Nearshore two-step typhoon wind-wave prediction using deep recurrent neural networks
Chih-Chiang Wei and Ju-Yueh Cheng

ABSTRACT
Because Taiwan is located within the subtropical high and on the primary path of western Pacific typhoons, the interaction of these two factors easily causes extreme climate conditions, with strong wind carrying heavy rain and huge wind waves. To obtain precise wind-wave data for weather forecasting and thus minimize the threat posed by wind waves, this study proposes a two-step wind-wave prediction (TSWP) model to predict wind speed and wave height. The TSWP model is further divided into TSWP1, which uses data attributes at the current moment as input values and TSWP2, which uses observations from a lead time and predicts data attributes from input data. The classical one-step wave height prediction (OSWP) approach, which directly predicts wave height, was used as a benchmark to test TSWP. Deep recurrent neural networks (DRNNs) can be used to construct TSWP and OSWP approach-based models in wave height predictions. To compare with the accuracy achieved using DRNNs, linear regression, multilayer perceptron (MLP) networks, and deep neural networks (DNNs) were tested as benchmarks. The Guishandao Buoy Station located off the northeastern shore of Taiwan was used for a case study. The results were as follows: (1) compared with the shallower MLP network, the DNN and DRNN demonstrated a lower prediction error. (2) Compared with OSWP, TSWP1 and TSWP2 provided more accurate results. Therefore, the TSWP approach using a DRNN algorithm can effectively predict wind-wave heights.

Key words | deep recurrent neural networks, prediction, typhoon, wave height

INTRODUCTION
Typhoons, weather systems that have a strong destructive power, are often formed in the intertropical convergence zone in the Western Pacific Region. After typhoons are formed, the influence from the peripheral airflows of the subtropical high causes most of them to move toward west or northwest. Approximately 79 typhoons will form in the world every year, and the greatest number (approximately 25.7/year) and most powerful ones are formed in the northwestern Pacific and the South China Sea (Neumann 2017). The geographical location of the Taiwan Island comprises 21°54’N–25°18’N and 120°E–122°E. Taiwan has a subtropical climate. Simultaneously under the influence of monsoon and marine climates, the weather in Taiwan is capricious (Wei 2014). Additionally, Taiwan is located on the major route of western Pacific typhoons. The interaction of these two factors easily causes an extreme climate, with strong wind carrying strong rain and huge wind waves.

Taiwan is located in the northwestern Pacific. Typhoons born in this part of the world move west to the northwest along the south bound of high pressure. According to the Central Weather Bureau (CWB) of Taiwan (http://rdc28.cwb.gov.tw/), from 1911 to 2017, a total of 365 typhoons (landed or not) caused damage to Taiwan. On average, three to four typhoons influence Taiwan each year; most come in August, followed by July and September. Typhoons cause damage through strong wind and heavy rain. The
influence of strong wind is related to the wind field structure, path, strength, and scope of the typhoon (Wei et al. 2018). Regardless of whether a typhoon near Taiwan lands, the strong wind and heavy rain it brings are great threats to coastal structures, transportation, and marine fishery and can even cause a casualty.

The Central Mountain Range (CMR) in the Taiwan Island runs from north to south; therefore, when typhoons approach Taiwan from east to west or northwest, western Taiwan is shielded by the protection of the CMR. By contrast, northeast regions of Taiwan are less sheltered by the CMR. When a typhoon heads toward the area off the northeastern coast of Taiwan, it generally creates immeasurable damage. Because northeastern Taiwan is often affected by the strong damaging force of typhoons, we were inspired to develop a model that could accurately predict wind-wave height; we hoped that with predictions of wave height for the upcoming hours, people in coastal areas and conducting marine activities would have comprehensive preparation, thereby reducing the threat typhoons pose to life and property. To increase the accuracy of wind-wave prediction, recent deep learning (DL) theory was used in this study to construct prediction models.

Early attempts to parameterize the tropical cyclone wave field have assumed that it would essentially mirror the wind field (Bretscher 1972; Ross 1976; Young 1988, 2017). However, increasingly, data from in situ measurements (Fox & Haskell 2001; Ashton et al. 2013; Meylan et al. 2014), remote sensing (He et al. 2006; Gao 2010; Klemas 2012), and numerical modeling (Roulston et al. 2005; Popinet et al. 2010; Reikard & Rogers 2011; Sánchez-Arcilla et al. 2014; Liu et al. 2017) have demonstrated a more complex spatial distribution of waves. Numerical models use wave governing equations with correction items generated from local differences to conduct complex computations. Because this method employs grid computing, the conversion from a global scale to a scale suitable for the target coast area, which involves complicated computations and grid selection, is often time-consuming (Gorrell et al. 2011; Tsai et al. 2013; Hu et al. 2015; Wei & Hsieh 2018).

Machine learning (ML) in the field of artificial intelligence can learn from past experiences and big data, find rules within them, and thereby make predictions. Artificial neural network (ANN) is an ML algorithm. Using ML, a machine can self-learn and find optimal functions. Several scholars have used ANNs to predict the wind speed and wave height caused by typhoons or storms (e.g., Zamani et al. 2008, 2009; Aminzadeh-Gohari et al. 2009; Mahjoobi & Mosabbeb 2005; Sylaios et al. 2009; Karimi et al. 2013; Akpinar et al. 2014; Zanganeh et al. 2015; Stefanakos 2016; Loridan et al. 2017; Zanganeh 2017). For example, Deo & Sridhar Naidu (1999), with actual wave height data as inputs, used a backpropagation neural network to predict wave height. Stefanakos (2016) employed adaptive network-based fuzzy inference systems coupled with a nonstationary time-series modeling for improved prediction of wind and wave parameters. Wei (2018) reported a comparative study of data-driven wave height prediction models using a principal component analysis during typhoon periods. As indicated by Chang et al. (2011), the advantages of an ANN include potentially high predictive power, especially for nonlinear problems, without requiring detailed geographic information. However, ANNs do not provide an insight into wave propagation processes provided by full-scale numerical models.

DL is a specific subfield of ML intended to enable machines to simulate the way the human brain thinks, and its operational model is based on neuroscience (Hinton et al. 2014). It takes on learning representations from data that puts an emphasis on learning successive layers of increasingly meaningful representations. In DL, these layered representations are learned via models called neural networks, structured in literal layers stacked on the top of each other (Chollet 2018). In recent years, deep neural networks (DNNs), a DL algorithm, have been widely implemented and created immeasurable influences in various industrial-level fields, such as weather forecasting, outer-space image recognition, smart transportation, and abnormal network intrusion. For example, Wang et al. (2016) used the nonlinear capability of DL to construct models that yielded favorable performance in predicting the nonlinear and nonstable characteristics of a wind speed time sequence and increased the potential of wind power in energy systems. Hu et al. (2016) used DNNs, support vector regression, and an extreme learning machine to predict wind speed. As indicated by Bengio et al. (2013), DL helps to disentangle abstractions and identify which features improve performance. When the training data were sufficient, DNNs demonstrated excellent prediction.
Initial research on recurrent networks was performed in the 1980s (Lipton et al. 2015). Hopfield (1982) introduced a family of recurrent neural networks (RNNs) with pattern-recognition capabilities. RNNs, which efficiently use temporal information for the sequence analysis of input data, are applied for classification and prediction because they use past and current inputs for predicting outputs (Graves 2012; Kim & Bowon 2019). Some researchers have used RNNs to predict wind speed and wave height during typhoons. For instance, Mandal & Prabaharan (2006), with actual wave height in Mormugao on the west coast of India, utilized an RNN to predict wave height. A traditional RNN is a network structure with a single hidden layer (i.e., shallow recurrent network). Prasad & Prasad (2014) indicated that RNNs can provide more efficient prediction if multiple recurrent layers are used in the networks along with longer backpropagation through time extent to create deep RNNs (DRNNs); these DRNNs can be used in modeling systems of higher orders than can traditional RNNs. Because DRNNs are seldom used for predicting wind-wave heights, this study adopted DRNNs to construct wind-wave models and make predictions.

MODEL DEVELOPMENT

The classical approach to wave height prediction involves directly using all input data attributes (e.g., Zamani et al. 2008; Malekmohamadi et al. 2011; Asma et al. 2012; Karimi et al. 2013; Stefanakos 2016; Wei 2018). This study denotes this approach as the one-step wave height prediction (OSWP) approach. The classical OSWP approach is illustrated in Figure 1.

In the classical OSWP approach, the data attributes at moment \( t \) to directly predict the significant wave height at hour \( i(H_{t+i}) \) are used; \( i \) ranges from 1 to \( L \) (lead time (in hours); here, let \( L \) be 6). The input data in OSWP were the time at that moment \( t \), significant wave height \( (H_t) \), and data attributes related to wave height that we assumed as \( (Y_{m,t}) \). The output was the predicted value of the wind speed. The equation is as follows:

\[
H_{t+i} = f_i(H_t, (Y_{m,t})_{m=1,M})
\]

where \( i \) is the index of lead time, \( H_t \) is the significant wave height at moment \( t \), \( Y_{m,t} \) is the \( m \)th data attributes related to wave height at hour \( t \), \( M \) is the number of data attributes related to significant wave height, and \( f_i \) is the function at the lead time \( i \). This study used an analysis of correlation to screen attributes for those that are more relevant to the target value as the input of the model. According to Taylor (1990), when a correlation coefficient \( (r) \) is over 0.3, it signifies moderate to high correlations, whereas when \( r \) is smaller than 0.3, it indicates low correlations. Therefore, \(|r| \geq 0.3 \) was used in this study as the threshold for screening \( Y_{m,t} \).

To increase prediction accuracy, this study proposes the two-step wind-wave prediction (TSWP) model. TSWP comprises two main steps. In Step 1, we used DL-based neural networks (i.e., DRNNs) to predict wind speed, and in Step 2, we used the predicted wind speed to predict wave height.

TSWP is divided between two strategies (subcases), namely TSWP1 and TSWP2. The procedures of these

---

**Figure 1** | Concept of the classical OSWP approach.
approaches are presented in Figure 2(a) and 2(b), respectively and are detailed as follows:

- TSWP1 approach (Figure 2(a)): In Step 1, the attributes related to wind speed at moment \( t(Z_{ni}) \) are first used to predict the wind speed at the future hour \( i \) (i.e., the \( (t + i) \) hour). The screening of \( Z_{ni} \) attributes is also based on the criterion of \(|r| \geq 0.5\). In Step 2, the predicted wind speed \( (V'_{t+1}) \) obtained from Step 1 and the significant wave height at moment \( t(H_t) \) are used as model input to predict the significant wave height at the future hour \( i(H'_{t+i}) \). The equations of Step 1 and Step 2 are as follows:

\[
\text{Step 1: } V'_{t+1} = g_i(V_t, (Z_{ni})_{n=1,N}) \quad (2)
\]

\[
\text{Step 2: } H'_{t+i} = h_i(H_t, V'_{t+i}) \quad (3)
\]

where \( V'_{t+i} \) is the predicted wind speed at the future hour \( i, V_t \) is the wind speed at moment \( t, Z_{ni} \) is the \( n \)th attribute at the moment \( t \) related to wind speed, \( N \) is the number of data attributes related to wind speed, \( g_i \) is the wind speed prediction model at the lead time \( i \), and \( h_i \) is the wave height prediction model at the lead time \( i \).

- TSWP2 approach (Figure 2(b)): first predicts wind speed and based on the predicted wind speed, predicts wave height. Other than the input data of the two steps that are identical to those of TSWP1, the wind speed prediction value and the wave height prediction value at the future hour \( k, (V'_{t+k}) \) and \( (H'_{t+k}) \), respectively, are added as input attributes in Step 1 and Step 2, where \( k \) is from the future hour 1 to \((i-1)\) hour. The equations for Step 1 and Step 2 are as follows:

\[
\text{Step 1: } V'_{t+i} = g_i(V_t, (Z_{ni})_{n=1,N}, (V'_{t+k})_{k=1,i-1}) \quad (4)
\]

\[
\text{Step 2: } H'_{t+i} = h_i(H_t, V'_{t+i}, (H'_{t+k})_{k=1,i-1}) \quad (5)
\]

**Model construction**

As mentioned, the DRNN algorithm has the ability of deep networks to extract high-level features and the ability of recurrent networks to make time-series inferences. We employed DRNN structures in the OSWP and TSWP approaches to test their predictions of wave heights during typhoons. Figure 3 illustrates the DRNN structures adopted for model construction as kernel algorithms for the OSWP, TSWP1, and TSWP2 approaches.

First, in Figure 3(a) of the OSWP approach, the multi-layers of the DRNN structure involved an input layer, multiple hidden layers (\( N_1 \) layers), and an output layer were depicted. When \( N_1 = 1 \), the DRNN is called a simple RNN. Here, the recurrent network is based on the networks developed by Elman (1990); hidden units are connected to context units, which feed back into the hidden units at the next time step. The hidden state at any time step can contain information from a nearly arbitrarily long context window (Lipton et al. 2015). Each layer receives connections from its previous layer. Through multiple hidden layers of neurons, DRNNs can find patterns in a huge amount of information and convert them into useful data to solve problems (Schmidhuber 2015).

As mentioned, the classical OSWP approach involves only one step using input data attributes to predict wave height. Unlike the OSWP approach, the TSWP first predicts wind speed and predicts wave height on the basis of the predicted wind speed. As shown in Figure 3(b), the DRNNs of both steps of the TSWP1 approach are identical to that in the OSWP approach, and as shown in Figure 3(c), the DRNNs in the TSWP2 approach are similar to those in the TSWP1 approach. The difference between TSWP1 and TSWP2 approaches is that the predicted outcomes can be maintained, allowing those as the inputs to perform the next time-step prediction. Regarding the length of multiple hidden layers \( N_1 \) in the OSWP approach, \( N_2 \) and \( N_3 \) in the TSWP1 approach, and \( N_4 \) and \( N_5 \) in the TSWP2 approach, the study determined these by using a trial-and-error method. The length of the hidden layers tested ranged from 1 to 12 (see subsequent section).

An activation function is a function in a neural network that delivers an output based on inputs. A node (neuron) receives inputs from connections, sums them, and passes the summation through an activation function to produce an output, which is delivered to nodes in the subsequent layer (Tao et al. 2016). In these approaches, we used the rectified linear unit (shorten as ReLU) as an activation function in the hidden layers. The ReLU is defined as the positive part of its argument: \( f(x) = \max(0, x) \), where \( x \) is the input to a neuron. A ReLU function has been recently used to replace the sigmoid function in neural networks, resulting in good performance and fast training times, as reported in Nair & Hinton (2010).
Figure 2 | Concept of proposed TSWP1 and TSWP2 approaches.
Here, to compare the accuracy achieved using DRNNs, linear regression (LR), a traditional ANN-based multilayer perceptron (MLP), and a DNN were tested as benchmarks. The MLP and DNN replaced the DRNNs in the OSWP, TSWP1, and TSWP2 approaches. For example, when the context units are removed from DRNNs, DNNs can be archived. Furthermore, when the number of hidden layers in a DNN is 1, the MLP can also be archived.

**STUDY AREA AND DATA**

The research area of this study was off the northeastern shore of Taiwan (Figure 4), an area through which typhoons in the northwestern Pacific may pass; their strong wind carries heavy rain and causes huge wind waves, severely threatening the safety of sailing in that area. To prevent typhoon damage, it is paramount to increase our ability to predict wind waves in the research area.

This study collected data from survey stations directly supervised by the CWB, including meteorology and marine meteorology data from weather stations and buoy stations. Typhoon warnings issued by the CWB during typhoon periods were recorded. The aforementioned numerical data were all recorded at 1-hour intervals. The studied location was the Guishandao Buoy Station. The meteorological stations in the sea off northeastern Taiwan included Keelung, Pengjiazu, and Yilan stations. In addition to the Guishandao Buoy Station, this sea area includes the Longdon Buoy Station. This study collected data regarding typhoon events between 2002 and 2017 (Figure 4) that had triggered the CWB to issue sea and land warnings, especially those with typhoons whose path or storm radius passing by or covering the area. Because buoy data in

---

**Figure 3**  | DRNN structures of (a) the OSWP approach, (b) the TSWP1 approach, and (c) the TSWP2 approach.
2014 and 2015 were scant, data from those years were excluded from the present study. After data organization, we included a total of 48 typhoon events (Table 1), with 5041 hourly data recordings.

Data collected in this study were divided into three datasets: typhoon dataset (A), buoy dataset (B), and weather dataset (C). First, the dataset (A) contains the following typhoon attributes: central minimum air pressure (hPa) ($A_1$), near-center maximum wind speed ($\text{ms}^{-1}$) ($A_2$), and storm radius for winds of at least category 7 (km) ($A_3$). The dataset (B) contains the following attributes: average wind speed ($\text{ms}^{-1}$) ($B_{i,1}$), gust speed ($\text{ms}^{-1}$) ($B_{i,2}$), average wind direction ($^\circ$) ($B_{i,3}$), and significant wave height (m) ($B_{i,4}$), where $i = 1$ denotes Longdon, and $i = 2$, Guishandao. The dataset (C) contains the following attributes: station air pressure (hPa) ($C_{j,1}$), sea level air pressure (hPa) ($C_{j,2}$), average wind speed ($\text{ms}^{-1}$) ($C_{j,3}$), average wind direction ($^\circ$) ($C_{j,4}$), and precipitation hours (h) ($C_{j,5}$), where $j = 1$ denotes Pengjiayu, $j = 2$ means Keelung, and $j = 3$ represents Yilan. The three sets contain a total of 26 characteristics. Tables 2 and 3 list attributes with the minimum, maximum, and mean values from typhoon, buoy, and ground stations.

### MODEL TRAINING

The proposed approaches were used to conduct experimental simulations of the sea off the northeastern coast of
Taiwan. This section details the processes of model training and parameter evaluation.

### Attribute screening

Attributes with moderate to high correlations to the prediction objectives were selected in this study as the input for the designed models. In the OSWP approach, the model input comprises attributes related to wave height ($Y_{m,t}$); thus, the data attribute of $Y_{m,t}$ must be determined first. To do so, the analysis of correlation was conducted on the three data sets \{A, B, C\} for data attribute screening, and attributes demonstrating a correlation coefficient $r$ with significant wave heights of the absolute value $\geq 0.3$ were selected as input. Figure 5(a) reveals that in the typhoon event dataset, \{A\}, the correlation coefficient of $A_3$, storm radius for winds of category 7, was 0.319, which was the highest; the typhoon central air pressure ($A_1$) and the near-center maximum wind speed ($A_2$) exhibited low correlations. In the buoy dataset, \{B\}, except for the correlations of the average wind direction at two buoy stations, ($B_{1,3}$, $B_{2,3}$), which were only $-0.109$ and $-0.006$, respectively, the following attributes all demonstrated at least moderate correlations to wave height at the Guishandao Buoy Station ($B_{2,4}$): average wind speed ($B_{1,1}$, $B_{2,1}$) and gust wind speed ($B_{1,2}$, $B_{2,2}$) of two buoy stations and Longdon station wave height ($B_{1,4}$). In the weather dataset, \{C\}, the sea level atmosphere ($C_{1,2}$, $C_{2,2}$, $C_{3,2}$) and average wind speed ($C_{1,3}$, $C_{2,3}$, $C_{3,3}$) of the three stations, the station atmosphere of Keelung and Yilan ($C_{2,1}$, $C_{3,1}$), and the precipitation hours of Pengjiayu and Yilan ($C_{1,5}$, $C_{3,5}$) demonstrated at least moderate correlations.

In the TSWP approach, because model input data include attributes related to wind speed ($Z_{n,t}$), the data attribute of $Z_{n,t}$ must be predetermined. Figure 5(b) shows that in the typhoon event dataset, \{A\}, the correlation coefficient of $A_3$, storm radius for winds of category 7, was 0.407, which was the highest; typhoon central air pressure ($A_1$) and near-center maximum wind speed ($A_2$) demonstrated low correlations. In the buoy dataset, \{B\}, except for the low correlations of the average wind direction at two buoy stations, ($B_{1,3}$, $B_{2,3}$), the following attributes all demonstrated at least moderate correlations to wind speed at the Guishandao Buoy Station ($B_{2,4}$): the average wind speed ($B_{1,1}$, $B_{2,1}$) and gust wind speed ($B_{1,2}$, $B_{2,2}$) of two buoy stations and Longdon station wave height ($B_{1,4}$). In the weather dataset, \{C\}, the sea level atmosphere ($C_{1,2}$, $C_{2,2}$, $C_{3,2}$) and average wind speed ($C_{1,3}$, $C_{2,3}$, $C_{3,3}$) of the three stations, the station atmosphere of Keelung and Yilan ($C_{2,1}$, $C_{3,1}$), and the precipitation hours of Pengjiayu and Yilan ($C_{1,5}$, $C_{3,5}$) demonstrated at least moderate correlations.

### Data categorization

All typhoon events in this study were separated into three sets: training, validation, and testing. There are numerous data-splitting methods available, such as random separation and chronological separation. Preliminarily, this study divides the data by using a chronological separation method. That is, the training set contained 25 typhoon events from 2002 to 2008 (2,466 data entries). The
validation set comprised 15 typhoon events from 2009 to 2013 (2,142 data entries). The testing set contained eight typhoon events from 2016 to 2017 (433 data entries). Table 1 lists the typhoon events in each dataset.

Table 1 | Hurricane wind scale and number of typhoon events in three sets

<table>
<thead>
<tr>
<th>Wind intensity scale</th>
<th>Ranges</th>
<th>Training set</th>
<th>Validation set</th>
<th>Testing set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saffir–Simpson wind category</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category 5</td>
<td>≥70 m/s</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Category 4</td>
<td>58–70 m/s</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Category 3</td>
<td>50–58 m/s</td>
<td>6</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Category 2</td>
<td>43–50 m/s</td>
<td>5</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Category 1</td>
<td>33–43 m/s</td>
<td>6</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Additional classifications</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tropical storm</td>
<td>18–33 m/s</td>
<td>8</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Tropical depression</td>
<td>&lt;18 m/s</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
tropical depression) depending on the intensities of their maximum sustained winds. Table 4 lists the classification of the typhoons into training, validation, and testing sets. Moreover, the percentage of these categories in each dataset is displayed in Figure 6. It shows that approximately these categories are equally distributed in training, validation, and testing sets. Therefore, these datasets are used in the subsequent analysis.

**Evaluation indicators**

This study used the following indicators to evaluate the error and degree of the dispersion of data to assess model performance, including the root-mean-square error (RMSE) and $R$:

$$\text{RMSE} = \frac{1}{n} \sqrt{\sum_{i=1}^{n} (h_{i}^{\text{pre}} - h_{i}^{\text{obs}})^2}$$

$$R = \frac{\sum_{i=1}^{n} (h_{i}^{\text{pre}} - \bar{h}^{\text{pre}}) \cdot \sum_{i=1}^{n} (h_{i}^{\text{obs}} - \bar{h}^{\text{obs}})}{\sqrt{\sum_{i=1}^{n} (h_{i}^{\text{pre}} - \bar{h}^{\text{pre}})^2 \cdot \sum_{i=1}^{n} (h_{i}^{\text{obs}} - \bar{h}^{\text{obs}})^2}}$$

where $n$ is the total number of data entries, $h_{i}^{\text{pre}}$ is the $i$th predicted value of the model, $h_{i}^{\text{obs}}$ is the $i$th observed value, $ar{h}^{\text{obs}}$ is the average observed value, and $\bar{h}^{\text{pre}}$ is the predicted average value.

**Parameter evaluation**

Three approaches (OSWP, TSWP1, and TSWP2) using DRNNs were employed in this study for prediction model construction. These models were implemented in the Waikato Environment for Knowledge Analysis (WEKA) which is a suite of an ML software written in Java (Bouckaert et al. 2010). Traditionally, neural network-based models are calibrated using trial-and-error approaches (Dai & Macbeth 1997). As indicated by Lo et al. (2015), parameters are typically adjusted separately during this type of calibration. Although this approach is simple and widely accepted, it can produce satisfactory and suboptimal results.

In a DRNN, the hyperparameters are momentum and learning rate. The adaptive moment estimation optimization algorithm (Adam optimizer) presented by Kingma & Ba (2015) can be used instead of the classical stochastic gradient descent procedure to iteratively update network weights iterative based on training data. The Adam optimizer is an extension of stochastic gradient descent and has recently seen as a broader adoption for DL applications in computer vision and natural language processing (Yao et al. 2017; Sharma 2018; Han et al. 2019). This study used the Adam optimizer to optimize the momentum and learning rate. Moreover, the number of neurons in the hidden layer was determined according to Trenn’s (2008) method: summing numbers of neurons in the input and output layers, subtracting 1 from this sum, and dividing this number by 2.

The length of hidden layers in the networks was adopted using the trial-and-error method and adjusted from one layer to 12 layers until the curve of RMSE of the predicted and the measured values reached a stable, horizontal line. Then the value corresponding to the minimum RMSE was obtained as the optimal solution for the number of hidden layers. This study analyzed three approaches, all of which undergo the aforementioned DRNN-based parameter evaluation when using DRNNs. Figures 7 and 8 illustrate the RMSE resulting from using different lengths of hidden layers for $N_{1}$–$N_{5}$. For TSWP1 with the lead time from 1 to 6 h as an example, Step 1 of wind speed prediction model evaluation, including the parameter sensitivity analysis process, is illustrated in Figure 7(a); when the number of hidden layers ($N_{A}$) was 2, 3, 4, and 6, the minimum RMSE was obtained at the lead times of 1, 2, 3, 4, and 6 h, respectively. The parameter sensitivity analysis process for Step 2 of wave height prediction model evaluation is illustrated in Figure 8(b); when the number of hidden layers ($N_{A}$) was 5, 7, 9, 7, 6, and 6, the minimum RMSE was obtained at the lead times of 1, 2, 3, 4, 5, and 6 h, respectively.

![Figure 6](https://i.imgur.com/346.png)

*Figure 6* | Percentage of various wind intensity categories in training, validation, and testing sets.
EXPERIMENTAL RESULTS

Result analysis

This section describes the testing results of the eight typhoons from 2016 to 2017 in the testing set from the classical OSWP and proposed TSWP approaches. The wave height prediction model used in this study is based on multilayer RNNs. As mentioned, the accuracy resulting from using LR, an MLP, and a DNN was compared with that from using DRNNs. We use hyphens to abbreviate and denote the combinations of three approaches and four
algorithms (e.g., OSWP-DRNN for the OSWP approach using DRNNs).

**OSWP approach**

Figure 9 displays the scatter diagrams of the predicted and observed wave height values with lead times of 1, 3, and 6 h obtained using OSWP-LR, OSWP-MLP, OSWP-DNN, and OSWP-DRNN. As shown in Figure 10, we evaluated RMSEs and $R^2$ values for wave height prediction models at various lead times (1 – 6 h). OSWP-DRNN prediction performance was superior to those of the OSWP-LR, OSWP-MLP, and OSWP-DNN models. In general, the prediction results obtained using the DRNNs were more consistent with the observed data than were those obtained using the LR, MLP, and DNN.

**TWSP1 approach**

We subsequently analyzed the proposed TWSP1 approach. Figure 11 illustrates scatter plots of the TWSP1 predicted and observed wave height values with lead times of 1, 3,
Figure 10 | RMSE and $R$ results for the OSWP approach-based wave height prediction model.

Figure 11 | Scatter diagram of predicted and measured wave height values using TSWP1-LR, TSWP1-MLP, TSWP1-DNN, and TSWP1-DRNN at the lead time: (a) $t + 1$, (b) $t + 3$, and (c) $t + 6$. 
According to the $R^2$ values, the TSWP1-DRNN prediction results were satisfactory and consistent with the observed data. For example, for the lead time of 3 h, the $R^2$ values are in ascending order as follows: TSWP1-LR (0.7195) < TSWP1-MLP (0.7575) < TSWP1-DNN (0.7987) < TSWP1-DRNN (0.8625; Figure 11(b)). Figure 12 depicts the performance values calculated for the wave height predictions using TSWP1. Figure 12(a) shows the variations in RMSE using four algorithms with lead times of 1–6 h. For the 3-h-ahead prediction, the RMSEs in ascending order were as follows: TSWP1-DRNN (0.672) < TSWP1-DNN (0.773) < TSWP1-MLP (0.831) < TSWP1-LR (0.896). The RMSE results for TSWP1-DRNN were lower than those for TSWP1-LR, TSWP1-MLP, and TSWP1-DNN. These results may imply that TSWP1-DRNN made relatively few prediction errors.

**TWSP2 approach**

The figures in this section present the scatter diagrams of the TSWP2 predicted and observed values with lead times of 1, 3, and 6 h (Figure 13). Figure 13(a) shows that with the lead time of 1 h, the prediction results of TSWP2 were identical to those of TSWP1 (Figure 11(a)); because Equations (4) and (5) of TSWP2 and Equations (2) and (3) of TSWP1 do not include prediction values from previous times, they yield the same results. Figure 14 depicts the RMSE and R results for TSWP2 approach-based wave height prediction model. For the 3-h-ahead prediction, the RMSEs in ascending order were those of TSWP2-DRNN (0.556) < TSWP2-DNN (0.694) < TSWP2-MLP (0.743) < TSWP2-LR (0.825; Figure 14(a)). The same order of RMSEs resulted from lead times of 1, 2, 4, 5, and 6 h. TSWP1-DRNN yielded smaller errors and thus more accurate predictions than did other models.

**Indicator comparisons**

To understand the overall performance of each approach, average indicators from predictions with 1–6 h of the lead time were compared. Figure 15(a) shows the average RMSE of the predictions of each approach for all lead times. Shallower networks (MLP; excluding LR) had higher RMSE than deeper networks (DNN and DRNN), with respective RMSE of 0.984 to 0.789 m using the OSWP approach, 0.853 to 0.677 m using the TSWP1 approach, and 0.764 to 0.615 m using the TSWP2 approach. Thus, the DRNN algorithm exhibited the lowest overall prediction error. As the models increased in complexity from a one-step approach (OSWP) to two-step approaches (TSWP1 and TSWP2), the RMSE correspondingly decreased from 0.984 to 0.853 m using the MLP, from 0.884 to 0.683 m using the DNN algorithm, and from 0.789 to 0.615 m using the DRNN algorithm. Thus, two-step predictions were more accurate than one-step predictions.

Figure 15(b) illustrates the correlation coefficients of all approaches with lead times of 1–6 h. Moreover, the correlation coefficient of the TSWP approach was larger than that of the OSWP approach; even when complex neural network structures were applied, the TSWP1 and TSWP2 approaches demonstrated more accurate than did the OSWP approach.

**SIMULATIONS OF EACH TYPHOON**

Among the eight typhoon events in the testing dataset, typhoons Megi in 2016 and Nesat in 2017 most substantially affected the research area. Among the 48 analyzed events,
the maximum wave height values of these two events ranked first and second. Therefore, this study used the two typhoon events to evaluate the simulation results of the models and determine the applicability of the models.

Figure 13 | Scatter diagram of predicted and measured wave height values using TSWP2-LR, TSWP2-MLP, TSWP2-DNN, and TSWP2-DRNN at the lead time: (a) $t + 1$, (b) $t + 3$, and (c) $t + 6$.

Figure 14 | RMSE and $R$ results for the TSWP2 approach-based wave height prediction model.
Typhoon tracks

The histories of the two typhoons are briefly described below:

- In 2016, typhoon Megi formed on the sea near Guam and gradually moved west–northwest (Figure 16(a)). At 06:00 UTC on September 27 of the same year, it landed near Hualien City. After the typhoon center passed over the CMR, it moved slightly SW. At 13:10 of the same day, its center reentered the sea at Wuqi in western Taiwan. During influence in Taiwan, its minimum central air pressure was 940 hPa, its near-center maximum wind speed was 45 m/s, and its storm radius for winds of category 7 was 250 km, making it a medium-strength typhoon. Figure 16(b) shows the historical records of the wind speed and wave height values at the Guishandao Buoy Station. The first hour on the timeline is 2016/9/25 15:00 UTC. At the 39th hour, the wind speed reached its maximum of 23.7 m/s, approximately an hour before the typhoon center landed in Taiwan. On the 38th hour, the wave height reached its maximum of 10.39 m. Subsequently, wave height slightly decreased, and at the 47th hour, its height reached a second local maximum of 9.20 m.

- In 2017, typhoon Nesat developed on the ocean southeast of the Philippines (Figure 17(a)). Nesat landed in Yilan County in eastern Taiwan on 2017/7/29 at 11:10 UTC. After landing, its speed increased, and at 14:30 UTC the same day, its center reentered the sea at Wuqi.
Miaoli County in western Taiwan. During the time over the Taiwan Island, its minimum central air pressure was 955 hPa, its near-center maximum wind speed was 40 m/s, and its storm radius for winds of category 7 was 180 km, making it a medium-strength typhoon. Figure 17(b) shows the historical records of the wind speed and wave height values at the Guishandao Buoy Station. The first hour on the timeline of the figure is
2017/7/28 00:00 UTC. At the 36th hour, the wind speed reached a maximum of 26.0 m/s, approximately the hour the typhoon center landed in Taiwan. At the 39th hour, the wave height reached a maximum of 9.72 m.

Simulation results

Figures 18–20 present the wave heights of typhoons Megi and Nesat at lead times of 1 to 6 h simulated using the OSWP, TSWP1, and TSWP2 approaches, respectively. The results showed that at the lead time \((t + 1)\), the simulation results predicted by OSWP (Figure 18(a)) at the Guishandao Buoy Station were already substantially underestimated. By contrast, wave height simulations using TSWP (Figures 19(a) and 20(a)) at all areas were closer to the observations. At lead times \((t + 2)\) and \((t + 3)\), the simulation results predicted by OSWP at the Guishandao Buoy Station were also substantially underestimated. By contrast, wave height simulations using TSWP at all areas were closer to the observation results, and the wave heights predicted by TSWP2 were higher than those of TSWP1. At the lead time \((t + 6)\), the estimation results of both OSWP and TSWP were substantially underestimated, but those of TSWP were closer to the observations at all areas.

Figure 21 demonstrates the results for wave height RMSEs of typhoons Megi (Figure 21(a)–21(c)) and Nesat (Figure 21(d)–21(f)). TSWP2 exhibited accuracy superior to that of TSWP1 and OSWP. The RMSEs for all lead times (1–6 h) of TSWP2 predictions were 1.006–1.206 and 0.969–1.220 m for typhoons Megi and Nesat (using MLP, DNN, or DRNN), respectively (Table 5). In addition, as previously mentioned, because the maximum wave heights of these two typhoon events were the largest and second largest among all events, the model training did not include these two typhoons, and this missed learning opportunity might have caused the underestimation of the maximum wave height in all the prediction models.
Figure 20 | TSWP2 predictions of typhoons Megi and Nesat at the Guishandao Buoy Station.

Figure 21 | Performance levels of RMSE: (a)–(c) Typhoon Megi in 2016 and (d)–(f) Typhoon Nesat in 2017.
CONCLUSIONS

To obtain precise wind-wave data for weather forecasting to minimize the threat posed by wind waves, this study proposed the TSWP approach, which first predicts wind speed followed by wave height. TSWP is divided into two subcases: TSWP1, which used the data attributes at the current moment as input values, and TSWP2, which uses observations from a lead time and predicts data attributes from input data. The classical OSWP approach, which uses only one step to directly predict wave height, was used as a benchmark.

A DRNN was used in this study to construct three approaches to predict significant wave heights at the Guishandao Buoy Station. To compare with the accuracy achieved using the DRNN algorithm, LR, an MLP, and a DNN were employed. Typhoon warning data, marine meteorology data, and meteorological data from the CWB from 2002 to 2017 were gathered. With data from a total of 48 typhoons from two buoy stations and three meteorological stations, we screened data for relevant attributes. Based on relevant data attributes, cases were designed, and typhoon events were categorized. Typhoon events from 2002–2008, 2009–2013, and 2016–2017 were categorized into the training set, validation set, and testing set, respectively. The training set was used to train the approach-based models, the validation set was used for model parameter evaluation, and the testing set was used to simulate typhoon influence and evaluate model fit.

This study predicted results 1–6 h in the future and arrived at the following conclusions. First, the prediction results of TSWP were substantially superior to those of the OSWP approach, and TSWP2 demonstrated slightly superior prediction to that of TSWP1. Second, the RMSE of all approach-based models increased as the prediction duration increased. As the prediction time increased, the error of OSWP increased most substantially. The errors of TSWP1 and TSWP2 also increased with time, but not as much. In addition, the average RMSEs for 1–6 h prediction were compared: (1) the shallower network (MLP) generated a higher error than deeper ones (DNN and DRNN), and (2) accuracy was higher with two processing steps than with one.

Table 5  Comparison of average RMSEs among approaches

<table>
<thead>
<tr>
<th>Typhoon</th>
<th>Approach</th>
<th>LR</th>
<th>MLP</th>
<th>DNN</th>
<th>DRNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Megi</td>
<td>OSWP</td>
<td>1.925</td>
<td>1.595</td>
<td>1.525</td>
<td>1.385</td>
</tr>
<tr>
<td></td>
<td>TSWP1</td>
<td>1.496</td>
<td>1.419</td>
<td>1.279</td>
<td>1.156</td>
</tr>
<tr>
<td></td>
<td>TSWP2</td>
<td>1.387</td>
<td>1.206</td>
<td>1.088</td>
<td>1.006</td>
</tr>
<tr>
<td>Nesat</td>
<td>OSWP</td>
<td>1.676</td>
<td>1.416</td>
<td>1.309</td>
<td>1.163</td>
</tr>
<tr>
<td></td>
<td>TSWP1</td>
<td>1.507</td>
<td>1.322</td>
<td>1.200</td>
<td>1.032</td>
</tr>
<tr>
<td></td>
<td>TSWP2</td>
<td>1.333</td>
<td>1.220</td>
<td>1.075</td>
<td>0.969</td>
</tr>
</tbody>
</table>

ACKNOWLEDGEMENTS

The support for this study under Grant Nos. MOST107-2111-M-019-001-CC3 and MOST107-2111-M-019-003 provided by the Ministry of Science and Technology, Taiwan is greatly appreciated. The authors acknowledge data provided by the Central Weather Bureau of Taiwan.

REFERENCES


