A Comparative Evaluation of Various Imputation Methods for LSTM-Based Climatological Time Series Forecasting

Dhia Rafifah Thifal[1], Aji Prasetya Wibawa[2*], Adelia Desyana Eka Putri [3], Adelia Khansa Ristiaputri[4], Adhelia Wida Khaidir[5], Agung Bella Putra Utama[6]

[1,2,3,4,5,6] *Department of Electrical Engineering and Informatics, Universitas Negeri Malang, Malang, Indonesia*

*corr-author: aji.prasetya.ft@um.ac.id

Abstract - Missing values substantially degrade the reliability of environmental time-series forecasting; however, prior studies largely evaluate imputation methods in isolation without systematically linking missingness mechanisms to deep learning forecasting performance. To address this gap, this study proposes a mechanism-aware comparative framework that evaluates deletion and six imputation methods (Mean, Median, Mode, LOCF, KNN, and MICE) across three environmental time-series datasets with naturally occurring missing values, using LSTM as the forecasting model. The novelty lies in jointly analyzing statistical error (MAPE, RMSE), goodness-of-fit (R^2), and statistical significance to identify structurally aligned imputation strategies under different missingness patterns. Experimental results show that deletion as baseline consistently produces the worst performance (MAPE: 5.91429; 7.35000; 2.84881), whereas imputation reduces proportional error by more than 70% on average ($p < 0.05$). LOCF performs best under temporal dependency (MAPE 0.73959; R^2 0.92757), KNN achieves the most balanced performance under MCAR-like behavior (R^2 0.94086), and Mean imputation yields the lowest error in MAR-structured data (MAPE 0.41560; R^2 0.97077). These findings demonstrate that imputation effectiveness depends on alignment with missingness structure rather than methodological complexity, providing evidence-based guidance for robust environmental.

Keywords: Imputation, missing values, LSTM, forecasting, climatology.

I. INTRODUCTION

Time series data is very important across various scientific fields because it is used to understand the dynamics of changes in natural phenomena in a sustainable manner [1]. In the context of climatology and environmental data science, the completeness of time-series data is crucial for capturing seasonal patterns, long-term trends, and anomalies that affect analysis and decision-making [2]. When data is lost, the temporal structure becomes incomplete, potentially leading to inaccurate interpretations by statistical analysis and prediction models. Data loss or missing values in a time series not only degrade data quality but also can lead to estimation bias and lower model stability [3], especially in deep learning-based models such as Long Short-Term Memory (LSTM), which are highly sensitive to temporal pattern continuity [4]. Since LSTM models learn sequential dependencies, disruptions in temporal consistency can significantly affect forecasting accuracy and robustness.

Handling missing values is an important stage before the analysis and forecasting process is carried out. Missing values due to data loss are generally categorized into three types: Missing Completely at Random (MCAR), Missing at Random (MAR), and Missing Not at Random (MNAR) [5]. Every missing-value category has different characteristics, requiring an appropriate handling approach to maintain the data structure and ensure stable prediction results. Various approaches have been developed to deal with missing values in time series data. Simple methods, such as removing rows with missing data, can be used; however, they may eliminate important information and increase bias [6]. Alternatively, statistical imputation methods such as mean, median, and mode imputation [7], as well as time-based approaches such as Last Observation Carried Forward (LOCF) [8, 9], have been widely applied. More advanced techniques include machine learning-based methods such as k-Nearest Neighbors (KNN) [10, 11] and Multiple Imputation by Chained Equations (MICE) [12, 13]. These methods aim to minimize data distortion while preserving temporal structure, ensuring the data remains suitable for predictive modeling.

Although various methods for handling missing values have been proposed, existing studies still exhibit several limitations. Many investigations focus only on a single dataset, limiting the generalizability of their findings across different missing data characteristics [14-

16]. Some studies evaluate imputation performance under only one missing-value mechanism, without systematically comparing MCAR, MAR, and MNAR scenarios [17-19]. Other research concentrates solely on comparing imputation accuracy without examining its impact on deep learning-based forecasting models such as LSTM [20, 21]. Furthermore, comparative studies integrating multiple environmental datasets with all three missingness types and evaluating forecasting robustness remain scarce. As a result, the relationship between missing-value mechanisms, imputation strategies, and LSTM forecasting performance in climatology and environmental data science remains incompletely understood.

To address these gaps, this study proposes a structured comparative framework to evaluate the performance of an LSTM-based forecasting model under different missing-value mechanisms across multiple environmental time-series datasets, including daily temperature, sunspot activity, and water temperature. The novelty of this research lies in its integrated evaluation framework, which combines multiple datasets, missing-value mechanisms, and imputation approaches within a unified robustness analysis of LSTM forecasting performance.

II. METHOD

This section describes the methodological stages used in research. Each stage is carried out to ensure data quality and the accuracy of the developed model. The complete sequence of the research procedure is presented in Fig. 1. At the same time, detailed explanations of each stage are provided in the subsequent subsections.

A. Data Acquisition

This study uses three time-series datasets from the fields of climatology and environmental science, obtained from Kaggle. The first dataset, Eighty Years of Canadian Climate Data, contains daily climate records spanning 80 years from across Canada. The second dataset, Daily Sun Spot Data, contains the number of daily sunspots from January 1, 1818, to October 31, 2019, as a representation of the Sun's activity that affects the variation in light and heat radiation reaching Earth. The third dataset, Sofia Weather Records, contains daily weather data for Sofia, Bulgaria, including air temperature records. The descriptive statistics of the three datasets are presented in Table I.

Table I presents a statistical summary of the three datasets used in the study, with a diverse distribution of values. Dataset 1 shows Canada's daily air temperature, with an average of −0.42°C and a wide range of −48.1°C to 23.9°C, and contains 1,691 missing values. Dataset 2 reflects sunspot activity, with an average of 79.25, an extreme range of 528, an anomaly at the minimum value (−1), and 3,204 missing values. Meanwhile, Dataset 3 depicts daily temperatures in Sofia, with an average of 1,647°C, a range of −36.9°C to 29.1°C, and 50 missing values. All three datasets exhibit variations in characteristics that need to be considered during data processing.

The target attributes for each dataset are selected based on their relevance to the purpose of time-series analysis and the availability of observation values [22]. This research requires data with missing values. In Dataset 1, the mean_temperature_whitehorse attribute is used as the target because it represents the stable daily average temperature and is best suited for modeling local climate dynamics in Whitehorse, Canada. In Dataset 2, the number_of_sunspot attribute was chosen as the target because it is a direct indicator of sunspot activity, which is highly variable in studies of solar radiation variation. Meanwhile, in Dataset 3, the air_temperature attribute is used as the target because it captures daily temperature changes and serves as a core variable for predicting weather context.

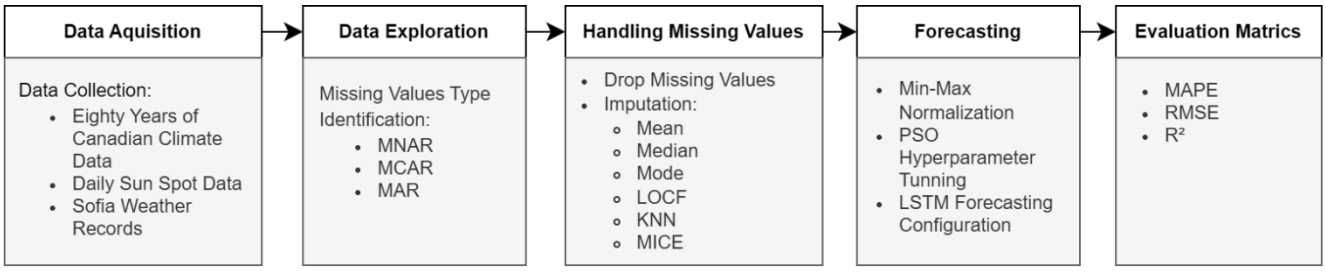

Fig. 1 Research methods

TABLE I
STATISTICS OF THE DATASET

Statistics	Dataset 1	Dataset 2	Dataset 3
Number of	27.530	73.718	87.599
Average	-0.42°C	79.25	1.647
Standard	12.87	77.47	9.695
Minimum	-48.1°C	-1	-36.9
Quartile 1	-8.4°C	15	-3.6
Median	1.9°C	58	1.7
Quartile 3	10.0°C	125	8.3
Maximum	23.9°C	528	29.1
Total	1.691	3.204 (4.35%)	50

B. Data Exploration

Based on Table I, each dataset has a different number of missing values in the target attributes, with varying proportions and characteristics. Dataset 1 contains 1,691 missing values, Dataset 2 contains 3,204 missing values, while Dataset 3 contains only 50 missing values. The presence of this missing value is important for evaluating the causes of data loss patterns, which can affect the quality of the forecasting model [23]. In addition, the differences in data types and the initial reasons for missing values in each dataset provide a methodological basis for testing different imputation methods [24]. A complete overview of the types of missing values in each dataset is shown in Table II.

In Table II, the mechanisms underlying missing data in each dataset can be formally described using probability notation [25]. Let Rdenote the missingness indicator (where $R = 1$ indicates a missing value and $R = 0$otherwise), and let Xrepresent the complete set of variables. MNAR mechanism in (1), the probability of missingness depends on both observed and unobserved values, expressed as $P(R \mid X) = P(R \mid X_{observed}, X_{missing})$. This condition is observed in Dataset 1, where missing values in the *mean_temperature_whitehorse* attribute tend to occur during extreme temperature periods, indicating that the temperature value itself directly influences the probability of data loss. In contrast, Missing Completely at Random (MCAR) in (2) is defined as $P(R \mid X) = P(R)$, meaning that the probability of missingness is independent of any variable in the dataset. Dataset 2 satisfies this assumption, as missing values in *number_of_sunspot* do not exhibit any systematic relationship with other attributes or with the target value, and are instead associated with archival inconsistencies. Meanwhile, Missing at Random (MAR) in (3) follows the formulation $P(R \mid X) = P(R \mid X_{observed})$, where the probability of missingness depends only on observed variables but not on the missing variable itself. This mechanism characterizes Dataset 3, in which missing values in *air_temperature* are associated with other observed attributes, such as dew_point_temperature and surface_air_pressure, without being directly influenced by air_temperature itself. These differing missing-data mechanisms justify evaluating multiple imputation strategies across datasets.

C. Handling Missing Values

Handling missing values is a crucial step in data preprocessing, especially in time-series analysis used for forecasting. Incompleteness or improper handling of data can degrade model accuracy and lead to biased interpretations [26]. There are two main approaches to handling missing values: deletion and imputation [27]. However, deletion techniques may lead to substantial information loss and disrupt temporal continuity, particularly in sequential datasets. Therefore, imputation methods are generally preferred in time-series forecasting because they aim to reconstruct missing observations while preserving the underlying data structure.

Imputation methods are being developed to maintain temporal structures and patterns over time series data [28]. This method has many variations, starting from simple statistical techniques to algorithmic approaches. In this study, six imputation methods were selected to represent basic and advanced approaches and evaluated for their effects on forecasting quality across datasets with different data-loss patterns (Table III) [27]. The selection of these methods aims to assess each technique's ability to preserve the characteristics of the original data and to support accurate forecasting. Each method in Table III may produce different impacts because the three datasets exhibit distinct missing-value patterns: MNAR, MCAR, and MAR.

To provide a clearer methodological understanding, each missing value handling method presented in Table III is formally expressed using a unified mathematical notation. Let $X = \{x_1, x_2, \ldots, x_n\}$ denote a time-series dataset, where $x_t$represents the observation at time t. Let $\mathcal{O} \subset \{1,2, \ldots, n\}$ be the set of observed indices and $\mathcal{M} \subset \{1,2, \ldots, n\}$ be the set of missing indices, with $n_{obs} = |\mathcal{O}|$.

TABLE II
OBSERVED MISSING DATA MECHANISMS ACROSS DATASETS

Dataset	Type of Missing Values	Formula		Description
1	Missing Not at Random (MNAR)	$P(R\|X) = P(R\|X_{observed}, X_{misssing})$	(1)	Data loss is directly affected by the value of the target attribute
2	Missing Completely at Random (MCAR)	$P(R\|X) = P(R)$	(2)	Completely random data loss
3	Missing at Random (MAR)	$P(R\|X) = P(R\|X_{observed})$	(3)	Data loss is affected by other attributes, but not by the value of the target attribute.

TABLE III
HANDLING MISSING VALUES

Method	Description	Advantages	Disadvantages
Drop Missing values	Delete instances that contain missing values.	Simple and ideal if missing < 1%.	Creating information loss.
Mean Imputation	Imputation based on the average attribute.	Quick and easy to implement.	Eliminate variation and bias from extreme data.
Median Imputation	Imputation based on the middle value of the data.	Robust against outliers.	Not considering temporal context.
Mode Imputation	Imputation based on the value that appears most often.	Stable if there is a repetition of the value.	Bias if the dominant value is not representative.
MICE Imputation	Imputation based on a model that considers other variables.	Taking into account other variables.	Bad for univariate but complex time-series.
LOCF Imputation	Imputation based on the last available observation value.	Maintaining temporal patterns.	Causes stagnation if missing in a row.
KNN Imputation	Based on the proximity of the feature value to the nearest neighbor of K.	Considering the similarities of observations.	Sensitive to K parameters and data distribution.

1) Drop Missing Values: The drop method removes all observations containing missing values in (4).

$$X^* = \{x_t \mid t \in \mathcal{O}\} \tag{4}$$

2) Mean Imputation: The mean of observed values is computed as in (5).

$$\bar{x} = \frac{1}{n_{obs}} \sum_{t \in \mathcal{O}} x_t \,, x_t = \begin{cases} \bar{x}, & t \in \mathcal{M} \\ x_t, & t \in \mathcal{O} \end{cases} \tag{5}$$

3) Median Imputation: The median of observed values is defined as in (6).

$$\tilde{x} = \text{median}(\{x_t \mid t \in \mathcal{O}\}),\ x_t = \begin{cases} \tilde{x}, & t \in \mathcal{M} \\ x_t, & t \in \mathcal{O} \end{cases} \tag{6}$$

4) Mode Imputation: The mode of observed values (x_{md}) is written in (7).

$$x_{mode} = \arg\max_v \text{freq}(v),\ x_t = \begin{cases} x_{mode}, & t \in \mathcal{M} \\ x_t, & t \in \mathcal{O} \end{cases} \tag{7}$$

5) MICE Imputation: For a target variable X predicted using auxiliary variables Z, the iterative regression model is defined as in (8). Where $f(\cdot)$ is a regression model, $\epsilon_t \sim N(0, \sigma^2)$, and k denotes the iteration step.

$$x_t^{(k+1)} = f^{(k)}(Z_t) + \epsilon_t^{(k)} \tag{8}$$

6) LOCF Imputation: In time-series data, missing values are replaced by the most recent observed value, as in (9).

$$x_t = \begin{cases} x_{t-1}, & t \in \mathcal{M} \\ x_t, & t \in \mathcal{O} \end{cases} \tag{9}$$

7) KNN Imputation: The Euclidean distance between observations t and s based on p auxiliary variables Z is defined as in (10) [29].

$$d(t,s) = \sqrt{\sum_{j=1}^{p} (z_{tj} - z_{sj})^2} \tag{10}$$

The missing value at time t is estimated using the average of its K nearest neighbors defined as in (11), where $N_K(t)$ denotes the set of K nearest neighbors of observation t.

$$x_t = \frac{1}{K} \sum_{s \in N_K(t)} x_s \tag{11}$$

All missing value handling methods in this study are formulated using a unified time-series notation, where $X = \{x_1, x_2, \ldots, x_n\}$ represents the observed sequence, $\mathcal{O}$ denotes the set of observed indices, and $\mathcal{M}$ denotes the set of missing indices. The drop method removes all observations in $\mathcal{M}$, producing a reduced dataset X^*.

D. Forecasting

Forecasting in this study used the Long Short-Term Memory (LSTM) method implemented in Python with TensorFlow and Keras. In contrast, Scikit-learn and Pandas were used for preprocessing and evaluation. The dataset was cleaned, the target variable selected, and then split into 80% for training and 20% for testing. A sliding-window approach with 24 time steps was applied, where the previous 24 observations were used to predict the next value, and the input was reshaped into a three-dimensional format suitable for LSTM. Model training used mini-batch learning, and performance was evaluated using RMSE, MAPE, and R^2 before further optimization using Particle Swarm Optimization.

Normalization plays an important role in time-series models that are sensitive to scale differences between values, such as LSTM [30]. Data normalization was performed using the MinMaxScaler to scale values to the range 0-1, thereby stabilizing the training process [30]. The Min-Max normalization formula is shown in (12).

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (12)$$

In the Min-Max Normalization formula, the calculation process is carried out by subtracting the original value (x) from the minimum value ($x_{\min}$). Then the result is divided by the difference between the maximum values ($x_{\max}$) and minimum values ($x_{\min}$) resulting in generating new value that is standardized and proportional to the initial distribution (x') [31]. This normalization step ensures that the input data are consistently scaled, thereby improving numerical stability and facilitating effective learning in the LSTM forecasting model described in the following section.

LSTM is a development of the Recurrent Neural Network (RNN) designed to overcome vanishing gradients and increase the ability to model long-term temporal dependencies in time series [32]. LSTM has three main gates (input, forget, and output gates) that govern the information stored or discarded, allowing the model to retain relevant internal memory. LSTMs' mechanism can capture nonlinear patterns, long-term and seasonal trends, and extreme fluctuations [33]. These characteristics make LSTM suitable for climatological data, sunspot activity, and weather recordings, and therefore appropriate as the primary forecasting model in this study. The mathematical formulation of the LSTM is described in (13-18).

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (13)$$

$$f_t = \sigma\big(W_f[h_{t-1}, x_t] + b_f\big) \quad (14)$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (15)$$

$$\tilde{c}_t = tanh(W_c[h_{t-1}, x_t] + b_c) \quad (16)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (17)$$

$$h_t = o_t \odot tanh(c_t) \quad (18)$$

Equations (13) to (18) describe the internal mechanisms of the LSTM in processing information at each step of time. Value of input gate (i_t), forget gate (f_t), and output gate (o_t) Calculated using the sigmoid activation function that received the previous hidden state merge (h_{t-1}) and current inputs (x_t), as a result, the three gates determine the proportion of information that enters, retains, or forwards. The candidate of cell state ($\tilde{c}_t$) It is then passed through a tanh activation function to generate a new representation ready to be combined. Next, the main cell state (c_t) updated by combining previous memory (c_{t-1}) which is maintained by the forget gate with new information filtered by the input gate. Lastly, the hidden state (h_t) is generated by multiplying the cell state transformed by tanh in the output gate, so that only the selected information is passed to the next time step [34]. Through this structure, an LSTM dynamically regulates the flow of information to meet the model's needs.

To achieve the best results, Particle Swarm Optimization (PSO) is used to find the optimal hyperparameter combination for the LSTM model. PSO works by forming a population of particles that move in the search space to maximize the quality of the model with minimal time [35]. In this study, each particle represents a specific configuration of LSTM hyperparameters, including the number of units, learning rate, batch size, and number of epochs. The particles iteratively update their positions based on both their individual best (pbest) and global best (gbest) performances, enabling the swarm to converge to an optimal solution efficiently. This optimization process aims to maximize forecasting accuracy while minimizing training time and prediction error. The detailed representation of hyperparameter configurations used in this study is presented in Table IV.

In this study, the final configuration of model parameters was automatically determined through the hyperparameter tuning process using PSO. The optimization procedure systematically searched the predefined parameter space to identify the combination that yielded the best forecasting performance, as measured by the selected evaluation metrics. Consequently, the parameter settings reported in this research are not the result of manual trial-and-error selection but are derived from the PSO-driven optimization process.

E. Evaluation Metrics

The evaluation of model performance in this study was carried out using three main metrics, namely Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and Coefficient of Determination (R^2). These three metrics were chosen because they provide a comprehensive picture of the prediction error rate, model stability, and the model's ability to explain temporal patterns in time series data [44]. To ensure balanced model assessment and avoid bias toward a single performance indicator, a multi-metric selection criterion was applied by jointly considering MAPE, RMSE, and R^2 in the model comparison process [45]. The combination of these three complementary indicators enables a more objective, multidimensional evaluation of how different imputation strategies affect downstream forecasting performance.

MAPE is an evaluation metric that measures the average percentage of absolute error between the actual and predicted values [46]. The calculation formula, as shown in (19) [44], calculates the absolute difference between the actual values (y_i) and prediction ($\hat{y}_i$), then divide it by the actual value to obtain the relative proportion of error. The entire proportion of the error is then averaged by the number of observations (n) and multiplied by 100%.

$$MAPE = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (19)$$

RMSE is an evaluation metric that measures the average magnitude of the squared error between the actual and predicted values and returns it to the original scale via the square root, as shown in (20) [47] [44]. By squaring the difference $(y_i - \hat{y}_i)$, RMSE provides a greater penalty for extreme error. The average value of the squared error obtained is then rooted so that the result remains in the same unit as the original data. RMSE is sensitive to outliers and large fluctuations. RMSE is an appropriate metric for evaluating the accuracy of forecasting models that require consistent performance under extreme-value conditions and sudden changes.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \quad (20)$$

R^2 is an evaluation metric that measures how much variation in the actual values the model explains, indicating the model's ability to capture the time-series archetype after the imputation process. Based on (21) [44], R^2 is calculated by comparing the total Sum of Squared Errors (SSE) to the total variation of actual data or Total Sum of Squares (TSS). A value close to 1 indicates that the model captures the data's variance well. By distinguishing between prediction errors and natural variations in the data, R^2 enables a more comprehensive assessment of the model's accuracy in understanding the data's structure. Therefore, this metric is useful for evaluating the extent to which forecasting models can explain complex time-series dynamics.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \quad (21)$$

To ensure robustness, MAPE evaluation is complemented by a paired t-test to determine whether performance differences between imputation strategies are statistically significant rather than due to random variation. The test assesses whether the mean difference in forecasting errors across identical test sets differs significantly from zero at a predefined significance level. In addition to quantitative metrics, graphical time-series analysis of actual versus predicted values is conducted to assess trend alignment, responsiveness to extreme fluctuations, and stability across different missing-data mechanisms.

III. RESULT AND DISCUSSION

In this section, the performance of the LSTM model in predicting three time-series datasets is analyzed, with various methods for handling missing values applied. A comparison was made between a baseline model that uses the drop-missing-values technique and a model that uses various imputation methods. Model performance was evaluated using three main metrics: MAPE, RMSE, and R^2. This analysis aims to assess the extent to which the imputation strategy improves prediction accuracy compared to the lost-data erasure approach. The full results are presented in Table V.

Across the three datasets from results in Table V, removing incomplete observations significantly reduces prediction accuracy and disrupts temporal dependencies required by the LSTM model. Most imputation techniques reduce forecasting error by more than 70%, with paired t-tests confirming that these improvements are statistically significant. However, no single method consistently dominates across all conditions: LOCF performs best in Dataset 1 with MNAR-like temporal continuity; MICE achieves the lowest MAPE in Dataset 2 under MCAR conditions but shows higher RMSE due to occasional large residuals; while KNN provides more stable absolute error and better variance preservation. In Dataset 3 with MAR characteristics, central tendency methods such as Mean and Median perform similarly because the dataset's distribution is stable. Overall, the combined MAPE, RMSE, and R^2 analyses indicate that imputation effectiveness depends on the missingness mechanism, temporal dependency, and distribution stability, highlighting the importance of adaptive imputation strategies for LSTM-based time-series forecasting.

TABLE IV
HYPERPARAMETER TUNING PSO

Hyperparameter	Search Space / Values	Information
Hidden Layers	2, 10 [36]	Consistently < 5 across multiple datasets.
Neurons per Layer	1, 100 [37]	The best balance between accuracy and complexity of the model.
Activation Function	{tanh, ReLU, sigmoid} [38]	Stable for forecasting time series.
Optimizer	{Adam, RMSprop, SGD} [39]	Selected by PSO to ensure convergence.
Loss Function	{MSE, MAE} [40]	Minimize forecasting errors.
Batch Size	32 – 64 [41]	Provide a stable training process on the dataset.
Epochs	50 – 100	Tuned to avoid underfitting or overfitting
Dropout	0.0 – 0.5 [42]	Reduces overfitting.
PSO Parameters	swarm=10, generations=5, velocity [0,1], pbest=1.5, gbest=2.0 [43]	Optimization process.

TABLE V
MODEL PERFORMANCE EVALUATION RESULTS

Method	MAPE			RMSE			R^2		
	Dataset 1	Dataset 2	Dataset 3	Dataset 1	Dataset 2	Dataset 3	Dataset 1	Dataset 2	Dataset 3
Drop Missing Values	5.91429	7.35000	2.84881	4.09995	15.70427	1.49142	0.88524	0.93480	0.97492
Mean Imputation	0.95311	4.61261	0.41560	3.52167	16.64167	1.60957	0.91533	0.93439	0.97077
Median Imputation	0.78635	4.01872	0.53360	3.31405	20.32393	1.50689	0.92513	0.90215	0.97439
Mode Imputation	1.00753	8.59048	0.78746	4.03448	20.06531	2.37675	0.88887	0.90462	0.93627
MICE Imputation	0.90673	2.44567	1.25159	3.56648	24.43312	4.40376	0.91316	0.85858	0.78186
LOCF Imputation	0.73959	6.61441	0.77787	3.25713	18.34903	2.15973	0.92757	0.92024	0.94740
KNN Imputation	0.82810	3.12849	0.92293	3.55424	15.80088	2.24950	0.91375	0.94086	0.94348

Across the datasets, MICE shows the greatest variability in R^2 performance, particularly in Dataset 3, indicating that its multivariate iterative estimation may introduce structural distortion when inter-variable dependencies are weak or moderate. Mode imputation also performs poorly because replacing continuous values with the most frequent observation disrupts natural variance patterns and weakens structural consistency. When considered alongside MAPE and RMSE results, the R^2 analysis highlights the importance of multi-metric evaluation: LOCF consistently performs best in Dataset 1 under strong temporal dependency; KNN provides the most stable structural performance in Dataset 2 despite MICE's proportional advantage; and central tendency methods remain competitive in Dataset 3. In addition to these quantitative metrics, visual inspection of forecasting graphs is used as complementary validation to assess pattern alignment, temporal consistency, and structural stability across Dataset 1 (Fig. 2), Dataset 2 (Fig. 3), and Dataset 3 (Fig. 4).

The forecasting graph for Dataset 1 (Fig. 2) shows that the predicted curve closely follows the cyclical pattern of the actual data, with aligned peaks and troughs and no noticeable phase shifts, indicating stable temporal continuity and supporting the metric-based results. In Dataset 2 (Fig. 3), the data exhibit higher variability, where the prediction line occasionally overshoots or undershoots local extremes but still follows the overall signal trajectory without structural divergence. For Dataset 3 (Fig. 4), the model successfully captures the gradual upward trend and subsequent decline, with only small, evenly distributed deviations between predicted and actual values. Overall, the graphical analysis confirms the quantitative evaluation, demonstrating that the selected imputation methods maintain structural alignment between predicted and actual patterns. To further assess robustness, the LSTM model is also compared with other recurrent architectures (RNN, GRU,

and BiLSTM) using the same optimal imputation strategy, with detailed results presented in Table VI.

From Table VI, for Dataset 1 with LOCF imputation, LSTM shows the most balanced performance, achieving the lowest MAPE and RMSE and the highest R^2 compared with GRU, Bi-LSTM, and standard RNN. GRU and Bi-LSTM produce results that are relatively similar but do not surpass LSTM. In contrast, RNNs perform significantly worse due to their limited ability to maintain temporal dependencies. In Dataset 2 with KNN imputation, LSTM again delivers the best overall results, showing lower prediction errors and higher explanatory power. At the same time, GRU and Bi-LSTM remain competitive but slightly less accurate. The gap between LSTM and RNN becomes more evident in this dataset, indicating that gated architectures handle volatile patterns more effectively. In Dataset 3 with Mean imputation, performance differences among GRU, LSTM, and Bi-LSTM become smaller. However, LSTM still maintains stable, competitive forecasting performance across varying data conditions.

The comparative evaluation indicates that LSTM remains the most consistent architecture across datasets and optimal imputation strategies. Although it is not always the absolute best in every case, it consistently ranks among the top-performing models. It demonstrates stable forecasting capability, confirming that the selected LSTM framework is numerically strong, visually coherent, and competitively reliable compared with alternative recurrent models. The findings also suggest that imputation effectiveness depends more on the underlying missingness mechanism and statistical structure of the dataset than on methodological complexity. For example, in Dataset 3 with MAR characteristics and stable variance, Mean imputation produced competitive results, indicating that simple statistical estimators can remain effective when distributional structure and temporal continuity are preserved. Overall, forecasting stability improves when imputation maintains statistical continuity and variance patterns. At the same time, methods that disrupt temporal adjacency or variance structure may amplify sequential prediction errors.

In the context of recent literature on environmental and climatological forecasting, many studies emphasize architectural innovation while giving limited attention to the interaction between missingness mechanisms and deep sequence models [48, 49]. However, emerging research in spatiotemporal modeling suggests that different missing-data mechanisms, such as MCAR, MAR, and MNAR, influence model generalization and stability. The present study contributes to this growing body of work by empirically demonstrating that identical neural architectures can produce substantially different forecasting outcomes depending on the imputation strategy applied. By explicitly linking missingness theory with recurrent learning dynamics, this study extends prior research that typically evaluates imputation and forecasting separately.

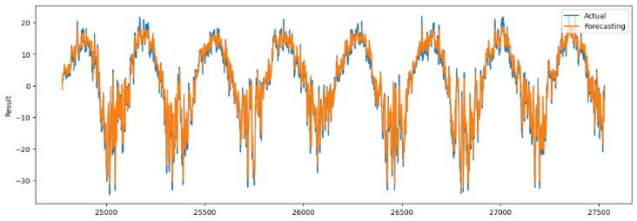

Fig. 2 Best result Graph Forecasting LSTM of dataset 1 (locf imputation)

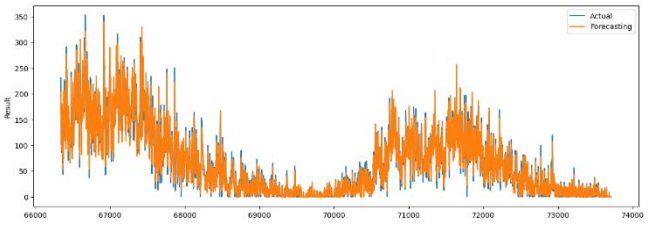

Fig. 3 Best result Graph Forecasting LSTM of dataset 2 (KNN Imputation)

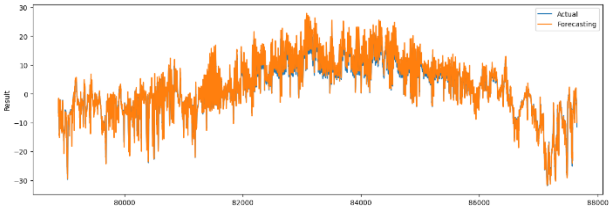

Fig. 4 Best result Graph Forecasting LSTM of dataset 3 (Mean Imputation)

TABLE VI
COMPARISON OF VARIOUS LOCF IMPUTATION RESULTS

Method	Dataset 1			Dataset 2			Dataset 3		
	MAPE	RMSE	R^2	MAPE	RMSE	R^2	MAPE	RMSE	R^2
RNN	2.07995	7.00709	1.55435	7.20033	67.68945	0.08539	2.85135	8.7261	0.14101
GRU	0.92128	3.63496	0.91079	6.50112	21.78677	0.88756	0.49631	1.36801	0.97889
LSTM	0.73959	3.25713	0.92757	3.12849	15.80088	0.94086	0.41560	1.60957	0.97077
Bi-LSTM	0.83314	3.37314	0.92318	5.28019	16.44733	0.93592	0.59817	1.57143	0.97214

Despite these contributions, several limitations should be acknowledged. The analysis is conducted within a predominantly univariate framework. It uses synthetically generated missingness scenarios, which may not fully capture the multivariate and spatiotemporal complexity of real environmental data. Additionally, although several recurrent architectures were compared, the study does not include newer attention-based or Transformer models, suggesting that future research should extend the analysis to multivariate datasets, incorporate advanced imputation approaches, and evaluate forecasting robustness under real-world missing-data conditions.

IV. CONCLUSION

This study shows that appropriate missing value imputation significantly improves LSTM forecasting performance compared to the Drop Missing Values approach, which can remove important temporal dependencies. The results indicate that the effectiveness of imputation methods depends on the underlying missingness mechanism of each dataset. Improvement forecasts are reflected in lower MAPE and RMSE values and higher R^2, indicating that reconstruction quality strongly influences model stability. However, the study is limited to a mainly univariate framework, synthetic missingness scenarios, and comparisons restricted to recurrent-based models. Future research should explore multivariate spatiotemporal datasets, real-world missing-data patterns, and advanced models, including attention-based, Transformer-based, and generative imputation approaches.

REFERENCES

[1] M. Patacca, M. Lindner, M. E. Lucas-Borja, T. Cordonnier, G. Fidej, B. Gardiner, Y. Hauf, G. Jasinevičius, S. Labonne, E. Linkevičius, M. Mahnken, S. Milanovic, G. Nabuurs, T. A. Nagel, L. Nikinmaa, M. Panyatov, R. Bercak, R. Seidl, M. Z. Ostrogović Sever, J. Socha, D. Thom, D. Vuletic, S. Zudin, and M. Schelhaas, "Significant increase in natural disturbance impacts on European forests since 1950," *Global Change Biology*, vol. 29, no. 5, pp. 1359–1376, Mar. 2023, doi: 10.1111/gcb.16531.

[2] U. A. Usmani, I. Abdul Aziz, J. Jaafar, and J. Watada, "Deep Learning for Anomaly Detection in Time-Series Data: An Analysis of Techniques, Review of Applications, and Guidelines for Future Research," *IEEE Access*, vol. 12, pp. 174564–174590, 2024, doi: 10.1109/ACCESS.2024.3495819.

[3] K. Zhang, Q. Yang, C. Li, X. Sun, and J. Chen, "Missing Data Recovery Methods on Multivariate Time Series in IoT: A Comprehensive Survey," *IEEE Commun. Surv. Tutorials*, vol. 28, pp. 149–180, 2026, doi: 10.1109/COMST.2025.3585962.

[4] Z. Li, J. F. L. Chavez, S. Pant, and D. Li, "Enhancing Ecological Forecasting with LSTM Models: The Impact of Partition-Based Data Shuffling on Predictive Accuracy," in *SoutheastCon 2025*, Concord, NC, USA: IEEE, Mar. 2025, pp. 314–319. doi: 10.1109/SoutheastCon56624.2025.10971684.

[5] D. Popovich, "How To Treat Missing Data In Survey Research," *Journal of Marketing Theory and Practice*, vol. 33, no. 1, pp. 43–59, Jan. 2025, doi: 10.1080/10696679.2024.2376052.

[6] A. D. Woods, D. Gerasimova, B. V. Dusen, J. Nissen, S. Bainter, A. Uzdavines, P. E. Davis-Kean, M. Halvorson, K. M. King, J. A. R. Logan, M. Xu, M. R. Vasilev, J. M. Clay, D. Moreau, K. Joyal-Desmarais, R. A. Cruz, D. M. Y. Brown, K. Schmidt, and M. M. Elsherif, "Best practices for addressing missing data through multiple imputation," *Infant and Child Development*, vol. 33, no. 1, pp. 1–37, Jan. 2024, doi: 10.1002/icd.2407.

[7] H. Wang, Z. Chen, Y. Shen, H. Zheng, D. Yang, D. Zhao, and B. Liang, "Robust Missing Value Imputation With Proximal Optimal Transport for Low-Quality IIoT Data," *IEEE Trans. Neural Netw. Learning Syst.*, vol. 37, no. 1, pp. 82–94, Jan. 2026, doi: 10.1109/TNNLS.2025.3601130.

[8] D. Mavridis, G. Salanti, T. A. Furukawa, A. Cipriani, A. Chaimani, and I. R. White, "Allowing for uncertainty due to missing and LOCF imputed outcomes in meta-analysis," *Statistics in Medicine*, vol. 38, no. 5, pp. 720–737, Feb. 2019, doi: 10.1002/sim.8009.

[9] L. Bourguignon, L. P. Lukas, J. D. Guest, F. H. Geisler, V. Noonan, A. Curt, S. C. Brüningk, and C. R. Jutzeler, "Studying missingness in spinal cord injury data: challenges and impact of data imputation," *BMC Med Res Methodol*, vol. 24, no. 1, p. 5, Jan. 2024, doi: 10.1186/s12874-023-02125-x.

[10] M. Rizki, A. Hermawan, and D. Avianto, "Optimization of Hyperparameter K in K-Nearest Neighbor Using Particle Swarm Optimization," *JUITA*, vol. 12, no. 1, p. 71, May 2024, doi: 10.30595/juita.v12i1.20688.

[11] A. Altamimi, A. A. Alarfaj, M. Umer, E. A. Alabdulqader, S. Alsubai, T. Kim, and I. Ashraf, "An automated approach to predict diabetic patients using KNN imputation and effective data mining techniques," *BMC Med Res Methodol*, vol. 24, no. 1, p. 221, Sep. 2024, doi: 10.1186/s12874-024-02324-0.

[12] K. P. Junaid, T. Kiran, M. Gupta, K. Kishore, and S. Siwatch, "How much missing data is too much to impute for longitudinal health indicators? A preliminary guideline for the choice of the extent of missing proportion to impute with multiple imputation by chained equations," *Popul Health Metrics*, vol. 23, no. 1, p. 2, Feb. 2025, doi: 10.1186/s12963-025-00364-2.

[13] Z. Jinbo, L. Yufu, and M. Haitao, "Handling missing data of using the XGBoost-based multiple imputation by chained equations regression method," *Front. Artif. Intell.*, vol. 8, p. 1553220, Apr. 2025, doi: 10.3389/frai.2025.1553220.

[14] M. Boukdire, Ç. A. İnan, G. Varra, R. Della Morte, and L. Cozzolino, "Interpolation and Machine Learning Methods for Sub-Hourly Missing Rainfall Data Imputation in a Data-Scarce Environment: One- and Two-Step Approaches," *Hydrology*, vol. 12, no. 11, p. 297, Nov. 2025, doi: 10.3390/hydrology12110297.

[15] Alfarelzi, M. A. Hakim, M. R. Ediananta, B. A. Pramudita, F. Corputty, and D. P. Setiawan, "Imputation of Missing Values for Generating Rainfall Data Using a Multi-Method Approach," in *2025 IEEE 16th Control and System Graduate Research Colloquium (ICSGRC)*, Shah Alam, Malaysia: IEEE, Aug. 2025, pp. 143–148. doi: 10.1109/ICSGRC65918.2025.11159759.

[16] W. Pangesti, N. Syukri, K. A. Notodiputro, Y. Angraini, and L. N. A. Mualifah, "Performance Evaluation of ARIMA and GRU Models for Forecasting Chili Price in East Jawa," *JUITA*, vol. 13, no. 2, pp. 209–218, Aug. 2025, doi: 10.30595/juita.v13i2.26445.

[17] N. Niako, J. D. Melgarejo, G. E. Maestre, and K. P. Vatcheva, "Effects of missing data imputation methods on univariate blood pressure time series data analysis and forecasting with ARIMA and LSTM," *BMC Med Res Methodol*, vol. 24, no. 1, p. 320, Dec. 2024, doi: 10.1186/s12874-024-02448-3.

[18] M. L. Hossain, S. M. N. Shams, and S. M. Ullah, "Time-series and deep learning approaches for renewable energy forecasting in Dhaka: a comparative study of ARIMA, SARIMA, and LSTM models," *Discov Sustain*, vol. 6, no. 1, p. 775, Aug. 2025, doi: 10.1007/s43621-025-01733-5.

[19] A. Rouabhi, "Century-Scale Climate Evolution in Semiarid High Plateaus, Algeria," *Journal of Applied Meteorology and Climatology*, vol. 65, no. 2, pp. 311–321, Feb. 2026, doi: 10.1175/JAMC-D-25-0181.1.

[20] M. A. De Sousa Guilhon Araujo, F. L. Cyrino Oliveira, and R. C. Souza, "Missing data imputation using climate data reanalysis in solar energy generation time series," *Energy Conversion and Management: X*, vol. 30, p. 101653, May 2026, doi: 10.1016/j.ecmx.2026.101653.

[21] C. Wongoutong, "Performance Comparison of Various Imputation Methods for Missing Data Mechanisms (MAR, MCAR, and MNAR) in a Nonstationary Time Series," *International Journal of Mathematics and Mathematical Sciences*, vol. 2025, no. 1, p. 3031708, Jan. 2025, doi: 10.1155/ijmm/3031708.

[22] F. T. Bahadur, S. R. Shah, and R. R. Nidamanuri, "Geospatial time series forecasting of air quality across Delhi using ARIMA and LSTM models," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 279, p. 106723, Feb. 2026, doi: 10.1016/j.jastp.2026.106723.

[23] W. Astuti and R. Govindaraju, "Effect of Data Gaps on Harmful Algal Bloom Prediction Models for Inland Lakes," *J. Hydrol. Eng.*, vol. 31, no. 1, p. 04025045, Feb. 2026, doi: 10.1061/JHYEFF.HEENG-6544.

[24] S. Khalili and H. Chulia, "Data Imputation in Large Datasets: A Comparative Study of PCA and Machine Learning Approaches," *Comput Econ*, Jan. 2026, doi: 10.1007/s10614-025-11201-x.

[25] H. Gong and K. W. Chan, "Sensitivity analysis for generalized estimating equation with non-ignorable missing data," *Scandinavian J Statistics*, p. sjos.70060, Feb. 2026, doi: 10.1111/sjos.70060.

[26] Y. Zhang and P. J. Thorburn, "A Dual-Head Attention Model for Time Series Data Imputation," *Computers and Electronics in Agriculture*, vol. 189, p. 106377, Oct. 2021, doi: 10.1016/j.compag.2021.106377.

[27] A. Palanivinayagam and R. Damaševičius, "Effective Handling of Missing Values in Datasets for Classification Using Machine Learning Methods," *Information*, vol. 14, no. 2, p. 92, Feb. 2023, doi: 10.3390/info14020092.

[28] A. Almeida, S. Brás, S. Sargento, and F. C. Pinto, "Focalize K-NN: an imputation algorithm for time series datasets," *Pattern Anal Applic*, vol. 27, no. 2, p. 39, Jun. 2024, doi: 10.1007/s10044-024-01262-3.

[29] M. Muhammad, T. Sutikno, and I. Riadi, "A Comparative Study of K-Means and KNN Imputation for Handling Missing Data in Scholarship Applicant Datasets," *JUITA*, vol. 13, no. 3, pp. 245–254, Nov. 2025, doi: 10.30595/juita.v13i3.26502.

[30] T. Mahmud, T. Akter, S. Anwar, M. T. Aziz, M. S. Hossain, and K. Andersson, "Predictive Modeling in Forex Trading: A Time Series Analysis Approach," in *2024 Second International Conference on Inventive Computing and Informatics (ICICI)*, Bangalore, India: IEEE, Jun. 2024, pp. 390–397. doi: 10.1109/ICICI62254.2024.00070.

[31] K. Mahmud Sujon, R. Binti Hassan, Z. Tusnia Towshi, M. A. Othman, M. Abdus Samad, and K. Choi, "When to Use Standardization and Normalization: Empirical Evidence From Machine Learning Models and XAI," *IEEE Access*, vol. 12, pp. 135300–135314, 2024, doi: 10.1109/ACCESS.2024.3462434.

[32] M. Waqas and U. W. Humphries, "A critical review of RNN and LSTM variants in hydrological time series predictions," *MethodsX*, vol. 13, p. 102946, Dec. 2024, doi: 10.1016/j.mex.2024.102946.

[33] P. Nguyen-Duc, H. D. Nguyen, Q.-H. Nguyen, T. Phan-Van, and H. Pham-Thanh, "Application of Long Short-Term Memory (LSTM) Network for seasonal prediction of monthly rainfall across Vietnam," *Earth Sci Inform*, vol. 17, no. 5, pp. 3925–3944, Oct. 2024, doi: 10.1007/s12145-024-01414-3.

[34] X. Wen and W. Li, "Time Series Prediction Based on LSTM-Attention-LSTM Model," *IEEE Access*, vol. 11,

pp. 48322–48331, 2023, doi: 10.1109/ACCESS.2023.3276628.

[35] R. Zhu, G.-Y. Zhong, and J.-C. Li, "Forecasting Price in A New Hybrid Neural Network Model with Machine Learning," *Expert Systems with Applications*, vol. 249, p. 123697, Sep. 2024, doi: 10.1016/j.eswa.2024.123697.

[36] E. Seidi, F. Kaviari, and S. F. Miller, "Hyperparameter Tuning of Artificial Neural Network-Based Machine Learning to Optimize Number of Hidden Layers and Neurons in Metal Forming," *JMMP*, vol. 9, no. 8, p. 260, Aug. 2025, doi: 10.3390/jmmp9080260.

[37] M. Jin and H. Emami-Meybodi, "Prediction of Flowing Bottomhole Pressure in Gas-Lifted Wells Using LSTM-ANN Models," *Energy Fuels*, vol. 39, no. 31, pp. 14992–15002, Aug. 2025, doi: 10.1021/acs.energyfuels.5c02240.

[38] F. H. Sezgin, Ö. Algorabi, G. Sart, and M. Güler, "Hyperparameter-Optimized RNN, LSTM, and GRU Models for Airline Stock Price Prediction: A Comparative Study on THYAO and PGSUS," *Symmetry*, vol. 17, no. 11, p. 1905, Nov. 2025, doi: 10.3390/sym17111905.

[39] I. Darmawan, D. P. Kemuning, A. Rahmatulloh, R. Rizal, E. Haerani, and L. Sutriani, "Optimizer Algorithms for Air Temperature Prediction using LSTM," in *2025 9th International Conference on Recent Advances and Innovations in Engineering (ICRAIE)*, Kaula Lumpur, Malaysia: IEEE, Aug. 2025, pp. 1–6. doi: 10.1109/ICRAIE65839.2025.11239081.

[40] N. N. Karima, S. Ahmad, A. Islam, A. S. N. Huda, L. J. Awalin, S. A. Halim, H. Mokhlis, M. S. Ali, and S. Mubarok, "Enhancing Short-Term Load Forecasting Using Hyperparameter-Optimized Deep Learning Approaches," *Energies*, vol. 19, no. 3, p. 705, Jan. 2026, doi: 10.3390/en19030705.

[41] H. Yüce and H. Arabacı, "Effect of Different Normalization Methods on Battery SoC Estimation Performance at the Optimal LSTM Batch Size," in *2025 16th International Conference on Electrical and Electronics Engineering (ELECO)*, Istanbul, Turkiye: IEEE, Nov. 2025, pp. 1–5. doi: 10.1109/ELECO69582.2025.11329352.

[42] B. Liang, J. Deng, F. Yuan, J. Gao, J. Yang, and P. Zhao, "A bio-inspired IVY metaheuristic for optimizing LSTM models in urban traffic flow prediction," *Computing*, vol. 108, no. 2, p. 29, Feb. 2026, doi: 10.1007/s00607-026-01623-2.

[43] P. Aduama, A. S. Al-Sumaiti, V. K. Saini, and R. Kong, "Explainable optimized voting ensemble for photovoltaic power forecasting," *Energy and AI*, vol. 24, p. 100690, May 2026, doi: 10.1016/j.egyai.2026.100690.

[44] B. Wang, Z. Yang, M. Ji, J. Shan, M. Ni, Z. Hou, J. Cai, X. Gu, X. Yuan, Z. Gong, Q. Du, Y. Yin, and K. Jiao, "Long short-term memory deep learning model for predicting the dynamic performance of automotive PEMFC system," *Energy and AI*, vol. 14, p. 100278, Oct. 2023, doi: 10.1016/j.egyai.2023.100278.

[45] A. Flores, H. Tito-Chura, J. Guzman-Valdivia, R. Morales-Gonzales, C. Rosado-Chavez, and C. Silva-Delgado, "Multi-Step Forecasting of Weekly Dengue Cases Using Data Augmentation and Convolutional Recurrent Neural Networks," *Computers*, vol. 15, no. 2, p. 114, Feb. 2026, doi: 10.3390/computers15020114.

[46] D. Ramadhani, A. M. Soleh, and E. Erfiani, "Characteristics of Machine Learning-based Univariate Time Series Imputation Method," *JUITA*, vol. 12, no. 2, p. 279–288, Nov. 2024, doi: 10.30595/juita.v12i2.23453.

[47] R. A. Putri, S. Surono, and A. Thobirin, "Evaluating LSTM Performance on Multivariate Time Series with One-Class SVM Outlier Detection," *JUITA*, vol. 13, no. 2, pp. 127–134, Aug. 2025, doi: 10.30595/juita.v13i2.26160.

[48] R. R. Salim and D. Adytia, "Deep Learning-Based Sea Level Forecasting Using Informer in Cilacap, Indonesia," *JUITA*, vol. 13, no. 3, pp. 317–327, Nov. 2025, doi: 10.30595/juita.v13i3.27176.

[49] S. Lang, M. Alexe, M. C. A. Clare, C. Roberts, R. Adewoyin, Z. Ben Bouallègue, M. Chantry, J. Dramsch, P. D. Dueben, S. Hahner, P. Maciel, A. Prieto-Nemesio, C. O'Brien, F. Pinault, J. Polster, B. Raoult, S. Tietsche, and M. Leutbecher, "AIFS-CRPS: ensemble forecasting using a model trained with a loss function based on the continuous ranked probability score," *npj Artif. Intell.*, vol. 2, no. 1, p. 18, Feb. 2026, doi: 10.1038/s44387-026-00073-7.