

Research Article

Development of Stacked Long Short-Term Memory Neural Networks with Numerical Solutions for Wind Velocity Predictions

Chih-Chiang Wei 

Department of Marine Environmental Informatics and Center of Excellence for Ocean Engineering,
National Taiwan Ocean University, Keelung City, Taiwan

Correspondence should be addressed to Chih-Chiang Wei; ccwei@ntou.edu.tw

Received 27 December 2019; Accepted 8 July 2020; Published 23 July 2020

Academic Editor: Theodore Karacostas

Copyright © 2020 Chih-Chiang Wei. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Taiwan, being located on a path in the west Pacific Ocean where typhoons often strike, is often affected by typhoons. The accompanying strong winds and torrential rains make typhoons particularly damaging in Taiwan. Therefore, we aimed to establish an accurate wind speed prediction model for future typhoons, allowing for better preparation to mitigate a typhoon's toll on life and property. For more accurate wind speed predictions during a typhoon episode, we used cutting-edge machine learning techniques to construct a wind speed prediction model. To ensure model accuracy, we used, as variable input, simulated values from the Weather Research and Forecasting model of the numerical weather prediction system in addition to adopting deeper neural networks that can deepen neural network structures in the construction of estimation models. Our deeper neural networks comprise multilayer perceptron (MLP), deep recurrent neural networks (DRNNs), and stacked long short-term memory (LSTM). These three model-structure types differ by their memory capacity: MLPs are model networks with no memory capacity, whereas DRNNs and stacked LSTM are model networks with memory capacity. A model structure with memory capacity can analyze time-series data and continue memorizing and learning along the time axis. The study area is northeastern Taiwan. Results showed that MLP, DRNN, and stacked LSTM prediction error rates increased with prediction time (1–6 hours). Comparing the three models revealed that model networks with memory capacity (DRNN and stacked LSTM) were more accurate than those without memory capacity. A further comparison of model networks with memory capacity revealed that stacked LSTM yielded slightly more accurate results than did DRNN. Additionally, we determined that in the construction of the wind speed prediction model, the use of numerically simulated values reduced the error rate approximately by 30%. These results indicate that the inclusion of numerically simulated values in wind speed prediction models enhanced their prediction accuracy.

1. Introduction

A typhoon is a severe natural disaster that affects tropical and subtropical coastal countries, and it occurs most frequently in the northwestern Pacific Ocean. Taiwan is to the east of the Eurasian Continent and at the western side of the Pacific Ocean; its climate is intermediate between tropical and subtropical climate. Thus, typhoons frequently occur in Taiwan, generally in summer and fall. Typhoons affecting Taiwan typically develop at the sea surface southeast of Taiwan, and most typhoons are accompanied by torrential rains and strong winds [1]. Such rain and wind add to the damage from typhoons, posing a great threat to the transportation, economic, agricultural, and fishery activities in

Taiwan even if the typhoon itself does not hit Taiwan. Therefore, typhoons constitute a serious natural disaster in Taiwan.

For example, the 2015 Typhoon Soudelor was the most destructive typhoon that occurred in Taiwan in recent history, with gust intensity exceeding 12 on the Beaufort wind force scale (32.7 m/s). Its strong gusts caused widespread damage to infrastructure, affecting gas supply, power and utilities, transportation and communication, and weather radar stations. Electricity was cut off in approximately 4.5 million households simultaneously during Typhoon Soudelor—the greatest recorded number in recent history. The economic loss from the typhoon was estimated to be as high as US\$76 million [2].

Therefore, we aimed to establish an accurate wind speed prediction model for future typhoons, allowing for better preparation to mitigate a typhoon's toll on life and property. In this study, cutting-edge machine learning (ML) techniques were used to improve predictive accuracy. In general, ML algorithms learn from a huge dataset, improving their ability to identify patterns in the data. Specifically, ML involves creating algorithms to make prediction from sets of unknown data. Given their ability to perform parameter adjustment and achieve optimization through self-learning, neural network algorithms in ML are particularly powerful [3]. Such algorithms have been extensively used in recent wind speed prediction models, and these models are increasingly data-driven due to developments in ML [4–11]. In the development of neural networks, multilayer perceptron (MLP) networks are a classic approach that are often used and compared with other neural network models. For example, Wei [12] compared the accuracy of MLP with that of adaptive network-based fuzzy-inference-system neural networks in the construction of typhoon wind speed prediction models.

Deep learning has become possible due to the exponential increase in computing power in recent years. This approach is the further derivation of multiple neural layers from the original neural layers of a model. Such derivation improves an algorithm's ability to learn, better approximating the complex neural network structure of a human being. For example, Hu et al. [13] formulated multilayer deep neural networks that were trained using data from data-rich wind plants. These networks extracted wind speed patterns, and the mapping was finely tuned using data from newly constructed wind plants. Tiancheng et al. [14] proposed a sandstorm prediction method that considered both the effect of atmospheric movement and ground factors on sandstorm occurrence, called improved naive Bayesian convolutional neural network classification algorithm.

In the field of neural networks, recurrent neural networks (RNNs), which can analyze sequential (or time-series) data, have recently been developed [15–20]. RNNs are connectionist models with the ability to selectively pass information across sequence steps while processing sequential data one element at a time [21]. Therefore, RNNs are important, especially in the analysis of sequential data. A particular type of RNNs is long short-term memory (LSTM), a class of artificial neural networks where connections between units form a directed cycle [22] introduced the LSTM primarily to overcome the problem of diminishing gradients. LSTM creates an internal network state that allows it to exhibit dynamic temporal behavior. Unlike feedforward neural networks, RNNs can use their internal memory to process arbitrary sequences of input [23]. To model the time series of wind speed data, Byeon et al. [24] developed the LSTM for the prediction of typhoon wind speeds. Hence, feature enhancement from RNNs has been explored in wind prediction.

The widespread application of ensembles in numerical weather prediction (NWP) has helped researchers improve weather forecasts [25–30]. Numerical models can be used to calculate all climate parameters through atmospheric dynamics and numerical methods, allowing for simulated

results to be explained in terms of physical relationships. NWP models simulate the atmosphere on user-defined grid scales as a moving fluid. Through several types of parameterizations, NWP models also account for the influence of subgrid physical processes on grid-resolved motions [31, 32]. NWP models, such as the Weather Research and Forecasting model (WRF), have been increasingly popular as a low-cost alternative source of data for such assessments of climate parameters [33]. WRF offers a wide variety of physical and dynamical elements to choose from; these elements must be put together to form model configurations, with which the model can be run [34]. However, because of imperfect models and uncertain initial boundary atmospheric conditions, errors exist in the NWP output [35, 36].

Recent studies have applied the NWP model to typhoons and tropical cyclones [37–39]. However, the prediction of severe meteorological phenomena (such as typhoons), which result from a multiplicity of mutually interacting multiscale processes, remains a major challenge for NWP systems [25, 40, 41]. These models are typically unable to predict wind intensity with satisfactory accuracy, even at short forecast times and a high horizontal resolution [42, 43]. Some studies have tried combining NWP with ML models to develop an integrated climate prediction model. For example, Zhao et al. [44] evaluated the performance and enhanced accuracy of a day-ahead wind-power forecasting system. The system comprised artificial neural networks and an NWP model.

Furthermore, to enhance the predictive accuracy of constructed ML models, other researchers have used numerically simulated results as input data for the construction of ML models. For example, Zhao et al. [45] developed the ARIMAX model, where wind speed results from the WRF simulation were chosen as an exogenous input variable. However, to the best of our knowledge, in the literature on short-term wind speed prediction for typhoons (or tropical cyclones), numerically simulated values have seldom been used as an input variable for ML prediction models. Therefore, in relation to the construction of a ML-based wind speed prediction model, we evaluated the improvements to predictive accuracy afforded by the use of numerically simulated values (by comparing between its use and nonuse).

Therefore, our study has two primary aims: (1) develop an ML- and neural network-based wind speed prediction model and compare the predictive accuracy of various neural network-based algorithms and (2) evaluate the improvements to predictive accuracy afforded by the use of numerical solutions obtained from NWP models (by comparing between its use and nonuse) in a typhoon-surface wind-speed prediction model.

2. Methodology and Algorithms

Figure 1 illustrates the flow of the construction, involving NWP numerical solutions, of our neural network-based typhoon wind-velocity prediction model. In the first stage, we collected data on the typhoon characteristics and ground wind speed of the research area's historical typhoon events.

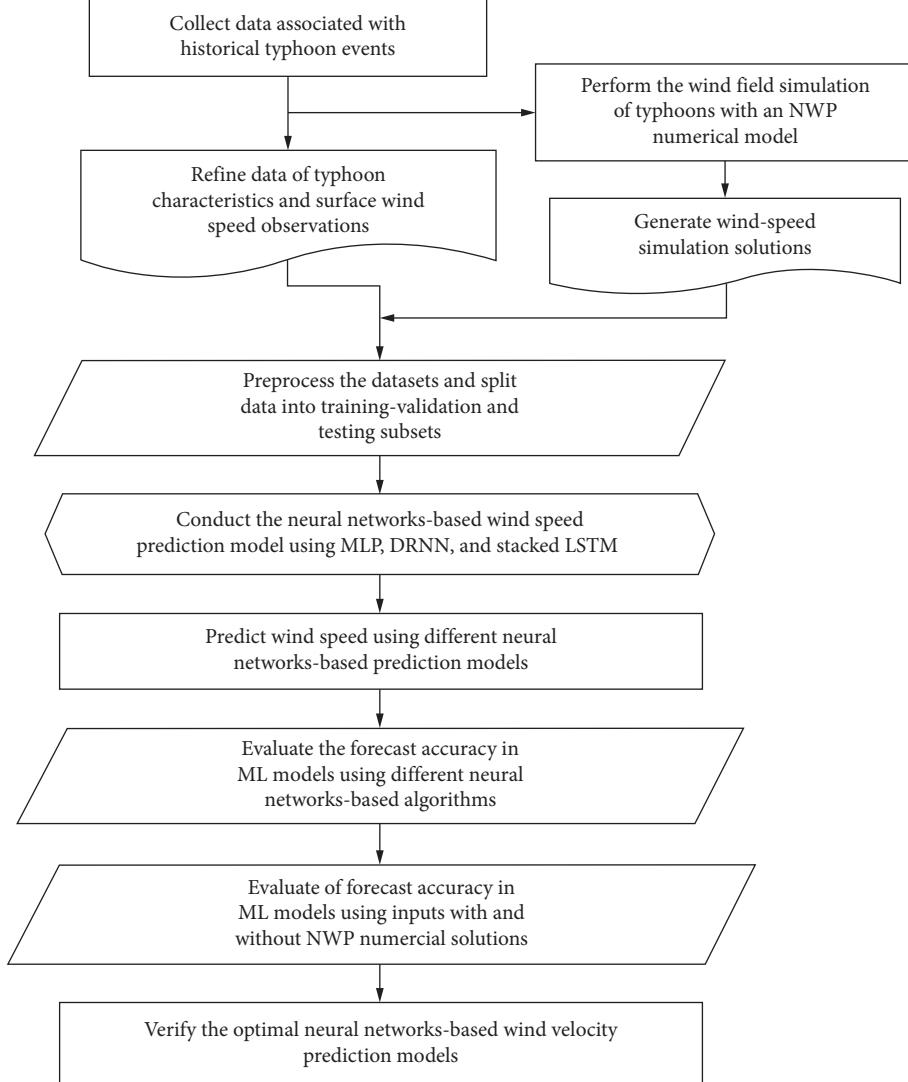


FIGURE 1: Flowchart of neural network-based typhoon wind velocity prediction model using NWP numerical solutions.

In the second step, wind speed solutions were obtained from a wind field simulation of typhoons, involving an NWP numerical model. In our study, we also employed a WRF numerical model to simulate circulation distribution, thus obtaining the wind speed values in the research area. Two datasets could be built, one comprising typhoon data and the other comprising wind simulation results from the NWP model. The datasets were split into testing and training-validation subsets. The training-validation set was used for the learning of several ML-based wind velocity prediction models, and the testing set was used for the identification of the optimal prediction model among these models.

The best model among a set of ML neural network-based models, involving MLP, DRNN, and stacked LSTM, was determined. According to [46], real-time dynamics constitute the most challenging aspect of wind speed forecasting. We determined the stacked LSTM model to be the best for wind speed forecasting due to its appropriate handling of long- and short-term time dependency. In the final stage, the

forecast accuracy of the stacked LSTM model was evaluated against other neural networks. We also evaluated whether forecast accuracy in ML-based models improved when NWP numerical solutions were used as input.

2.1. Frameworks Underlying the Proposed Neural Networks. In this section, we describe the neural network-based architectures, which used the MLP, DRNN, and stacked LSTM algorithms, that were adopted for model construction. As illustrated in Figure 2(a), the MLP is a typical type of feedforward backpropagation neural network that uses processing units placed in the input, hidden, and output layers [47–49]. Each unit (with an associated weight) in a layer is connected to the units in adjacent layers [50, 51]. In the study, to enhance learning efficacy—and by implication, approximation, and prediction accuracy—we added hidden layers to a simple MLP neural network; the MLP was trained through backpropagation. Generally, the weight updates

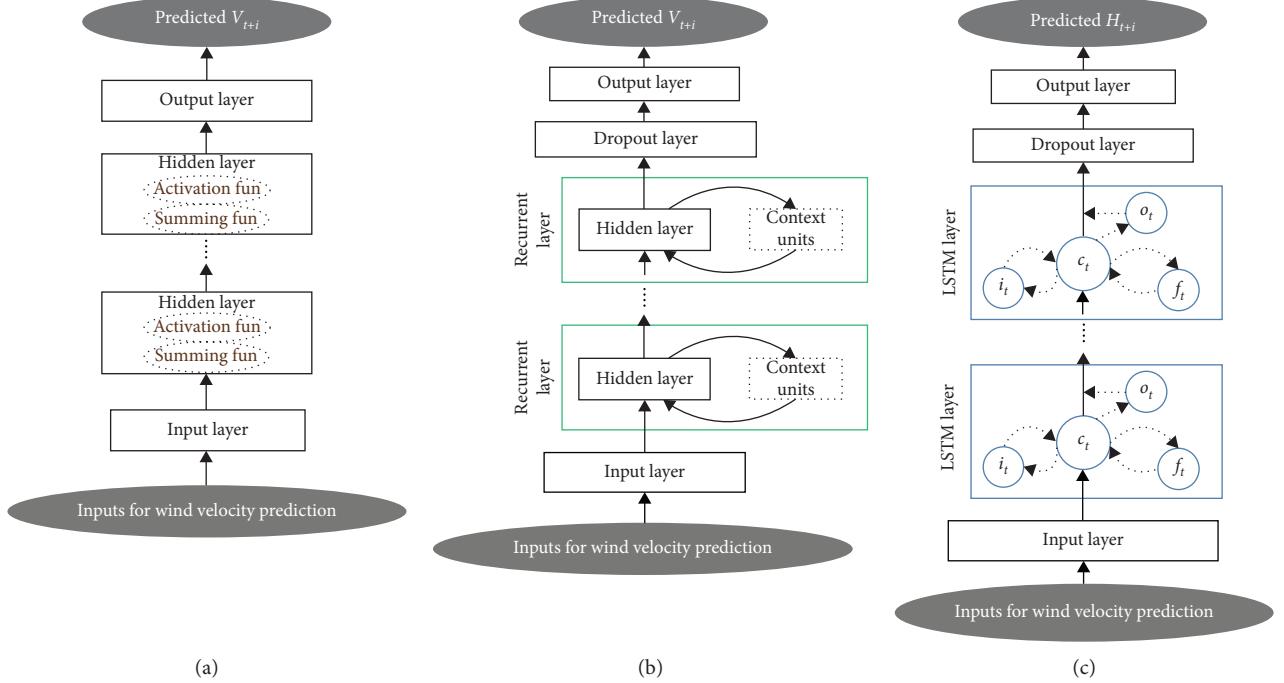


FIGURE 2: Architecture of (a) MLP, (b) DRNN, and (c) stacked LSTM neural networks.

between layers are calculated in terms of the stochastic gradient descent [52]. Specifically,

$$\Delta w_{ij}(t+1) = \beta \Delta w_{ij}(t) + \eta \frac{\partial E}{\partial w_{ij}}, \quad (1)$$

where $w_{ij}(t)$ is the weight set connecting layers i and j at time t , Δw is the weight correction, η is the learning rate, β is a momentum coefficient, and E is a cost function that indicates the difference between the target and predicted values. In particular, η and β are hyperparameters for adjusting the spacing of weight correction.

As illustrated in Figure 2(b), the multilayers of the RNN structure comprise an input layer, multiple recurrent layers, and an output layer. When the length of recurrent layer is 1, the framework is a simple RNN. Here, the recurrent network is based on the networks developed by [53]. In the RNN, the hidden units are connected to context units; in the successive time step, the units feed back into the hidden units. The hidden state at any time step can contain information from an (almost) arbitrarily long context window [21]. The DRNN model framework has multiple recurrent layers before the forwarding to a dropout layer and output layer at the final output. In this paper, the dropout layer excludes 10% of the neurons to avoid overfitting.

A stacked LSTM architecture is defined as an LSTM model that comprises multiple LSTM layers (Figure 2(c)). The stacked LSTM, also known as deep LSTM, was first formulated by [54] and was applied to speech recognition problems. Similar to the framework underpinning the DRNN model, the stacked LSTM model uses multiple LSTM layers that are stacked before the forwarding to a dropout layer and output layer at the final output. In a stacked LSTM,

the first LSTM layer produces sequence vectors used as the input of the subsequent LSTM layer. Moreover, the LSTM layer receives feedback from its previous time step, thus allowing for the capturing of data patterns. The dropout layer also excludes 10% of the neurons to avoid overfitting.

The basic structure of LSTM, as illustrated in the LSTM layer in Figure 2(c), comprises an input gate i_t , output gate o_t , forget gate f_t , and memory cell c_t . A single LSTM layer has a second-order RNN architecture that excels at storing sequential short-term memories and retrieving them at many time steps later [55]. An LSTM network is identical to a standard RNN, with the exception of the summation units in the hidden layer being replaced by memory blocks [56]. Equations (2)–(6) describe how output values are updated at each step [22, 46, 57]. Specifically,

$$f_t = \sigma(W_f \cdot x_t + U_f \cdot h_{t-1} + b_f), \quad (2)$$

$$i_t = \sigma(W_i \cdot x_t + U_i \cdot h_{t-1} + b_i), \quad (3)$$

$$o_t = \sigma(W_o \cdot x_t + U_o \cdot h_{t-1} + b_o), \quad (4)$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \sigma(W_c \cdot x_t + U_c \cdot h_{t-1} + b_c), \quad (5)$$

$$h_t = o_t \cdot \sigma(c_t), \quad (6)$$

where x_t is the input vector; σ is the activation function; W_f , W_i , W_o , U_f , U_i , U_o , and U_c are the weight vector terms; b_f , b_i , b_o , and b_c are the corresponding bias terms; and h_t and h_{t-1} are the current and previous hidden vectors, respectively.

In this study, because the function converges quickly and has no problems with a vanishing gradient, we used the

ReLU activation function in the middle layers (the hidden, recurrent, and LSTM layers) of the aforementioned three models. The ReLU is defined as the positive part of its argument; that is, $f(x) = \max(0, x)$, where x is the input to a neuron. The ReLU function has been recently used to replace the sigmoid function in neural networks, resulting in good performance and fast training times [58]. Moreover, in these proposed models, the input layers receive the observed and simulated meteorological values as their input.

3. Study Area and Data

Taiwan is located at the western Pacific Ocean, a region frequently in the path of typhoons. Typhoons in Taiwan are often accompanied by torrential rains, which severely endanger life and property. Thus, more accurate regional wind speed prediction is needed to improve typhoon protection. In this study, we chose northeastern Taiwan as the study area (Figure 3), wherein Taipei City and Yilan City are the two major cities.

3.1. Data from Gauge and Typhoon Weather Stations. According to typhoon data from the Central Weather Bureau (CWB), between 2000 and 2018, 29 typhoons have directly struck Taiwan (Table 1; 2003, 2011, and 2018 were excluded from the table because typhoons did not directly affect Taiwan in those years). The tracks of these historical typhoons are illustrated in Figure 4. These typhoons can be classified according to the Saffir–Simpson wind scale depending on the intensity of their maximum sustained winds. A tropical storm has the wind speed range of 18–33 m/s, and category 1, 2, and 3 typhoons have the wind speed ranges of 33–43 m/s, 43–50 m/s, and 50–58 m/s, respectively. In these collected typhoons, the numbers for the category 1, 2, and 3 typhoons were 9, 5, and 8, respectively, and the number for the tropical storm was 7.

The data included two types of datasets: one on typhoon dynamic characteristics and another (comprising 2621 hourly records) on surface wind speed observations. Surface wind speed data were released by the CWB and measured from eight gauge stations: Anbu, Banqiao, Keelung, Pengjiayu, Su-ao, Taipei, Tamsui, and Yilan (Figure 3). In addition, data on typhoon dynamics were released by the CWB. The data comprised six variables: pressure at the typhoon center, latitude of the typhoon center, longitude of the typhoon center, typhoon radius (i.e., the distance from typhoon center for points with winds greater than 15.5 m/s), moving speed of the typhoon, and maximum wind speed at the typhoon center.

3.2. Data from Numerical Solutions. WRF model simulations using typhoon data were used to generate wind speed values at each meteorological station. To set the initial field and boundary conditions, we used data from the Final Operational Global Analysis dataset. The dataset is a part of the Global Data Assimilation System of the US Government's National Center for Environmental Protection. The grid was set to be a two-domain nested grid. The horizontal

grid spacing for the coarser domain (118°E – 126°E , 21°N – 29°N) was 15 km, and that of the finer domain (121°E – 123°E , 24°N – 26°N) was 3 km. Both domains had 32 vertical levels. This study used the WRF physical parameters recommended by [59, 60]. They studied the tracks and rainfall of typhoons that have invaded Taiwan in addition to other physical parameters, and they conducted wind forecasts using the WRF model. They determined the following physical parameters to be suitable for wind speed forecasts: with regard to the planetary layer, those from the Yonsei University (YSU) scheme; with regard to micro-physics, those from the WRF Single-Moment 5-Class (WSM5) scheme; for cumulus parameterization, those from the Kain–Fritsch scheme; and with regard to long-wave radiation, those from the Rapid Radiative Transfer Model scheme.

After our simulation of a large number of typhoon events, the WRF model generated wind velocity simulations at eight ground gauge stations. In the verification of subsequent WRF wind outcomes, the output objective for the WRF model was the simulated wind value at an altitude of 10 m, because the ground gauge-station wind speed reached the observation value at an altitude of 10 m. Figure 5 displays the scatter plot of the WRF model simulation values and observed gauge-station values. According to Figure 5, several observation stations (Keelung, Anbu, Yilan, and Su-ao) have consistently higher wind speeds. These higher wind speeds are caused by their surrounding topographies—which cause them to face windwards—in conjunction with the anticlockwise circulatory flow characteristic of typhoons. Pengjiayu Station is similar to the four stations. However, being located on an island on the sea around northeastern Taiwan, the station is directly affected only by a typhoon's circulatory flow and not the station's surrounding topography; this station also records higher wind speed values. The Taipei, Banqiao, and Tamsui Stations are located within the Tamsui Basin and are thus fully enclosed by mountains. Because the circulatory flow of typhoons can be easily disrupted by the surrounding topography, the wind speed values observed at these stations are slightly lower.

For the four windward observation stations, wind speed simulation values were lower and approximately equal to actual values for high and low wind speeds, respectively. For the three observation stations in the Tamsui Basin, wind speed simulation values were approximately equal to the actual values, for all wind speeds.

To evaluate the simulation outcomes, this study used the mean absolute error (MAE) and root mean squared error (RMSE) that are defined as follows:

$$\begin{aligned} \text{MAE} &= \frac{1}{n} \sum_{i=1}^n |O_i - Y_i|, \\ \text{RMSE} &= \sqrt{\frac{1}{n} \sum_{i=1}^n (O_i - Y_i)^2}, \end{aligned} \quad (7)$$

where Y_i is the estimated value of record i , O_i is the observation of record i , and n is the total number of records.

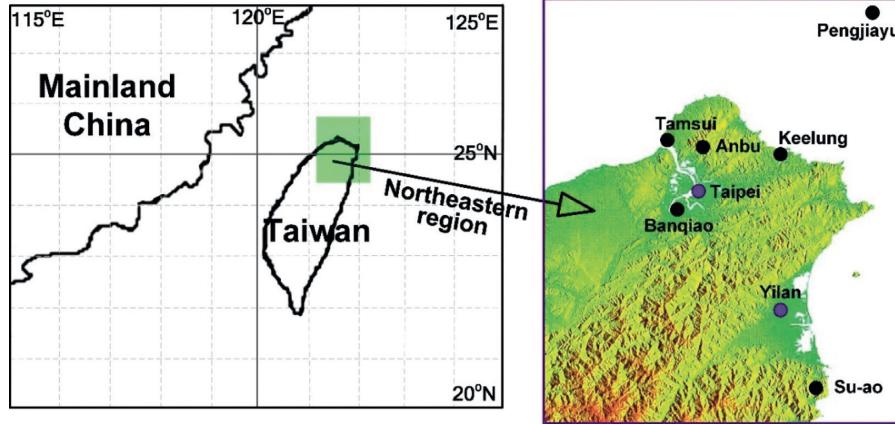


FIGURE 3: Map of Taiwan’s northeastern region and the meteorological stations therein.

TABLE 1: Typhoons affecting the study area from 2000 to 2017 (29 incidents).

Year	Typhoon	Period (UTC)
2000	Kai-Tak	6–10 Jul
2001	Toraji	28–31 Jul
2001	Nari	10–20 Sep
2002	Nakri	9–11 Jul
2004	Mindulle	29 Jun–3 Jul
2004	Aere	23–26 Aug
2004	Haima	12–13 Sep
2004	Nock-Ten	24–26 Oct
2005	Haitang	17–19 Jul
2005	Longwang	1–2 Oct
2006	Bilis	12–14 Jul
2007	Sepat	16–19 Aug
2007	Krosa	5–8 Oct
2008	Kalmaegi	16–19 Jul
2008	Fung-Wong	27–29 Jul
2008	Sinlaku	10–16 Sep
2008	Jangmi	27–30 Sep
2009	Morakot	6–9 Aug
2010	Namtheun	30–31 Aug
2012	Saola	30 Jul–3 Aug
2013	Soulik	10–14 Jul
2013	Trami	16–22 Aug
2013	Kong-Rey	27–30 Aug
2014	Matmo	21–23 Jul
2014	Fung-Wong	19–22 Sep
2015	Soudelor	6–9 Aug
2015	Dujuan	25–29 Sep
2016	Megi	25–28 Sep
2017	Nesat	26–30 Jul

The performance of the estimation indicators (MAE and RMSE) in terms of errors is presented in Figure 6. The estimation indicators for the Pengjiayu Station had the highest error values ($MAE = 3.547 \text{ m/s}$ and $RMSE = 5.139 \text{ m/s}$), followed by those of the four observation stations (Keelung, Anbu, Yilan, and Su-ao) facing windward—their MAE and RMSE values had errors in the ranges $1.623\text{--}2.020 \text{ m/s}$ and $2.405\text{--}2.728 \text{ m/s}$, respectively. The estimation indicators for the three observation stations (Taipei, Banqiao, and Tamsui) in the Tamsui Basin had the

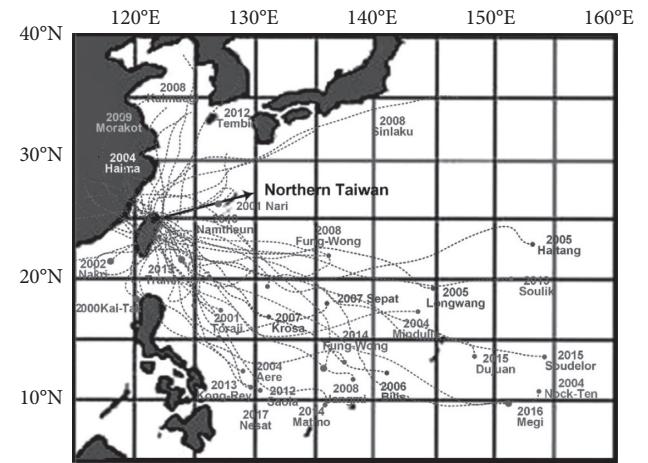


FIGURE 4: Tracks of historical typhoon incidents that affected Taiwan.

lowest error values; their MAE and RMSE values had errors in the ranges of $1.092\text{--}1.362 \text{ m/s}$ and $1.422\text{--}1.654 \text{ m/s}$, respectively.

4. Modeling

We used ML-based models to construct our wind speed prediction model for our study area. Observation stations at Taipei and Yilan were chosen as the test location. When constructing the model, the attribute data entered into the model included data on typhoon dynamics, data from ground meteorological observation stations, and data obtained from the aforementioned simulation. In this study, we performed data splitting for all typhoon episodes. Our training-validation set comprised data on 23 typhoon episodes between 2000 and 2013 (2093 records in total); our testing set comprised data on six typhoon episodes between 2014 and 2017 (528 records in total). Model training and validation were performed through 10-fold cross-validation, in which the training set was divided into 10 subsamples, one of which was retained for model verification, and the other nine were used for model

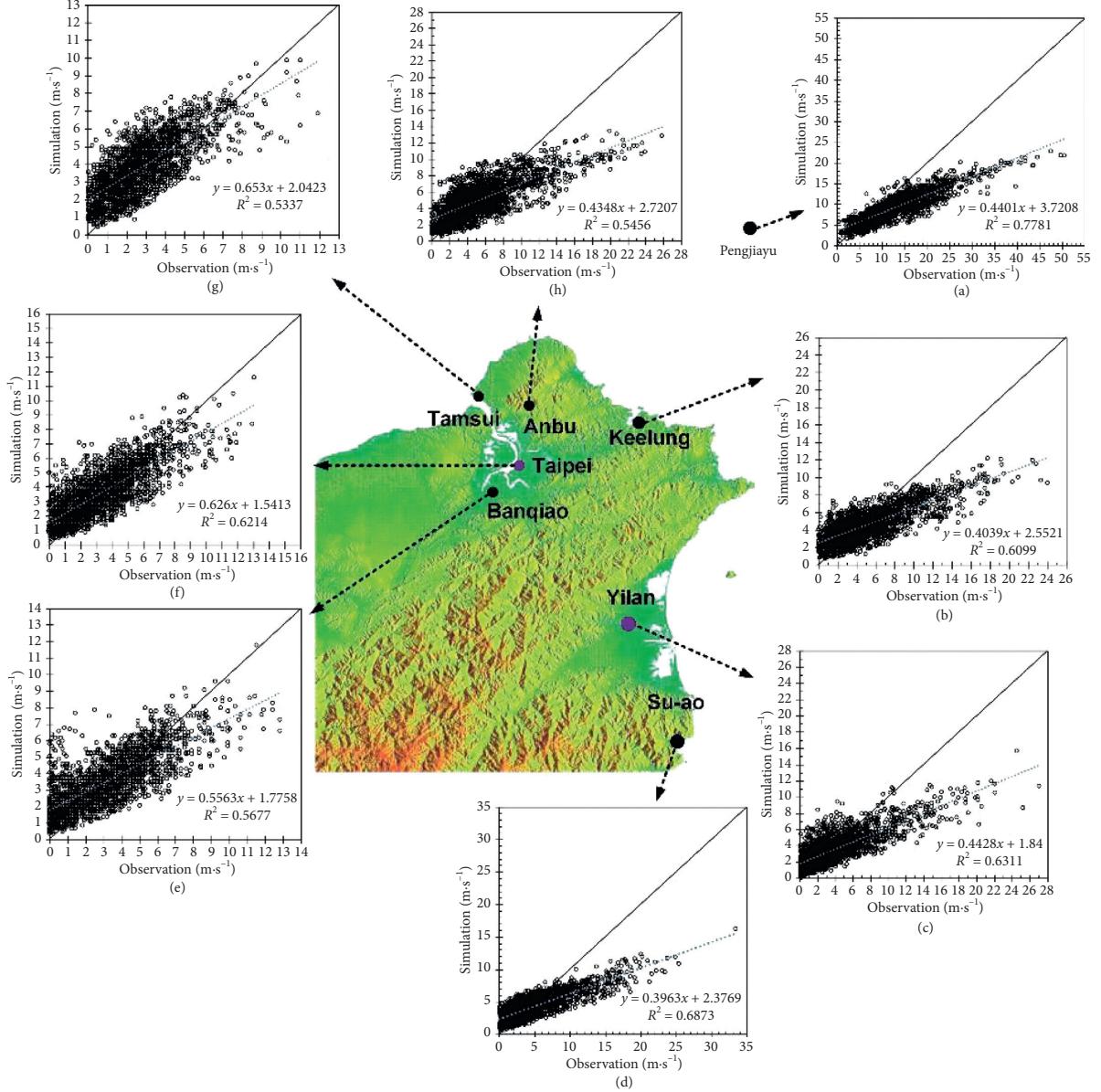


FIGURE 5: Scatterplots depicting observations vs. WRF simulations at eight stations: (a) Pengjiayu, (b) Keelung, (c) Yilan, (d) Su-ao, (e) Banqiao, (f) Taipei, (g) Tamsui, and (h) Anbu.

training. In the verification process, each subsample must be verified.

As mentioned in Section 2.1, multilayer neural networks were used to construct our ML neural network-based models. We used the adaptive moment estimation optimization algorithm (also known as the Adam optimizer) to optimize the momentum and learning rate [61]. Generally, the Adam optimizer is more broadly applied in neural networks [62–64]. The Adam optimizer can be used instead of the classical stochastic gradient descent procedure to iteratively update network weights based on training data. We also calibrated the hyperparameters, specifically, the number of neurons in a middle layer and the length of the middle layers in the MLP, DRNN, and stacked LSTM models.

These hyperparameters were evaluated through trial and error. First, the number of neurons in a middle layer was adjusted from 10 to 100 until the curves of the RMSE of the RMSE errors were approximated. The middle layers of the MLP, DRNN, and stacked LSTM models are the hidden, recurrent, and LSTM layers, respectively. Subsequently, the prediction errors corresponding to the minimum RMSE were obtained in the form of the optimal solution for the number of neurons in the middle layer. When the lead time was 1 h, for the Taipei Station, the optimal numbers of neurons were 30, 50, and 50 for the MLP, DRNN, and LSTM models, respectively; for the Yilan Station, those numbers for the three models were 40, 30, and 30, respectively (Figures 7(a) and 8(a)). Subsequently, we calibrated the length of the middle layers in

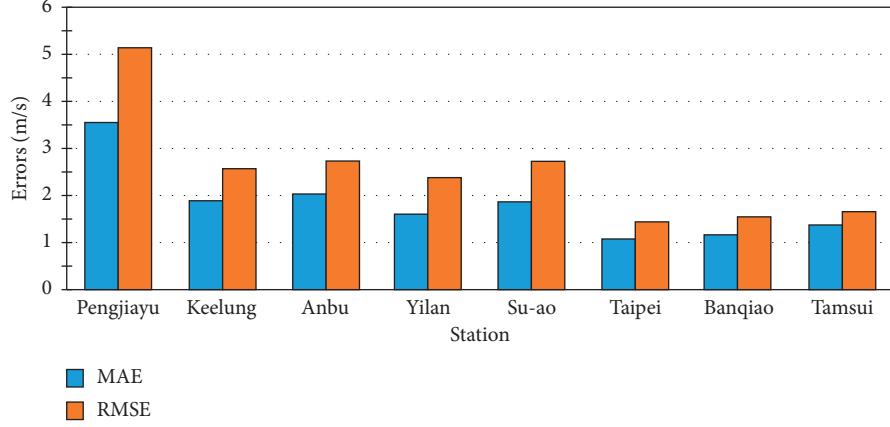


FIGURE 6: Simulation performance in terms of prediction errors.

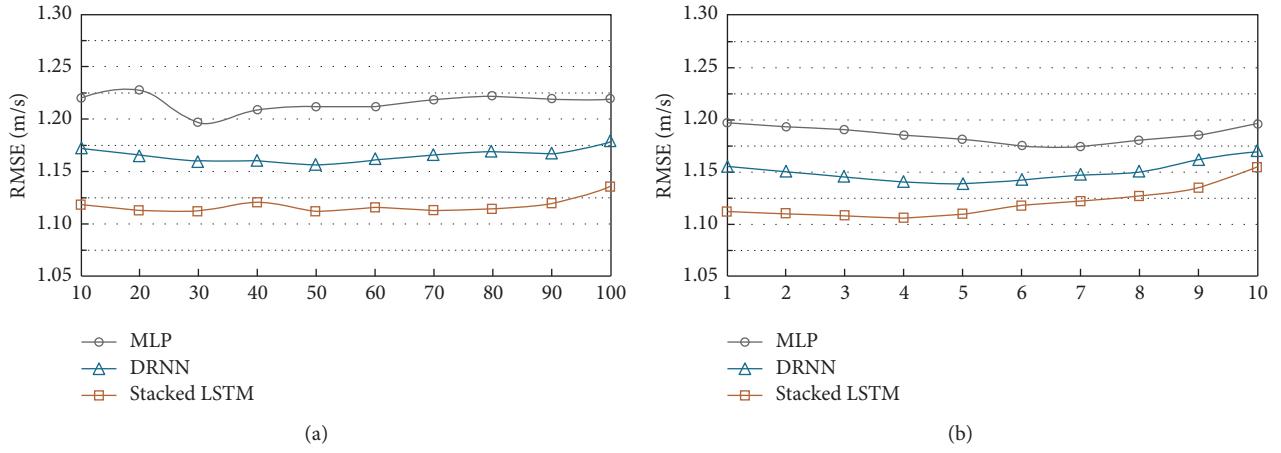


FIGURE 7: Sensitivity of model parameters on the Taipei gauge: (a) number of neurons in a middle layer and (b) length of middle layers.

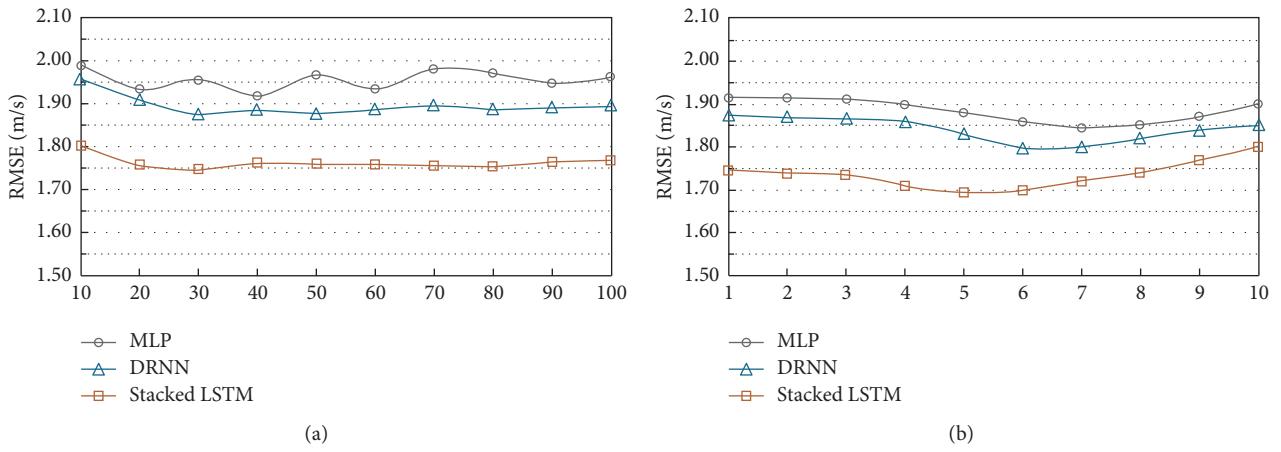


FIGURE 8: Sensitivity of model parameters on the Yilan gauge: (a) number of neurons in a middle layer and (b) length of middle layers.

the networks, adjusting them to be between 1 and 10 layers. For the Taipei Station, the optimal lengths of the middle layers were 7, 5, and 4 for the MLP, DRNN, and LSTM models, respectively; for the Yilan Station, those for

the three models were 7, 6, and 5, respectively (Figures 7(b) and 8(b)).

Using the aforementioned method, we conducted parameter testing for the prediction models. Tests were

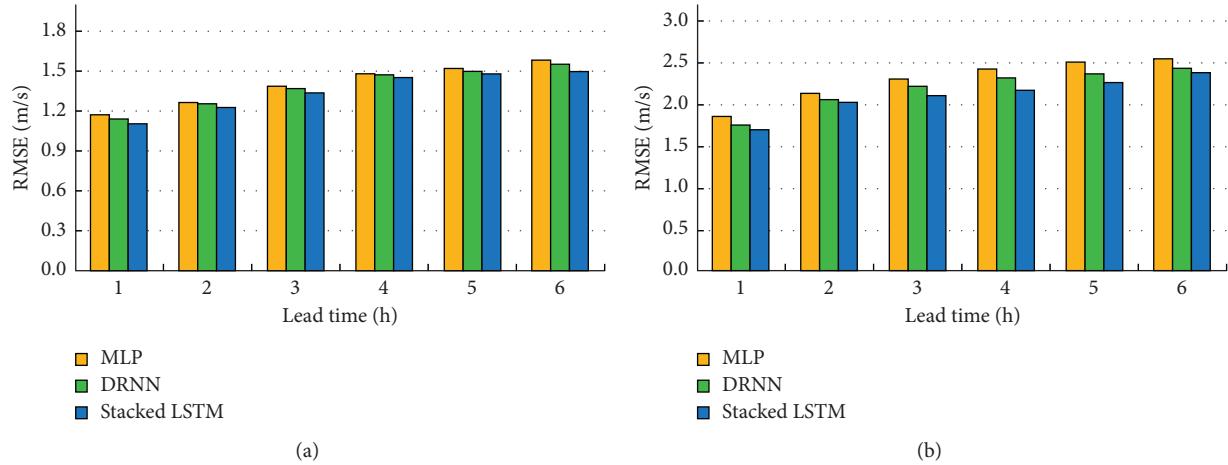


FIGURE 9: Performance levels of neural network-based predictions for lead times between 1 and 6 h using a training-validation set for the (a) Taipei Station and (b) Yilan Station.

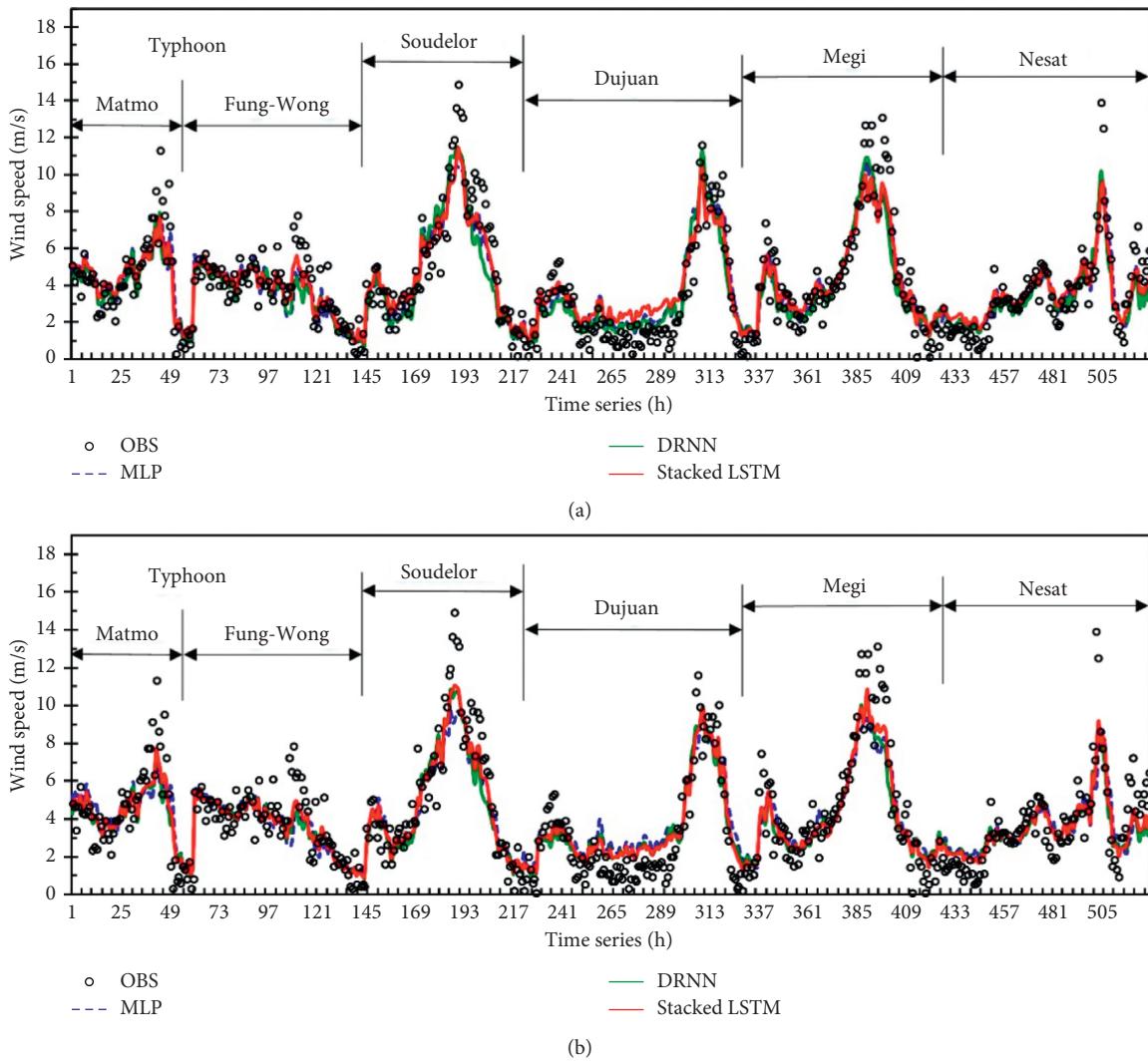


FIGURE 10: Continued.

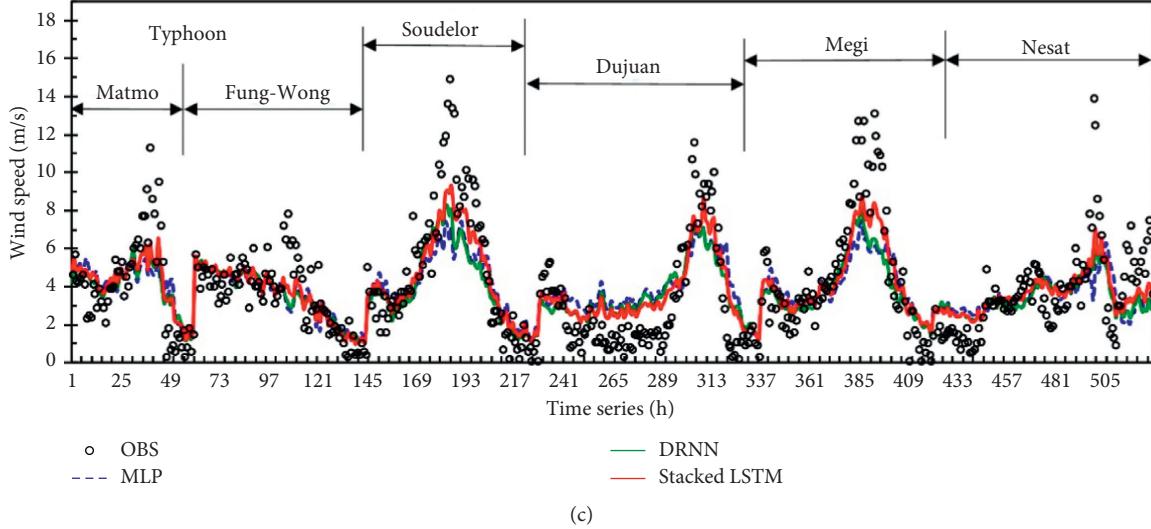


FIGURE 10: Predicted results of typhoons (2014–2017) in the testing set on the Taipei gauge for the lead times of (a) 1 h, (b) 3 h, and (c) 6 h.

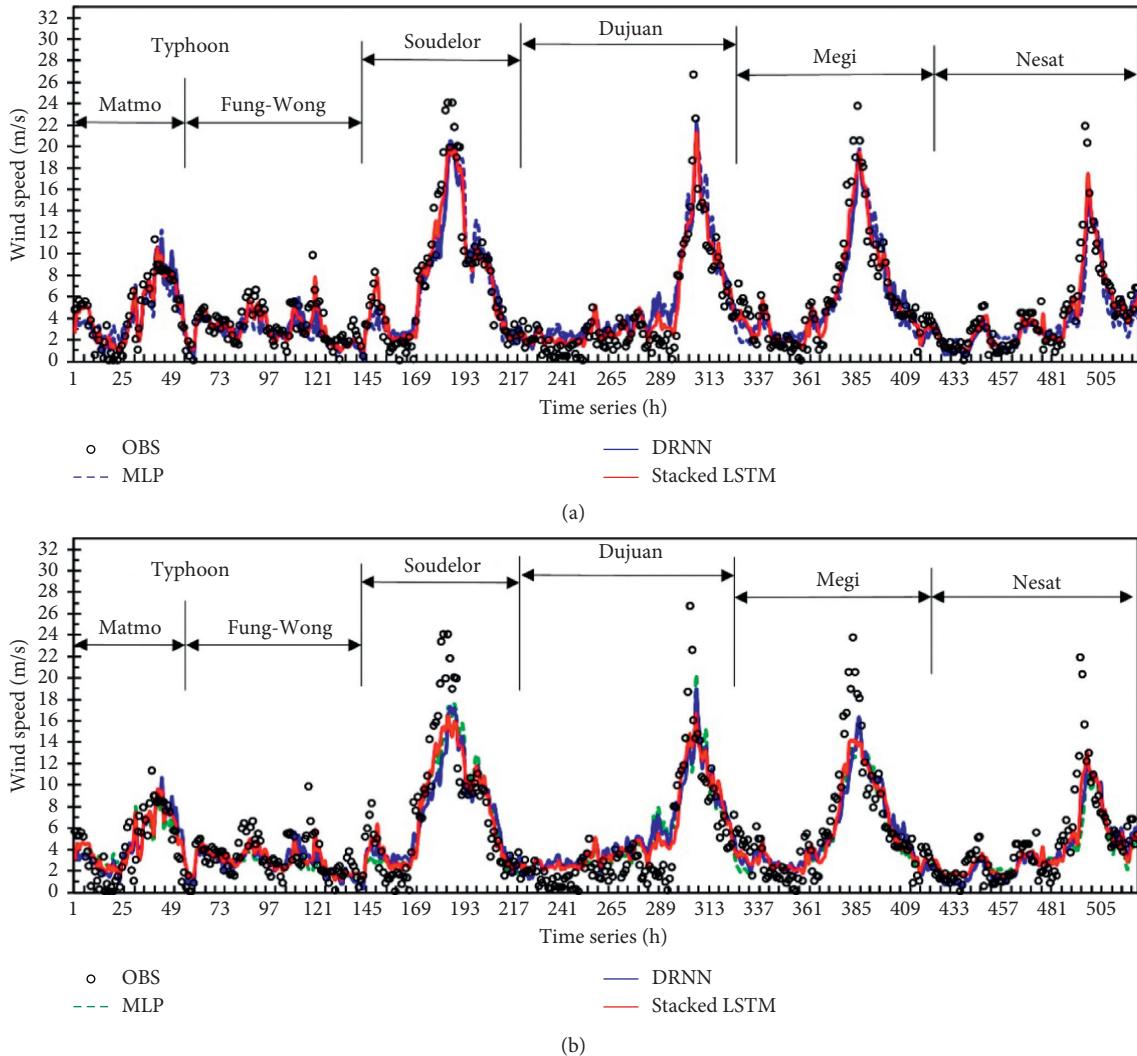


FIGURE 11: Continued.

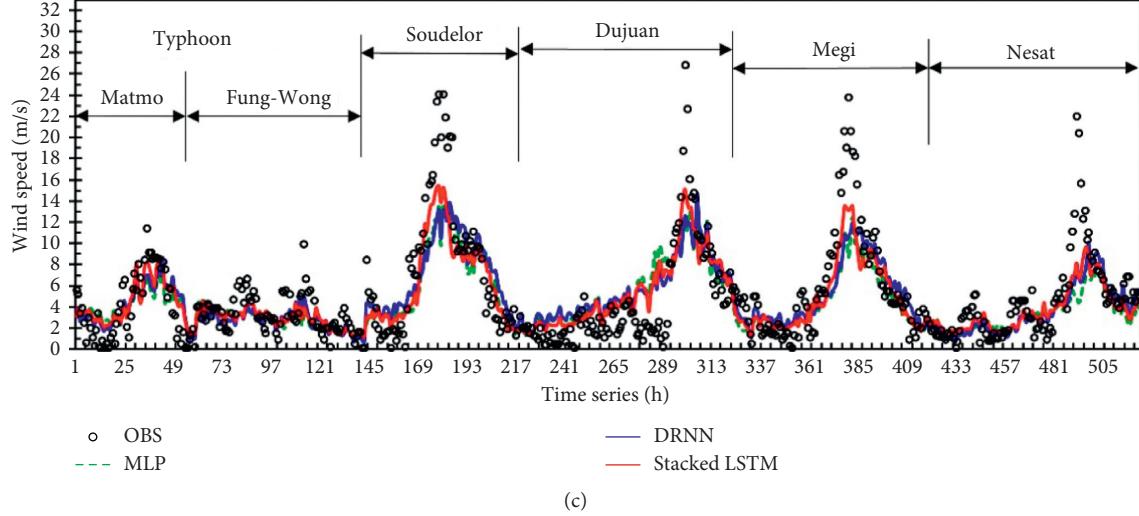


FIGURE 11: Predicted results of testing typhoons (2014–2017) on the Yilan gauge for the lead times of (a) 1 h, (b) 3 h, and (c) 6 h.

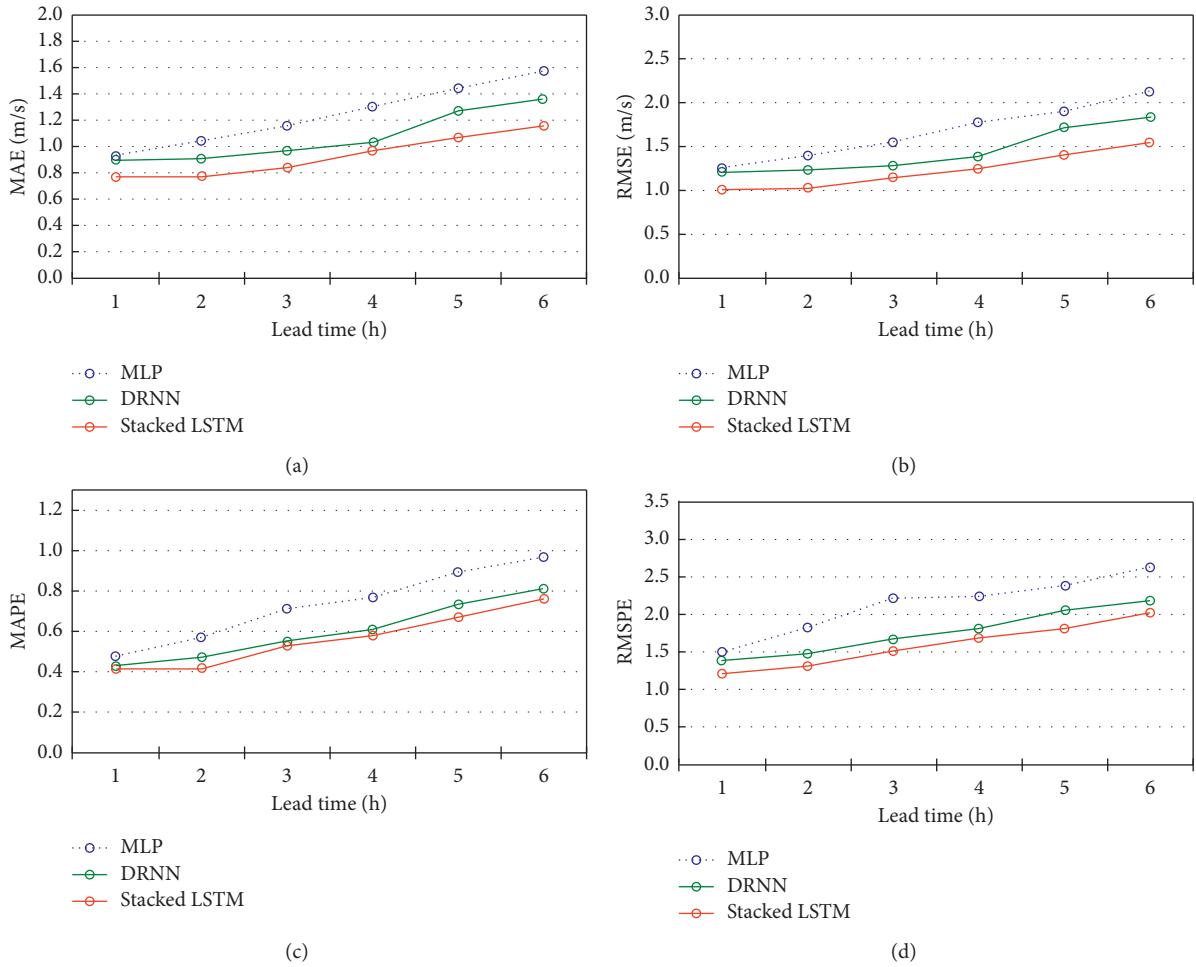


FIGURE 12: Performance measures at lead times from 1 to 6 h on the Taipei gauge: (a) MAE, (b) RMSE, (c) MAPE, and (d) RMSPE.

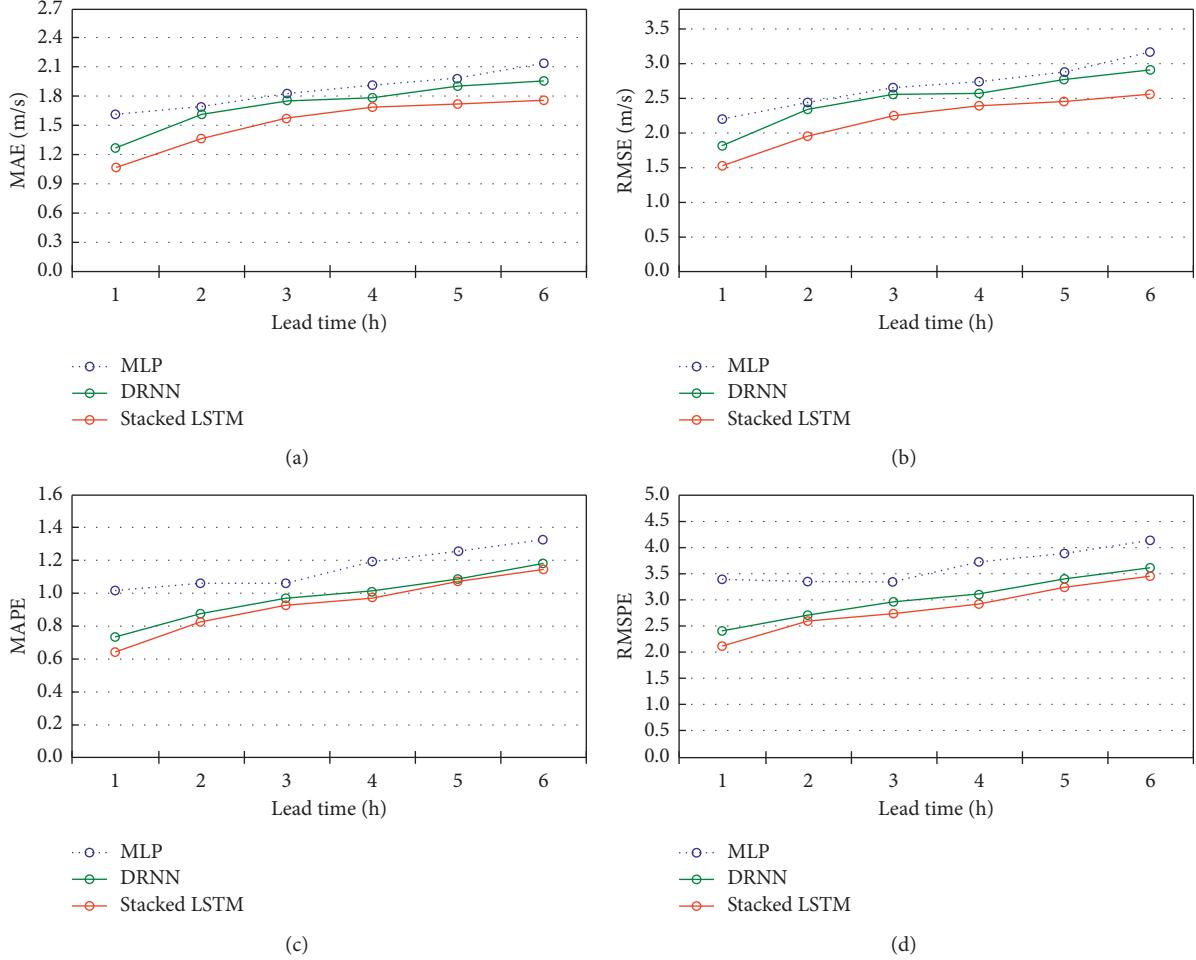


FIGURE 13: Performance measures at lead times from 1 to 6 h on the Yilan gauge: (a) MAE, (b) RMSE, (c) MAPE, and (d) RMSPE.

conducted using the training-validation set at six hourly lead times between 1 and 6 h. The RMSE performance values (for all lead times) for the three models are presented in Figure 9. The stacked LSTM model outperformed the MLP and DRNN models for all lead times and for both the Taipei and Yilan Stations. We further tested and fine-tuned these models using the testing set to confirm the accuracy and feasibility of each model.

5. Evaluation

5.1. Predictions of Forecasting Horizons. We tested and evaluated the MLP, DRNN, and stacked LSTM models using the testing set. The testing set comprised data (measured at the Yilan and Taipei Stations) on six typhoon episodes that occurred between 2014 and 2017. Figures 10 and 11 illustrate the time-series charts of the simulated and observed wind speed values for these typhoon episodes. The six typhoon episodes were Matmo in 2014, Fung-Wong in 2014, Soudelor in 2015, Dujuan in 2015, Megi in 2016, and Nesat in 2017. According to Figure 10(a), the maximum observed wind speeds for the Taipei Station were between 7.8 m/s (Fung-Wong) and 14.9 m/s (Soudelor). According to Figure 11(a), those for

the Yilan Station were between 9.9 m/s (Fung-Wong) and 26.8 m/s (Soudelor). As mentioned, the windward Yilan Station is more susceptible to the circular flow of the typhoon, unlike the Taipei Station, which is shielded by the mountains of the Tamsui Basin. Therefore, wind speed was consistently higher at the Yilan station. Figures 10(a)–10(c) illustrate the prediction results for the Taipei Station at the lead times of 1, 3, and 6 h, respectively; for all prediction models, prediction accuracy was inversely related to the lead time. The data for the Yilan Station exhibited a similar trend (Figures 11(a)–11(c)). Such a situation is not atypical for prediction models. Intuitively, the further a model predicts into the future, the harder it is to obtain useful, real-time features for prediction. Therefore, to evaluate predictive accuracy, we used the error evaluation indicators to quantify each model's rate of error in predicting the subsequent hour's wind speed.

To compute term-by-term comparisons of the relative error in the prediction with respect to the actual value of the variable, the mean absolute percentage error (MAPE) and root mean square percentage error (RMSPE) were calculated. MAPE and RMSPE can be calculated using the following formulae:

TABLE 2: Average performance measures of absolute error terms (MAE and RMSE) and relative error terms (MAPE and RMSPE) for 1–6 h predictions.

Station	Measure	MLP	DRNN	Stacked LSTM
Taipei	MAE (m/s)	1.241	1.070	0.928
	RMSE (m/s)	1.666	1.439	1.228
	MAPE	0.730	0.599	0.560
	RMSPE	2.129	1.764	1.593
Yilan	MAE (m/s)	1.863	1.714	1.528
	RMSE (m/s)	2.680	2.489	2.185
	MAPE	1.151	0.974	0.927
	RMSPE	3.643	3.030	2.847

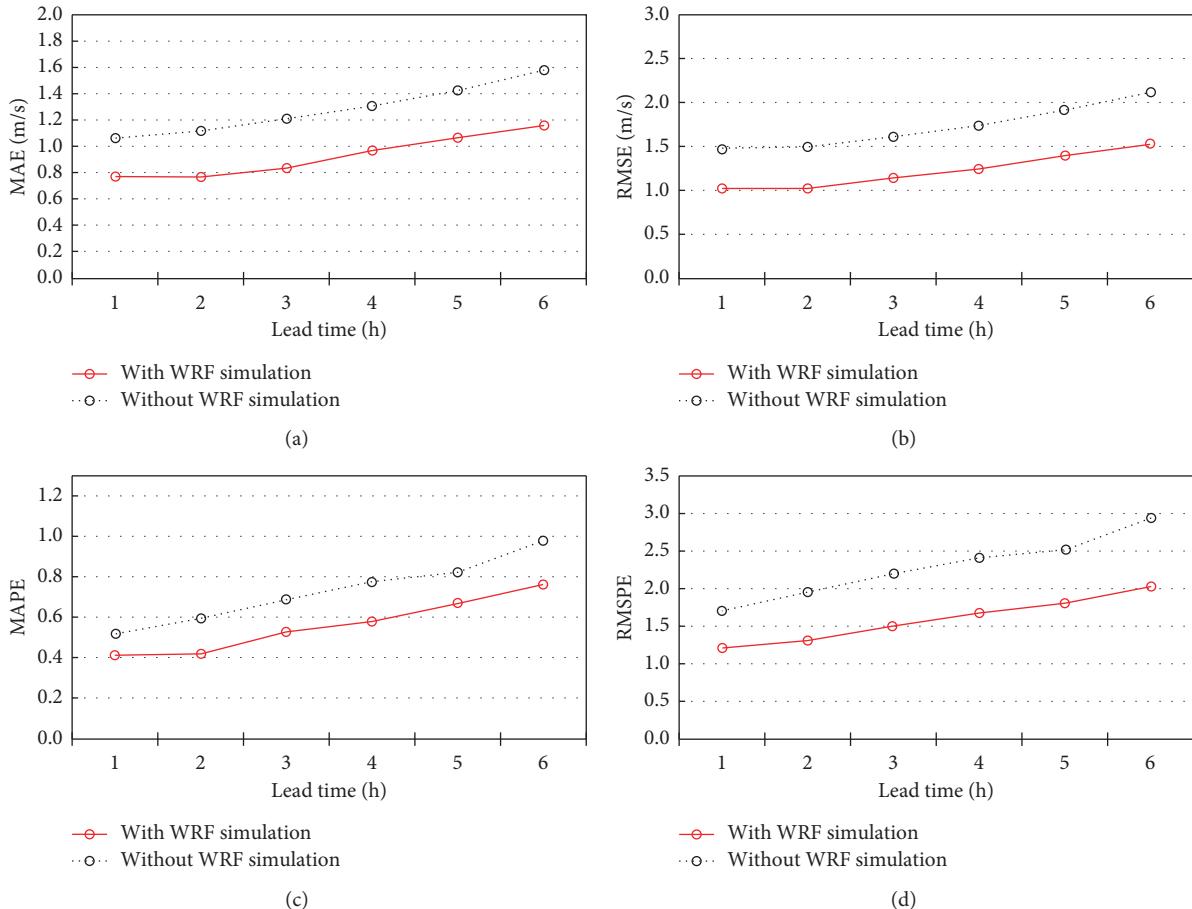


FIGURE 14: Performance measures of predictive accuracy for the use and nonuse of WRF simulation values on the Taipei gauge: (a) MAE, (b) RMSE, (c) MAPE, and (d) RMSPE.

$$\begin{aligned} \text{MAPE} &= \frac{1}{n} \sum_{i=1}^n \left| \frac{O_i - Y_i}{O_i} \right|, \\ \text{RMSPE} &= \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{O_i - Y_i}{O_i} \right)^2}. \end{aligned} \quad (8)$$

Generally, MAPE is used to express the MAE as a percentage of the observations. MAPE is an unbiased statistic for measuring the predictive capability of a model [65]. RMSPE has the same properties as the RMSE but is expressed as a percentage [66].

Figures 12 and 13 illustrate the performance of each model in terms of the four error indicators (MAE, RMSE,

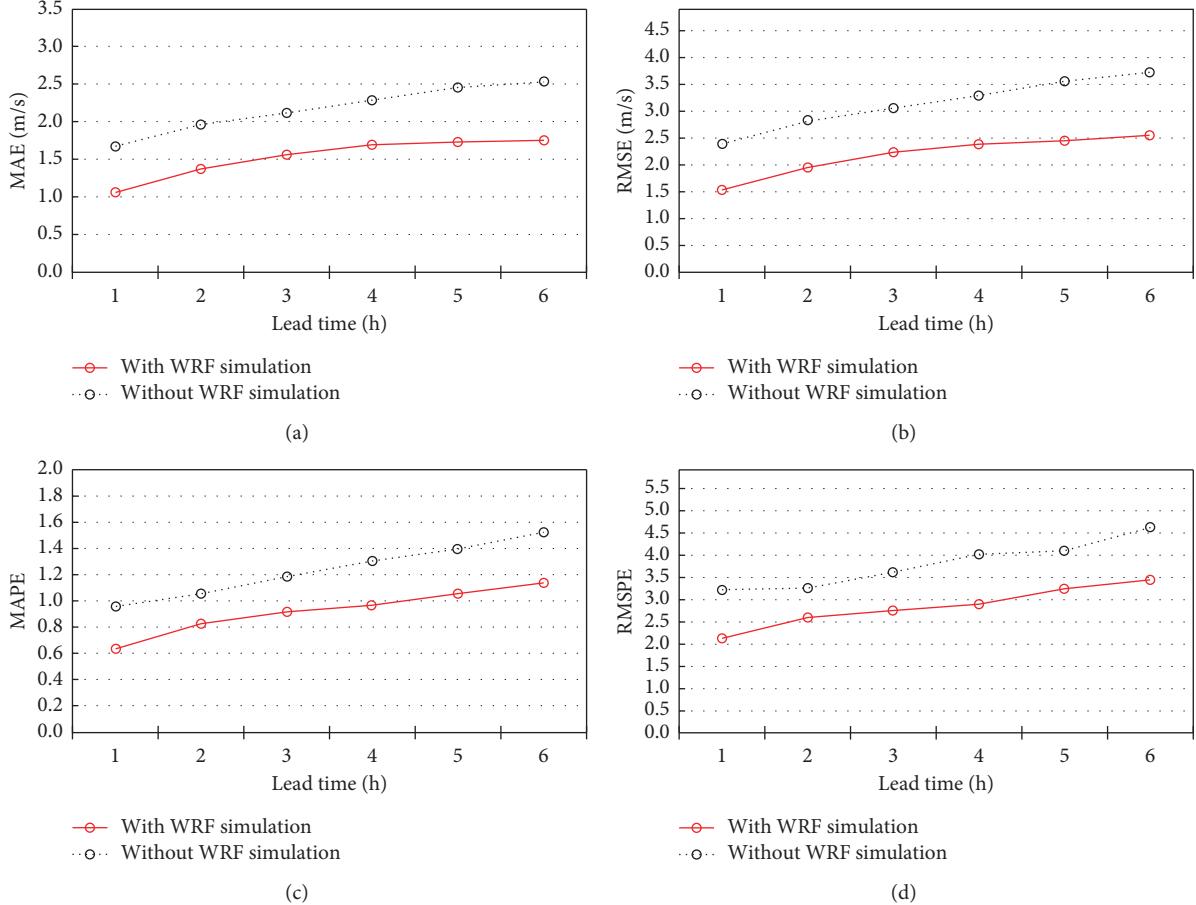


FIGURE 15: Performance measures of predictive accuracy for the use and nonuse of WRF simulation values on the Yilan gauge: (a) MAE, (b) RMSE, (c) MAPE, and (d) RMSPE.

MAPE, and RMSPE) for all lead times and for both Taipei and Yilan Stations. For the MLP model, the error rate was steeply and positively related to the lead time. However, for the DRNN and stacked LSTM models, this relation was positive but slighter. To explain the results using the RMSPE error as an example, the RMSPE error increased as the lead time increased. For the Taipei Station and for any given lead time, the MLP model exhibited the steepest increase in error, followed by the DRNN and RMSPE models. Specifically, when the lead time increased from 1 to 6 h, RMSPE increased from 1.494 to 2.616, from 1.386 to 2.181, and from 1.205 to 2.030 for the MLP, DRNN, and stacked LSTM models, respectively. Similar results were obtained for the Yilan Station. When the lead time increased from 1 to 6 h, RMSPE increased from 3.391 to 4.142, from 2.409 to 3.613, and from 2.127 to 3.457 for the MLP, DRNN, and stacked LSTM models, respectively.

Table 2 presents the average measurement values of the four error indicators for all lead times. In terms of the two unbiased statistical indicators, MAPE and RMSPE, the stacked LSTM and MLP models exhibited the most and least favorable performance, respectively. Stacked LSTM's superiority is attributable to its model structure. Specifically, LSTM is a neural network that contains LSTM blocks. LSTM blocks can be described as a type of smart

network unit that can memorize numerical values of different time lengths (to determine the quantity of useful information). Additionally, a gate in the LSTM blocks can help determine if the input is important enough to be remembered and if the input can be exported as outputs. Because the data involved in wind speed prediction are sequential, the decision on how far back the retention of memory data should go becomes consequential for predictive accuracy. By contrast, because the MLP model has no memory capacity, its predictions are based on limited and currently available information, thus decreasing its predictive accuracy. For the DRNN model, although it can receive memory information from long ago, the lack of a gate filter system can cause the receiving of too much (i.e., overly long) memory information, potentially undermining predictive accuracy.

5.2. Evaluation of Forecast Efficiency with and without Numerical Solutions. We evaluated whether the use of WRF simulation values as input affects the model's accuracy in predicting wind speed. We focused on the stacked LSTM model because it was the best performing model. Figures 14 and 15 illustrate the performance of the stacked LSTM model in terms of the four error indicators (MAE, RMSE,

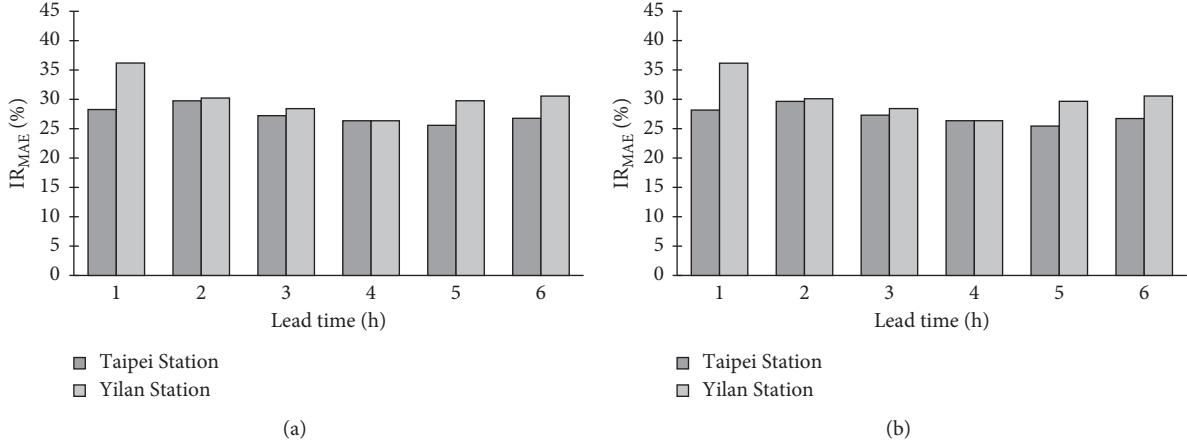


FIGURE 16: Improvement rate of predictive accuracy in terms of MAE and RMSE, comparing the use and nonuse of WRF simulation values for the Taipei and Yilan Stations.

MAPE, and RMSPE) for all lead times. We compared forecast efficiency with and without WRF simulated values. Data for the Taipei and Yilan stations are presented in Figures 14 and 15, respectively.

We determined that the use of WRF simulation values as input increased the model's predictive accuracy. Therefore, the use of numerically simulated values as a part of the input data aids in the reduction of predictive error.

We define the improvement rates for the MAE and RMSE (denoted IR_{MAE} and IR_{RMSE}, respectively) as follows:

$$\begin{aligned} \text{IR}_{\text{MAE}} &= \frac{(\text{MAE}_{\text{with}} - \text{MAE}_{\text{without}})}{\text{MAE}_{\text{without}}}, \\ \text{IR}_{\text{RMSE}} &= \frac{(\text{RMSE}_{\text{with}} - \text{RMSE}_{\text{without}})}{\text{RMSE}_{\text{without}}}, \end{aligned} \quad (9)$$

where MAE_{with} and MAE_{without} are the MAE results on the use and nonuse of WRF simulation values, respectively, and RMSE_{with} and RMSE_{without} are the RMSE results on the use and nonuse of the WRF simulation values, respectively.

Figure 16 details the improvement rates for the Taipei and Yilan Stations for all lead times. For the Taipei Station, IR_{MAE} ranged between 25.5% and 29.7%, and IR_{RMSE} ranged between 27.0% and 30.8%. For the Yilan Station, IR_{MAE} ranged between 26.4% and 36.3%, and IR_{RMSE} ranged between 28.4% and 35.7%. Generally, the average IR_{MAE} (for all lead times) was 27.3% and 30.3% for the Taipei and Yilan Stations, respectively, and the average IR_{RMSE} (for all lead times) was 28.7% and 31.1% for the Taipei and Yilan Stations, respectively. For both the Taipei and Yilan Stations, improvement was demonstrated in predictive accuracy, although the improvement rate was higher for the Yilan Station. In summary, although wind speed was difficult to predict accurately at the Yilan Station, the use of WRF simulation values increased the predictive accuracy of the model at the Yilan Station.

6. Conclusions

To accurately predict the wind speed of future typhoons, we constructed a typhoon wind speed prediction model using cutting-edge ML techniques. RNNs have been recently developed as a type of neural network that can analyze sequential data. The structure of such networks facilitates the effective processing of wind speed-relevant climate data over an extended period. That is, the structure imbues RNN models with long-term memory capacity. Therefore, such networks are suitable for predicting typhoon wind speeds. LSTM is a type of RNN that allows the user to decide the memory time's length. Additionally, LSTM gives users the option to filter output results, increasing LSTM's predictive accuracy. According to current developments in deep learning, learning performance is enhanced when the layers of neural networks are deepened. Therefore, we used deep learning neural networks in this study. Additionally, we compared the performance of three types of RNNs—MLP, DRNN, and stacked LSTM—in predicting wind speed values. These three types of model structure differ by their memory capacity: MLPs are model networks with no memory capacity, whereas DRNNs and stacked LSTM are model networks with memory capacity.

We chose northeastern Taiwan as the study area, and the observation stations at Taipei and Yilan were selected as the study subjects. The results indicated that for both the Taipei and Yilan Stations and for the MLP, DRNN, and stacked LSTM models, prediction error is positively related to the prediction lead time (of which there were six, one for each hour between 1 and 6 h). In other words, the lead time is inversely related to predictive accuracy. Model networks with memory capacity (DRNN and stacked LSTM) are more accurate than those without memory capacity (MLP). Stacked LSTM was also discovered to be more accurate than DRNN. Stacked LSTM is better than the traditional RNN method. Because stacking the LSTM's hidden layers deepens

the model, model accuracy is increased; LSTM is closer to being a deep learning technique than its counterparts. Thus, stacked LSTMs is a stable technique for overcoming problems in wind velocity prediction.

Finally, we examined if model accuracy is increased by the use of WRF simulation values as an input variable in the wind speed prediction model. The stacked LSTM model was used in the evaluation, given its best performance in most evaluation tests of the current study. Evaluating predictive accuracy in terms of the four error indicators, namely, MAE, RMSE, MAPE, and RMSPE, verified that the inclusion of the numerically simulated input greatly increased model accuracy. Specifically, for all lead times, the use of numerically simulated values reduced the MAE and RMSE error rates of the Taipei Station by 27.3% and 28.7%, respectively. Similarly, the MAE and RMSE rates for the Yilan Station were reduced by 30.3% and 31.1%, respectively. Therefore, the inclusion of numerically simulated values in wind speed prediction models enhanced their prediction accuracy.

Data Availability

The Final Operational Global Analysis data were obtained from the Global Data Assimilation System of the US Government's National Center for Environmental Protection, which is available at <https://rda.ucar.edu/datasets/ds083.2/>. The typhoon data were obtained from the Central Weather Bureau of Taiwan, which are available at <https://rdc28.cwb.gov.tw/> and <https://e-service.cwb.gov.tw/wdps/>.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This study was supported by the Ministry of Science and Technology, Taiwan, under Grant no. MOST107-2111-M-019-003. The author acknowledges the data provided by the Central Weather Bureau of Taiwan and the Research Data Archive (RDA) at NCAR.

References

- [1] C.-C. Wei, "Soft computing techniques in ensemble precipitation nowcast," *Applied Soft Computing*, vol. 13, no. 2, pp. 793–805, 2013.
- [2] C. H. Chang, U. T. Wang, H. C. Fu, and U. C. Lin, *Disaster Survey of Typhoon Soudelor in 2015*, pp. 1–48, National Science and Technology Center for Disaster Reduction, Investigation Report, Taipei, Taiwan, 2015.
- [3] C. C. Wei, P. C. Peng, C. H. Tsai, and C. L. Huang, "Regional forecasting of wind speeds during typhoon landfall in Taiwan: a case study of westward-moving typhoons," *Atmosphere*, vol. 9, 2018.
- [4] E. Cadenas, W. Rivera, R. Campos-Amezua, and C. Heard, "Wind speed prediction using a univariate ARIMA model and a multivariate NARX model," *Energies*, vol. 9, 2016.
- [5] M. Kim, M. Park, J. Im, S. Park, and M. I. Lee, "Machine learning approaches for detecting tropical cyclone formation using satellite data," *Remote Sensing*, vol. 11, 2019.
- [6] M. E. Kozar, V. Misra, and M. D. Powell, "Hindcasts of integrated kinetic energy in Atlantic tropical cyclones: a neural network prediction scheme," *Monthly Weather Review*, vol. 144, no. 12, pp. 4591–4603, 2016.
- [7] M. Madhiarasan and S. N. Deepa, "A novel criterion to select hidden neuron numbers in improved back propagation networks for wind speed forecasting," *Applied Intelligence*, vol. 44, no. 4, pp. 878–893, 2016.
- [8] J. Naik, P. Satapathy, and P. K. Dash, "Short-term wind speed and wind power prediction using hybrid empirical mode decomposition and kernel ridge regression," *Applied Soft Computing*, vol. 70, pp. 1167–1188, 2018.
- [9] I. P. Panapakidis, C. Michailides, and D. C. Angelides, "A data-driven short-term forecasting model for offshore wind speed prediction based on computational intelligence," *Electronics*, vol. 8, 2019.
- [10] C.-C. Wei, "Conceptual weather environmental forecasting system for identifying potential failure of under-construction structures during typhoons," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 168, pp. 48–59, 2017.
- [11] G. Pang, J. He, Y. Huang, and L. Zhang, "A binary logistic regression model for severe convective weather with numerical model data," *Advances in Meteorology*, vol. 2019, Article ID 6127281, 15 pages, 2019.
- [12] C.-C. Wei, "Surface wind nowcasting in the Penghu Islands based on classified typhoon tracks and the effects of the Central Mountain Range of Taiwan," *Weather and Forecasting*, vol. 29, no. 6, pp. 1425–1450, 2014.
- [13] Q. Hu, R. Zhang, and Y. Zhou, "Transfer learning for short-term wind speed prediction with deep neural networks," *Renewable Energy*, vol. 85, pp. 83–95, 2016.
- [14] L. Tiancheng, R. Qing-dao-er-ji, and Q. Ying, "Application of improved naive Bayesian-CNN classification algorithm in sandstorm prediction in inner Mongolia," *Advances in Meteorology*, vol. 2019, Article ID 5176576, 13 pages, 2019.
- [15] B. Cortez, B. Carrera, Y.-J. Kim, and J.-Y. Jung, "An architecture for emergency event prediction using LSTM recurrent neural networks," *Expert Systems with Applications*, vol. 97, pp. 315–324, 2018.
- [16] L. Guo, N. Li, F. Jia, Y. Lei, and J. Lin, "A recurrent neural network based health indicator for remaining useful life prediction of bearings," *Neurocomputing*, vol. 240, pp. 98–109, 2017.
- [17] T. Liu, S. Yu, B. Xu, and H. Yin, "Recurrent networks with attention and convolutional networks for sentence representation and classification," *Applied Intelligence*, vol. 48, no. 10, pp. 3797–3806, 2018.
- [18] A. S. Qureshi, A. Khan, A. Zameer, and A. Usman, "Wind power prediction using deep neural network based meta regression and transfer learning," *Applied Soft Computing*, vol. 58, pp. 742–755, 2017.
- [19] X. Yuan, C. Chen, M. Jiang, and Y. Yuan, "Prediction interval of wind power using parameter optimized Beta distribution based LSTM model," *Applied Soft Computing*, vol. 82, 2019.
- [20] Y. Zhang, X. Wang, and H. Tang, "An improved Elman neural network with piecewise weighted gradient for time series prediction," *Neurocomputing*, vol. 359, pp. 199–208, 2019.
- [21] Z. C. Lipton, J. Berkowitz, and C. Elkan, "A critical review of recurrent neural networks for sequence learning," 2015, <http://arxiv.org/abs/1506.00019>.
- [22] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

- [23] Q. Zhang, H. Wang, J. Dong, G. Zhong, and X. Sun, "Prediction of sea surface temperature using long short-term memory," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 10, pp. 1745–1749, 2017.
- [24] W. Byeon, M. Liwicki, and T. M. Breuel, "Scene analysis by mid-level attribute learning using 2D LSTM networks and an application to web-image tagging," *Pattern Recognition Letters*, vol. 63, pp. 23–29, 2015.
- [25] E. Fiori, A. Parodi, and F. Siccardi, "Uncertainty in prediction of deep moist convective processes: turbulence parameterizations, microphysics and grid-scale effects," *Atmospheric Research*, vol. 100, no. 4, pp. 447–456, 2011.
- [26] F. Molteni, R. Buizza, T. N. Palmer, and T. Petroliagis, "The ECMWF ensemble prediction system: methodology and validation," *Quarterly Journal of the Royal Meteorological Society*, vol. 122, no. 529, pp. 73–119, 1996.
- [27] A. H. Murphy, "What is a good forecast? An essay on the nature of goodness in weather forecasting," *Weather and Forecasting*, vol. 8, no. 2, pp. 281–293, 1993.
- [28] X. Shen, B. Jiang, Q. Cao, W. Lin, and L. Zhang, "Sensitivity of precipitation and structure of typhoon Hato to bulk and explicit spectral bin microphysics schemes," *Advances in Meteorology*, vol. 2019, Article ID 2536413, 2019.
- [29] B. Storm and S. Basu, "The WRF model forecast-derived low-level wind shear climatology over the United States Great Plains," *Energies*, vol. 3, no. 2, pp. 258–276, 2010.
- [30] K. Trachte, "Atmospheric moisture pathways to the highlands of the tropical Andes: analyzing the effects of spectral nudging on different driving fields for regional climate modeling," *Atmosphere*, vol. 9, no. 11, p. 456, 2018.
- [31] D. Carvalho, A. Rocha, M. Gómez-Gesteira, and C. Silva Santos, "WRF wind simulation and wind energy production estimates forced by different reanalyses: comparison with observed data for Portugal," *Applied Energy*, vol. 117, pp. 116–126, 2014.
- [32] X. T. Chadee, N. R. Seegobin, and R. M. Clarke, "Optimizing the weather research and forecasting (WRF) model for mapping the near-surface wind resources over the southernmost caribbean islands of Trinidad and Tobago," *Energies*, vol. 10, 2017.
- [33] N. K. Awan, H. Truhetz, and A. Gobiet, "Parameterization-induced error characteristics of MM5 and WRF operated in climate mode over the Alpine region: an ensemble-based analysis," *Journal of Climate*, vol. 24, no. 12, pp. 3107–3123, 2011.
- [34] D. E. K. Dzebre, A. A. Acheampong, J. Ampofo, and M. S. Adaramola, "A sensitivity study of surface wind simulations over coastal Ghana to selected time control and nudging options in the weather research and forecasting model," *Heliyon*, vol. 5, 2019.
- [35] E. Kalney, *Atmospheric Modeling, Data Assimilation, and Predictability*, Cambridge University Press, Cambridge, UK, 2003.
- [36] Y. Chu, C. Li, Y. Wang, J. Li, and J. Li, "A long-term wind speed ensemble forecasting system with weather adapted correction," *Energies*, vol. 9, no. 11, p. 894, 2016.
- [37] Z. Di, W. Gong, Y. Gan, C. Shen, and Q. Duan, "Combinatorial optimization for WRF physical parameterization schemes: a case study of three-day typhoon simulations over the Northwest Pacific Ocean," *Atmosphere*, vol. 10, no. 5, p. 233, 2019.
- [38] C.-C. Wei, "Wavelet support vector machines for forecasting precipitation in tropical cyclones: comparisons with GSVM, regression, and MM5," *Weather and Forecasting*, vol. 27, no. 2, pp. 438–450, 2012.
- [39] Z. Wu, C. Jiang, B. Deng, J. Chen, and X. Liu, "Sensitivity of WRF simulated typhoon track and intensity over the South China Sea to horizontal and vertical resolutions," *Acta Oceanologica Sinica*, vol. 38, no. 7, pp. 74–83, 2019.
- [40] F. Cassola, F. Ferrari, and A. Mazzino, "Numerical simulations of Mediterranean heavy precipitation events with the WRF model: a verification exercise using different approaches," *Atmospheric Research*, vol. 164–165, pp. 210–225, 2015.
- [41] S. Davolio, D. Mastrangelo, M. M. Miglietta, O. Drofa, A. Buzzi, and P. Malguzzi, "High resolution simulations of a flash flood near Venice," *Natural Hazards and Earth System Science*, vol. 9, no. 5, pp. 1671–1678, 2009.
- [42] M. L. Weisman, C. Davis, W. Wang, K. W. Manning, and J. B. Klemp, "Experiences with 0–36-h explicit convective forecasts with the WRF-ARW model," *Weather and Forecasting*, vol. 23, no. 3, pp. 407–437, 2008.
- [43] M. M. Miglietta and R. Rotunno, "Application of theory to simulations of observed cases of orographically forced convective rainfall," *Monthly Weather Review*, vol. 140, no. 9, pp. 3039–3053, 2012.
- [44] P. Zhao, J. Wang, J. Xia, Y. Dai, Y. Sheng, and J. Yue, "Performance evaluation and accuracy enhancement of a day-ahead wind power forecasting system in China," *Renewable Energy*, vol. 43, pp. 234–241, 2012.
- [45] E. Zhao, J. Zhao, L. Liu, Z. Su, and N. An, "Hybrid wind speed prediction based on a self-adaptive ARIMAX model with an exogenous WRF simulation," *Energies*, vol. 9, 2016.
- [46] S. Liang, L. Nguyen, and F. Jin, "A multi-variable stacked long-short term memory network for wind speed forecasting," in *Proceedings of the 2018 IEEE International Conference on Big Data*, pp. 4561–4564, Seattle, WA, USA, December 2018.
- [47] A. Rostami, A. Baghban, A. H. Mohammadi, A. Hemmati-Sarapardeh, and S. Habibzadeh, "Rigorous prognostication of permeability of heterogeneous carbonate oil reservoirs: smart modeling and correlation development," *Fuel*, vol. 236, pp. 110–123, 2019.
- [48] H. R. Saghafi, A. Rostami, and M. Arabloo, "Evolving new strategies to estimate reservoir oil formation volume factor: smart modeling and correlation development," *Journal of Petroleum Science and Engineering*, vol. 181, p. 106180, 2019.
- [49] A. Rostami, M. Arabloo, A. Kamari, and A. H. Mohammadi, "Modeling of CO₂ solubility in crude oil during carbon dioxide enhanced oil recovery using gene expression programming," *Fuel*, vol. 210, pp. 768–782, 2017.
- [50] S. S. Behzadi, C. Prakasvudhisarn, J. Klocker, P. Wolschann, and H. Viernstein, "Comparison between two types of artificial neural networks used for validation of pharmaceutical processes," *Powder Technology*, vol. 195, no. 2, pp. 150–157, 2009.
- [51] C.-C. Wei, "Comparing lazy and eager learning models for water level forecasting in river-reservoir basins of inundation regions," *Environmental Modelling & Software*, vol. 63, pp. 137–155, 2015.
- [52] D.-C. Lo, C.-C. Wei, and E.-P. Tsai, "Parameter automatic calibration approach for neural-network-based cyclonic precipitation forecast models," *Water*, vol. 7, no. 12, pp. 3963–3977, 2015.
- [53] J. L. Elman, "Finding structure in time," *Cognitive Science*, vol. 14, no. 2, pp. 179–211, 1990.
- [54] A. Graves, "Generating sequences with recurrent neural networks," 2013, <http://arxiv.org/abs/1308.0850>.

- [55] F. Weninger, J. Geiger, M. Wöllmer, B. Schuller, and G. Rigoll, “Feature enhancement by deep LSTM networks for ASR in reverberant multisource environments,” *Computer Speech & Language*, vol. 28, no. 4, pp. 888–902, 2014.
- [56] A. Graves, *Supervised Sequence Labelling with Recurrent Neural Networks*, Springer, Berlin, Germany, 2012.
- [57] F. A. Gers, J. Schmidhuber, and F. Cummins, “Learning to forget: continual prediction with LSTM,” *Neural Computation*, vol. 12, no. 10, pp. 2451–2471, 2000.
- [58] V. Nair and G. Hinton, “Rectified linear units improve restricted Boltzmann machines,” in *Proceedings of Machine Learning Research*, pp. 807–814, Sardinia, Italy, May 2010.
- [59] T. C. Yeh, S. H. Chen, and J. S. Hong, *The Forecast Technique Development Studies on the Typhoon Track, Rainfall and Winds Forecast Over Taiwan Area: A Study on the Implementation of WRF Typhoon Forecasting Component in the Operational Environment of CWB*, Funded by Ministry of Science and Technology of Taiwan, Taipei, Taiwan, No. NSC96-2625-Z-052-003 in Chinese, 2008.
- [60] T. C. Yeh, C. T. Terng, C. S. Lee, and M. J. Yang, “The forecast technique development studies on the typhoon track, rainfall and winds forecast over taiwan area: a study on the implementation of WRF typhoon forecasting component in the operational environment of CWB (II),” Funded by Ministry of Science and Technology of Taiwan, No. NSC97-2625-M-052-002 (in Chinese), 2009.
- [61] D. P. Kingma and J. L. Ba, “Adam: a method for stochastic optimization,” in *Proceedings of the International Conference on Learning Representations*, pp. 1–13, San Diego, CA, USA, May 2015.
- [62] F. Han, J. Jiang, Q.-H. Ling, and B.-Y. Su, “A survey on metaheuristic optimization for random single-hidden layer feedforward neural network,” *Neurocomputing*, vol. 335, pp. 261–273, 2019.
- [63] A. Sharma, “Guided stochastic gradient descent algorithm for inconsistent datasets,” *Applied Soft Computing*, vol. 73, pp. 1068–1080, 2018.
- [64] C. Yao, D. Cai, J. Bu, and G. Chen, “Pre-training the deep generative models with adaptive hyperparameter optimization,” *Neurocomputing*, vol. 247, pp. 144–155, 2017.
- [65] W.-C. Wang, K.-W. Chau, C.-T. Cheng, and L. Qiu, “A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series,” *Journal of Hydrology*, vol. 374, no. 3-4, pp. 294–306, 2009.
- [66] D. A. Swanson, J. Tayman, and T. M. Bryan, “MAPE-R: a rescaled measure of accuracy for cross-sectional subnational population forecasts,” *Journal of Population Research*, vol. 28, no. 2-3, pp. 225–243, 2011.