

Hybrid CNN-BiLSTM-Attention Modeling for Air Temperature Prediction Using LOF-iForest-RFE Multi-Stage Preprocessing on Soekarno-Hatta Meteorological Station AWS Data (2021–2025)

Muhamad Nur Fuat ^{1*}, Sajarwo Anggai ^{2*}, Murni Handayani ^{3*}

* Department of Computer Engineering, Pamulang University, South Tangerang, Indonesia
alkanif0208@gmail.com ¹, sajarwo@gmail.com ², dosen02710@unpam.ac.id ³

Article Info

Article history:

Received 2026-08-01

Revised 2026-09-11

Accepted 2026-10-02

Keyword:

*Air Temperature Prediction,
Anomaly Detection,
Attention Mechanism,
Deep Learning,
Flight Safety.*

ABSTRACT

Air temperature is a critical meteorological parameter that dynamically affects density altitude and Maximum Take-Off Weight (MTOW) calculations in aviation operations. Although the 10-minute resolution Automatic Weather Station (AWS) data at Soekarno-Hatta International Airport effectively captures micrometeorological fluctuations, it is chaotic and prone to anomalies, often causing prediction failures in conventional models. This study aims to propose an artificial intelligence-based short-term framework to predict air temperature for the next 10-minute horizon by integrating multi-level preprocessing with a hybrid CNN-BiLSTM-AM (Convolutional Neural Network, Bidirectional Long Short-Term Memory, Attention Mechanism) architecture. Utilizing a Data-Centric AI approach on observational data from 2021–2025, the methodology begins with local and global anomaly purification via Local Outlier Factor (LOF) and Isolation Forest (iForest) algorithms. Multicollinearity is then mitigated through hybrid feature selection using Pearson Correlation and Recursive Feature Elimination (RFE). In the modeling phase, the 1D-CNN layer extracts spatial correlations among variables, the BiLSTM evaluates bidirectional temporal dependencies, and the attention mechanism adaptively distributes probability weights to prevent lag response during extreme weather oscillations. The findings demonstrate that data purification is strictly necessary to avoid model generalization failures, and RFE successfully isolated the four most optimal physical predictors: maximum temperature, relative humidity, air pressure, and solar radiation. Comparatively, the proposed CNN-BiLSTM-AM model proved to be the most superior, achieving the highest accuracy with an RMSE of 0.8151 °C, MAE of 0.6329 °C, and R² of 0.9106. The addition of attention and convolution mechanisms was shown to reduce the root mean square error (RMSE) by 16.44% and the mean absolute error (MAE) by 13.64% compared to the pure BiLSTM baseline model. In conclusion, achieving an extreme error deviation (RMSE) consistently below the 1.0 °C operational tolerance confirms that this architecture is highly feasible, robust, and has strong potential as a highly reliable meteorological reference to support Density Altitude and MTOW estimations in equatorial climates.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. INTRODUCTION

In the era of modern aviation, which constantly demands high levels of operational efficiency and absolute safety, meteorological parameters play a fundamental and

indispensable role in every phase of flight operations. From flight planning through to landing, weather conditions are the primary determining factor. Among the various weather parameters, air temperature is often viewed merely as a static

variable in the context of daily public weather forecasts. However, in aviation operations, air temperature is a highly critical dynamic variable because it directly dictates an aircraft's aerodynamic performance [1]. Soekarno-Hatta International Airport, located in Cengkareng and serving as Indonesia's primary air gateway with the busiest traffic volume in the Southern Hemisphere, faces complex geographical and meteorological challenges. Located in an equatorial tropical region with intense solar radiation exposure throughout the year, the airport experiences extreme daily air temperature fluctuations and long-term warming trends that significantly affect density altitude values [2]. Rising air temperatures directly impact aircraft takeoff performance by altering the air density. Quantitatively, every 1.0 °C increase in surface temperature can raise the density altitude by approximately 120 feet, which significantly reduces air density and aircraft lift capability. Consequently, inaccuracies in temperature predictions can lead to critical errors in calculating the maximum take-off weight (MTOW), potentially endangering lives or forcing airlines into economic losses due to unplanned payload or fuel reductions. To mitigate these risks, the World Meteorological Organization (WMO) and the International Civil Aviation Organization (ICAO) have established strict operational tolerance standards, requiring aerodrome temperature forecast accuracy to be within a margin of ± 1.0 °C [3], [4]. Therefore, a crucial challenge in mitigating these aviation risks is the provision of highly accurate air temperature forecast data with high temporal resolution. Efforts to predict atmospheric dynamics have traditionally faced significant challenges. Conventional weather forecasting models, including numerical approaches, have often proven inefficient at detecting sudden temperature fluctuations and lack sufficient sensitivity to capture specific micro-meteorological phenomena in airport areas [5]. Significant computational limitations make it difficult for conventional methods to accurately represent complex nonlinear atmospheric dynamics [6], [7]. As a form of advanced monitoring, Automatic Weather Stations (AWS) installed at meteorological stations generate surface observation data with very high temporal resolution—every ten minutes [8]. This high-resolution time-series data actually contains information on highly nonlinear and chaotic short-term temperature fluctuation patterns, which are regulated by complex interactions between solar radiation, cloud cover, humidity, and local wind direction. However, this massive volume of data introduces new problems in the form of anomalies and noise. Meteorological sensors are highly susceptible to external disturbances and internal errors, producing extreme values that can disrupt the learning process of prediction algorithms [8], [9]. Simple statistical approaches often fail to distinguish anomalies caused by sensor malfunctions from valid spikes in extreme weather. Furthermore, meteorological data consists of many closely correlated parameters, so using all features without strict

selection will lead to overfitting and computational inefficiency in the model.

To address the complexity of modeling meteorological time series data, artificial intelligence paradigms—particularly deep learning—have been empirically proven to outperform classical statistical models, which tend to be rigid and rely on the assumption of linear stationarity [9]. Current sequential data processing is generally standardized using the Long Short-Term Memory (LSTM) algorithm due to its reliability in addressing the vanishing gradient problem and its capability to store long-term memory [10]. However, recent literature has identified a fundamental weakness in this unidirectional architecture: the model is only capable of absorbing information from the past toward the future, resulting in a loss of comprehensive understanding of the data transition context [11]. To address this limitation, architectural transformations have focused on the use of Bidirectional LSTM (BiLSTM), which processes information sequences in parallel from both forward and backward directions [12], [13]. Previous research has demonstrated that this bidirectional architecture consistently reduces the absolute error margin compared to standard unidirectional models when processing atmospheric parameters [18]. Although superior in the temporal domain, this bidirectional memory model remains less sensitive in independently extracting spatial feature representations across variables, even though weather data inherently possesses multivariate spatiotemporal characteristics [14], [15].

To address this limitation, the integration of a Convolutional Neural Network (CNN) architecture into the early layers of the model has been widely adopted as a feature extractor to map deep connections among multivariate weather elements before the data is processed sequentially [5], [6], [15]. Although the fusion of convolutional and bidirectional memory architectures has demonstrated improved performance, these standard hybrid models tend to distribute attention weights evenly across all time steps. Consequently, in very long high-resolution input sequences, crucial information is prone to being overshadowed by redundant data. Furthermore, the 10-minute micro-temporal resolution generates prolonged observation sequences, causing standard recurrent memory to fade over time. Therefore, incorporating an Attention Mechanism becomes imperative to dynamically prioritize weights at the most deterministic weather transition points. Various contemporary studies in the fields of electrical and hydrological forecasting show that the incorporation of an Attention Mechanism can mimic the selective human cognitive system by dynamically prioritizing weights at the most relevant time steps [12], [14], [19]. Alongside such predictive modeling, the computational integrity of advanced models heavily depends on the quality of the input data. Recent preprocessing studies recommend a multi-level anomaly detection approach that combines local density detection—such as the Local Outlier Factor (LOF)—with global anomaly isolation using Isolation Forest (iForest) to

efficiently address noise from meteorological instruments [8], [20]. In addition to removing duplicate anomalies, the feature selection process—which employs a linear correlation approach combined with Recursive Feature Elimination (RFE)—has proven crucial in filtering out sets of mutually correlated weather variables (multicollinearity) to identify the most informative subset of features [16], [17].

Based on a comprehensive literature review, it is clear that the development of layered hybrid models can provide superior accuracy in predicting environmental parameters. While previous studies have applied standalone models like CNN, LSTM, or standard CNN-LSTM hybrids for meteorological temperature prediction [5], [6], they often struggle with extreme micro-temporal fluctuations and high error margins due to the lack of dynamic attention weighting. However, a fundamental research gap has been identified; the majority of previous studies tend to focus exclusively on the engineering of prediction algorithm architectures and often overlook the crucial role of automated input data integrity cleaning. Very little literature has been found that comprehensively integrates multi-level unsupervised outlier detection methods and algorithm-based feature selection techniques into a single, cohesive workflow alongside deep learning modeling. Furthermore, the specific implementation of such predictive frameworks on air temperature data with micro-temporal resolution (ten minutes) in the tropical equatorial zone which is specifically intended to support the safety of density altitude calculations for commercial flight takeoffs remains largely unexplored [1], [20]. Previous models have not provided a preprocessing framework robust enough to tolerate extreme noise levels at high-frequency airport observation stations.

In response to this academic gap and practical urgency, this study proposes and develops a comprehensive end-to-end framework that models and compares short-term air temperature forecasts at the Soekarno-Hatta Meteorological Station. This study aims to implement a state-of-the-art deep learning architecture in the form of a hybrid CNN-BiLSTM with the addition of an Attention Mechanism (AM). As the foundation of the system's innovation, this model is preceded by an integrated noise-cleaning stage using LOF and iForest, as well as rigorous feature selection using Pearson Correlation and RFE to address multicollinearity issues. Specifically, the expected contribution of this research is the development of a hybrid predictive model that is significantly more robust, free from anomalies, and highly accurate in detecting tropical weather fluctuations. The results obtained are expected to serve as a valid tool for refining Maximum Takeoff Weight (MTOW) calculations, thereby potentially supporting the efficiency and safety parameters in aviation operations in Indonesia.

II. METHOD

This study was designed with a strict sequential workflow to address the challenge of high-resolution air temperature prediction, which is crucial for aviation operations. The

system workflow proceeds through key stages, including data preprocessing, feature engineering, and modeling using a hybrid CNN-BiLSTM-AM architecture. The flowchart of the overall research stages is shown in Figure 1.

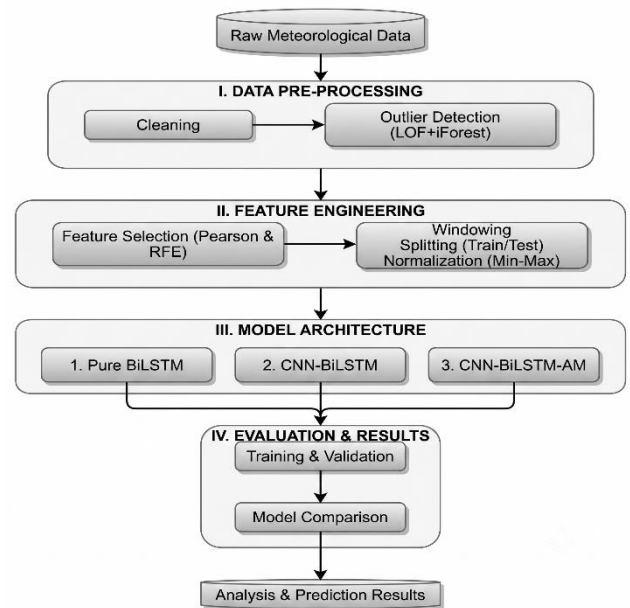


Figure 1. Flowchart of the Research Stages

A. Data Collection

The secondary data used in this study were sourced from the Automatic Weather Station (AWS) at the Soekarno-Hatta Meteorological Station over a five-year period (January 1, 2021, to December 31, 2025). These observational data have a high temporal resolution of every 10 minutes, resulting in a raw dataset consisting of 254,306 observation rows and 11 variables covering hydrological, thermodynamic, and atmospheric kinematic parameters, as summarized in (Table 1). The primary target variable is mean air temperature (°C), supported by 10 meteorological parameters as predictor variables.

TABLE 1
LIST OF INITIAL WEATHER PARAMETERS

Meteorological Variables	Description / Parameter Unit
rr	Daily Rainfall (mm)
ws_avg	Average Wind Speed (m/s)
ws_max	Maximum Wind Speed (m/s)
wd_avg	Average Wind Direction (Degrees)
tt_air_max	Maximum Air Temperature (°C)
tt_air_min	Minimum Air Temperature (°C)
rh_avg	Average Humidity (%)
pp_air	Air Pressure (hPa)
sr_avg	Average Solar Radiation (W/m ²)
ev_pan	Open Pan Evaporation (mm)
tt_air_avg	Average Air Temperature (°C) – Target

B. Data Preprocessing

Raw data from AWS often contains missing values and instrument noise, so preprocessing is carried out in stages. Short-duration data gaps (less than 1 hour) are handled using linear interpolation to maintain the continuity of the time-series data trend. Conversely, data gaps with long durations (more than 1 hour) are dropped to avoid information bias. The next step is a multi-level outlier detection strategy applied sequentially to address two distinct types of meteorological noise. First, the Local Outlier Factor (LOF) algorithm is utilized specifically to detect local or contextual anomalies, such as sudden and unnatural micro-meteorological spikes compared to their immediate neighbors. Subsequently, the data is passed to the Isolation Forest (iForest) algorithm, which is designed to isolate systemic global anomalies that deviate significantly from the overall historical data distribution. The workflow diagram for this hybrid outlier detection process is shown in Figure 2.

As shown in Figure 2, the first filtering layer uses LOF to evaluate the local reachability density of a data point relative to its 20 nearest neighbors ($n_neighbors = 20$), thereby identifying contextual anomalies. The remaining data is then passed to the iForest algorithm to detect global anomalies through an isolation forest structure ($n_estimators = 100$, $random_state = 42$), with the anomaly tolerance threshold set at 5%. Given that these high-resolution meteorological data observations are unsupervised meaning they lack inherent anomaly classification labels determining the contamination threshold parameter in the Isolation Forest algorithm is crucial.

To ensure objectivity, the tolerance threshold is set based on the Elbow Method applied to the distribution of anomaly scores (decision function) to identify the most effective parameter. Through a comparison of three contamination threshold scenarios (1%, 5%, and 10%) as visualized in Figure 3, the identification of the elbow point indicates that the most significant decrease in anomaly scores stops precisely around the 5th percentile. "The selection of 5% is mathematically the most optimal because the 1% scenario (underfitting) falls within a steep curve region that could potentially allow extreme instrument noise to pass through. Conversely, the 10% scenario (overfitting) results in a curve that is too flat, posing the risk of excessively discarding valid extreme weather events and natural data variance, such as legitimate daily maximum temperature peaks. By strictly identifying the elbow point at the 5th percentile, the algorithm ensures that only the steepest deviations which structurally represent true instrument noise or sensor errors rather than valid meteorological phenomena are isolated and removed.

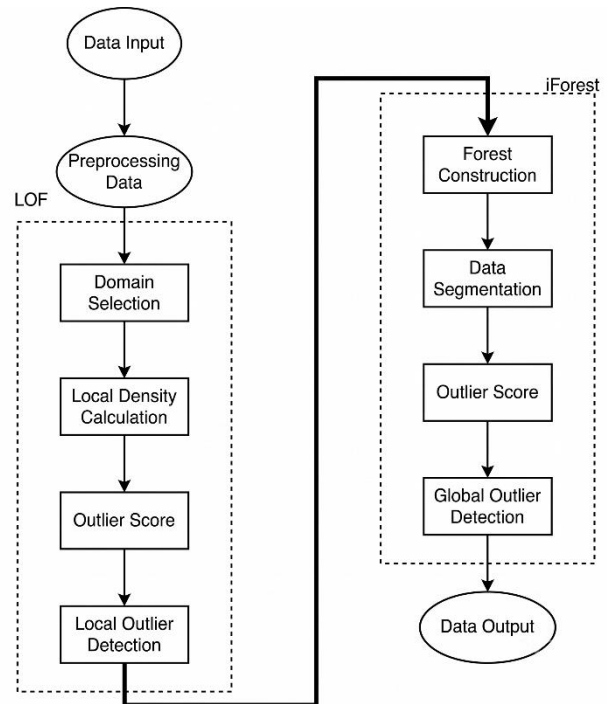


Figure 2. Hybrid Outlier Detection Flowchart

To preserve the integrity and sequential nature of the time-series data, detected anomalous observations were converted to NaN values and reconstructed using two-way linear interpolation (limit direction = both). This ensures a noise-free dataset without distorting the temporal dimension.

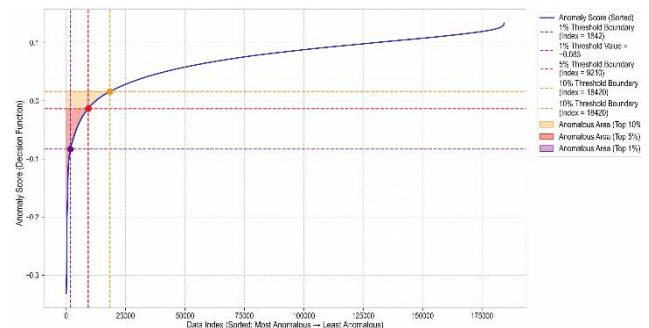


Figure 3. Comparison of Contamination Parameters

C. Feature Selection

The feature selection stage aims to reduce the dimensionality of the dataset and eliminate redundancy using a hybrid Filter-Wrapper approach. Starting from the initial set of 10 meteorological predictor variables, the process begins with a Pearson correlation analysis. Predictor features with very low absolute linear correlation to the target ($|r| < 0.1$) are excluded, and feature pairs showing extreme multicollinearity ($|r| > 0.95$) have one of them eliminated. Following this linear filtering, the remaining predictors are further evaluated using the Recursive Feature Elimination (RFE) algorithm based on a Random Forest Regressor (100

decision trees, random_state = 42). The RFE algorithm iteratively calculates feature importance scores and eliminates the least significant variables step-by-step. The elimination process is strictly configured to stop only when exactly the four most optimal predictor features remain from the original set of predictors. To prepare the filtered data for modeling, the dataset was first split into 80% training data and 20% testing data. It is imperative to note that this data splitting was performed strictly in chronological (temporal) order. Random splitting was completely avoided to preserve the sequential integrity of the weather time-series and prevent temporal data leakage. Subsequently, the data was normalized to the [0, 1] range using the Min-Max Scaling method. To rigorously prevent any data leakage that could bias the model's accuracy, the scaling parameters were fitted exclusively on the training dataset. Finally, the normalized data was restructured using the sliding window technique with a look-back parameter of 24, conditioning the model to extract historical patterns from the last 240 minutes of observations as predictors.

D. CNN-BiLSTM-AM Model Architecture

The model architecture is designed to be modular and accepts data representations in the form of 3D Tensors of size (Batch_Size, 24, 4). A diagram of the proposed architecture is shown in Figure 4.

Preliminary processing is performed by a 1D Convolution block configured with 64 filters, a kernel size of 3, and a padding of 1, acting as a local spatial feature extractor across the multivariate inputs. To stabilize the gradient propagation and prevent internal covariate shifts, this layer is equipped with Batch Normalization and a non-linear ReLU activation function, followed by a Dropout layer (ratio 0.3) to mitigate overfitting, and a 1D Max Pooling layer (kernel size 2) to reduce the temporal dimensions of the dominant signal.

The extracted spatial feature map is then passed to a BiLSTM recurrent block consisting of 1 hidden layer with 64 hidden units per direction (forward and backward). To address concerns regarding potential data leakage in real-time operational scenarios, it is critical to clarify the BiLSTM's bidirectional mechanism in this architecture. The 'future' context processed by the backward LSTM layer is strictly confined within the boundaries of the 24-step (240-minute) historical look-back window. The network never accesses the actual future target ($t + 1$) during the sequential processing, ensuring absolute validity and zero data leakage for real-time inference. This bidirectional processing produces a total of 128 feature representations.

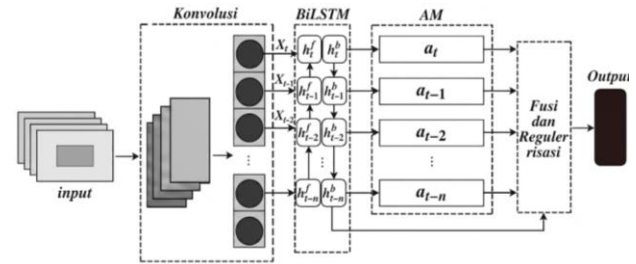


Figure 4. Structural Diagram of the CNN-BiLSTM-AM Model

The next stage applies a Transformer-style Scaled Dot-Product Attention Mechanism, where the output of the last time step from the BiLSTM is projected as the Query (Q) vector, while all historical time steps form the Key (K) and Value (V) vectors (all projected to a dimension of 128). Attention scores are calculated via the dot product between the scaled Q and K vectors and passed through a Softmax function, resulting in a final context vector from the weighted sum of the V values. To prevent overfitting and simplify the combination of information, this context vector is directly added to the output of the last time step via a Residual Connection technique. This ensures the model retains the baseline global temperature trend while applying dynamic attention weights to the anomalies. The result is then normalized using Layer Normalization to ensure the combined feature representation is highly stable. In the final stage, the tensor is passed through Dropout and mapped by a linear (Dense) Output Layer to predict a single continuous air temperature value. A summary of the Hybrid CNN-BiLSTM-AM model architecture configuration in this study can be seen in Table 2.

TABLE 2
SUMMARY OF THE CNN-BiLSTM-AM MODEL ARCHITECTURE

Type Layer	Parameter Konfigurasi
Input Layer	Shape: (Batch_Size, Look_Back, N_Features) Shape: (Batch_Size, 24, 4)
Conv1D	Filters: 64, Kernel Size: 3, Padding : 1
Max Pooling 1D	Kernel 2
BatchNorm1d & ReLU	Features: 64
Dropout (CNN)	Rate: 0.3
BiLSTM	Hidden Dim: 64, Layers: 1
Scaled Dot-Product Attention	Proyeksi Linier: Q, K, V (Dim = 128), Scale = $\sqrt{128}$
Residual Connection & LayerNorm	Operasi: Penjumlahan, Features: 128
Dropout (FC)	Rate: 0.3
Dense (Output Layer)	Units: 1, Activation: Linear

The training scenario was configured using the Adam optimization algorithm with an initial learning rate of 0.001, weight decay of 1×10^{-4} , and the Mean Squared Error (MSE) loss function. The computational scenario runs for a maximum of 150 epochs (batch size 64), combined with the ReduceLROnPlateau learning rate scheduler (50% reduction every 10 epochs if validation loss plateaus) and an early stopping mechanism to lock in the parameter configuration at the optimal generalization point. The implementation was fully executed on NVIDIA GPU infrastructure (CUDA-enabled) using the PyTorch library.

III. RESULTS AND DISCUSSION

A. Preprocessing and Feature Selection

The initial data used consisted of observation records from automatic weather stations, comprising 254,306 observation rows and 11 meteorological variables. An initial profile of the raw dataset revealed an extremely high rate of data gaps in the open-pan evaporation (ev_pan) instrument, reaching 59.38%.

Based on an empirical comparison, retaining the ev_pan variable would result in the loss of 58.52% of the total observations due to the linear interpolation cutoff rule for long gaps. To provide a more comprehensive picture, the comparison of the impact of data loss due to the presence of the ev_pan column, as visualized in Figure 5, proved effective in preserving the time-series representation. This decision successfully preserved 184,207 rows (72.44%) of net observations a significantly higher figure than if the `ev\_pan` column had been retained, which would have left only 105,485 rows (41.48%) of data.

High-resolution meteorological data are highly susceptible to different types of noise, necessitating a sequential multi-level filtering approach. In the first stage, the Local Outlier Factor (LOF) algorithm was applied to detect contextual local anomalies, which successfully identified 3,843 anomalous observations (approximately 2.08% of the data). In the second stage, the remaining data was processed using the Isolation Forest (iForest) algorithm to isolate systemic global anomalies. Using an optimal contamination threshold of 5%, iForest identified an additional 9,211 anomalies (~5.00%). Details of the comparison of anomaly filtering results from this hybrid approach are presented in Table 3. By employing these two methods sequentially, this hierarchical approach addresses different anomaly characteristics and successfully identified a total of 13,054 anomalous observations (approximately 7.08% of the entire dataset). To preserve the sequential structure of the 10-minute time intervals, the rows containing these outliers were not removed row-wise but were instead set to NaN and reconstructed using two-way linear interpolation.

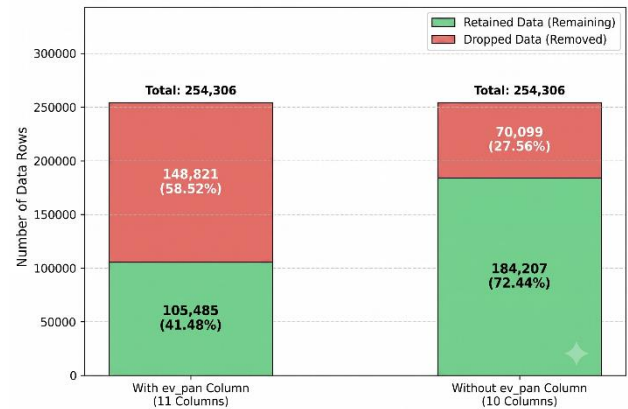


Figure 5. Comparison of Data Volumes After the Cleaning Process

TABLE 3
COMPARISON OF HYBRID OUTLIER DETECTION RESULTS

Process Description	Total Number of Rows	Number of Detected Outliers	Anomaly Presentation
Initial Data (Pre-Outlier)	184.207	-	-
Local Detection (LOF)	184.207	3.843	~2,08%
Global Detection (iForest)	184.207	9.211	~5,00%
Total Cleaned Data	184.207	13.054	~7,08%

To prevent the curse of dimensionality, Pearson's correlation analysis was applied as an initial step in feature selection. The goal was to evaluate the linear relationships between the set of predictor variables and the target variable (tt_air_avg), as well as to detect signs of multicollinearity among the independent variables. Feature selection at this stage is based on two criteria: eliminating features with an absolute correlation value relative to the target below 0.1, and excluding redundant features if the correlation between predictors exceeds 0.95.

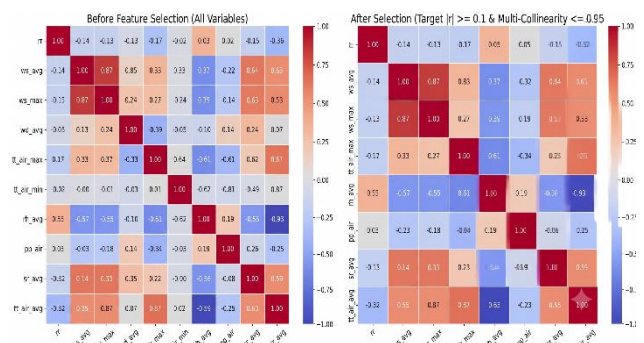


Figure 6. Visualization of the Pearson Correlation Matrix

Based on the correlation matrix in Figure 6, the variables for average wind direction (wd_avg, $r = 0.07$) and minimum

air temperature (tt_air_min, $r = 0.02$) were excluded because their correlation levels were not significant (close to random). Meanwhile, in terms of multicollinearity evaluation, the highest correlation among independent variables was only 0.87 (between ws_avg and ws_max). Since this value is still below the 0.95 threshold, no features were eliminated due to redundancy. The remaining predictor variables, which were free from this high multicollinearity issue, were then further evaluated using the Recursive Feature Elimination (RFE) algorithm.

TABLE 4.
RESULTS OF FEATURE SELECTION USING THE RFE ALGORITHM

Feature Status	List of Weather Variables	Description
Target Features	tt_air_avg (Average Air Temperature)	Dependent variables
Eliminated Features	rr, ws_avg, ws_max	Rainfall and wind metrics were excluded due to their low predictive power and multicollinearity.
Selected Features (Top-Ranked)	tt_air_max, rh_avg, pp_air, sr_avg	Maximum temperature, relative humidity, air pressure, and solar radiation were found to contribute the highest variance to the target.

Based on the evaluation using the RFE algorithm shown in Table 4, the four features retained are maximum temperature (tt_air_max), relative humidity (rh_avg), air pressure (pp_air), and solar radiation (sr_avg). These four features were maintained because they empirically contributed the highest variance to the target variable, average air temperature (tt_air_avg). Furthermore, this selection is theoretically validated by atmospheric thermodynamics: solar radiation acts as the primary energy source, while air pressure has a direct physical correlation with temperature (Gay-Lussac's Law) [1]. Conversely, other meteorological variables such as precipitation (rr) and wind speed metrics (ws_avg and ws_max) were explicitly eliminated from the modeling process. Although these variables may have complex temporal relationships with temperature, the RFE algorithm identified that their direct predictive contribution was significantly low and they exhibited indications of multicollinearity, which would otherwise introduce noise and hinder the model's learning convergence.

B. Training and Convergence Results

The modeling was performed using the PyTorch framework on Graphics Processing Unit (GPU)-based hardware via the CUDA interface. The network accepted input with a look-back window of 24 time steps, representing

the historical data from the past 4 hours, with 5 features (4 selected predictors and 1 historical target). Experimental reproducibility is ensured by setting the global random seed to 42, using the Mean Squared Error (MSE) loss function, the Adam optimizer, and a maximum of 150 epochs.

Observation of the loss curve (MSE) reveals contrasting convergence patterns between the use of raw data and preprocessed data. On the raw dataset, the network fails to generalize and experiences vanishing gradients from the early stages of iteration. The dominance of extreme anomalies forces the model's weight search space to get trapped in local minima, causing the Train Loss and Val Loss metrics to stagnate, which ultimately triggers the Early Stopping mechanism prematurely.

In contrast, as illustrated in Figure 7, the convergence profile on the cleaned dataset demonstrates a stable, efficient, and representative learning process. The slope transitions and gradient changes recorded in the curve reflect the network's direct response to adjustments in the dynamic learning rate parameter. The BiLSTM baseline architecture maintained its rate of decline until triggering Early Stopping at epoch 49 (Train Loss 0.001205, Val Loss 0.006854). A performance improvement is evident in the CNN-BiLSTM, which smooths out the curve and stops at epoch 50 (Train Loss 0.001339, Val Loss 0.004315), proving the effectiveness of the initial convolution operations in eliminating spatial redundancy.

The proposed CNN-BiLSTM-AM architecture demonstrated the most optimal convergence performance. The integration of the attention mechanism prevents the diffusion of historical information and facilitates direct gradient propagation, allowing training to be extended to the 82nd epoch. This model successfully reduced the training loss to 0.001021 and the validation loss to 0.004232, with a learning rate reaching a microscopic value of 0.000063. The tight asymptotic convergence between the training and validation losses at this termination point serves as an empirical indicator of the absence of high-variance error, ensuring that the model has fully extracted the intrinsic patterns of weather physics.

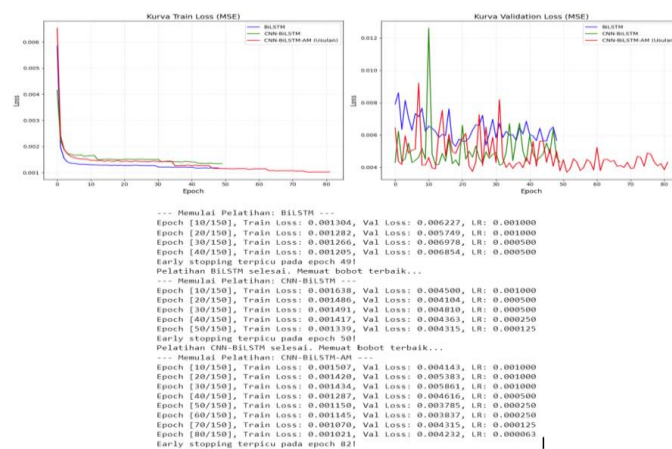


Figure 7. Model Convergence Curve After Data Preprocessing

C. Model Performance Evaluation

Inference testing was conducted using an independent test dataset (20% of the dataset). Performance was evaluated using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the Coefficient of Determination (R^2). The results of the comparison of inference performance between models trained using raw data (before preprocessing) and cleaned data (after preprocessing) are fully illustrated in Figure 8.

Based on the results of an ablation study on the dataset before preprocessing (blue panel in Figure 8), a drastic degradation in predictive capacity was observed across all model variants. The presence of anomalies and extreme noise distorts the spatial and temporal feature mapping, thereby crippling the model's inference. Despite this decline, the CNN-BiLSTM-AM architecture demonstrated the highest level of resilience. This variant was the only model capable of maintaining an R^2 value at a marginally positive level of 0.1315 and keeping the MAE surge at 90.7405 °C. This confirms the effectiveness of the Attention Mechanism module in autonomously mitigating observational noise by reducing the attention weights on vectors of extreme fluctuations.

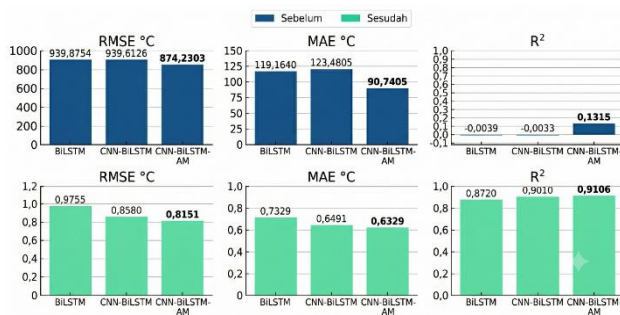


Figure 8. Comparative Analysis of Model Performance

The inferential capabilities of the representative model can only be fully assessed after testing is applied to the preprocessed dataset (green panel in Figure 8). In this scenario, the pure BiLSTM baseline model showed fairly adequate estimates with an R^2 value of 0.8720 and an MAE of 0.7329 °C, although the RMSE deviation remained high at 0.9755 °C. The application of a 1D-CNN layer to the CNN-BiLSTM successfully extracted spatial correlations, resulting in a reduction of the RMSE error to 0.8580 °C and an increase in R^2 to 0.9010.

The most optimal levels of precision and accuracy, in quantitative terms, were achieved by the main proposed architecture, CNN-BiLSTM-AM. The integration of the Attention Mechanism successfully reduced the mean absolute error (MAE) to 0.6329 °C and lowered the RMSE to 0.8151 °C. These significant improvements in error metrics are directly proportional to the model's enhanced explanatory capacity, as evidenced by a surge in the R^2 value to 0.9106.

D. Discussion

A comprehensive evaluation of analytical metrics, along with their visual exploration, confirms the functional

superiority of the primary proposed architecture. The integration of the local spatial feature extraction module (1D-CNN), bidirectional recurrent processing (BiLSTM), and the Attention Mechanism as an adaptive computational weighting system has been empirically proven to be complementary in optimizing prediction reliability. The synthesis of this series of tests provides strategic implications ranging from the validation of the feasibility of artificial intelligence architectures and the demonstration of the effectiveness of hybrid approaches to tangible contributions to the safety aspects of applied aeronautical instruments.

Meanwhile, a performance comparison between the use of raw datasets and preprocessed data reaffirms the validity of the “Garbage-In, Garbage-Out” principle in intelligent computational modeling. The complexity of the matrices in the deep learning architecture proved insufficient to tolerate parameter saturation caused by anomalies in Automatic Weather Station (AWS) recordings. Regarding the data purification process, the implementation of the LOF algorithm successfully eliminated micro-scale anomalies without compromising the structure of essential features. This cleaning was then refined by iForest to isolate macro-scale fluctuation biases in the five-year observation records. RFE-based dimensionality reduction subsequently played a crucial role by selecting 11 initial variables to become four primary physical predictors: daily maximum temperature, relative humidity, barometric pressure, and shortwave radiation exposure. This variable space reduction proved effective in suppressing noise that risks hindering the initial phase of gradient convergence. The absence of this structured, multi-stage preprocessing step not only marginally reduces the model's accuracy but also has the strong potential to trigger fatal inference failures.

Furthermore, the evaluation results reveal the empirical limitations of a pure BiLSTM architecture operating exclusively in the temporal domain. Although trained on a clean dataset, this network remains vulnerable to convergence fluctuations due to the high spatial heterogeneity of variable distributions. Concurrent physical interactions—such as a decrease in humidity coinciding with a surge in solar radiation flux from morning through midday—absolutely require a separate feature filtering component. Through the intervention of a one-dimensional convolutional layer (1D-CNN), micro-resolution temporal distortions can be filtered out and converted into a dense feature map. This representation is capable of synthesizing the dynamics of local meteorological relationships into a more structured abstract signal, which in turn successfully reduces the computational and memory load on the BiLSTM recurrent layer. The functional integration of these two modules confirms that network architecture design must be aligned with the principles of weather thermodynamics in tropical climates.

The most substantial performance leap was achieved by the proposed CNN-BiLSTM-AM architecture. This model's most notable capability is its ability to reduce the root mean

square error (RMSE) to as low as 0.8151 °C. This computational success is closely tied to the role of the Scaled Dot-Product Attention module. Unlike standard BiLSTM models, which compress the entire time series chronology into a static dimensional space making them susceptible to the vanishing sequence phenomenon the attention mechanism enables the network to dynamically calculate probability scores for the relevance distribution. By assigning maximum exponential weights to historical periods showing strong correlations with current physical anomalies, the attention mechanism effectively circumvents the phase response lag typically observed in standard recurrent networks during rapid thermal drops. In addition to reducing the mean absolute error (MAE) to 0.63 °C, this dynamic selectivity system offers a form of transparency into the black-box nature of deep learning modeling. To clarify this mechanism, it is highly recommended that visual representations using heatmaps which illustrate the transition of the attention intensity spectrum when processing specific weather events be included in subsequent research evaluation stages.

From a technical standpoint, the outputs of this computational synthesis have practical implications for operational efficiency at the Soekarno-Hatta International Airport Meteorological Station (WIII). In the context of air navigation, the system's 10-minute resolution outputs are highly relevant to the reporting intervals required for synoptic surface weather updates for aviation (Meteorological Aerodrome Report or METAR). From an aviation perspective, a deviation in the daily average temperature of 1.0 to 1.5 °C can compromise the precision of take-off roll distance calculations. Temperature variability that is not well predicted will proportionally affect the determination of the Maximum Take-Off Weight (MTOW) due to fluctuations in density altitude values in the equatorial latitudes. With an RMSE of 0.8151 °C, the CNN-BiLSTM-AM model demonstrates a highly accurate estimation capability, keeping the error margin strictly below 1.0 °C, which is a critical threshold for minimizing density altitude miscalculations.

Furthermore, a deeper residual analysis reveals the model's behavior under different weather conditions, which is visually corroborated by the time-series trajectory plots in Figure 9 and Figure 10. In these plots, the actual temperature data (ground truth) is represented by the black line, the baseline model's predictions are indicated by the blue line, while the predictions of the proposed CNN-BiLSTM-AM model are denoted by the red line.

During normal diurnal cycles, the baseline BiLSTM model performs adequately. However, during extreme micro-meteorological events—such as rapid temperature drops preceding localized convective storms, evidently captured in various panels of the provided visual plots (Part 1, Part 3, and Part 5)—the baseline model exhibits a distinct pattern of overprediction, lagging behind the rapid drop.

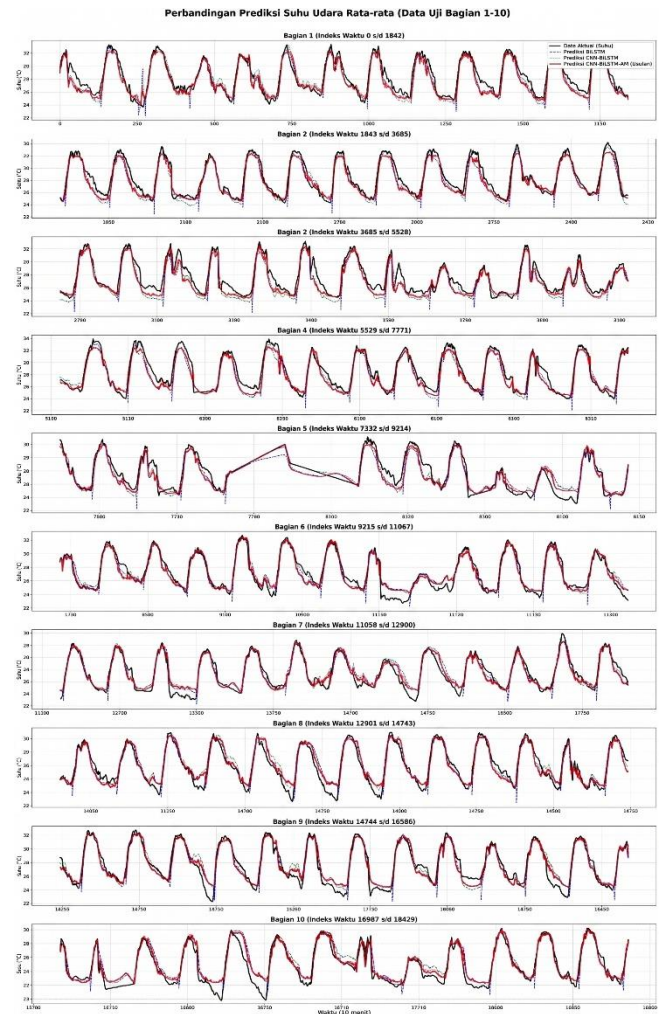


Figure 9. Comparison of Predictions on the Entire Test Data (Parts 1-10)

The CNN-BiLSTM-AM model significantly mitigates this residual error. As observed in the plots, where the red trajectory of the proposed model closely aligns with the fluctuations of the black ground truth line, the attention weights autonomously shift to prioritize historical vectors with similar sharp gradient changes. This mechanism effectively resolves the lag response delay and traces the extreme temperature spikes much more accurately than the baseline models.

Finally, for operational deployment in an airport environment, the computational cost and hardware requirements are highly feasible. The training and inference phases in this study were executed on a consumer-grade NVIDIA GPU (CUDA-enabled) with 4GB of VRAM and an Intel Core i7 processor. The integration of the 1D-CNN and

Attention modules only marginally increased the training time

Perbandingan Prediksi Suhu Udara Rata-rata (Data Uji Bagian 11-20)

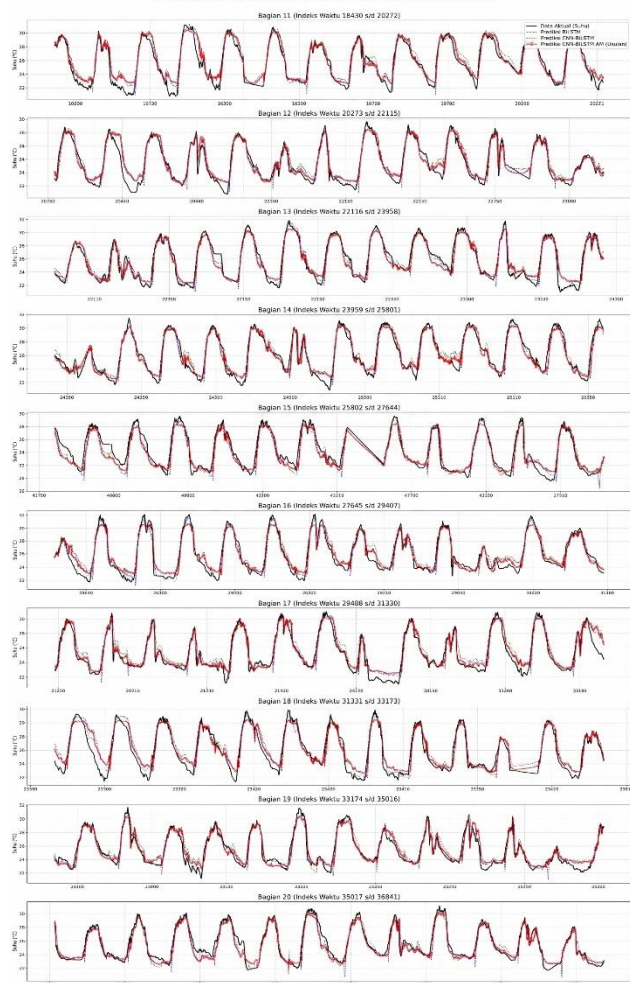


Figure 10. Comparison of Predictions on the Entire Test Data (Parts 11-20)

compared to the pure BiLSTM architecture. More importantly, the real-time inference time for a single prediction ($t + 1$) is executed in the microsecond range. This low computational footprint confirms that the proposed predictive framework can be seamlessly integrated into the airport's existing Automatic Weather Station (AWS) logging system, which updates every 10 minutes, without requiring expensive high-performance computing (HPC) clusters.

IV. CONCLUSION

Based on a comprehensive analysis of the CNN-BiLSTM-AM hybrid model for high-resolution air temperature prediction at the Soekarno-Hatta Meteorological Station, this study confirms that data integrity plays a role as crucial as the neural network's complexity. The implementation of a hybrid preprocessing framework proved highly effective as an anomaly filtration layer for Automatic Weather Station (AWS) data. The Local Outlier Factor (LOF) algorithm successfully reduced local micro-meteorological anomalies

by 2.08%, which was further refined by Isolation Forest (iForest) with a 5% contamination threshold to isolate global systemic anomalies. This multi-level detection successfully prevented generalization failure (vanishing gradient) caused by noise saturation. Furthermore, the hybrid feature selection approach optimally reduced the 10 initial predictors to 4 key features. By utilizing Pearson Correlation and Recursive Feature Elimination (RFE), features with insignificant linear correlations and low predictive contributions were eliminated. The four retained features—maximum air temperature, average humidity, air pressure, and solar radiation—were physically and computationally validated as the predictors carrying the strongest variance signal.

During the modeling phase, the proposed architecture successfully captured the complete spatio-temporal data patterns using a 24-step sliding window. The 1D-CNN layer effectively extracted local spatial correlations across variables, which were then processed by the BiLSTM network to extract temporal memory dependencies from both forward and backward directions. This construction was perfected by integrating a Scaled Dot-Product Attention Mechanism, which adaptively distributed maximum probability weights at critical time points, effectively resolving the lag response (phase delay) issue commonly experienced by standard recurrent models during sharp temperature oscillations. Consequently, the CNN-BiLSTM-AM model outperformed all tested variants. Compared to the pure BiLSTM baseline, it significantly reduced the RMSE by 16.44% and the MAE by 13.64%, culminating in a final RMSE of 0.8151 °C, an MAE of 0.6329 °C, and the highest explanatory capacity (R^2) of 0.9106. With an RMSE consistently suppressed below the 1.0 °C tolerance threshold, the model demonstrates strong potential as a highly reliable meteorological reference to support Density Altitude calculations, thereby refining Maximum Take-Off Weight (MTOW) estimations. However, it should be noted that the generalizability of this model is currently bounded by its reliance on single-station point data from Soekarno-Hatta International Airport. For future research, it is highly recommended to validate this predictive framework using multi-station AWS datasets from other airports or to integrate spatial regional models (such as Spatio-Temporal Graph Convolutional Networks) to capture the dynamics of air mass movement across different climatic zones.

ACKNOWLEDGMENTS

We would like to express our gratitude to all parties who supported the conduct of this research, particularly the Indonesian Agency for Meteorology, Climatology, and Geophysics (BMKG) at the Soekarno-Hatta Class I Meteorological Station for providing historical data from the Automatic Weather Station (AWS), which was crucial to the completion of this modeling study. We would also like to express our appreciation to the university and the academic

advisor who provided invaluable guidance and academic direction throughout the research process.

REFERENCES

- [1] A. Fadholi, "Study Pengaruh Suhu Dan Tekanan Udara Terhadap Operasi Penerbangan Di Bandara H.A.S. Hananjoeddin Buluh Tumbang Belitung Periode 1980-2010," *J. Penelit. Fis. dan Apl.*, vol. 3, no. 1, pp. 1–10, 2013.
- [2] S. Siswanto, J. Van Oldenborgh, and G. Van Der Schrier, "Temperature , extreme precipitation , and diurnal rainfall changes in the urbanized Jakarta city during the past 130 years," *int. J.Climatol*, 2015, doi: 10.1002/joc.4548.
- [3] International Civil Aviation Organization, "Annex 3 to the Convention on International Civil Aviation: Meteorological Service for International Air Navigation," Montreal, Canada, 2018.
- [4] World Meteorological Organization, "Technical Regulations (WMO-No. 49), Volume II - Meteorological Service for International Air Navigation," Geneva, Switzerland, 2021.
- [5] A. Utku, "Hybrid CNN-LSTM Model for Accurate Long-Term and Short-Term Temperature Prediction : A Case Study for Bingöl and Tunceli," *Int. J. PURE Appl. Sci.*, vol. 10, no. 2, pp. 550–562, 2024, doi: 10.29132/ijpas.
- [6] Q. Guo, Z. He, and Z. Wang, "Monthly Climate Prediction Using Deep Convolutional Neural Network and Long Short-Term Memory," 2024. doi: 10.1038/s41598-024-68906-6.
- [7] S. Melnyk, K. Vasiutynska, I. Korduba, Y. Trach, R. Trach, and D. Butenko, "Modeling and Forecasting of the Local Climate of Odesa Using CNN-LSTM and the Statistical Analysis of Time Series," *Sustainability*, pp. 1–21, 2025.
- [8] D. R. Wibawanty, Wandayantolis, and Ishak, "Verifikasi Kinerja Alat Automatic Weather System (AWS) dan Termometer Digital terhadap Observasi Manual di Stasiun Klimatologi Palembang Verification of The Performance of Automatic Weather Station (AWS) and Digital Thermometer Instruments towards Man," vol. 6, no. 2, pp. 151–163, 2022.
- [9] L. Bai, J. Wang, and Y. Zhou, "Outlier Detection and Explanation Method Based on FOLOF Algorithm," *Entropy*, vol. 27, pp. 1–21, 2025, doi: 10.3390/e27060582.
- [10] R. N. Fitrianiingsih and I. Kharisudin, "Time Series Modeling of Stock Price Using CNN-BiLSTM with Attention," vol. 13, no. 1, pp. 11–20, 2024.
- [11] S. Octaviana and Y. M. I. Indah, "Perbandingan Metode LSTM dan BiLSTM untuk Prediksi Data Curah Hujan," vol. 1, no. Februari, pp. 45–52, 2025.
- [12] X. Hao, Y. Liu, L. Pei, W. Li, and Y. Du, "Atmospheric Temperature Prediction Based on a BiLSTM-Attention Model," *Symmetry (Basel)*, vol. 14, p. 2470, 2022, [Online]. Available: <https://api.semanticscholar.org/CorpusID:253824824>
- [13] G. Airlangga, E. Faculty, and A. J. Catholic, "Advanced Machine Learning Techniques For Seismic Anomaly Detection In Indonesia : A Comparative Study Of Lof , Isolation Forest , And One-Class SVM University of Indonesia Situated on the volatile Pacific ‘ Ring of Fire ,’ Indonesia presents a contrast of," vol. 5, no. 1, pp. 49–61, 2024.
- [14] X. Liu, J. Song, H. Tao, P. Wang, H. Mo, and W. Du, "Quarter-Hourly Power Load Forecasting Based on a Hybrid CNN-BiLSTM-Attention Model with CEEMDAN , K-Means , and VMD," pp. 1–29, 2025.
- [15] J. Zhang, Y. Peng, B. Ren, and T. Li, "PM2.5 Concentration Prediction Based on CNN-BiLSTM and Attention Mechanism," 2021.
- [16] A. Kristyanto, C. Chairani, and S. Sriyanto, "Analisis Performa Model BiLSTM dan CNN-LSTM Dalam Prediksi Sea Water Level Pada Pelabuhan Berdasarkan Data Historis," vol. 7, no. 1, pp. 444–452, 2025, doi: 10.47065/bits.v7i1.7364.
- [17] S. Benkacali and G. Notton, "Comparative Study of Feature Selection Techniques for Machine Learning-Based Solar Irradiation Forecasting to Facilitate the Sustainable Development of Photovoltaics : Application to Algerian Climatic Conditions," *Sustainability*, vol. 17, 2025, doi: 10.3390/su17146400.
- [18] S. Shubham, R. Nirmalkar, T. Wadhvani, V. Shivani, and A. Pandey, "International Journal of Research Publication and Reviews Time Series Forecasting and Anomaly Detection with ARIMA and SARIMAX," vol. 5, no. 12, pp. 475–479, 2024.
- [19] A. El-sayed, L. Becker, and M. Thompson, "Comparative Study of Feature Selection Algorithms in Machine Learning Models for Rainfall Prediction," 2025.
- [20] L. Guo, Y. Pu, and W. Zhao, "CNN-BiLSTM Daily Precipitation Prediction Based on Attention Mechanism," *Atmosphere (Basel)*, vol. 16, pp. 1–13, 2025, doi: 10.3390/atmos16030333.
- [21] Z. Cheng, C. Zou, and J. Dong, "Outlier Detection using Isolation Forest and Local Outlier Factor," no. September 2019, 2025, doi: 10.1145/3338840.3355641.