ISSN 1648-7788 / eISSN 1822-4180

VILNIUS TECH
Vilnius Gediminas Technical University

AVIATION

2026
Volume 30
Issue 2
Pages 143–155
https://doi.org/10.3846/aviation.2026.26809

# APPLICATION OF LSTM NEURAL NETWORKS WITH MULTIVARIATE NUMERICAL ANALYSIS TO AVIATION WIND GUST FORECASTING

Chuen-Jyh CHEN

*Department of Aviation and Maritime Transportation Management, Chang Jung Christian University, Tainan, Taiwan, R.O.C.*

**Article History:**
- received 21 August 2025
- accepted 23 January 2026

**Abstract.** This paper presents a long short-term memory (LSTM) framework developed for predicting wind gusts 1 h in advance at Taiwan Taoyuan International Airport (RCTP) during typhoons. Hourly surface observations were collected from 12 landfalling typhoons (2010–2020) and used to compare three feature-selection strategies: Pearson correlation, recursive feature elimination with cross validation, and random-forest importance. Models were trained on 12-h multivariate histories. A leave-one-typhoon-out cross-validation scheme revealed that the LSTM model with random-forest selection achieved a mean root-mean-square error of 2.33 m/s and mean absolute percentage error of 21.12%. Although these statistics are comparable to those of a 1-h persistence baseline model on average, the proposed model considerably outperformed the persistence baseline model during rapid intensification and decay phases, reducing errors by approximately 45%. Forecast errors generally remained within the ±5 m/s operational advisory threshold. The results of this case study for RCTP suggest that feature selection can be combined with sequence-based deep learning to provide robust decision support for aviation operations during extreme weather events.

**Keywords:** airport meteorology, air traffic management, runway operations, gust forecasting, long short-term memory, feature selection, decision support.

Corresponding author. E-mail: *chuenjyh@mail.cjcu.edu.tw*

## 1. Introduction

Extreme weather events can have severe effects on aviation operations (Anaman et al., 2017). Sudden increases in wind speed on a runway can compromise the safety of aircraft takeoffs and landings. Wind gusts, which are brief increases in wind speed, increase the difficulty of handling aircraft. During a typhoon, wind gusts on runways affect airport operations and flight safety. Understanding how extreme weather events can affect aviation requires the use of advanced observational systems and instruments (Ummenhofer & Meehl, 2017). Alves et al. (2025) used deep learning to generate high-resolution wind speed and direction forecasts for airport approach/departure corridors, improving runway safety decisions in gusty conditions. Khattak et al. (2025) developed a hybrid Variational Mode Decomposition, Tree-structured Parzen Estimator, and Temporal Convolutional Network model for minute-scale wind-shear prediction near runways, giving airlines added lead time for typhoon-related hazards. The long short-term memory (LSTM) model is suitable for processing time-series data, such as weather patterns and energy demand data.

Extensive numerical research has been conducted on extreme weather events. Gao et al. (2018) proposed an LSTM model and used it to produce 6–24 h nowcasts of typhoon tracks. Chen et al. (2019) developed a hybrid convolutional neural network (CNN)-LSTM model to predict typhoon formation and intensity. Moreover, Yuan et al. (2021) proposed an LSTM model to predict a typhoon center's maximum wind speed and minimum pressure. Li et al. (2021) developed an LSTM model for flood prediction; this model uses synchronized data sequences as its input and produces synchronized data sequences as its output. In addition, Chen et al. (2021) proposed an LSTM model for predicting the water level for the next 1 h during storm surge periods. Wei (2021) developed an alert system based on an LSTM model to evaluate the stability of scaffolding under wind stress. Caseri et al. (2022) proposed a CNN-LSTM model for predicting the rate of rainfall during heavy rainfall events. Li et al. (2022) designed a CNN-LSTM model for forecasting water availability and providing flood alerts. Finally, Zhou et al. (2022) developed a modified LSTM model for predicting the rapid intensification of typhoons. Modé et al. (2025) introduced a dual-output LSTM that classifies and regresses extreme winds, yielding superior 0–3 h gust forecasts during tropical cyclones. Su et al. (2025) improved storm-surge prediction by coupling ML-derived hybrid wind fields with physics-based surge models, highlighting the value of fusing meteorological and oceanographic features.

Current literature presents extensive numerical analyses of extreme weather phenomena; however, it is marred by several significant shortcomings. Many deep learning studies omit a formal specification for lead time ($L$) and history length ($k$). This oversight obscures their operational usefulness. Without these parameters, benchmarking performance or replicating results becomes almost impossible. Validation fragility remains a persistent concern. Evaluations often gravitate toward single-event case studies, bypassing the rigors of multi-event cross-validation. This narrow focus leaves a fundamental question unanswered: can a model calibrated for a single typhoon truly withstand the structural variability of another? Existing research often isolates weather features in isolation. This approach overlooks the intricate, multivariate interplay required to understand the rapid evolution of wind gusts. A disconnect exists. Single-variable models may offer theoretical insights, but they often fail to capture the nonlinear dynamics of a landfalling storm. Closing these gaps necessitates a framework that explicitly defines the forecasting task and integrates multivariate feature selection across a spectrum of diverse events.

Feature selection methods enhance the accuracy of weather prediction models and are thus suitable for weather forecasting. Li et al. (2015) developed an ensemble model that integrates neural networks, the wavelet transform, feature selection, and partial least-squares regression to forecast wind power for energy generation. Zhao et al. (2016) proposed a neural network with a feature selection method for air quality forecasting. Wolff et al. (2017) used three feature selection methods – support vector regression (SVR), random forest, and linear regression – to enhance the accuracy of photovoltaic power forecasting. Das et al. (2018) proposed a hybrid artificial neural network (ANN) that can predict rice yield from long-term weather data. For turbulence modeling, Yin et al. (2020) designed a modified approach involving feature selection. Chen and Chang (2021) suggested using an LSTM model based on Pearson correlation coefficients to predict photovoltaic power. Ren et al. (2022) designed a hybrid LSTM model that performs the Granger causality test to select multiple features for typhoon track prediction. Yildirim et al. (2022) developed a hybrid CNN model with a feature selection method for classifying weather images. Tao et al. (2022) proposed a machine learning model with a feature selection method for forecasting relative humidity at two meteorological stations in their study area. Lv and Wang (2023) showed that multi-objective evolutionary feature selection cuts 1–12 h wind-speed forecast error by over 20%. Kartal et al. (2023) used a decision-tree-based measure-correlate-predict method on global reanalysis data to estimate peak gusts, with tree-derived feature importance capturing key nonlinear interactions. The aforementioned literature indicates that suitable feature selection can considerably improve the accuracy of weather predictions. The Pearson correlation coefficient, which is leveraged in the LSTM model of Chen et al. (2021), is not suitable for predicting wind gust speed because of the characteristics of data on this parameter. Therefore, this study bridges the identified gaps by synthesizing random-forest feature importance with a long short-term memory (LSTM) architecture. This hybrid approach hones the accuracy of short-term gust forecasts during the complex dynamics of landfalling typhoons. Our contributions are threefold:

1. Task Formalization. A precise mathematical architecture for the forecasting problem is established. By anchoring the task to a fixed 1-h lead time ($L$ = 1 h) and a 12-h antecedent history ($k$ = 12 h), operational transparency and exact reproducibility is prioritized.
2. Robust Cross-Validation. Our evaluation protocol bypasses the inherent limitations of single-event analysis. A leave-one-typhoon-out cross-validation scheme across 12 distinct historical events was implemented. This rigorous testing ensures the framework withstands the structural variability of diverse storm tracks and intensities.
3. Operational Benchmarking. We demonstrate the framework's clear superiority over a persistence baseline. The performance gains are most acute during rapid intensification and decay. These critical phases require precise timing for runway management; our model delivers that precision where standard persistence falls short. It offers tangible, data-driven support for ground operations.

This research focuses exclusively on Taiwan Taoyuan International Airport (RCTP). These findings establish a strong, localized proof of concept. However, universal optimality remains untested; diverse airport topographies and distinct climatic regimes demand site-specific validation. The remainder of this paper follows a systematic progression. Section 2 delineates the deep learning architectures and feature selection logic. Section 3 details the experimental framework and data provenance. Section 4 presents a comprehensive analysis of the results. Finally, Section 5 synthesizes the core conclusions and projects future research directions.

## 2. Deep learning neural networks and feature selection

Compared with traditional machine learning methods, such as linear regression and support vector machine, artificial intelligence methods, such as ANN-based methods, can handle larger quantities of data and achieve higher learning performance. However, ANN models exhibit limitations in processing data with strong correlations in time or space, such as solar radiation data (Pang et al., 2020). Recurrent neural networks (RNNs) address this limitation by efficiently processing sequence data, such as time-series, text documents, or audio information.

The structure of an RNN enables it to learn the dynamic behavior reflected in time-series data. In an RNN, input data are processed by one or more hidden layers to extract and store historical information and contextual correlations. Compared with RNN models, LSTM models exhibit better performance in the processing of long sequences

without exhibiting long-term memory problems. There are three main gates: forget gate $f_t$, input gate $i_t$, and output gate $o_t$ in LSTM to determine the data storage and loading information of past events as shown in Figure 1. LSTM can add or delete information in the cell state $C_t$, which is processed by the gate structure and selectively controls information passage. The forget gate $f_t$ is the first operation step that uses the sigmoid (σ) activation function to decide whether information from the previous time step should be kept or discarded. The sigmoid function takes in the output of the previous time step and the current input and outputs a value between [0,1] for each value in the cell state of the previous time step $C_{t-1}$. 1 means "completely keep" and 0 means "completely discard". The equation of forget gate $f_t$ can be written as:

$$f_t = \sigma\left((w_f \cdot [h_{t-1}, x_t] + b_f\right), \tag{1}$$

where $w_f$ and $b_f$ are the weight and offset value of forget gate $f_t$, $h_{t-1}$ represents the output, $x_t$ is the input of the current time step, and + denotes scalar addition. The input gate $i_t$ is the next operation step to choose the new information stored in the cell state $C_t$ and input the same as the forget gate $f_t$,

$$i_t = \sigma\left(w_i \cdot [h_{t-1}, x_t] + b_i\right), \tag{2}$$

where $w_i$ and $b_i$ are the weight and offset value of the input gate, the new information can be updated by the sigmoid function (σ), and a *tanh* layer creates a vector of new candidate value $\tilde{C}_t$ to the state.

$$\tilde{C}_t = \tanh\left(w_c \cdot (h_{t-1}, x_t) + b_c\right), \tag{3}$$

where $w_c$ and $b_c$ are the weight and offset value of the new candidate cell state $\tilde{C}_t$. Next, combine $f_t$, $i_t$ and $\tilde{C}_t$ to update the old cell state $C_{t-1}$ into the new cell state $C_t$,

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t, \tag{4}$$

where $*$ denotes convolution operation. The output gate $o_t$ is the last operation to decide whether to output the information,

$$o_t = \sigma\left(w_o \cdot [h_{t-1}, x_t] + b_o\right); \tag{5}$$

$$h_t = o_t * \tanh\left(C_t\right), \tag{6}$$

where $w_o$ and $b_o$ are the weight and offset value of output gate $o_t$, $h_t$ is the hidden state, and $C_t$ is the cell state for representing long-term memory.

An LSTM model contains three main gates for the storage and processing of data: a forget gate $f_t$, an input gate $i_t$, and an output gate $o_t$ (Figure 1). The indices $t-1$, $t$, and $t+1$ denote the previous, current, and subsequent time steps, respectively. Pointwise multiplication ⊗ and pointwise addition ⊕ facilitate the flow of data across the cell. The sigmoid function σ regulates the degree of information passage through three primary gates. The forget gate $f_t$ determines which data from the preceding step to discard. The input gate $i_t$, manages the integration of new observations into the memory. $\tilde{c}_t$ represents the candidate cell state at the current time step, serving as the bridge between historical context and future predictions. An LSTM model utilizes the aforementioned three gates to decide which information should be retained or forgotten. The forget gate $f_t$ controls whether to retain the output of the previous unit in the previous time step and the input of the current unit. The input gate $i_t$ controls whether to input the value for this unit and calculate the value. Finally, the output gate $o_t$ controls whether to output the value calculated by this unit. The aforementioned gates can adaptively decide how much information should be forgotten. These decisions are made on the basis of weight values learned by hidden layers and input functions to mitigate the vanishing gradient problem (Graves, 2012).

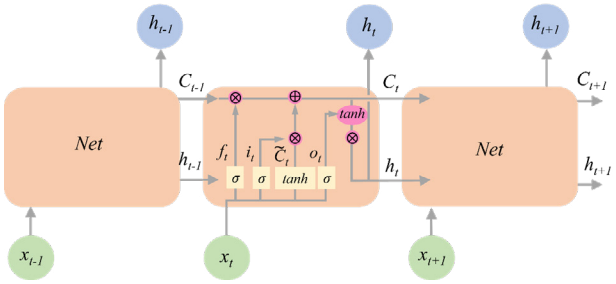



**Figure 1.** Architecture of the LSTM cell

Large and complex data sets often contain redundant features or noise. Therefore, feature selection must be conducted on such data sets to eliminate redundant features and noise and thereby improve the model's accuracy and interpretability. In general, three types of feature-selection methods exist: filter, wrapper, and embedded methods (Figure 2). In the present study, one feature-selection method from each of the aforementioned categories for improving the accuracy and reliability of an LSTM model for forecasting maximum wind gust speed was selected.

Among filter methods, the chi-squared test (Greenwood & Nikulin, 1996) is the most appropriate technique for categorical problems because it indicates the dependence between two features. Spearman's rank correlation coefficient (Spearman, 1904) indicates the association between two features on the basis of their coefficient ranks rather than their raw data. Despite its historical origins, it remains a highly effective and stable filter-based feature selection technique in modern high-dimensional data analytics (Salman et al., 2026). Moreover, the Pearson product-

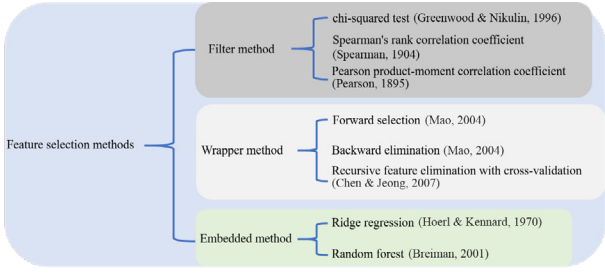



**Figure 2.** The overview map using three methods to achieve the feature selection: filter, wrapper, and embedded methods

moment correlation coefficient (Pearson, 1895) reflects the linear association between two features; this coefficient is used to measure the strength and direction of a linear relationship between two features under the assumption that the two features have normal data distributions and are linearly related. The equation for the Pearson product-moment correlation coefficient $\rho$ can be written as:

$$\rho(A,B) = (1/N-1)\sum_{i=1}^{N}\left((A_i-\mu_A)/\beta_A\right)\left((B_i-\mu_B)/\beta_B\right), \quad (7)$$

where $N$ is the number of observed data, $\mu_A$ and $\beta_A$ are the mean and standard deviation of feature $A$, respectively, and similarly for feature $B$. The value of the Pearson product-moment correlation coefficient ranges between [–1, 1]. 1 indicates a strong positive linear relationship between the two features. This means that as one feature increases, the other feature also tends to increase. –1 indicates a strong negative linear relationship between the two features. This means that the other feature tends to decrease as one feature increases. 0 indicates no linear relationship between the two features. This study aims to predict the maximum wind gust speed, and the Pearson product-moment correlation coefficient is suitable to measure the strength and direction of a linear relationship between two features. This method will be appropriate for selecting extreme weather features.

Wrapper methods, such as forward selection (Mao, 2004) and backward elimination (Mao, 2004), are commonly used for feature selection. Forward selection involves adding features one at a time to a model on the basis of the strength of their association with the outcome. This method begins with an empty set of features, and features are added individually until the model's performance no longer improves significantly. By contrast, backward elimination involves removing features one at a time from a model on the basis of their association with the outcome. This method begins with a full set of features, and the feature with the weakest effect on model performance is removed in each step, with feature removal being conducted until a stopping criterion is met.

Recursive feature elimination with cross-validation (RFECV) (Chen & Jeong, 2007) combines the concepts of recursive feature elimination (RFE) and cross validation. In RFE, features are repeatedly removed from a model, and the feature set with the highest cross-validation accuracy is retained. Each step of this process involves eliminating the feature contributing the least to accuracy. In RFECV, cross-validation is employed to estimate generalization performance and select features. The process begins with all features and involves the elimination of the feature with the weakest effect on the cross-validation accuracy in each step, with feature removal being conducted until a stopping criterion is met.

Although forward selection and backward elimination are simple and computationally efficient, they are prone to overfitting and underfitting, respectively. Compared with these methods, RFECV is more robust but also requires more computational resources; it involves the use of a prediction model and the cross-validation technique to evaluate feature importance. In the present study, RFECV was used to select extreme weather features for predicting maximum wind gust speed, and SVR (Smola & Schölkopf, 2004) was employed to train and test the developed model. The method involved training the model, evaluating its performance through cross-validation, and removing the least important feature from the feature set. The aforementioned process was repeated until the performance of the model no longer improved.

Ridge regression, originally introduced by Hoerl and Kennard (1970), remains a widely used regularization method for mitigating multicollinearity and overfitting in regression-based feature selection (Saputro et al., 2025). The L2 regularization term encourages the use of smaller weights in the model. The cost function evaluates the disparity between the predicted and true output values. This function is the sum of the residuals and L2 regularization term. A residual is the difference between a pair of predicted and true values. Ridge regression can be employed for the selection of features based on their parameter values.

Random forest (Breiman, 2001) is an extension of decision trees in which the predictions of multiple decision trees are combined, leading to more accurate and stable predictions. Random forest can be used for feature selection. First, a random subset of features is selected for each decision tree while a random-forest model is trained. This process is repeated multiple times, and the feature most frequently selected across all decision trees is considered the most important feature. Feature importance is determined by training the random-forest algorithm on the entire data set. Subsequently, the most crucial features are selected.

Ridge regression and random forest can be used for feature selection; however, they offer different advantages depending on the nature of the data and the goal of feature selection. Ridge regression is computationally efficient, and its results are easy to interpret; however, the ridge regression algorithm may struggle to handle high-dimensional data sets with many features. The random-forest algorithm measures feature importance on the basis of each feature's contribution to the model's performance; features with higher feature importance scores are considered more important. Random forest is robust to outliers and noisy data and can effectively handle high-dimensional data sets with many features. Extreme weather data contain outliers and noise; therefore, the random-forest algorithm is appropriate for selecting extreme weather features.

In random forest, multiple decision trees are trained on different data samples, and the impurity of the data in each split is evaluated. Thus, this algorithm can provide robust and accurate measurements of feature importance. In the random-forest algorithm, data impurity is measured in terms of the mean square error (MSE), which refers to the average squared difference between the actual and predicted values in a data set. A feature with a lower MSE contributes more to reducing the data impurity. The

feature with the lowest MSE is selected as the splitting feature, and this process is continued recursively until the decision trees are expanded to the point where each leaf node contains at least the minimum number of samples. The splitting feature is used to divide the data into subnodes at each internal node, with the objective being to split the data such that similar instances are grouped and dissimilar instances are separated to reduce data impurity. The MSE can be written as:

$$\mathrm{MSE} = \sum_{i=1}^{N} \left(y_i - \widehat{y_i}\right)^2 / N, \tag{8}$$

where $N$ is the number of observed data, $y_i$ is the real value, and $\widehat{y_i}$ is the forecast value. MSE takes the difference between the actual and predicted values for each data point, squares it to ensure positive values, and then calculates the average of these squared differences. The resulting value indicates how well the predicted values align with the actual values. Smaller MSE values indicate a better fit of the model to the data. This provides a more robust and reliable estimate of the feature's importance, as it accounts for the variability and uncertainty in the predictions made by individual trees in the forest.

## 3. Methods

### 3.1. Forecasting task formulation

Let $y_t$ denote the hourly maximum wind gust speed (m/s) observed at Taoyuan International Airport (RCTP) at time $t$, and let $x_t$ be the column vector of near-surface predictors at time $t$ (sea-level pressure, temperature, dew-point temperature, relative humidity, 10 m wind speed and direction, gust direction, precipitation amount and duration, and related variables described in Section 2). The goal is to predict the future gust at a short lead time $L$ using only locally observed quantities over a finite historical window of length $k$. The forecasting task as a nonlinear sequence-to-scalar mapping was formulated:

$$y_{t+L} = f\left(x_{t-k+1}, \ldots, x_t\right), \tag{9}$$

where $y_{t+L}$ is the hourly maximum gust at time $t+L$, and $x_{t-k+1}, \ldots, x_t$ denote the multivariate input sequence over the previous $k$ hours. $L$ is the forecast lead time in hours, $k$ is the history length in hours, and $f(\cdot)$ is an unknown nonlinear function learned by the LSTM network. In this case study, it is focused on a one-hour-ahead aviation application, setting $L$ = 1 h and $k$ = 12 h, so that the model utilizes the previous 12 hours of multivariate near-surface observations to predict the maximum gust one hour ahead. This choice balances the need to capture pre-landfall trends in pressure, wind, and precipitation against the limited amount of data available for each typhoon event.

### 3.2. Data preprocessing

Raw meteorological records frequently harbor gaps or temporal inconsistencies. Rigorous preprocessing is therefore mandatory prior to model integration. In this study, ten core variables were archived: sea-level pressure (SP, hPa), temperature (T, °C), dew point temperature (DP, °C), relative humidity (RH, %), wind speed (WS, m/s), wind direction (WD, rad), maximum wind gust speed (GS, m/s), wind gust direction (GD, rad), precipitation (P, mm), and precipitation duration (PD, h). Gust intensities were further indexed via the Beaufort wind force scale (Table 1). Raw feature magnitudes proved prohibitive for efficient LSTM convergence. This necessitated global normalization. Scaling within the [−1, 1] interval effectively dampens outliers in these high-variance datasets. In contrast, [0, 1] normalization maps all observations to a consistent range based on observed historical minima and maxima. Weather variables exhibit disparate dynamic ranges. Left unscaled, these differences would skew the model's weight distribution and bias the learning process. Normalization neutralizes this risk. It accelerates training and ensures the LSTM resolves subtle gust fluctuations with high fidelity. The normalization can be shown as:

$$X' = (X - X_{\min}) / (X_{\max} - X_{\min}), \tag{10}$$

where $X$ is the raw data point, $X'$ is the new data point after feature scaling, $X_{\min}$ and $X_{\max}$ represent the minimum and maximum value, respectively. This normalization of raw data enables LSTMs to learn diverse features in deep learning by standardizing all data to the same magnitude. Consistent with the cross-validation scheme described below, all normalization parameters are computed using only the training portion of each fold and then applied to the validation and test data to avoid information leakage.

Data gaps exceeding 6 h in any predictor warranted immediate exclusion. Artificial artifacts must not contaminate the temporal sequence. Hourly records for the 12 landfalling typhoons listed in Table 2 were chronologically ordered and partitioned into continuous segments.

**Table 1.** The list of the Beaufort wind force scale

| Beaufort scale | Knot (kt) | Metric system (m/s) | |
|---|---|---|---|
| 0 | 0~1 | 0 ~ | 0.2 |
| 1 | 1~3 | 0.3 ~ | 1.5 |
| 2 | 4~6 | 1.6 ~ | 3.3 |
| 3 | 7~10 | 3.4 ~ | 5.4 |
| 4 | 11~16 | 5.5 ~ | 7.9 |
| 5 | 17~21 | 8.0 ~ | 10.7 |
| 6 | 22~27 | 10.8 ~ | 13.8 |
| 7 | 28~33 | 13.9 ~ | 17.1 |
| 8 | 34~40 | 17.2 ~ | 20.7 |
| 9 | 41~47 | 20.8 ~ | 24.4 |
| 10 | 48~55 | 24.5 ~ | 28.4 |
| 11 | 56~63 | 28.5 ~ | 32.6 |
| 12 | 64~71 | 32.7 ~ | 36.9 |
| 13 | 72~80 | 37.0 ~ | 41.4 |

**Table 2.** The list of 12 typhoons that made landfall in Taiwan

| Year | Typhoon | Coordinated Universal Time (UTC) |
|---|---|---|
| 2010 | Namtheun | 30–31 Aug. |
| 2012 | Saola | 30 Jul–3 Aug. |
| 2013 | Soulik | 11–13 Jul. |
| 2013 | Trami | 20–22 Aug. |
| 2013 | Kong-Rey | 27–29 Aug. |
| 2014 | Matmo | 21–23 Jul. |
| 2014 | Fung-Wong | 19–22 Sep. |
| 2015 | Soudelor | 6–9 Aug |
| 2015 | Dujuan | 27–29 Sep. |
| 2016 | Megi | 25–28 Sep. |
| 2017 | Haitang | 28–31 Jul. |
| 2019 | Bailu | 21–25 Aug. |

These events underpinned the leave-one-typhoon-out cross-validation framework detailed in Section 3.6. Within this rotation, each typhoon served as an isolated test case. The objective remains clear: evaluate 1-h gust forecasts against truly independent meteorological profiles.

## 3.3. Feature-selection methods

Predictive skill often hinges on a lean subset of physical drivers. To quantify this, three feature-selection strategies prior to LSTM training were deployed: (i) Pearson correlation ranking, (ii) recursive feature elimination with cross-validation (RFECV-SVR), and (iii) random-forest importance. These techniques reduce the input space to its most essential elements. Each process runs independently in complete isolation. Every procedure functions solely within its respective cross-validation fold, utilizing only the training typhoons. By excluding the test event from the selection process, the possibility of information leakage is eliminated and the framework's integrity during blind testing is maintained.

### *3.3.1. Filter-based selection: Pearson correlation*

The first strategy quantifies the Pearson product-moment correlation coefficient ρ between each candidate predictor and the target gust. These coefficients are computed across the full training fold for each event. Variables are then ranked by absolute magnitude (|ρ|), with only the top four retained as inputs for the LSTM. This filtering technique guarantees high transparency with minimal computational cost. It reliably isolates key variables such as local wind speed, sea-level pressure, and precipitation intensity during tropical cyclone events. Its effectiveness is clear. However, the method has limitations. It overlooks possible collinearity among predictors and cannot capture the nonlinear interactions present in complex weather data.

### *3.3.2. Wrapper-based selection: RFECV-SVR*

RFECV acts as an effective wrapper method, focusing on optimizing predictive performance for different predictor subsets. For each training fold, a linear Support Vector Regression (SVR) model is trained using all features. The algorithm then systematically removes the least impactful variables – those whose removal results in the smallest decline or the greatest improvement in cross-validated RMSE. After each removal, the SVR is refitted. This process continues until further pruning no longer reduces the error. The final selected subset, usually consisting of four variables, is used as input for the LSTM. While this approach reduces redundancy and refines the input space, it requires considerably more computational resources compared to simple correlation-based filters.

### *3.3.3. Embedded selection: Random-forest importance*

The Random Forest (RF) combines multiple regression trees to model the complex, nonlinear relationships between potential predictors and gusts. We assess variable importance using the Mean Decrease in Impurity (MDI), which is computed by averaging impurity reductions across all trees. These scores are used to rank the predictors, with the top four chosen as input features for the LSTM. This selection process is inherently driven by the data. In practice, it results in a concise, physically meaningful subset – mainly wind speed, gust direction, and precipitation features – effectively ignoring less informative predictors. As explained in Section 4, the RF-based method outperformed other approaches, providing the most reliable skill across the 12-typhoon dataset.

## 3.4. LSTM architecture and training

The LSTM network processes sequences of length $k$ of multivariate inputs and outputs a single scalar prediction $y_{t+L}$. For a given feature set, each input sample consists of a $k \times d$ matrix, where $d$ is the number of selected predictors and $k$ = 12 h is the history length. The architecture employed in this study comprises two stacked LSTM layers, followed by a fully connected output layer. The first LSTM layer has 32 hidden units and returns the full sequence of hidden states; a dropout layer with a dropout rate of 0.1 is applied to reduce overfitting. The second LSTM layer has 6 hidden units and returns only its final hidden state, which is passed to a dense layer that maps to a single output representing the forecast gust $y_{t+L}$. The network is trained to minimize the mean squared error (MSE) between predictions and observations.

We employ the Adam optimizer (Kingma & Ba, 2015) with default hyperparameters and train the network on each training fold using mini-batch gradient descent with a batch size of 12 and a maximum of 150 epochs. A validation set is carved out from the training typhoons in each fold, and a model-checkpoint mechanism retains the weights corresponding to the lowest validation loss. All normalization parameters and feature-selection steps

are computed using only the training portion of each fold and then applied to the corresponding validation and test data, thereby avoiding any leakage of information from the held-out typhoon into the training process.

### 3.5. Persistence baseline

As an operationally transparent benchmark, a 1-h persistence forecast that assumes the future gust equals the most recent observation is included:

$$\hat{y}_{t+L}^{pers} = y_t \text{, with } L = 1 \text{ h.} \quad (11)$$

The persistence model adopts the identical target variable and verification window as the LSTM framework. No exogenous predictors or training phases are required. The logic is purely temporal. Performance hinges on atmospheric stability; the model excels when gust trends evolve gradually and recent conditions remain representative. Conversely, it falters when recent observations lose their predictive power, specifically during the volatile windows of rapid intensification or decay. We subject the persistence forecast to the same cross-validation regime and error metrics (RMSE and MAPE) used for the LSTM. This consistency isolates the added value of the deep-learning architecture. By benchmarking against this naive reference, a transparent performance floor is established. It ensures that the reported gains represent a genuine improvement over the simplest operational heuristic.

### 3.6. Leave-one-typhoon-out cross-validation

Model robustness across diverse storm trajectories and intensities using a leave-one-typhoon-out cross-validation scheme is validated. Continuous hourly records from the 12 landfalling typhoons listed in Table 2 are partitioned into discrete events. Boundaries are defined by gaps in observation timestamps exceeding 1.5 h. This partitioning preserves meteorological continuity. Each resulting segment encapsulates a single typhoon passage, typically spanning 48 to 120 h of high-resolution data.

Each iteration segregates the 12-event dataset into eleven training typhoons and one independent test case. Testing remains entirely blind. Preprocessing, normalization, and feature-selection parameters are derived solely from the training pool and then mapped onto the held-out typhoon to prevent information leakage. The LSTM ingests 12-hour sliding-window sequences (k = 12 h) to solve the 1-h-ahead prediction task. Evaluation relies on RMSE and MAPE across each fold. This event-based design provides a stringent assessment of the model's capacity to generalize across diverse meteorological profiles. It ensures that performance reflects the model's true architectural ability to withstand distinct landfall scenarios. The model must resolve universal gust dynamics rather than merely mastering the specific characteristics of a single event, such as Typhoon Dujuan.

Records were chronologically arranged at a 1-hour temporal resolution. Hourly observations from the 12 landfalling typhoons were categorized into discrete and continuous events. This partitioning ensures experimental integrity. These clusters underpinned the leave-one-typhoon-out cross-validation framework. Within each fold, the LSTM and its associated feature-selection logic were trained on 11 typhoons and verified against the 12th. This iterative rotation ensures that every storm functions as an entirely independent validation set for 1-h-ahead gust predictions. Such a design prevents the model from relying on specific storm characteristics, forcing it instead to resolve the broader atmospheric dynamics governing extreme wind events at RCTP.

## 4. Aviation weather prediction

The data collected in this study were sorted by time, with the data interval being 1 h. Hourly weather records associated with each of the 12 landfalling typhoons listed in Table 2 were grouped into continuous typhoon events. These 12 events were then used in a leave-one-typhoon-out cross-validation scheme: in each fold, the LSTM and feature-selection procedures were trained on 11 typhoons and evaluated on the remaining typhoon, so that every storm in turn served once as an independent test case for 1-h-ahead gust forecasts.

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \widehat{y_i}\right)^2}; \quad (12)$$

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\left(y_i - \widehat{y_i}\right) \middle/ y_i\right| \times 100\%, \quad (13)$$

where $N$ is the number of observed data, $y_i$ is the real value, and $\widehat{y_i}$ is the forecast value. The MAPE is a measure of the difference between a target feature's forecast and its actual values, expressed as a percentage. It is calculated as the average of the absolute differences between predicted and actual values, divided by the actual values. MAPE provides an intuitive way to interpret the relative error between forecast and actual values. Following the conventional interpretation of Lewis (1982), a MAPE between 10% and 20% is considered good, between 20% and 50% is considered reasonable, and greater than 50% is considered weak. However, because MAPE has known statistical limitations as highlighted by Hyndman and Koehler (2006), it is interpreted alongside RMSE in this study to ensure a balanced and robust error assessment.

The LSTM framework (Figure 3) integrates three feature-selection strategies to refine forecasting accuracy: Pearson correlation ranking, RFECV, and random-forest importance. The model consists of a multi-input sequence layer (receiving $k$ hours of history for $d$ features), two stacked LSTM layers (32 and 6 units), a fully connected layer, and a regression output layer to predict the single scalar value $y_{t+L}$. $N$ denotes the number of observations, and $t$ represents the time step. For the initial screening, the Pearson product-moment correlation coefficient between nine candidate weather features and the target gust speed

was computed.This metric quantifies linear dependence on a scale from –1 to 1. Coefficients of 1 and –1 indicate perfect positive and negative linear relationships, respectively, while a value of 0 denotes no linear coupling. We adopt Cohen's (2013) established thresholds for interpretation. Correlation magnitudes matter. Values between 0.1 and 0.3 indicate weak associations. Moderate strength lies between 0.3 and 0.5. Robust statistical coupling is characterized by coefficients exceeding 0.5. These benchmarks allow us to isolate the most influential linear drivers before passing them to the nonlinear LSTM layers. Table 3 quantifies the Pearson product-moment correlation coefficients ($\rho$) between each meteorological candidate and the observed gust speed. Associations vary across the feature set. The four variables with the strongest statistical coupling were isolated: sea-level pressure, wind speed, precipitation, and precipitation duration. These drivers anchor the forecasting logic. Table 4 presents the hierarchy of the nine candidate features based on their correlation magnitudes, providing a transparent ranking of their relative predictive weights.

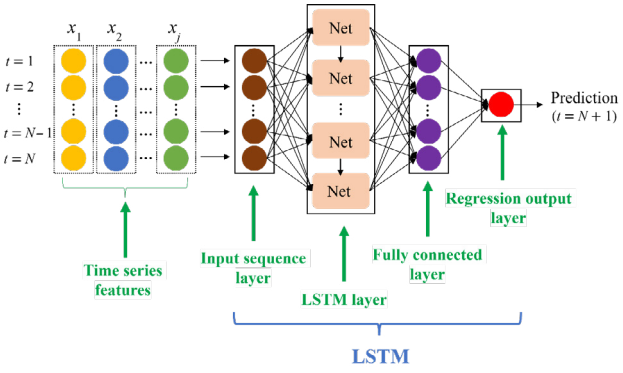



**Figure 3.** The multivariate numerical forecasting architecture using an LSTM network

**Table 3.** RMSE and MAPE of the maximum wind gust speed forecasting by using the remaining 8 features for training (Feature Importance Test)

| Deleted feature | RMSE (m/s) | MAPE (%) |
|---|---|---|
| Sea level pressure (SP) | 2.7859 | 26.6517 |
| Temperature (T) | 2.6641 | 25.9875 |
| Dew point (DP) | 2.6849 | 26.1608 |
| Relative humidity (RH) | 2.5825 | 25.6950 |
| Wind speed (WS) | 6.3094 | 68.8539 |
| Wind direction (WD) | 2.3416 | 23.8011 |
| Wind gust direction (GD) | 2.6936 | 23.5597 |
| Precipitation (P) | 2.6731 | 24.4937 |
| Precipitation duration (PD) | 2.8388 | 24.1610 |

Raw datasets often contain noise irrelevant to gust dynamics. We systematically isolated and removed features to quantify their impact on predictive skill. Optimal performance emerged with a four-feature configuration: sea-level pressure (SP), temperature (T), dew point (DP), and wind speed (WS). This combination achieved an RMSE of 2.0914 m/s and a MAPE of 20.9884%, as detailed in Table 5 and Figure 4. The peak forecast discrepancy reached 8.0252 m/s. This spike coincided with the landfall of Typhoon Dujuan at hour 41. The error aligns with the storm's peak intensity. Meteorological conditions at this juncture diverged sharply from the broader training distribution. Such volatility underscores the inherent difficulty of resolving abrupt gust surges within a deterministic framework. The RFECV procedure independently corroborated this four-variable limit (Figure 5). Consequently, a subset comprising sea-level pressure, wind speed, wind direction, and precipitation duration were finalized. Precision remains the priority. Table 6 delineates the definitive hierarchy of these predictors, ranking the nine candidate features by their importance within the RFECV wrapper.

The Civil Aeronautics Administration [CAA] (2021) mandates formal reporting for any predicted gust fluctuation exceeding 5 m/s. This threshold is critical for flight safety. Consider a 28.4 m/s gust – magnitude 10 on the Beaufort scale. At this intensity, a forecast with 24.58% MAPE produces an absolute error of 6.98 m/s. This discrepancy eclipses the regulatory limit. Such inaccuracies jeopardize pilot judgment and stall the execution of essential safety protocols. Conversely, precise gust data empower aviation

**Table 4.** The Pearson product-moment correlation coefficients matrix showing the linear relationship between features

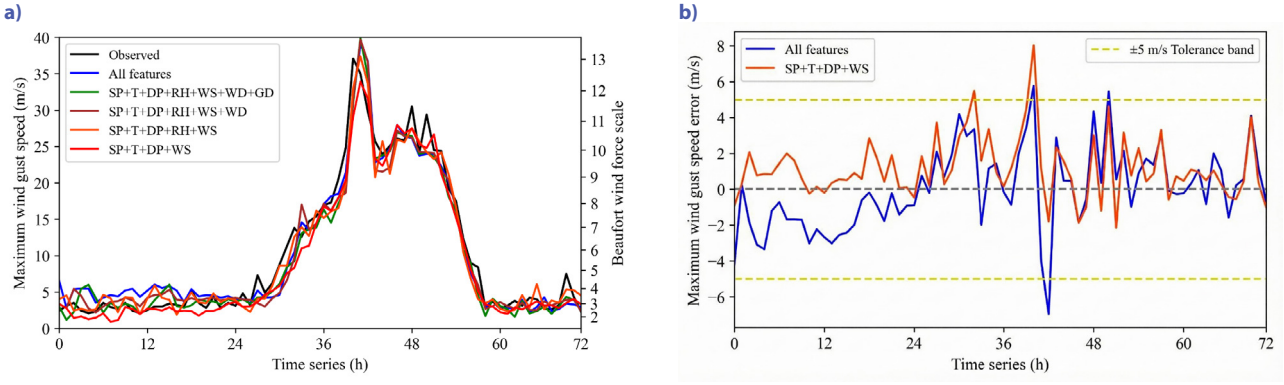
| Feature | SP | T | DP | RH | WS | WD | GS | GD | P | PD |
|---|---|---|---|---|---|---|---|---|---|---|
| SP | 1.00 | 0.23 | –0.13 | –0.30 | –0.55 | 0.02 | –0.65 | 0.05 | –0.46 | –0.56 |
| T | 0.23 | 1.00 | 0.19 | –0.82 | 0.11 | –0.18 | 0.03 | –0.24 | –0.35 | –0.56 |
| DP | –0.13 | 0.19 | 1.00 | 0.40 | 0.11 | –0.07 | 0.13 | –0.05 | 0.06 | 0.01 |
| RH | –0.30 | –0.82 | 0.40 | 1.00 | –0.02 | 0.13 | 0.06 | 0.20 | 0.39 | 0.54 |
| WS | –0.55 | 0.11 | 0.11 | –0.02 | 1.00 | –0.32 | 0.93 | –0.38 | 0.35 | 0.26 |
| WD | 0.02 | –0.18 | –0.07 | 0.13 | –0.32 | 1.00 | –0.27 | 0.65 | 0.05 | 0.13 |
| GS | –0.65 | 0.03 | 0.13 | 0.06 | 0.93 | –0.27 | 1.00 | –0.33 | 0.42 | 0.35 |
| GD | 0.05 | –0.24 | –0.05 | 0.20 | –0.38 | 0.65 | –0.33 | 1.00 | 0.04 | 0.16 |
| P | –0.46 | –0.35 | 0.06 | 0.39 | 0.35 | 0.05 | 0.42 | 0.04 | 1.00 | 0.45 |
| PD | –0.56 | –0.56 | 0.01 | 0.54 | 0.26 | 0.13 | 0.35 | 0.16 | 0.45 | 1.00 |

*Note*: *SP – sea pressure, T – temperature, DP – dew point temperature, RH – relative humidity, WS – wind speed, WD – wind direction, GS – wind gust speed, GD – wind gust direction, P – precipitation, PD – precipitation duration.

**Table 5.** RMSE and MAPE of maximum wind gust speed forecasting using different feature combinations

| Data Description | Training Feature combinations | Deleted Feature(s) | RMSE (m/s) | MAPE(%) |
|---|---|---|---|---|
| Raw data | All | -- | 9.8311 | 143.0817 |
| Preprocessing | All | -- | 2.1244 | 24.5849 |
| | SP+T+DP+RH+WS+GD+PD | WD, P | 2.8392 | 29.3101 |
| | SP+T+DP+RH+WS+GD+P | WD, PD | 2.6450 | 26.9682 |
| | SP+T+DP+RH+WS+WD+PD | GD, P | 2.8508 | 24.0557 |
| | SP+T+DP+RH+WS+WD+P | GD, PD | 2.6861 | 25.9122 |
| | SP+T+DP+RH+WS+P+PD | WD, GD | 2.6350 | 27.8233 |
| | SP+T+DP+RH+WS+WD+GD | P, PD | 2.2038 | 24.8399 |
| | SP+T+DP+RH+WS+GD | WD, P, PD | 2.5761 | 26.6489 |
| | SP+T+DP+RH+WS+WD | GD, P, PD | 2.5025 | 24.4957 |
| | SP+T+DP+RH+WS+P | WD, GD, PD | 2.5132 | 24.3585 |
| | SP+T+DP+RH+WS+PD | WD, GD, P | 2.1938 | 22.9731 |
| | SP+T+WS+WD+GD | RH, DP, P, PD | 2.1267 | 23.9753 |
| | SP+T+DP+WS+GD | RH, WD, P, PD | 2.2864 | 22.8674 |
| | SP+T+DP+RH+WS | WD, GD, P, PD | 2.1710 | 23.2207 |
| | SP+T+DP+WS | RH, WD, GD, P, PD | 2.0914 | 20.9984 |
| | SP+DP+RH+WS | T, WD, GD, P, PD | 2.2243 | 22.5426 |
| | T+DP+RH+WS | SP, WD, GD, P, PD | 2.5765 | 22.7458 |
| | DP+RH+WS | SP, T, WD, GD, P, PD | 2.6954 | 23.7854 |

*Note:* *SP – sea pressure, T – temperature, DP – dew point temperature, RH – relative humidity, WS – wind speed, WD – wind direction, GS – wind gust speed, GD – wind gust direction, P – precipitation, PD – precipitation duration.




*Note:* Panel (a) shows the time series for the best feature combination (SP+T+DP+WS), and Panel (b) compares the forecast errors. The horizontal (yellow) dashed lines at ±5 m/s indicate the operational tolerance threshold used by the Civil Aeronautics Administration (CAA); errors within this band are considered acceptable for runway operations.

**Figure 4.** Wind gust forecasts for Typhoon Dujuan: a – time series; b – forecast errors

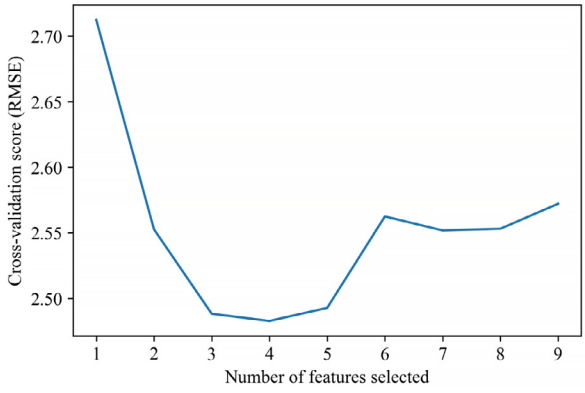



**Figure 5.** The result of the cross-validation scores at each step

**Table 6.** Comparison of feature rankings by different selection methods

| Features | Pearson Product-Moment Rank | Recursive Feature Elimination with Cross-Validation Rank | Random-forest Rank |
|---|---|---|---|
| Sea level pressure (SP) | 2 | 2 | 2 |
| Temperature (T) | 9 | 6 | 7 |
| Dew point (DP) | 7 | 7 | 9 |
| Relative humidity (RH) | 8 | 8 | 8 |
| Wind speed (WS) | 1 | 1 | 1 |
| Wind direction (WD) | 6 | 4 | 6 |
| Wind gust direction (GD) | 5 | 9 | 4 |
| Precipitation (P) | 3 | 5 | 3 |
| Precipitation duration (PD) | 4 | 3 | 5 |

personnel to recalibrate flight plans and optimize runway throughput. Accuracy directly translates to enhanced passenger experience and operational efficiency. The proposed LSTM framework consistently maintains errors within the ±5 m/s margin. Reliability is its core advantage. By adhering to these strict advisory tolerances, the model demonstrates genuine potential as a decision-support tool in high-stakes aviation environments.

### 4.1. Cross-validation performance across 12 typhoons

The leave-one-typhoon-out scheme (Section 3.6) evaluates the framework by training on 11 events and testing on the remaining storm, iterating until all 12 typhoons have served as the validation set. This rigorous partitioning tests how well the LSTM generalizes across diverse trajectories, intensities, and background environments. It also enforces a strict separation between training and testing, ensuring each evaluation period remains entirely unseen. Table 7 summarizes the error distributions for these held-out cases. The optimized configuration achieved a mean RMSE of 2.33 m/s and a mean MAPE of 21.12%. Variability remains tight, with standard deviations of 0.63 m/s and 4.79%, suggesting the model maintains stable skill across heterogeneous landfall scenarios rather than relying on a few favorable cases.

These metrics must be viewed alongside the 1-h persistence baseline (Section 3.5), which yielded a mean RMSE of 2.28 m/s and a mean MAPE of 20.66%. Such a competitive baseline underscores the difficulty of short-range gust prediction at RCTP. While both the LSTM and persistence models exhibit moderate errors relative to Civil Aeronautics Administration (CAA) crosswind advisory thresholds, the persistence results provide a transparent performance floor. Benchmarking against this baseline validates whether the added complexity of deep learning translates into tangible operational advantages during the most critical phases of a storm.

**Table 7.** The resulting distributions of RMSE and MAPE across the 12 held-out typhoons (Cross-validation results)

| Typhoon | Year | RMSE (m/s) | MAPE (%) |
|---|---|---|---|
| Namtheun | 2010 | 1.4249 | 27.6978 |
| Saola | 2012 | 3.0742 | 24.8856 |
| Soulik | 2013 | 2.6180 | 23.0843 |
| Trami | 2013 | 2.9221 | 26.6141 |
| Kong-Rey | 2013 | 1.6112 | 18.8294 |
| Matmo | 2014 | 2.5513 | 14.7786 |
| Fung-Wong | 2014 | 1.7605 | 14.8805 |
| Soudelor | 2015 | 2.8461 | 21.0383 |
| Dujuan | 2015 | 1.5559 | 14.0684 |
| Megi | 2016 | 2.5709 | 19.1536 |
| Haitang | 2017 | 3.0714 | 22.9045 |
| Bailu | 2019 | 1.9267 | 25.5021 |
| Mean (Std Dev) | – | 2.33 (0.63) | 21.12 (4.79) |

### 4.2. Effect of feature-selection strategy

Evaluation focused on four LSTM configurations: a full-predictor baseline and three feature-selected variants (Pearson, RFECV-SVR, and RF importance). Streamlining the input space did more than prune noise; it stabilized performance. Although all selection schemes matched or slightly exceeded the full-model accuracy, they did so with significantly fewer variables. Random-forest (RF) selection proved the most robust. By isolating a compact, physically coherent subset – anchored by wind speed, gust direction, and precipitation – the RF-based model maintained precision without redundant computational overhead.

Typhoon Dujuan (2015) exemplifies this predictive edge. As shown in Figure 6, forecast errors spiked precisely at landfall (hour 41), yet the RF-selected LSTM showed remarkable resilience. Panel (a) shows the time series of observed gusts versus LSTM forecasts using features selected by Pearson Correlation, RFECV, and Random-forest. Panel (b) displays the forecast errors. The Random-forest method (red line) consistently maintains errors within the ±5 m/s operational tolerance band (yellow dashed line) during the rapid intensification phase. While the Pearson and RFECV variants struggled with peak errors of 6.87 m/s and 6.51 m/s, the RF model capped the error at 4.32 m/s.

The contrast with the persistence baseline is even starker. For Dujuan, the RF-LSTM achieved an RMSE of 1.56 m/s and a MAPE of 14.1%, nearly halving the persistence model's error (RMSE 2.82 m/s; MAPE 25.9%). This 45% reduction in error validates a core premise of this study: parsimonious, data-driven feature selection unlocks the predictive potential latent in local surface observations,

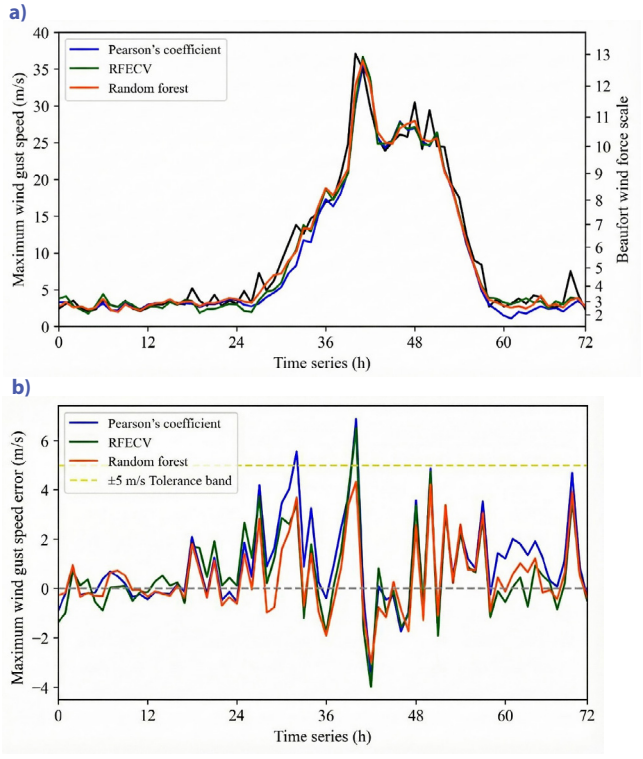



**Figure 6.** Comparison of feature-selection methods for wind gust forecasting: a – time series; b – forecast errors

providing a critical buffer during the most volatile phases of a landfalling storm.

### 4.3. Skill relative to persistence

Persistence serves as a formidable benchmark during quasi-steady phases or when typhoons remain far offshore. It falters, however, during rapid transitions, consistently lagging the amplitude and timing of peak gusts. In contrast, LSTM architectures – specifically those integrated with RF-based feature selection –demonstrate superior skill across most of the 12-event dataset. These gains sharpen during volatile windows: pre-landfall intensification and post-landfall decay. In these regimes, surface pressures and precipitation patterns shift abruptly. While persistence merely echoes the immediate past, the LSTM parses the 12-hour antecedent signal to anticipate sudden surges or drops in the wind field.

Operationally, the LSTM's ability to surpass a robust persistence baseline validates the shift toward model complexity. It delivers actionable improvements at the critical thresholds required for runway crosswind management and ground safety. At the same time, persistence remains an essential sanity check – a transparent floor for performance as these models scale to diverse airports and storm regimes.

## 5. Conclusions

This study developed and evaluated a long short-term memory (LSTM) framework for predicting the hourly maximum speed of wind gusts during the landfall of tropical cyclones at Taiwan Taoyuan International Airport (RCTP). Routine surface observations associated with 12 typhoons between 2010 and 2020 were used to examine how the combination of multivariate near-surface predictors with three feature-selection methods within a sequence model can provide 1-h-ahead gust forecasts for aviation operations.

The forecasting task was formulated as a 1-h lead-time problem to be solved using the 12-h history of several local predictors, including sea-level pressure, temperature, dew point, relative humidity, wind speed and direction, gust direction, and precipitation amount and duration. Three feature-selection strategies were compared: a filter method involving Pearson correlation, a wrapper method involving recursive feature elimination with cross validation and support vector regression, and an embedded method involving random-forest feature importance. These procedures were applied within each training fold to prevent information leakage, and the resulting feature subsets were used as inputs to the LSTM network.

A leave-one-typhoon-out cross-validation scheme was applied to assess robustness across the various storm tracks and intensities. The model was trained on 11 typhoons and tested on the remaining typhoon, with this process performed for all 12 events. For the best-performing configuration (LSTM with random-forest feature selection), the mean root-mean-square error (RMSE) across the 12 training–testing scenarios was 2.33 m/s (standard deviation: 0.63 m/s), and the mean absolute percentage error (MAPE) was 21.12% (standard deviation: 4.79%). These results indicated reasonably stable performance across a range of landfall scenarios rather than dependence on a single favorable case.

To contextualize these numbers, a 1-h persistence forecast as a simple and operationally transparent baseline was introduced. Across the 12 typhoons, the persistence forecast achieved a mean RMSE of 2.28 m/s and a MAPE of 20.66%, confirming its strength as a benchmark for short-range gust prediction at RCTP. The LSTM models performed comparably to the persistence forecast on average but demonstrated clear gains during periods of rapid change, such as the intensification and decay phases of Typhoon Dujuan (2015), for which both the RMSE and MAPE were 45% lower when the LSTM models were used. These improvements are particularly relevant near the ±5 m/s tolerance threshold commonly used for gust advisories at airports overseen by the Civil Aeronautics Administration.

From an operational perspective, the proposed framework demonstrates that feature selection can be combined with sequence-based deep learning to yield short-range gust forecasts that require only routine surface observation data and are competitive with or even superior to a benchmark persistence baseline. Such forecasts can support runway crosswind management, threshold-exceedance advisory decisions, and short-horizon ATM capacity decisions when typhoons reach land near RCTP.

Despite the advantages of the proposed approach, this work should be viewed as a case study rather than a general solution. Analysis was limited to a single airport, 12 landfalling typhoons, and a 1-h deterministic lead time, and only near-surface measurements, which cannot represent larger-scale typhoon structure or upstream conditions, were employed. In future research, this framework could be extended to multiple airports, account for additional radar or numerical weather prediction variables, produce probabilistic and multi-lead-time forecasts, and incorporate specialized methods for extreme-event learning. Given that wind and gust directions were processed as normalized scalar values in this study, scholars should investigate the use of trigonometric encodings (sine and cosine transformations) to model the circular nature of these directional variables, which could improve the forecast precision. These extensions can clarify the degree to which the study findings are transferable to other coastal airports exposed to tropical cyclone hazards.

## Acknowledgements

The authors would like to express their sincere gratitude to the Editor and the anonymous reviewers for their valuable and constructive comments. Their insightful suggestions – particularly regarding the implementation of the

leave-one-out cross-validation scheme and the inclusion of persistence benchmarks – have significantly improved the robustness and clarity of this study. We also thank P.-R. Wang for his help with data curation and analysis, and Professor S.-M. Yang for his valuable advice and guidance on the organization of the study's framework and the manuscript writing.

## Funding

This work was supported in part by the National Science and Technology Council, Taiwan, ROC under contract NSTC 111-2410-H-309-001.

## Data availability

The input weather information is collected from the database: https://opendata.cwb.gov.tw/index.

## Disclosure statement

The authors declared that they have no conflicts of interest in this work.

## AI Acknowledgements

During the preparation of this manuscript the authors used *Gemini* (*Gemini 3 Flash*, Paid Tier) for improving the grammar and readability of the manuscript. The tool was used solely to enhance the linguistic quality, ensure grammatical correctness, and improve the overall flow of the text. The authors have reviewed and edited the output to ensure accuracy and take full responsibility for the content of the manuscript.

## References

Alves, D., Mendonça, F., Mostafa, S. S., & Morgado-Dias, F. (2025). Deep learning enhanced wind speed and direction forecasting for airport regions. *Weather and Forecasting*, *40*(1), Article 207221. https://doi.org/10.1175/WAF-D-24-0069.1

Anaman, K. A., Quaye, R., & Owusu-Brown, B. (2017). Benefits of aviation weather services: A review of international literature. *Research in World Economy*, *8*(1), 45–58. https://doi.org/10.5430/rwe.v8n1p45

Breiman, L. (2001). Random forests: A machine learning approach. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Caseri, A. N., Santos, L. B. L., & Stephany, S. (2022). A convolutional recurrent neural network for strong convective rainfall nowcasting using weather radar data in Southeastern Brazil. *Artificial Intelligence in Geosciences*, *3*, 8–13. https://doi.org/10.1016/j.aiig.2022.06.001

Chen, H., & Chang, X. (2021). Photovoltaic power prediction of LSTM model based on Pearson feature selection. *Energy Reports*, *7*(7), 1047–1054. https://doi.org/10.1016/j.egyr.2021.09.167

Chen, K., Kuang, C., Wang, L., Chen, K., Han, X., & Fan, J. (2021). Storm surge prediction based on long short-term memory neural network in the East China Sea. *Applied Sciences*, *12*(1), Article 181. https://doi.org/10.3390/app12010181

Chen, R., Wang, X., Zhang, W., Zhu, X., Li, A., & Yang, C. (2019). A hybrid CNN–LSTM model for typhoon formation forecasting. *Geoinformatica*, *23*(3), 375–396. https://doi.org/10.1007/s10707-019-00355-0

Chen, X. W., & Jeong, J. C. (2007). Enhanced recursive feature elimination. In *Sixth International Conference on Machine Learning and Applications* (pp. 429–435). IEEE Xplore. https://doi.org/10.1109/ICMLA.2007.35

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic Press. https://doi.org/10.4324/9780203771587

Das, B., Nair, B., Reddy, V. K., & Venkatesh, P. (2018). Evaluation of multiple linear, neural network and penalised regression models for prediction of rice yield based on weather parameters for West Coast of India. *International Journal of Biometeorology*, *62*(10), 1809–1822. https://doi.org/10.1007/s00484-018-1583-6

Gao, S., Zhao, P., Pan, B., Li, Y., Zhou, M., Xu, J. L., Zhong, S., & Shi, Z. W. (2018). A nowcasting model for the prediction of typhoon tracks based on a long short-term memory neural network. *Acta Oceanologica Sinica*, *37*(5), 8–12. https://doi.org/10.1007/s13131-018-1219-z

Google. (2025). *Gemini 3 Flash* [Large language model]. https://gemini.google.com

Graves, A. (2012). Long short-term memory. In *Supervised sequence labelling with recurrent neural networks*. *Studies in computational intelligence* (Vol. 385, pp. 37–45). Springer. https://doi.org/10.1007/978-3-642-24797-2_4

Greenwood, P. E., & Nikulin, M. S. (1996). *A guide to Chi-Squared testing*. John Wiley & Sons.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting, 22*(4), 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001

Kartal, S., Basu, S., & Watson, S. J. (2023). A decision-tree-based measure–correlate–predict approach for peak wind gust estimation from a global reanalysis dataset. *Wind Energy Science*, *8*(10), 1533–1551. https://doi.org/10.5194/wes-8-1533-2023

Khattak, A., Chan, P.-W., Chen, F., & Almaliki, A. H. (2025). A hybrid VMD-TPE-TCN framework for wind shear prediction near an airport runway. *Atmosphere*, *16*(4), Article 381. https://doi.org/10.3390/atmos16040381

Kingma, D. P., & Ba, J. L. (2015). *ADAM: A Method for stochastic optimization*. ArXiv.

Lewis, C. D. (1982). *Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting*. Butterworth Scientific.

Li, P., Zhang, J., & Krebs, P. (2022). Prediction of flow based on a CNN–LSTM combined deep learning approach. *Water*, *14*(6), Article 993. https://doi.org/10.3390/w14060993

Li, S., Wang, P., & Goel, L. (2015). Wind power forecasting using neural network ensembles with feature selection. *IEEE Transactions on Sustainable Energy*, *6*(4), 1447–1456. https://doi.org/10.1109/TSTE.2015.2441747

Li, W., Kiaghadi, A., & Dawson, C. (2021). Exploring the best sequence LSTM modeling architecture for flood prediction. *Neural Computing and Applications*, *33*(11), 5571–5580. https://doi.org/10.1007/s00521-020-05334-3

Lv, S.-X., & Wang, L. (2023). Multivariate wind speed forecasting based on multi-objective feature selection approach and hybrid deep learning model. *Energy*, *263*, Article 126100. https://doi.org/10.1016/j.energy.2022.126100

Mao, K. Z. (2004). Orthogonal forward selection and backward elimination algorithms for feature subset selection. *IEEE Transactions on Systems, Man, and Cybernetics*, *34*(1), 629–634. https://doi.org/10.1109/TSMCB.2002.804363

Modé, P., Demartino, C., Georgakis, C. T., & Lagaros, N. D. (2025). Short-term extreme wind speed forecasting using dual-output LSTM-based regression and classification model. *Journal of Wind Engineering and Industrial Aerodynamics*, *259*, Article 106035. https://doi.org/10.1016/j.jweia.2025.106035

Pang, Z., Niu, F., & O'Neill, Z. (2020). Solar radiation prediction using recurrent neural network and artificial neural network: A case study with comparisons. *Renewable Energy*, *156*, 279–289. https://doi.org/10.1016/j.renene.2020.04.042

Pearson, K. (1895). Note on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London*, *58*(1895), 240–242. https://doi.org/10.1098/rspl.1895.0041

Ren, J., Xu, N., & Cui, Y. (2022). Typhoon track prediction based on deep learning. *Applied Sciences*, *12*(16), Article 8028. https://doi.org/10.3390/app12168028

Salman, R., Alzaatreh, A., & Sulieman, H. (2026). Stability analysis using Spearman's rank correlation coefficient. *Journal of Big Data*.

Saputro, D. R. S., Wahyu, N. L., & Widyaningsih, Y. (2025). Performance of Ridge Regression, Least absolute shrinkage and selection operator, and elastic net in overcoming multicollinearity. *Journal of Multidisciplinary Applied Natural Science, 5*(2), 370–382. https://doi.org/10.47352/jmans.2774-3047.251

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, *14*(3), 199–222. https://doi.org/10.1023/B:STCO.0000035301.49549.88

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, *15*(1), 72–101. https://doi.org/10.2307/1412159

Su, C., Sahoo, B., Mao, M., & Xia, M. (2025). Machine learning techniques for predicting typhoon-induced storm surge using a hybrid wind field. *Journal of Geophysical Research: Machine Learning and Computation*, *2*(2), Article e2024JH000507. https://doi.org/10.1029/2024JH000507

Tao, H., Awadh, S. M., Salih, S. Q., Shafik, S. S., & Yaseen, Z. M. (2022). Integration of extreme gradient boosting feature selection approach with machine learning models: Application of weather relative humidity prediction. *Neural Computing and Applications*, *34*, 515–533. https://doi.org/10.1007/s00521-021-06362-3

The Civil Aeronautics Administration. (2021). *Aviation meteorological regulations* (110-TS-3(8)). CAA.

Ummenhofer, C. C., & Meehl, G. A. (2017). Extreme weather and climate events with ecological relevance: A review. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1723), Article 20160135. https://doi.org/10.1098/rstb.2016.0135

Wei, C. C. (2021). Collapse warning system using LSTM neural networks for construction disaster prevention in extreme wind weather. *Journal of Civil Engineering and Management*, *27*(4), 230–245. https://doi.org/10.3846/jcem.2021.14649

Wolff, B., Kramer, O., & Heinemann, D. (2017). Selection of numerical weather forecast features for PV power predictions with random forests. In W. Woon, Z. Aung, O. Kramer, & S. Madnick (Eds), *Data analytics for renewable energy integration* (Vol. 10097, pp. 78–91). Springer. https://doi.org/10.1007/978-3-319-50947-1_8

Yildirim, M., Çinar, A., & Cengil, E. (2022). Classification of the weather images with the proposed hybrid model using deep learning, SVM classifier, and mRMR feature selection methods. *Geocarto International*, *37*(9), 2735–2745. https://doi.org/10.1080/10106049.2022.2034989

Yin, Y., Yang, P., Zhang, Y., Chen, H., & Fu, S. (2020). Feature selection and processing of turbulence modeling based on an artificial neural network. *Physics of Fluids*, *32*(10), Article 105117. https://doi.org/10.1063/5.0022561

Yuan, S., Wang, C., Mu, B., Zhou, F., & Duan, W. (2021). Typhoon intensity forecasting based on LSTM using the rolling forecast method. *Algorithms*, *14*(3), Article 83. https://doi.org/10.3390/a14030083

Zhao, C., van Heeswijk, M., & Karhunen, J. (2016). Air quality forecasting using neural networks. In *IEEE Symposium Series on Computational Intelligence* (pp. 1–7). IEEE. https://doi.org/10.1109/SSCI.2016.7850128

Zhou, G., Xu, J., Qian, Q., Xu, Y., & Xu, Y. (2022). Discriminating technique of typhoon rapid intensification trend based on artificial intelligence. *Atmosphere*, *13*(3), Article 448. https://doi.org/10.3390/atmos13030448