732	8.3068	8.2920	8.2602	8.2510	8.2310	8.1906
AR 2	8.3561	8.2967	8.2847	8.2480	8.2346	8.2190	8.1909
AR 3	8.3355	8.2668	8.2561	8.2376	8.2286	8.2121	8.1840
AR 4	8.3265	8.2426	8.2305	8.2236	8.2198	8.1990	8.1741
AR 5	8.3067	8.2270	8.2173	8.2108	8.1971	8.1885	8.1729
AR 6	8.2398	8.1857	8.1845	8.1788	8.1732	8.1726	8.1705

The spatial weights used in the GSTARX$(1, 1)$ model were inverse-distance weights, assuming that rainfall at each location on Madura Island is influenced by the distance to other locations. Greater distances between locations result in smaller weights compared to closer distances.

Table 6: RMSE Evaluation of GSTARX Model at Four Locations on Madura Island

Location	RMSE
Bangkalan	1.514
Sampang	0.256
Pamekasan	0.477
Sumenep	0.127

Table 6 presents the RMSE values of the GSTARX model. The GSTARX$(1, 1)$ model demonstrates excellent performance in predicting rainfall across the four locations, with low RMSE values: Bangkalan (1.514), Sampang (0.256), Pamekasan (0.477), and Sumenep (0.127). These small RMSE values indicate that the predictions are very close to the actual data. Supported by the comparison graphs, this model effectively captures spatio-temporal patterns, making it a suitable choice for weather data analysis involving both temporal and spatial influences.

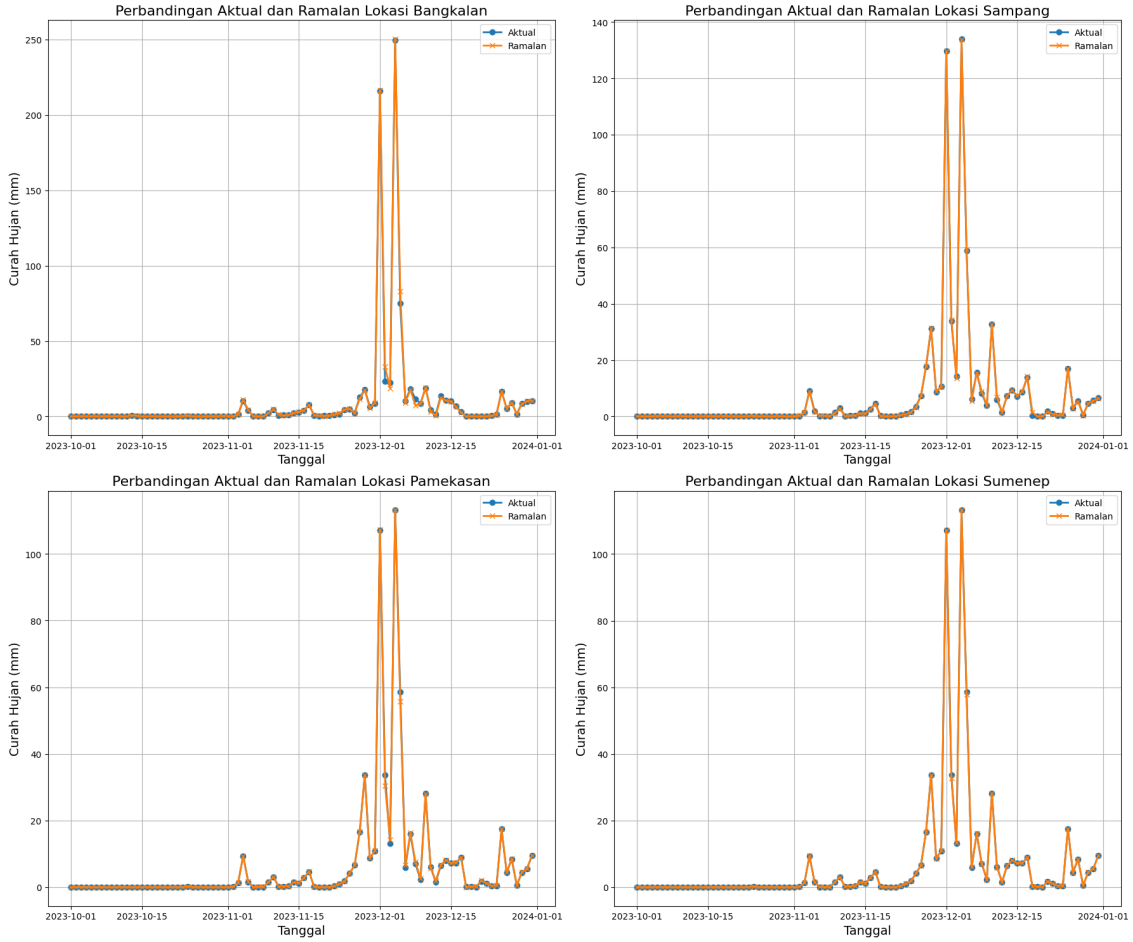


Fig. 6: Comparison Between Actual and Forecasted Rainfall Using the GSTARX Model at Four Locations on Madura Island

3.3. Hybrid GSTARX-FFNN Modeling

The input variables for the hybrid GSTARX-FFNN model were derived from the GSTARX model output, consisting of eight variables, including lagged residuals and spatial weights. The hidden layer used the ReLU (Rectified Linear Unit) activation function, which enables the model to capture non-linear relationships among input variables. Meanwhile, the output layer used a linear activation function, as the goal was to produce continuous outputs in the form of rainfall forecasts. The FFNN architecture is illustrated in the following figure:

Based on the architecture in Figure 7, the hybrid GSTARX-FFNN model was developed and evaluated through a comparison between the actual and forecasted data, as shown in Figure 8.

These results demonstrate that the hybrid model improved forecasting accuracy at all four locations on Madura Island.

The graph shows that the hybrid GSTARX-FFNN model successfully captures the non-linear patterns that could not be modeled by GSTARX alone. The RMSE values of the hybrid model for each location are as follows:

Table 7: RMSE Evaluation of the Hybrid GSTARX-FFNN Model at Four Locations on Madura Island

Location	RMSE
Bangkalan	1.359
Sampang	0.392
Pamekasan	0.679
Sumenep	0.412

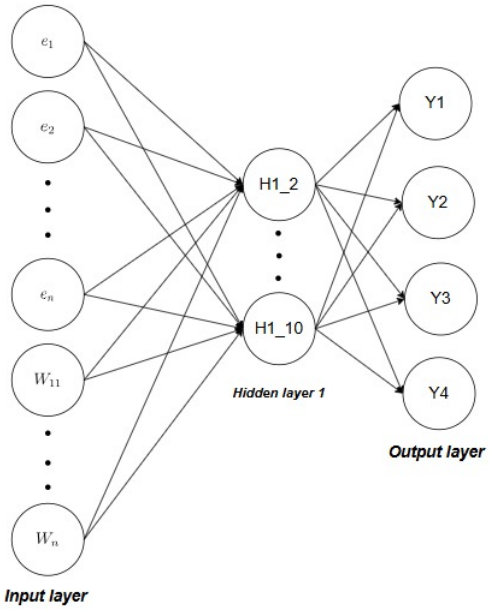


Fig. 7: Neural Network Model Architecture

3.4. Best Model Selection

The best model was selected based on the lowest RMSE value on the testing data (October 2023 – December 2023), which reflects the model's predictive performance. The model comparison plot and evaluation table are presented below:

Table 8: Comparison of RMSE Values

Location	GSTARX	Hybrid GSTARX-FFNN
Bangkalan	1.514	1.359
Sampang	0.256	0.392
Pamekasan	0.477	0.679
Sumenep	0.127	0.412

Based on Table Table 8 and Figure 9, the GSTARX model demonstrates better rainfall prediction performance compared to the Hybrid GSTARX-FFNN, with lower RMSE values across all locations. This indicates that GSTARX is more accurate and consistent in producing rainfall predictions with smaller errors.

4. Conclusion

Rainfall patterns in Madura Island exhibit significant fluctuations influenced by seasonal factors and extreme weather events. Although the average rainfall across the four locations (Bangkalan, Sampang, Pamekasan, and Sumenep) is relatively similar, there are substantial differences in variability, as indicated by high standard deviation values. The peak of rainfall typically occurs in December–January, while the driest period is observed from March to August, reflecting the region's characteristic rainy and dry seasons.

The results of this study indicate that rainfall in Madura Island exhibits a clear seasonal pattern with significant interannual fluctuations, influenced by extreme weather events. The GSTARX model proved effective in capturing the spatio-temporal patterns of rainfall, yielding low RMSE values across all locations and demonstrating strong linear predictive capability. Meanwhile, the hybrid GSTARX-FFNN model enhanced accuracy, particularly in capturing complex non-linear patterns, as observed in the regions of Sampang, Pamekasan, and Sumenep. Although the hybrid model showed a slight decrease in performance in Bangkalan, overall, this combined approach provided more accurate forecasting results. Therefore, the GSTARX-FFNN

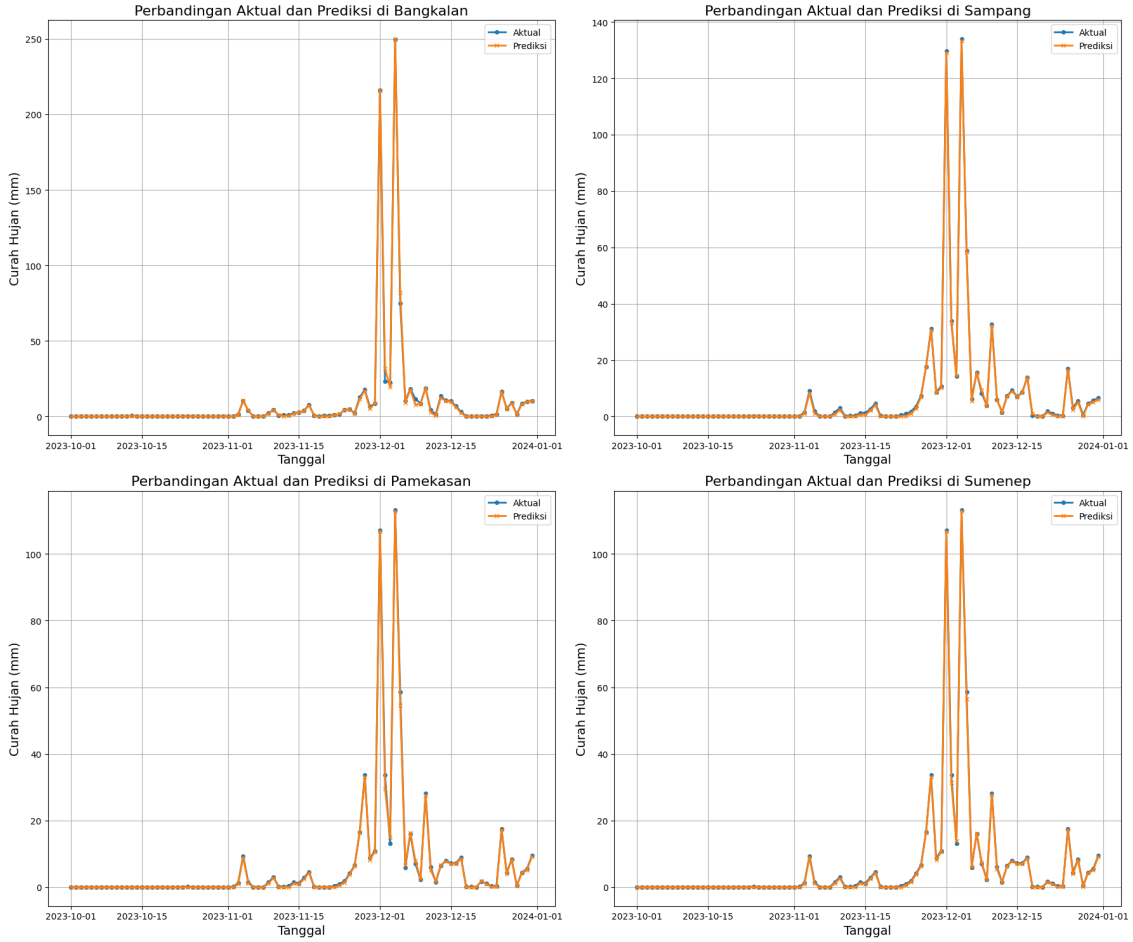


Fig. 8: Comparison Between Actual and Forecasted Rainfall Using the Hybrid GSTARX-FFNN Model at Four Locations on Madura Island

model can serve as a promising alternative for rainfall forecasting in regions characterized by spatio-temporal and non-linear data patterns.

CRediT Authorship Contribution Statement

Ryanta Meylinda: Led the study conceptualization, model development, and data analysis using the GSTARX-FFNN approach. She also performed data preprocessing, visualization, and prepared the original manuscript draft. **Regita Putri:** Supervised the research process, provided guidance in modeling and analysis, and contributed to manuscript review and feedback as the primary advisor. **Amalia Nur Alifah:** Supervised and supported the research methodology and validation of results. She also provided critical input during the review and refinement of the manuscript as co-advisor. **Yohanes Setiawan:** Contributed to the methodological and analytical review, offering suggestions to strengthen the modeling framework and manuscript clarity. **Adzanil Rachmadhi Putra:** Reviewed the research alignment with objectives and assisted in refining the overall structure of the paper and providing editorial feedback.

Declaration of Generative AI and AI-assisted technologies

I acknowledge the use of ChatGPT in assisting with grammar refinement and paraphrasing during the preparation of this manuscript. All AI-generated content was reviewed and edited under human supervision to ensure accuracy and appropriateness, and no data, references, or images produced by AI tools were included in the final version of the manuscript.

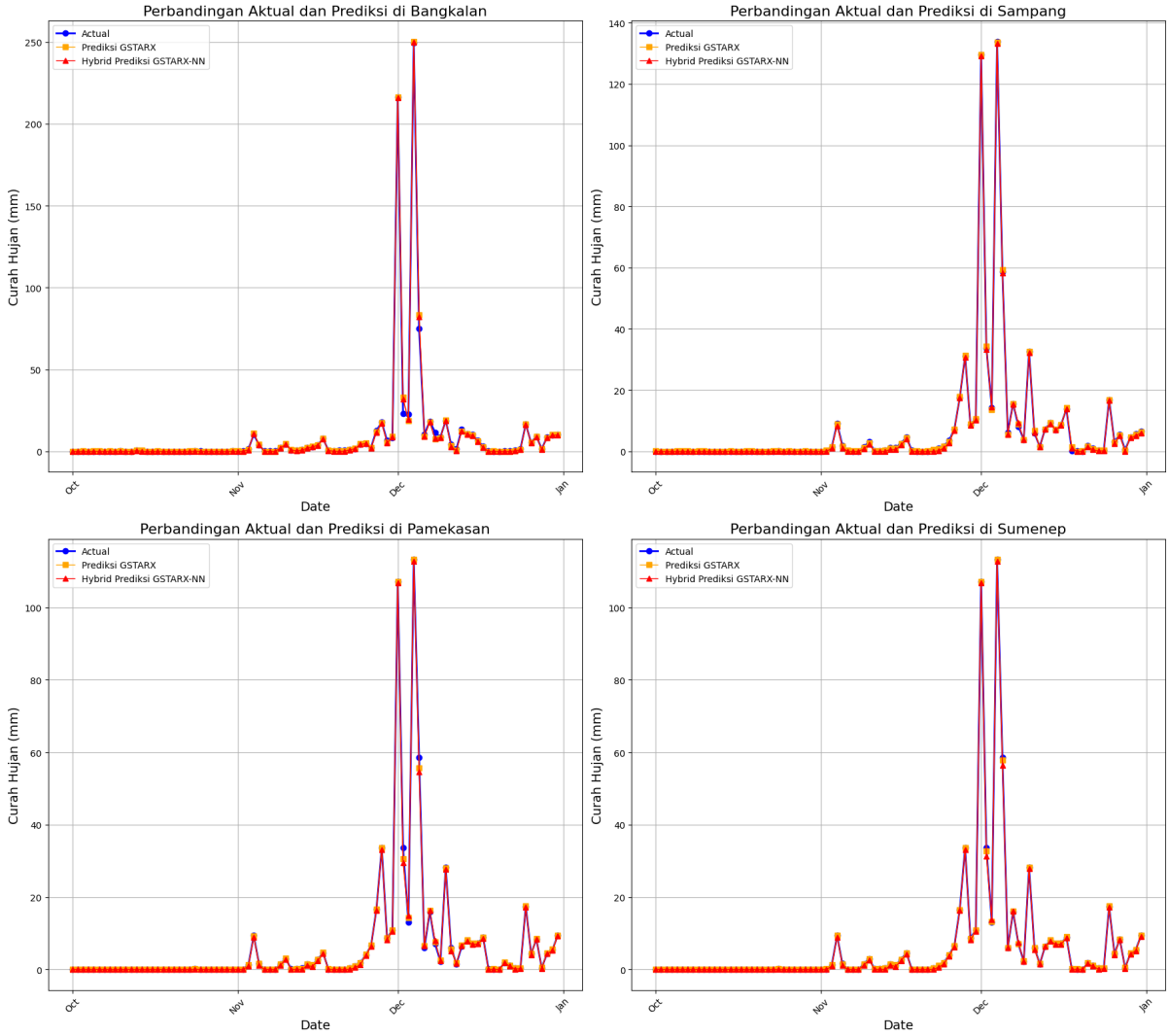


Fig. 9: Comparison Between Actual and Forecasted Values from GSTARX and Hybrid GSTARX-FFNN Models at Four Locations on Madura Island

Declaration of Competing Interest

The authors declare no competing interests.

Funding and Acknowledgments

Thank you to LPPM Telkom University rpvided funding for this research. We are sincerely grateful for the financial support and assistance that they have provided. Activity through Assignment Letter No.088/LIT06LPPM-LIT/2025.

Data and Code Availability

The rainfall dataset and forecasting code used in this study are available upon reasonable request. Interested researchers may contact the corresponding author at ryantmeylindasavira@gmail.com to request access.

References

[1] Ritesh Pandey, Maneesh Upadhya, and Manvendra Singh. “Rainfall Prediction Using Logistic Regression and Random Forest Algorithm”. In: *2024 IEEE International Conference*

on Computing, Power and Communication Technologies (IC2PCT). Vol. 5. 2024, pp. 663–668. DOI: 10.1109/IC2PCT60090.2024.10486681.

[2] Nana Kofi Ahoi Appiah-Badu, Yaw Marfo Missah, Leonard K. Amekudzi, Najim Ussiph, Twum Frimpong, and Emmanuel Ahene. "Rainfall Prediction Using Machine Learning Algorithms for the Various Ecological Zones of Ghana". In: *IEEE Access* 10 (2022), pp. 5069–5082. DOI: 10.1109/ACCESS.2021.3139312.

[3] Kasarapu Ramani, Maddireddy Srilekha Reddy, Kelavathi Bhavani, Syed Feeza, and Vellampalli Sai Bavesh. "Optimization of Rainfall Prediction Using Satellite Data Through Machine Learning and Deep Learning Algorithms". In: *2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS)*. 2024, pp. 1–5. DOI: 10.1109/ICITEICS61368.2024.10625624.

[4] Reza Asadi and Amelia C. Regan. "A spatio-temporal decomposition based deep neural network for time series forecasting". In: *Applied Soft Computing* 87 (2020), p. 105963. DOI: https://doi.org/10.1016/j.asoc.2019.105963. https://www.sciencedirect.com/science/article/pii/S1568494619307446.

[5] Toni Toharudin, Rezzy Eko Caraka, Hasbi Yasin, and Bens Pardamean. "Evolving Hybrid Generalized Space-Time Autoregressive Forecasting with Cascade Neural Network Particle Swarm Optimization". In: *Atmosphere* 13.6 (2022). DOI: 10.3390/atmos13060875. https://www.mdpi.com/2073-4433/13/6/875.

[6] Ashraf Mohamed Hemeida, Somaia Awad Hassan, Al-Attar Ali Mohamed, Salem Alkhalaf, Mountasser Mohamed Mahmoud, Tomonobu Senjyu, and Ayman Bahaa El-Din. "Nature-inspired algorithms for feed-forward neural network classifiers: A survey of one decade of research". In: *Ain Shams Engineering Journal* 11.3 (2020), pp. 659–675. DOI: https://doi.org/10.1016/j.asej.2020.01.007. https://www.sciencedirect.com/science/article/pii/S2090447920300095.

[7] Emmanuel Gbenga Dada, Hurcha Joseph Yakubu, and David Opeoluwa Oyewola. "Artificial Neural Network Models for Rainfall Prediction". In: *European Journal of Electrical Engineering and Computer Science* 5.2 (Apr. 2021), pp. 30–35. DOI: 10.24018/ejece.2021.5.2.313. https://www.ejece.org/index.php/ejece/article/view/313.

[8] A Iriany, D Rosyida, A D Sulistyono, and B N Ruchjana. "Precipitation forecasting using neural network model approach". In: *IOP Conference Series: Earth and Environmental Science* 458.1 (Feb. 2020), p. 012020. DOI: 10.1088/1755-1315/458/1/012020. https://dx.doi.org/10.1088/1755-1315/458/1/012020.

[9] Bestari Safitri, Atiek Iriany, and Ni Wayan Surya Wardhani. "Perbandingan Akurasi Peramalan Curah Hujan dengan menggunakan ARIMA, Hybrid ARIMA-NN, dan FFNN di Kabupaten Malang". In: *Seminar Nasional Official Statistics* 2021.1 (Nov. 2021), pp. 245–253. DOI: 10.34123/semnasoffstat.v2021i1.853. https://prosiding.stis.ac.id/index.php/semnasoffstat/article/view/853.

[10] E Setyowati, Suhartono, and D D Prastyo. "A Hybrid Generalized Space-Time Autoregressive-Elman Recurrent Neural Network Model for Forecasting Space-Time Data with Exogenous Variables". In: *Journal of Physics: Conference Series* 1752.1 (Feb. 2021), p. 012012. DOI: 10.1088/1742-6596/1752/1/012012. https://dx.doi.org/10.1088/1742-6596/1752/1/012012.

[11] Anik Naha Biswas, Yee Hui Lee, Ding Yu Heh, and Shilpa Manandhar. "Study of Temporal and Spatial Correlation of Precipitable Water Vapor with Rainfall for Tropical Region". In: *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*. 2022, pp. 6464–6467. DOI: 10.1109/IGARSS46834.2022.9884477.

[12] A D Sulistyono, Hartawati, A Iriany, N W Suryawardhani, and A Iriany. "Rainfall forecasting in agricultural areas using GSTAR-SUR model". In: *IOP Conference Series: Earth and Environmental Science* 458.1 (Feb. 2020), p. 012041. DOI: 10.1088/1755-1315/458/1/012041. https://dx.doi.org/10.1088/1755-1315/458/1/012041.

[13] Adina Astasia, Septie Wulandary, Ahid Nur Istinah, and Istiqomatul Fajriyah Yuliati. "Peramalan Tingkat Profitabilitas Bank Syariah dengan Menggunakan Model Fungsi Transfer Single Input". In: *Jurnal Statistika dan Aplikasinya* 4.1 (July 2020), pp. 11–22. DOI: 10.21009/JSA.04102. https://journal.unj.ac.id/unj/index.php/statistika/article/view/15538.

[14] Regita Putri Permata, Rifdatun Ni'mah, and Andrea Tri Rian Dani. "Daily rainfall forecasting with ARIMA exogenous variables and support vector regression". In: *Jurnal Varian* 7.2 (2024), pp. 177–188. DOI: 10.30812/varian.v7i2.3202.

[15] K.W. Ng, Y.F. Huang, C.H. Koo, K.L. Chong, Ahmed El-Shafie, and Ali Najah Ahmed. "A review of hybrid deep learning applications for streamflow forecasting". In: *Journal of Hydrology* 625 (2023), p. 130141. DOI: https://doi.org/10.1016/j.jhydrol.2023.130141. https://www.sciencedirect.com/science/article/pii/S0022169423010831.

[16] Regita Putri Permata, Amri Muhaimin, and Sri Hidayati. "Rainfall Forecasting With An Intermittent Approach Using Hybrid Exponential Smoothing Neural Network". In: *BAREKENG: Jurnal Ilmu Matematika dan Terapan* 18.1 (2024), pp. 0457–0466. DOI: 10.30598/barekengvol18iss1pp0457-0466.

[17] Sam-Quarcoo Dotse, Isaac Larbi, Andrew Manoba Limantol, and Liyanage C De Silva. "A review of the application of hybrid machine learning models to improve rainfall prediction". In: *Modeling Earth Systems and Environment* 10.1 (2024), pp. 19–44. DOI: 10.1007/s40808-023-01835-x.

[18] Amol Ashok Patil and Kedar Kulkarni. "A Hybrid Machine Learning - Numerical Weather Prediction Approach for Rainfall Prediction". In: *2023 IEEE India Geoscience and Remote Sensing Symposium (InGARSS)*. 2023, pp. 1–4. DOI: 10.1109/InGARSS59135.2023.10490397.