

Received February 24, 2020, accepted April 27, 2020, date of publication May 4, 2020, date of current version June 4, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2992083

# A Novel Data-Driven Tropical Cyclone Track Prediction Model Based on CNN and GRU With Multi-Dimensional Feature Selection

JIE LIAN<sup>1</sup>, PINGPING DONG<sup>1</sup>, YUPING ZHANG, JIANGUO PAN, AND KEHAO LIU

College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai 200234, China

Corresponding authors: Yuping Zhang (yp\_zhang@shnu.edu.cn) and Jianguo Pan (panjg@shnu.edu.cn)

This work was funded by National Natural Science Foundation of China (61572326, 61802258, 61702333), Natural Science Foundation of Shanghai (18ZR1428300), the Shanghai Committee of Science and Technology (17070502800, 16JC1403000), and Shanghai Sailing Program (Grant No.19YF1436900).

**ABSTRACT** Strong tropical cyclones have made a drastic effect on human life and natural environment. As large amounts of meteorological data and monitoring data continue to accumulate, traditional methods for predicting tropical cyclone tracks face numerous challenges regarding their prediction efficiency and accuracy. Deep learning methods recently have been proven to be able to learn both spatial and temporal features from a large amount of dataset and be extremely efficient and accurate for forecasting data in complex structures. In this paper, we propose a novel data-driven deep learning model to predict tropical cyclone tracks using the spatial locations and multiple meteorological factors. This model comprises a multi-dimensional feature selection layer, a CNN layer and a GRU layer. The proposed model was trained using a dataset of real-world tropical cyclones from the years 1945 to 2017. Through the comparison experiments, the results verify that the proposed model outperforms the traditional forecasting methods, including a climatologically aware forecasting technique, the Sanders Barotropic technique and a numerical weather prediction (NWP) model. In addition, the proposed model has better accuracy than some deep learning methods, including RNN, GRU, CNN, AE-RNN, CNN-RNN, and CNN-GRU without the proposed feature selection layer.

**INDEX TERMS** Tropical cyclones tracks prediction, deep learning, feature selection, typhoon.

## I. INTRODUCTION

Tropical cyclones are a type of mesoscale or synoptic warm cyclone generated on the ocean surface in tropical and subtropical regions. A strong tropical cyclone may become a typhoon (hurricane), an extremely destructive and unpredictable natural disaster that is responsible for the loss of life and property each year. For example, in the year 2017 alone, there were 8 typhoons landing in China, which affected 5.879 millions of people and caused about 5 billions of dollars financial damages [1]. These days, the frequency of severe typhoons has increased, which makes the prediction of tropical cyclone tracks become even more important [4], allowing citizens and governments to be better prepared when faced with such a disaster. However, when a tropical cyclone forms, it can be affected by numerous factors such as the

meteorological environment, thermodynamics, and kinetics of the tropical cyclone system. After a tropical cyclone lands, its track will be affected by complex marine, the shoreline of the coastal areas, and the topography of the inland areas [2]. All of these complex problems make the tropical cyclone track prediction being a significant challenge. Therefore, given the impact of tropical cyclones on society and the complexity involved in their prediction, it is important to explore and apply new tropical cyclone track prediction techniques.

Traditional track prediction methods are mainly based on a numerical method and statistical formulas [3]. Although these methods have been widely used along with the development of computer technology and monitoring methods, such an approach continues to suffer from high complexity and relatively low prediction accuracy [6]. And these traditional methods are limited by their own characteristics, such that they cannot analyze the implicit data features, and rely too much on the rules summarized by the historical

The associate editor coordinating the review of this manuscript and approving it for publication was Jenny Mahoney.

data without considering the complex correlation of meteorological variables and the tropical cyclone tracks on both attribute and temporal dimensions. In addition, as meteorological satellites, ocean observation stations and ground stations are gradually being established, the amount of data available is accumulating, which is changing the prediction of tropical cyclone tracks into a big data problem. Recently, deep learning models have been applied to image processing, natural language processing, and object detection successfully [7], [8], [49]. They have shown great advantages in the processing of large amounts of complex time series data in two ways. First, deep learning models can extract implicit features from a dataset within multiple variables and improve the generalization capability, e.g., through the use of a convolutional neural network (CNN). Second, a deep learning model is more efficient in processing large-volume datasets as a time series problem, e.g., using a Gated recurrent unit (GRU) network to extract temporal correlation from a time series dataset [9]. For example, in [5], a CNN model is used to understand the spatial relationship of typhoon features and typhoon formation. In [69], a GRU model is used to examine the temporal sequence of relations in typhoon progression. Therefore, a deep learning method is an appropriate approach for our research into track prediction.

In this study, we proposed a novel fusion tropical cyclones track prediction method based on CNN and GRU models. The method is composed of three layers. In the first layer, a multi-dimensional feature selection process is proposed, which can select the meteorological variables that make the biggest influence on the tropical cyclone tracks based on Granger Causality analysis, and select the most correlated time range based on the autocorrelation analysis and partial autocorrelation analysis. In the second layer, the selected meteorological variables with tropical cyclone tracks in the correlated time range are processed as a series of two-dimensional matrices. It means the two-dimensional matrix will be generated for each timestamp by a sliding window technique. Then, the matrix of each timestamp is used as input to the CNN model. In the CNN model, a customized convolutional kernel can extract implicit features of the input matrix and learn the correlation from the meteorological variables and the tropical cyclone tracks in the correlated time range. In the third layer, a GRU-based prediction model is proposed. It can take the extracted features from the second layer in time series format, and learn the temporal features from it. The GRU model is a variant of the RNN model that can establish connections between neurons within layers. Not only that, the GRU model can solve the long-term dependency problems in RNN by adding gate structures and memory cells [10]. Then, the learned temporal features are used to predict the center location of the target tropical cyclone at the next timestamp. Generally speaking, our proposed model can effectively select the most correlated meteorological variables and time range that affect the tropical cyclones tracks from the multi-dimensional feature selection layer, which can eliminate unimportant data and improve the prediction

efficiency. The CNN layer and the GRU layer can learn the temporal features, and make an accurate prediction of the cyclone track.

Therefore, our study aims at providing a fusion deep learning model that can deal with the big data problem of traditional prediction methods and improve the accuracy of tropical cyclone track prediction. Hereinafter we unfold our main contributions as follows:

(1) A multi-dimensional feature selection layer is proposed in our method to choose the most correlated meteorological variables and time range to tropical cyclone tracks from the perspective of attribute correlation analysis and temporal correlation analysis. This layer is essential for evaluating the impacts of meteorological variables on tropical cyclone track forecasting in this paper.

(2) The deep feature extraction is considered in our work, where a CNN layer of our proposed model tends to learn and extract implicit features of the meteorological variables and tropical cyclone tracks in a correlated time range, and a GRU layer, taking output of the CNN layer as input, tends to mine the deep temporal features.

(3) The proposed model was validated using a real-world tropical cyclones dataset, and the experimental results suggest that our proposed model can achieve a more accurate prediction result than some existing traditional tropical cyclone forecasting methods and deep learning methods.

## II. RELATED WORK

The four traditional methods to predict typhoon tracks contain numerical model, statistical model, dynamic model and integrated model [11]. The numerical model needs a massive computing power to deal with the complex dynamic formula [12]. With this model, a grid system needs to be generated to model the interior structure of the tropical cyclone and simulate it in real-time. However, the numerical model is computationally intensive. Compared to a numerical model, the computations of a statistical model are lightweight, requiring only the tropical cyclone behavior patterns to be calculated from the historical data [13]. But the prediction accuracy is relatively low. The dynamical model can use massive variables in a meteorological dataset as a series of prediction factors to improve accuracy [3]. In addition, the integrated model can combine multiple models, physical parameters and initial conditions into a single prediction model, which can achieve a better prediction result than a single model [14]. However, with an increasing number of weather satellites, ocean stations, and ground stations, the amount of meteorological data is increasing. Using the traditional models to find non-linear typhoon patterns from the big spatio-temporal dataset is difficult.

The rapid development of deep learning opened up another way in climate field including precipitation estimation, extreme climate detection, weather classification, tropical cyclone intensity estimation and track prediction [15]–[17], [41]–[44], [54]. A Recurrent Neural Network (RNN) is one of the most widely used deep learning methods and it can be

used to simulate human memory cells [18]. It has shown great promise in tackling sequential data, including image segmentation [19], sound recognition [20], genes engineering [21] and stock prediction [22]. RNN models have been applied to typhoon and hurricane predictions as well. For example, Xu *et al.* [23] generated a novel method based on an RNN and an attention mechanism to predict typhoon tracks. The results of this model indicate that the deep learning method can be used to improve the prediction accuracy to a certain extent. However, naive RNNs face a technological challenge, namely, a long-term dependencies problem, in which as the time interval increases, the learned information cannot be connected to far-reaching information, which will lead to vanishing gradients. Therefore, the Long Short-Term Memory (LSTM) [24] was proposed to solve this problem and can represent dynamic temporal behavior for sequential data, which is similar to an RNN although the hidden layer in RNN is replaced by a long short-term memory (LSTM) cell. However, to reduce the training parameters in LSTM, a GRU model that is a variant of LSTM [10] was developed. There are only two gates in a GRU model, namely an update gate and a reset gate, making the GRU model being a simpler structure than an LSTM network. Meanwhile, a CNN model help to capture spatial features from weather phenomena and meteorological variables [37], which has been widely used in image recognition, natural language processing and so on [38]–[40].

Previously, many experts and scholars have studied the trajectory tracking by videos or images with the help of object tracking [55]–[57], [59]–[62], where a CNN structure is used in position tracking based on image data. In view of the temporal dimension of tropical cyclones, the tropical cyclone track prediction is a task of time sequence prediction. Specifically, Gao *et al.* [25] proposed a typhoon prediction model based on an LSTM model and trained the model by using a typhoon dataset from the years 1949–2012, which can predict 6 to 24 hours of typhoon tracks. Shen *et al.* [57] proposed a fused neural network that contained one neural network using past trajectory data and one convolutional neural network using reanalysis atmospheric wind fields images. The network was trained to forecast 6-h longitude and latitude of hurricanes. The results suggested that the average error distance was significantly lower than mainstream approaches. Kim *et al.* [26] proposed a ConvLSTM-based tracking and forecasting model to predict hurricane trajectories by using satellite images. Additionally, Su [31] proposed a variant of GRU model that including a convolutional structure to predict cloud movement. In this model, both the input and output were spatio-temporal sequences that consisted of ordered satellite images. Compared to ConvLSTM model, their proposed model had fewer parameters, and the experiment results indicated that this model performed well on GOES satellite data. However, most of the convolutional-based deep learning models took two-dimensional images as model input, and the LSTM-based models needed to deal

with the excessive parameters in the training process. There still exist some difficulties in learning the implicit features from time series data of tropical cyclone trajectories and meteorological variables efficiently and accurately.

Although there were many encouraging results obtained by deep learning methods, most of the methods neglected a key factor in prediction, namely, the feature selection process that was the process of selecting the most correlated features to a typhoon track. The feature selection is important to improve the performance of deep learning models. The first reason is that it can select the most important variables to train the deep learning model, which can increase the model accuracy. Second, it reduces the data dimensions, which can improve the training speed. Carta *et al.* [32] performed the feature selection process in wind speed prediction by using correlation analysis in meteorological data. Seo *et al.* [33] used feature selection for short-term heavy rainfall prediction based on evolutionary computation. The results of these papers proved that the prediction accuracy was significantly improved by feature selection. Granger causality test was a technique to determine whether a time series is useful in prediction, which was originally proposed to analyze causality relationships among economic variables in 1960s by Granger [45]. Recently, Granger causality test has been used in meteorological areas to analyze the complexity of the causality between meteorological factors [34]. For example, Feng *et al.* [35] proposed a multi-model methodology for short-term wind forecasting, the Granger causality test was implemented for feature selection before model training. Wang and Song [36] applied the Granger causality test to conduct dynamic effect analysis among meteorological conditions on air pollution. However, currently, there is limited work that analyzing both the most correlated meteorological variables that affect the tropical cyclone tracks, and the most correlated time range to the next timestamp before training the deep learning model in order to make an accurate track prediction.

### III. METHOD

#### A. PROBLEM FORMULATION

Given a series of tropical cyclones  $X = \{tc_1, tc_2, \dots, tc_n\}$ , where each tropical cyclone  $tc_i \in X$  has a track  $L_i = \{l_{i,t1}, l_{i,t2}, \dots, l_{i,tm}\}$  that is characterized by a series of spatial coordinates indicating the center locations of tropical cyclone  $tc_i$  from timestamp  $t1$  to timestamp  $tm$ , where each  $l_{i,tj}$  can be defined by a set of spatial coordinates  $(lat_{i,tj}, lon_{i,tj})$ ,  $tj \in T = \{t1, t2, \dots, tm\}$ , where  $t1$  is the beginning timestamp and  $tm$  is the end timestamp. Therefore,  $l_{i,tj}$  can represent the center location  $(lat_{i,tj}, lon_{i,tj})$  of the tropical cyclone  $tc_i$  at a corresponding timestamp  $tj$ . Also, each tropical cyclone  $tc_i$  has a number of meteorological factors that affect the tropical cyclone track at timestamp  $tj$ , denoted by a vector  $F_{i,tj} = \{f_{i,tj,1}, f_{i,tj,2}, \dots, f_{i,tj,c}\}$ , where  $c$  represents the number of meteorological factors. Therefore, each tropical cyclone  $tc_i$

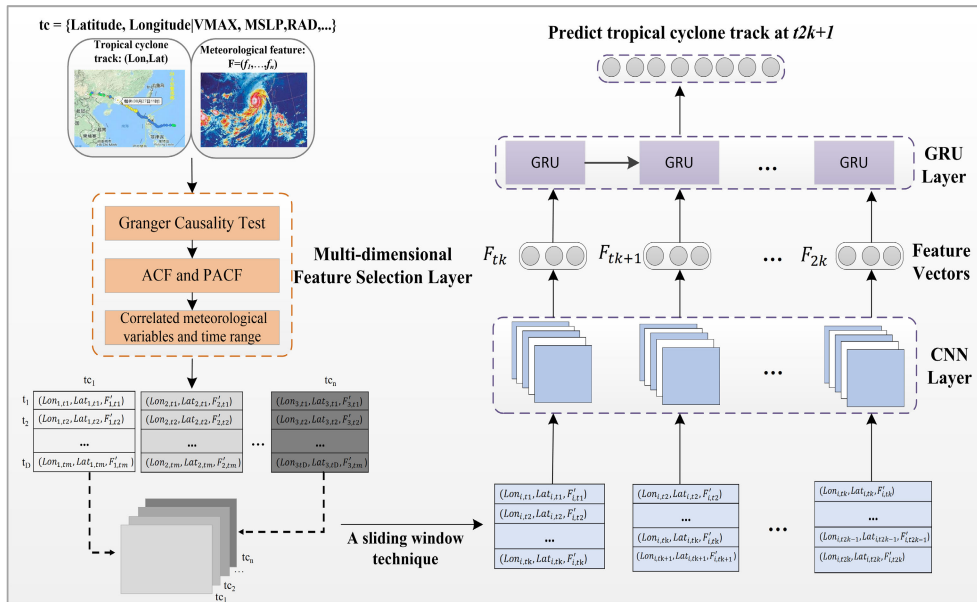


FIGURE 1. The structure of the tropical cyclone track prediction method.

can be represented by a two-dimensional matrix.

$$tc_i = \begin{pmatrix} lat_{i,t1}, lon_{i,t1}, F_{i,t1} \\ lat_{i,t2}, lon_{i,t2}, F_{i,t2} \\ \vdots \\ lat_{i,tm}, lon_{i,tm}, F_{i,tm} \end{pmatrix} \quad (1)$$

where the row in  $tc_i$  represents the latitude, longitude and meteorological factors, and the column represents an ordered set of timestamps  $T = \{t1, t2, \dots, tm\}$ .

Given a series of tropical cyclones  $X = \{tc_1, tc_2, \dots, tc_n\}$ , our main objective is to construct a data-driven prediction method to predict tropical cyclone track based on deep learning networks. In order to achieve this objective, the first step is to analyze the attribute relationship between different meteorological variables and tropical cyclone tracks based on Granger causality, and then analyze the temporal relationship based on autocorrelation analysis and partial autocorrelation analysis. This step can select the most correlated  $F'$  meteorological variables and the most correlated  $k$  time range to track prediction. Then, these variables will be chosen as the input dataset to our proposed deep learning model. The deep learning model is composed of a Convolutional Neural Network (CNN) layer that can extract feature vectors from the meteorological data and track data, and a Gated recurrent unit (GRU) layer that can learn the temporal features from the  $k$  timestamps and predict the center location of the tropical cyclone at the next timestamp.

### B. THE PROPOSED METHOD

The structure of the proposed method consists of three parts, shown in Figure 1. Given the tropical cyclone tracks and meteorological variables, the first part is to select the most

correlated meteorological variables and time range by a multi-dimensional feature selection layer. From the attribute dimension, the Granger causality test is conducted to choose the variables that make a big effect on tropical cyclone tracks, following by the autocorrelation analysis and partial autocorrelation analysis to choose the correlated time range from the temporal dimension. Then, the tracks and selected meteorological variables are normalized and processed as a series of two-dimensional matrices using a sliding window technique. In the second part, these matrices are used as input to a CNN layer for compressing and extracting implicit features between meteorological variables and tropical cyclone tracks. In the third part, these extracted features vectors in time series from the CNN layer are used as input to a GRU forecasting layer that can learn the temporal features and predict the tropical cyclone center location at the next timestamp.

#### 1) THE MULTI-DIMENSIONAL FEATURE SELECTION LAYER

The tropical cyclone tracks are affected by some complicated atmospheric environments, such as maximum sustained wind speed, central pressure. And the performance of a data-driven model highly depends on its input dataset. Therefore, the multi-dimensional feature selection layer is proposed as the first layer of our model. From the attribute dimension, this layer can select the most correlated variables to the tropical cyclone track. From the temporal dimension, this layer can analyze the most correlated time range to the target timestamp. The objective of this layer is to reduce the dimension of input variables and remove irrelevant noise data, which can improve the prediction accuracy and efficiency in the following layers.

The process of the multi-dimensional feature selection layer is illustrated in Algorithm 1, where first, the Granger

**Algorithm 1** Multi-Dimensional Feature Selection Layer

**Input:** Meteorological factor  $F = [f_1, f_2, \dots, f_c, ]$ , Trajectories:  $L$   
**Output:** Related meteorological factor  $F' = \{f_1, f_2, \dots, f_d\}$  ( $d \leq c$ ), Correlated time range  $K$   
1: Given tropical cyclone trajectories and meteorological variables values in the dataset.  
2: Stationarity = Augmented dickey-fuller unit root test (ADF)  
3: **if** Stationarity is False **then**  
4:     Cointegration = Johansen cointegration test  
5:     **if** Stationarity is True **then**  
6:         Granger causality test  
7:     **else**  
8:         End  
9:     **end if**  
10: **else**  
11:     Granger causality test  
12: **end if**  
13: Autocorrelation and partial autocorrelation analysis  
**Output:** Related meteorological factor  $F' = \{f_1, f_2, \dots, f_d\}$  ( $d \leq c$ )  
**Output:** Correlated time range  $K$

causality test is performed to select the most correlated meteorological variables from the attribute dimension, and then the autocorrelation analysis and partial autocorrelation analysis are performed to find the most correlated time range from the temporal dimension.

The traditional correlation analysis methods, such as covariance and Maximum Information Coefficient (MIC) can only analyze the linear correlation or non-linear correlation between variables, and then remove the features that are irrelevant to the target variable. However, not all of the relevant variables analyzed by covariance and MIC are useful for track forecasting. In order to further explore the correlations between the meteorological variables and tropical cyclone tracks, the Granger causality analysis was used in our model. The Granger Causality Test (GCT) between variables  $X$  and  $Y$  is defined as following: if the prediction result using the jointing past information of variables  $X$  and  $Y$  is superior to use variable  $Y$  only, namely, variable  $X$  helps to explain the change in the future, then  $X$  is the granger cause of  $Y$ . The process of GCT is shown on lines 2-12 of Algorithm 1. First, the prerequisite for Granger causality analysis is that the time series data must be stationary, otherwise false regression problems may occur. Therefore, the unit root test should be conducted for the stationarity of each time series variables. In this study, the Augmented Dickey-Fuller unit root test (ADF) [20] is used for the stationary test. The null hypothesis of the ADF means that if the unit root exists, then the time series variable is not stable. If the statistics obtained by ADF test are significantly less than the critical statistics of 3 confidence levels (1%, 5%, 10%) and the p-value is very

close, it indicates the null hypothesis is rejected and time series are stable. Otherwise, the co-integration test needs to be implemented, shown on line 4. If there is a co-integration relationship between the two sequences, then the GCT can be conducted. The specific implementation process of GCT is to estimate the unconstrained regression model using Equation (2), then estimate the constrained regression model using Equation (3).

$$Y_t^u = \alpha_0 + \sum_{i=1}^p \alpha_i Y_{t-i} + \sum_{i=1}^q \beta_i X_{t-i} + \varepsilon_t \quad (2)$$

$$Y_t^r = \alpha_0 + \sum_{i=1}^p \alpha_i Y_{t-i} + \varepsilon_t \quad (3)$$

Through GCT, the meteorological variables that affect the tropical cyclone tracks most are selected. However, the meteorological variables at one timestamp will make an influence on tropical cyclone track in the following timestamps. Therefore, determining the best correlated time range can improve the prediction accuracy. In this study, autocorrelation function (ACF) and partial autocorrelation function (PACF) are conducted to determine the best time range. In statistics, the autocorrelation of a random process is the Pearson correlation (PCC) between values of the process at different times. The PCC of the first  $N - k$  values  $L_1 = \{l_{t1}, l_{t2}, \dots, l_{t(N-k)}\}$  and the last  $N - k$  values  $L_2 = \{l_{tk}, l_{t(k+1)} \dots, l_{tN}\}$  is calculated by Equation (4).

$$\rho_1 = \frac{\sum_{i=1}^{N-k} (l_i - \bar{L}_1) (l_{i+k} - \bar{L}_2)}{\sqrt{\sum_{i=1}^{N-k} (l_i - \bar{L}_1)^2} \sqrt{\sum_{i=1}^{N-k} (l_{i+k} - \bar{L}_2)^2}} \quad (4)$$

where  $\bar{L}_1$  and  $\bar{L}_2$  are mean of the first  $N - k$  values and the last  $N - k$  values, respectively. If ignoring the difference between  $\bar{L}_1$  and  $\bar{L}_2$ , the formula of autocorrelation coefficient at time lag  $k$  can be shown in Equation (5).

$$ACF_1 = \frac{\sum_{i=1}^{N-k} (l_i - \bar{L}_1) (l_{i+k} - \bar{L}_1)}{\sqrt{\sum_{i=1}^{N-k} (l_i - \bar{L}_1)^2}} \quad (5)$$

Given the tropical cyclone tracks  $L = \{l_{t1}, l_{t2}, \dots, l_{tm}\}$ , the partial autocorrelation coefficient of time lag  $k$  refers to the influence of  $l_{t-k}$  on  $l_t$  after removing  $k - 1$  random variables. The math formula of PACF is shown in Equation (6):

$$\rho_{l_t, l_{t-k} | l_{t-1}, l_{t-2}, \dots, l_{t-k+1}} = \frac{E \left[ (l_t - \hat{E}l_t) (l_{t-k} - \hat{E}l_{t-k}) \right]}{E \left[ (l_{t-k} - \hat{E}l_{t-k})^2 \right]} \quad (6)$$

where  $\hat{E}l_t = E[l_t | l_{t-1}, \dots, l_{t-k+1}]$ ,  $\hat{E}l_{t-k} = E[l_{t-k} | l_{t-1}, \dots, l_{t-k+1}]$ . And the larger values of ACF and PACF indicate a more relevant relationship between the temporal sequence of  $l_{t-k}$  on  $l_t$ .

After implementing the multi-dimensional feature selection layer, the original features  $F = f_1, f_2, \dots, f_c$  will become  $F' = \{f_1, f_2, \dots, f_d\}$ , where ( $d \leq c$ ), and the most correlated time range  $k$  is selected as well.

## 2) THE CNN-GRU LAYER

In the second layer, the tropical cyclone tracks and selected meteorological variables are used as input to the CNN model for feature extraction. The input data is two-dimensional matrices in time series format, shown in Equation (7), where the row represents the latitude, longitude and selected meteorological factors  $F' = \{f_1, f_2, \dots, f_d\}$ , ( $d \leq c$ ) at a timestamp, and the column represents an ordered set of timestamps  $tm \in T = \{t1, t2, \dots, tm\}$ . Moreover, to improve the training efficiency and accuracy, data normalization is used to process the input data before training the CNN model. The detailed formulas of normalization are shown in Equations (8-10).

$$M = \begin{pmatrix} \text{lat}_{i,t1}, \text{lon}_{i,t1}, F'_{i,t1} \\ \text{lat}_{i,t2}, \text{lon}_{i,t2}, F'_{i,t2} \\ \vdots \\ \text{lat}_{i,tm}, \text{lon}_{i,tm}, F'_{i,tm} \end{pmatrix} \quad (7)$$

$$\mu_d = \frac{1}{n} \sum_{i=1}^n x_i \quad (8)$$

$$\sigma_d^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu_d)^2 \quad (9)$$

$$\hat{x}_i = \frac{x_i - \mu_d}{\sqrt{\sigma_d^2}} \quad (10)$$

where  $x_i$  is the input value of location or meteorological variables,  $n$  refers to the dataset size,  $\mu_d$  is the mean of one variable,  $\sigma_d^2$  is the variance of one variable and  $\hat{x}_i$  is normalized data. Then, these normalized matrices are used as input to the CNN layer.

CNN is a kind of feedforward neural network, in which neurons can respond to a portion of the surrounding cells. The difference between CNN and normal neural networks is that CNN uses the concept of weight sharing, which means using the same convolutional kernel of each layer [46]. Therefore, CNN is relatively easy to train and can extract deeper implicit features. The CNN model contains convolution layer and pooling layer. The convolution layer is to extract deep features from the input data and generate a feature map, which will be transferred to the pooling layer for feature selection and information filtering. Two-dimensional convolution kernels are used in this study. When the training process is completed, the deep features between meteorological variables and tropical cyclone tracks of a correlated time range can be extracted by the two convolution layers and two pooling layers of the CNN model. Namely, the trained convolution layer and pooling layer are used to generate feature vectors  $\{F_{tk}, F_{tk+1}, \dots, F_{tm}\}$  for training the GRU layer.

After extracting feature vectors from the CNN layer, the GRU layer is used to output the tropical cyclone location at the next timestamp. The GRU layer consists of two gates: one is an update gate that combines the input gate with the forget gate of LSTM, and the other is a reset gate. The input gate of LSTM determines how much information of the network input at the current time is saved to the cell state.

And the forget gate determines how much information of the cell state at the previous moment is retained to the current cell state. The update gate is used to control how much of the history state is retained in the output state of the current moment. The other gate is a reset gate, which acts directly on the hidden state of the previous moment and it determines how the new input information is combined with the previous memory. Therefore, these two gates can hold information in a long-term sequence. Based on this, GRU is introduced for time series prediction. The input of the GRU layer are feature vectors  $\{F_{tk}, F_{tk+1}, \dots, F_{tm}\}$  that generated from the CNN layer. The output is the predicted latitude and longitude of the target tropical cyclone at the next timestamp.

In summary, to predict the tropical cyclone track based on historical track data and meteorological data, first, the attribute correlation and temporal correlation are analyzed in the multi-dimensional feature selection layer. Then, the selected variables along with the tracks are used as input data to extract deep features in the CNN layer. The extracted feature vectors in time series format are used as input to the GRU layer, which can learn temporal features and predict the center location of a tropical cyclone at the next timestamp.

## C. TRAINING OF THE CNN-GRU MODEL

As shown in Figure 1, our proposed tropical cyclone track prediction model has three layers. Since the multi-dimensional feature selection layer need not be trained, the training process is divided into two stages, the CNN layer training and the global model training.

### 1) THE CNN LAYER TRAINING

The CNN layer training process is shown in Figure 2. For timestamp 1, the input is a two-dimension matrix that contains tropical cyclone tracks and meteorological variables from timestamp 1 to timestamp  $k$ , where  $k$  is the correlated time range determined from the multi-dimensional feature selection layer. For the next timestamp, the input data is from timestamp 2 to timestamp  $k + 1$  based on the sliding window mechanism. The rest process is analogous until the training is completed. When the training is completed, the layer can extract deep features between meteorological variables and tropical cyclone tracks at each timestamp.

After taking the two-dimension matrix as input, firstly the convolution layer can extract features from the two-dimensional matrix, where the convolution operation can maintain the sequential relationship of variables. The mathematical formula of the convolution operation in layer  $l$  is shown in Equation (11). The matrix obtained by the convolution operation is called °Feature Map.

$$C_j^l = f \left( \sum_{m_i \in M} m_i \times K_j^l + b_j^l \right) \quad (11)$$

where  $K_j^l$  ( $j = 1, 2, \dots, F_1$ ) are the convolution kernel and bias of  $j - th$  convolution kernel, respectively. And  $f$  is the activation function.  $F_1$  is the convolution kernel size of this

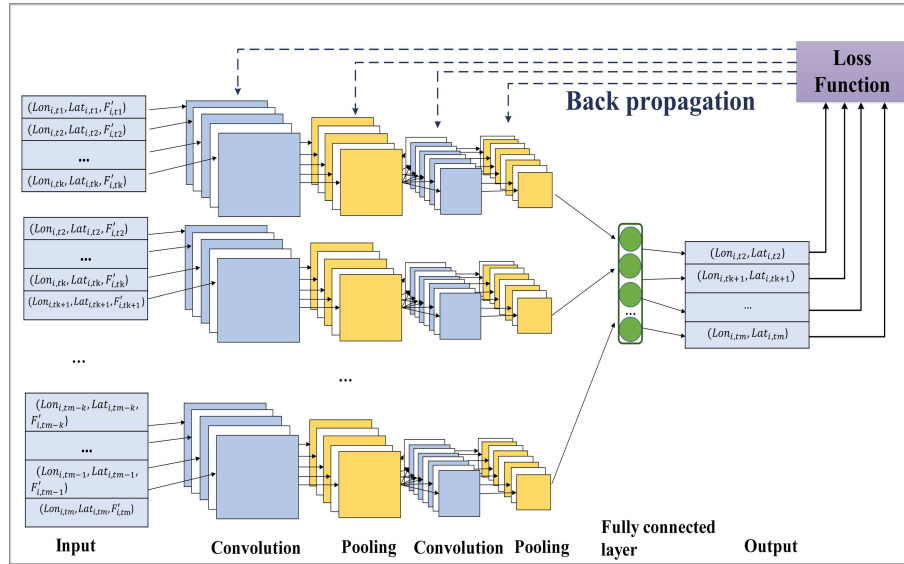


FIGURE 2. The training process of CNN layer.

convolution layer.  $F_1$  convolution kernels can produce  $F_1$  feature maps. Then, the feature maps are compressed by the pooling layer.

The pooling layer, also named subsampling layer, does not change the depth of the generated feature map matrixes, but it can reduce the size of the matrix. In image processing, the pooling operation can be considered as converting a high-resolution image into a low-resolution image. Through the pooling layer, the number of nodes in the following fully connected layer can be reduced. Therefore, the parameters of the entire neural network can be reduced, which can speed up the computation and prevent the overfitting problem. The common pooling methods contain maximum pooling and the average pooling, where the maximum pooling is used in this study. The pooling process in layer  $l$  is shown in Equation (12).

$$S_j^l = \beta_j \text{down} (C_j^l) + b_j^l \quad (12)$$

where  $\beta_j$  ( $j = 1, 2, \dots, F_1$ ) is the multiplicative bias of  $j$ -th pooling,  $\text{down}()$  is the subsampling operation and  $b_j^l$  is the bias. By establishing multiple convolution layers and pooling layers, the complex feature matrixes which represent information at each timestamp will be extracted for prediction. Finally, the feature matrix will be flattened into feature vectors, which will be the input of the fully connected layer. In our proposed model, the CNN model contains two convolution layers and two pooling layers. And only the feature vectors are used as input to the following layer.

The loss function of the CNN model uses the mean squared error with  $L_2$  normal, shown in Equation (13), where  $L_{i,tk} = \{Lon_{i,tk}, Lat_{i,tk}\}$  and  $\hat{L}_{i,tk} = \{\widehat{Lon}_{i,tk}, \widehat{Lat}_{i,tk}\}$  are observed values and fitted values of CNN at timestamp  $tk$  of tropical cyclone  $i$  respectively.  $M$  is the amount of training data. And  $L_2$  regularization constraint is added to prevent the

overfitting problem.  $\lambda$  is the regularization coefficient and  $\omega$  is a set of parameters including weights and biases of convolutional kernel and neuron. The back-propagation algorithm and minimizing loss function are used to update weights and optimize network prediction performance. As a result, the two convolution layers and two pooling layers of the CNN model can transform the input data into feature vectors for each timestamp accurately. When the network meets the expectation, the first stage of the network training is finished, and the next stage begins.

$$L = \frac{1}{M} \sum_{i=1}^M (L_{i,tk} - \hat{L}_{i,tk})^2 + \lambda \|\omega\|_2^2 \quad (13)$$

## 2) THE GLOBAL MODEL TRAINING

In the second training stage, the global CNN-GRU model needs to be trained, which is shown in Figure 3. As we can see, the feature vectors  $\{F_{tk}, F_{tk+1}, \dots, F_{t2k}\}$  that generated from the fully connected layer of CNN are input to the GRU layer. The output is the center location of the tropical cyclone at timestamp  $t2k + 1$ , where  $k$  is the time range calculated by the multi-dimensional feature selection layer. The training process of the GRU layer is demonstrated in the following steps.

i. As mentioned above, the GRU layer consists of two gates that one is an update gate. The formula of the update gate is shown in Equation (14). In our work, each tropical cyclone track is assumed to be independent, therefore, the appropriate hidden state will be reset for each tropical cyclone.

$$z_t = \sigma (W_z F_t + U_z h_{t-1} + b_z) \quad (14)$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (15)$$

In Equation (14),  $\sigma$  is a sigmoid function calculated by Equation (15).  $F_t$  denotes the feature vector at timestamp  $t$ ,  $h_{t-1}$  denotes the output vector from the last GRU cell that stores the information at timestamp  $t - 1$ , and  $W$ ,  $U$ ,  $b$  are parameter matrices and vectors, which contribute to making linear transformations of  $F_t$  and  $h_{t-1}$ . The update gate adds up this linearly transformed information and puts it into the sigmoid activation function, which can compress the result between 0 and 1.

ii. The other gate is the reset gate, which is used to determine how much the candidate state at the current moment will inherit from the previous moment, the formula is shown in Equation (16).

$$r_t = \sigma(W_r F_t + U_r h_{t-1} + b_r) \quad (16)$$

where  $\sigma$  is sigmoid function calculated by Equation (15) as well,  $F_t$  denotes the feature vector at timestamp  $t$ ,  $h_{t-1}$  denotes the output vector from last GRU cell. And  $W_r$ ,  $U_r$ ,  $b_r$  are parameters matrices and vectors of reset gate.

iii. After getting the output of the update gate  $z_t$ , which is then multiplied by the historical state  $h_{t-1}$ ,  $(1 - z_t)$  is multiplied by candidate state  $\tilde{h}_t$ , and then added them up. By doing this calculation, we can get the final memory information of the current timestamp  $t$ . The mathematical formula is shown in Equation (17).

$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot \tilde{h}_t \quad (17)$$

$$\tilde{h}_t = \tanh(W_c F_t + U_c (r_t \odot h_{t-1}) + b_c) \quad (18)$$

where  $\tilde{h}_t$  denotes the candidate hidden state, calculated by Equation (18). In this equation, make a Hadamard product of the reset gate output at timestamp  $t$  and the hidden state at timestamp  $t - 1$ . Then, the product is concatenated with the input at timestamp  $t$ . Finally,  $\tilde{h}_t$  is calculated by the fully connected layer of the tanh activation function to make the range of all elements to  $[-1, 1]$ .

iv. At last, the predicted location at the next timestamp of this particular tropical cyclone is the output by using the fully connected layer. To reach the goal of parameter optimization, the GRU model uses back-propagation algorithm to adjust the parameters during the training process. Since the CNN layer has been trained to achieve the best performance before training the global CNN-GRU model, the parameters of CNN are retained in the global training. And the loss function is also mean squared error with  $L_2$  regularization constraint, which is similar to Equation (13), where  $\omega$  is a set of parameters including weights and bias in GRU network. When the training process completes, the model can be used to predict tropical cyclone track.

Then, taking the historical data of meteorological variables and tropical cyclone tracks as input, the CNN-GRU model repeats the training and tuning process, and then optimizes the performance of the model gradually. During the training process, the CNN layer can realize the function of extracting implicit features between meteorological variables and tropical cyclone tracks. And the GRU layer adds the time series

prediction function to the model and avoids the complexity and unnecessary calculations since the input of the GRU layer is one-dimensional vectors. In general, the model can consider temporal features comprehensively and fully utilize the historical dataset to produce a more accurate prediction value.

#### IV. EXPERIMENT

To demonstrate the accuracy of our proposed model, we implemented some experiments to predict the tropical cyclones tracks. The experiments were performed on a workstation with Intel(R) Core(TM) i7-6498DU CPU@2.50GHz and 256GB of main memory. And all algorithms were implemented with python 3 and open-source machine learning packages, including Scikit-learn 0.20.3, keras 2.2.4, and TensorFlow 1.13.1.

##### A. DATASET

The proposed model was evaluated on the Western North Pacific (WNP) Ocean Best Track Data provided by Joint Typhoon Warning Center (JTWC) [47]. The dataset contained 61089 records of 2194 tropical cyclones from the year 1945 to 2017. Each record contained the 6-hourly center locations and some meteorological attributes, that including maximum sustained wind speed (VMAX), minimum sea level pressure (MSLP), Wind intensity (kts) for the radii defined in this record (RAD), radius of specified wind intensity (RAD1-RAD4), pressure in millibars of the last closed isobar (RADP), radius of the last closed isobar (RRP), radius of max winds (MRD), eye diameter (EYE) and gusts (GUSTS). The dataset was divided into three sets, where 80% of the total tropical cyclones records were used for training, 10% were used for validation, and 10% were used for testing. In other words, the training set contained 48871 tropical cyclones data tuples of 1758 tropical cyclones. The number of validation set and test set were both 6108 tuples. The experimental dataset in detail is shown in Table 1.

TABLE 1. The details of the experimental dataset.

Parameters	Value
Datasets span/amount	From the year 1945 to 2017 /61089 records
Training set	48871 tuples
Testing set	6108 tuples
Validation set	6108 tuples
Temporal dimension	6-hour

To evaluate the accuracy of our proposed model in a real-world dataset, some super typhoons, including Typhoon Meranti, Haiyan, Tip, and Djuan were selected in the testing set to visualize the predicted typhoon track results. Among them, Typhoon Haiyan was one of the strongest typhoons in 2013, which was also the strongest dependent tropical cyclone based on the records in the WNP dataset. This typhoon caused 16232 fatalities and up to 7.1 million financial damages, affecting people in the Philippines, mainland China, Taiwan



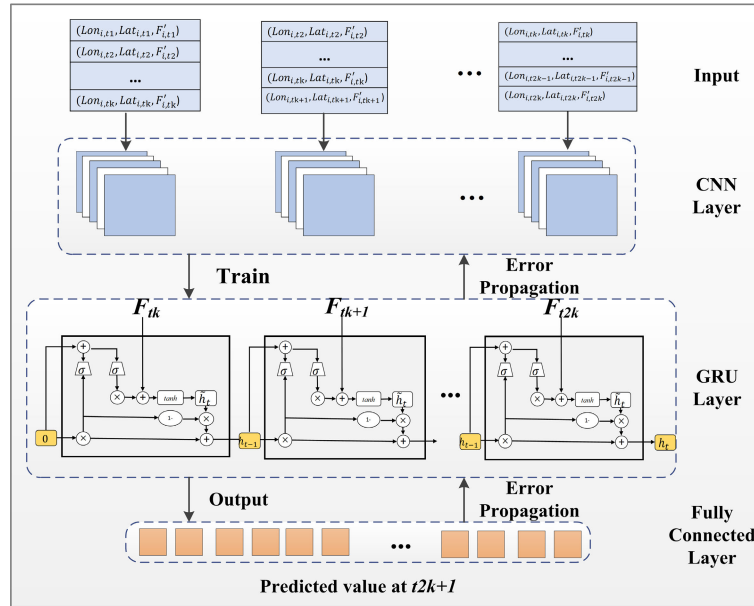


FIGURE 3. The training process of global model.

CY	YYYYMMDDHH	TAU	LatN/S	LonE/W	VMAX	MSLP	TY	RAD1	RAD2...
30	2013110106	0	70N	1410E	15	1010	DB	0	0 ...
30	2013110112	0	71N	1396E	15	1010	DB	0	0 ...
30	2013110118	0	70N	1383E	20	1007	DB	0	0 ...
30	2013110206	0	71N	1370E	20	1007	DB	0	0 ...
30	2013110200	0	72N	1359E	20	1007	DB	0	0 ...
...	...	...	...	...	...	...	...	...	...
30	2013112000	0	128N	599E	20	1007	DB	0	0 ...
30	2013112006	0	128N	594E	20	1007	DB	0	0 ...
30	2013112012	0	130N	583E	20	1007	DB	0	0 ...
30	2013112018	0	128N	571E	20	1007	DB	0	0 ...
30	2013112100	0	125N	560E	20	1007	DB	0	0 ...

FIGURE 4. Data representation of Typhoon Haiyan.

Province and Vietnam [48]. Figure 4 shows 80 historical records of Typhoon Haiyan in detail. Each row represents a record of the typhoon landing longitude, latitude, corresponding timestamp, and the meteorological attributes.

**B. MULTI-DIMENSIONAL FEATURE SELECTION**

In the dataset, there were eleven meteorological factors to be analyzed by Granger causal analysis to determine the causality between each meteorological factor and the tropical cyclone track. Since there are many missing values before the year 2000 in the JTWC dataset, the data used in the multi-dimensional feature selection layer was from 2000-2017 in order to fully consider the impact of each variable on the trajectory. First, ADF was applied to test stability. If the variable was stable, Granger causal analysis was implemented directly. Otherwise, the Johansen cointegration test (JTC) was conducted to determine whether the linear combination of a group of non-stationary series had a stable equilibrium relationship. If the relation existed, the Granger causal test can be implemented, otherwise, it cannot. The test results

TABLE 2. The ADF results of meteorological factors and tropical cyclone tracks.

Variables	ADF test statistic	Test critical values(1% level)	Stationary(Y/N)
LAT	-19.24	-3.430659	Y
LON	-25.17		Y
VMAX	-27.19		Y
MSLP	-4.21		Y
RAD	-15.02		Y
RAD1	-16.88		Y
RAD2	-16.60		Y
RAD3	-16.38		Y
RAD4	-16.88		Y
RADP	-3.81		Y
RRP	-8.23		Y
MRD	-11.51		Y
GUSTS	-6.74		Y

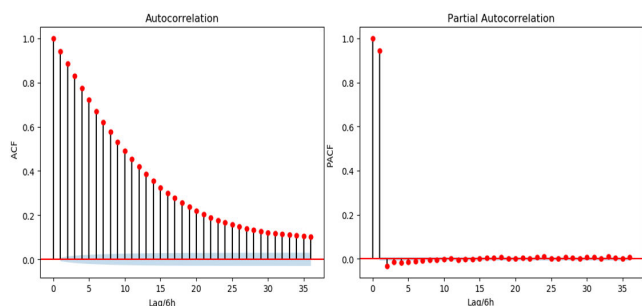
of ADF are shown in Table 2, where the critical value of the unit root test at 1% significance level is  $-3.430659$ . And under the null hypothesis that variable has a unit root, t-test statistics of the unit roots of all variables were less than  $-3.430659$ . Therefore, the null hypothesis was rejected. Then, these meteorological factors had no unit root and they were stationary series.

In the following, GCT was conducted to judge causality between the meteorological factors and tracks. In this study, the lag periods were set as 2. And the null hypothesis means the meteorological factor does not Granger cause latitude and longitude. The critical value of F at 95% confidence level was 3.0078 [34]. Assuming that the F-statistic of a variable was greater than 5, the null hypothesis was rejected in this study. The F-statistic results between meteorological factors and track are shown in Table 3, where the values within box mean that these corresponding meteorological factors

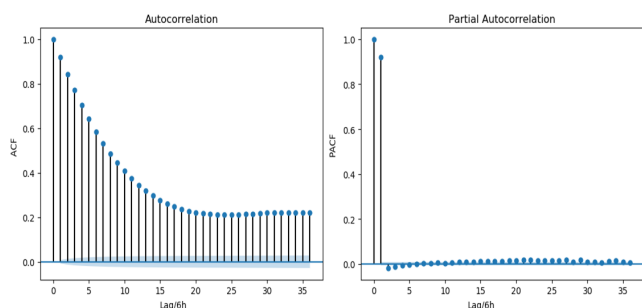
**TABLE 3. Pairwise Granger Causality tests results between meteorological factors and tropical cyclone tracks.**

F-Statistic	LAT	LON
LAT	/	76.707
LON	17.9308	/
VMAX	465.833	109.722
MSLP	4.7361	2.42685
RAD	155.994	42.0773
RAD1	196.953	35.6378
RAD2	199.612	30.5469
RAD3	174.4	31.1325
RAD4	176.247	33.5621
RADP	1.50444	0.03467
RRP	171.391	32.7972
MRD	25.2325	7.26623
GUSTS	11.1564	1.59121
EYE	51.496	3.04083

are discarded, and the rest factors combining with tropical cyclone track are used as input to the deep learning model, which can predict the longitude and latitude of the tropical cyclone center respectively. As we can see, meteorological factors VMAX, RAD, RAD1-RAD4, RRP, MRD, GUSTS and EYE were causalities of latitude, and meteorological factors VMAX, RAD, RAD1-RAD4, RADP RRP and MRD were causalities of longitude.



**FIGURE 5. The ACF and PACF of Latitude.**



**FIGURE 6. The ACF and PACF of Longitude.**

At last, ACF and PACF were implemented, which were employed to determine the temporal dependence of meteorological variables. The results are shown in Figure 5 and Figure 6. The ACF results demonstrated that meteorological variables had obvious correlation on both the latitude and longitude from lag 1 to lag 10. And the PACF results showed

that after removing the values of one time lag and two time lags after timestamp  $t$ , where one time lag was 6 hours, the PACF values were still bigger. That is to say, latitude and longitude of lag 1 and lag 2 had a high correlation with the meteorological values at the current timestamp. Therefore, when time lag was set to 2, namely time step was 2, the model can get the most accurate forecasting results. Thus, we implemented the experiments in the following processes by using the previous 12 hours data.

**C. EXPERIMENT SETTING**

To compare our proposed model with traditional tropical cyclone track prediction methods and deep neural network models, three traditional methods and six deep neural network models were chosen for the experiments. The traditional methods consist of a statistical forecasting method, dynamical and numerical cyclone forecasting technique, and a numerical weather prediction (NWP) model [63]. The statistical forecasting method developed by JTWC, relies on a climatologically aware forecasting technique, which is a climatology forecasting technique that produces 24-, 48- and 72-h forecasts through an unweighted averaging of the motion of previous cyclones in the same basin [64]. The Sanders Barotropic technique is a typical dynamical and numerical cyclone forecasting technique that can produce 12-, 24-, 48- and 72-h forecasts [65]. The best results from Goerss *et al.* [66] research on the predictive effects of various NWP models over the Western North Pacific Ocean since 1992 were chosen as the NWP model results.

The deep learning models consist of RNN, GRU, CNN, AE-GRU, CNN-RNN, CNN-GRU without a multi-dimensional feature selection layer and our proposed model. RNN is a naive recurrent neural network [18], in which nodes are oriented to form a loop. The internal state of the network can show dynamic temporal behaviors. The CNN model and GRU model have the same structure as the corresponding part of our proposed model. The AE-GRU model is a neural network model based on Auto-Encoder (AE) and GRU, in which an AE is used for extracting features between tropical cyclone tracks and meteorological variables, and a GRU is used for forecasting time series data. The CNN-RNN model is a combination of CNN and RNN, in which GRU is replaced to RNN for comparing the performance of RNN-based model and GRU-based model. The structure of CNN and RNN in CNN-RNN model are the same as the single CNN and single RNN models. In addition, the pure CNN-GRU model refers to our proposed model but without the multi-dimensional feature selection layer. All of these models used the same input dataset as our proposed model.

The aim of the experiment is to predict the tropical cyclone track using the previous 12 hours data. The input data of each model contained longitude, latitude and the meteorological features. The output was the center location of the target tropical cyclone at the next timestamp. The experiment of each model was run 10 times, and the maximum epoch was 100. In order to get the optimal prediction results, multi-layer

TABLE 4. Models parameters.

Model	parameters	value
The proposed model	Training method (for prediction model)	Adaptive Moment Estimation (Adam)
	Kernel size of convolution layer	2×1
	Kernel size of pooling layer	2×1
	Number of convolution layer	2
	Number of pooling layer	2
	Number of GRU layer	3 (Latitude); 2 (Longitude)
	Number of GRU nodes	100 (Latitude); 50 (Longitude)
	Learning rate	0.001
	Batch size	64
	Loss function	Mean squared error (MSE)
CNN	Kernel size of convolution layer	2×1
	Kernel size of pooling layer	2×1
	Number of convolution layer	2
	Number of pooling layer	2
GRU	Number of GRU layers	3
	Number of each GRU nodes	100
RNN	Number of RNN layers	3
	Number of each RNN nodes	100
AE-GRU	Number of AE layers	3
	Number of hidden layer nodes of AE	2
	Number of GRU layers	3
	Number of each GRU nodes	100
CNN-RNN	Same as the single CNN model and single RNN model	
CNN-GRU (Without multi-dimensional feature selection layer)	Same as our proposed model	

GRU and RNN were tested respectively, the number of neurons and learning rate were adjusted correspondingly. Ultimately, the best parameters setting for all models were listed in Table 4, where the batch-size, learning rate, training method (for prediction model) and loss function were the same as the parameters setting in our proposed model.

D. RESULTS

The Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) were used as assessment index between the normalized prediction value and the observation value. MAE denotes the mean absolute error between the predicted track and the observation track, the mathematical formula is shown in Equation (19).

$$MAE = \frac{\sum_{i=1}^n |P_i - O_i|}{n} \tag{19}$$

where  $n$  denotes the size of the testing data,  $P_i$  is the predicted value and  $O_i$  is the observation value. RMSE is the average of the root mean of squared error between the predicted track and the observation track, which is calculated

by Equation (20):

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}} \tag{20}$$

MAPE denotes the average error ratio of the correct values, which considers not only the error between the predicted value and the observation value, but also the ratio between the error and the observation value. The formula is in Equation (21).

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{P_i - O_i}{O_i} \right| \tag{21}$$

In general, the smaller value of these measurements indicates a more accurate prediction result.

TABLE 5. The 12-h, 24-h, 48-h,72-h forecast errors for the proposed method and traditional methods.

Methods	12-hour	24-hour	48-hour	72-hour
Climatologically-aware forecasting technique	-	> 200 km	>400 km	>600 km
Sanders barotropic technique	102 km	215 km	463 km	741 km
NWP model	100 km	170 km	300 km	400 km
Our proposed method	102.32 km	129.77 km	162.22 km	207.40 km

To compare the performance of three traditional methods and our proposed model, the performance measurement results are listed in Table 5, which indicates the average APE of 12-, 24-, 48- and 72-h tropical cyclone forecasts in the northwestern Pacific. The average absolute position error (APE) was used since traditional methods [64]–[66] used distance in kilometers to measure the error between the predicted location and the real location. The calculation of APE is shown in Equation (22), where  $R$  is the earth radius,  $Lat_{real}, Lon_{real}$  indicate the coordinates of real tropical cyclone position, and  $Lat_{pred}, Lon_{pred}$  is the predicted location. A larger value of APE indicates a more significant prediction error. According to the result of the multi-dimensional feature selection layer, the trajectories and meteorological variables data of previous 12 hours is used for 6-h forecasting. Using the same method, the previous 24-h, 36-h, 60-h, and 80-h data were used to predict the location at the next 12-h, 24-h, 48-h and 72-h, respectively. And there was no record of the 12-h forecast of the Climatologically-aware forecasting technique. From Table 5, it was obviously that our model outperformed the three traditional methods on 24-h, 48-h and 72-h forecast. Moreover, with an increase in the forecast time, the performances of the traditional methods decreased rapidly. In comparison, our proposed method showed a relatively stable performance for a longer forecast, the largest error was no more than 210 km, where the error of Sanders Barotropic technique had exceed to 700 km.

$$APE = R \times \text{Arccos} (\sin (Lat_{real}) \times \sin (Lat_{pred}) \times \cos (Lon_{real} - Lon_{pred}) + \cos (Lat_{real}) \times \cos (Lat_{pred})) \times \pi / 180 \tag{22}$$

TABLE 6. The prediction results using evaluation metrics of RMSE, MAE, and MAPE.

Algorithms	Dataset	lat			lon		
		MAE	RMSE	MAPE	MAE	RMSE	MAPE
RNN	Train	0.005312	0.006717	0.007452	0.006403	0.010114	0.025302
	Validation	0.004991	0.006062	0.006848	0.006283	0.089503	0.020700
	Test	0.005611	0.007734	0.008132	0.006526	0.010314	0.020780
GRU	Train	0.004112	0.004999	0.005140	0.004276	0.004981	0.016520
	Validation	0.004434	0.005799	0.005589	0.006031	0.007835	0.021936
	Test	0.004898	0.007551	0.006535	0.006407	0.008310	0.0216414
CNN	Train	0.022812	0.013241	0.009752	0.037701	0.050357	0.073201
	Validation	0.021634	0.010564	0.009631	0.033949	0.051892	0.069315
	Test	0.020651	0.011452	0.008658	0.032131	0.049201	0.067620
AE-GRU	Train	0.002265	0.003434	0.003084	0.005603	0.007857	0.022672
	Validation	0.002332	0.005383	0.003166	0.005718	0.011287	0.023257
	Test	0.002168	0.002835	<b>0.002977</b>	0.005734	0.007591	0.023257
CNN-RNN	Train	0.003461	0.004752	0.004552	0.006843	0.009238	0.019254
	Validation	0.003892	0.004172	0.004951	0.006634	0.008835	0.018976
	Test	0.003633	0.004043	0.004011	0.006490	0.008595	0.018762
CNN-GRU (without multi-dimensional feature selection layer)	Train	0.002198	0.003901	0.003159	0.004759	0.007891	0.021079
	Validation	0.002520	0.004078	0.003161	0.004631	0.009132	0.022447
	Test	0.002159	0.002818	0.003146	0.004432	0.007724	0.019371
The proposed model	Train	0.002131	0.003531	0.003242	0.003850	0.007383	0.013099
	Validation	0.002196	0.003391	0.003518	0.003542	0.007167	0.013050
	Test	<b>0.002101</b>	<b>0.002746</b>	0.003011	<b>0.003937</b>	<b>0.006789</b>	<b>0.014473</b>

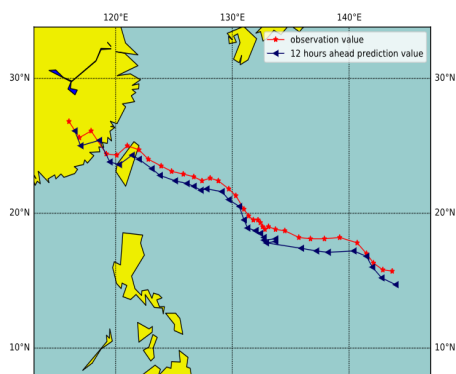


FIGURE 7. The track prediction of Typhoon Tip.

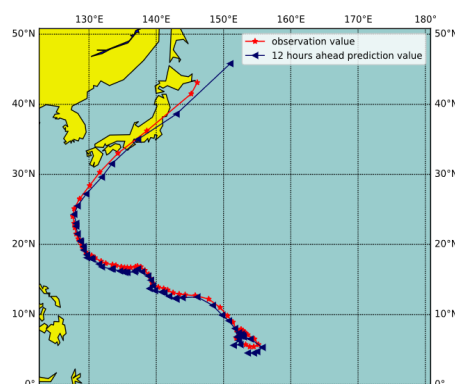


FIGURE 9. The track prediction of Typhoon Dujan.

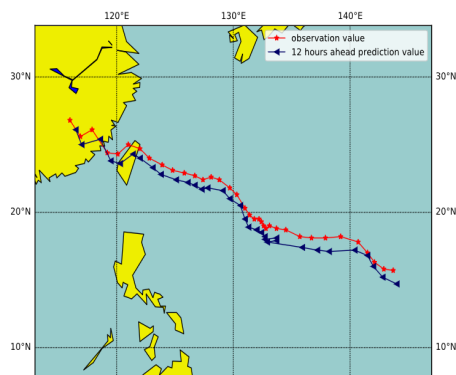


FIGURE 8. The track prediction of Typhoon Meranti.

To compare the performances of RNN, GRU, CNN, AE-GRU, CNN-RNN, CNN-GRU without multi-dimensional feature selection layer and our proposed model, we list the performance measurements results in Table 6, which contains the training results, validation results and

testing results to predict the 6-hour longitude value and latitude value, respectively. The predicted results were based on the previous 12 hours tropical cyclones records, which means the time step was 2 in the experiment. The best performance of the testing results were emphasized in bold fonts. From Table 6, it was obvious that our proposed model can achieve the best performance of almost all measurements, the predicted longitude and latitude were most close to the real values. The single CNN model had the worst performance because it had no capability of extracting temporal features in time series data. And when comparing the GRU-based models with the RNN-based models, it was evident that the GRU-based model can outperform the RNN-based model, which proved that the RNN model was poor in dealing with long-term dependency of tropical cyclone tracks. Additionally, comparing the GRU model with the CNN-GRU model without feature selection, the measurement results showed that the CNN-GRU model without feature selection had a better result than the single GRU model.

This indicated that a CNN layer was significant in tropical cyclone track prediction because the CNN model can help to extract implicit features between meteorological variables and trajectories. And comparing AE-GRU model with the pure CNN-GRU model (without multi-dimensional feature selection layer), the CNN-GRU model performed slightly better, which demonstrated that a CNN layer contributed more than an AE layer in tropical cyclone track prediction. At last, comparing the pure CNN-GRU model (without multi-dimensional feature selection layer) with our proposed model (with multi-dimensional feature selection layer), we can see that our proposed model had less error, which proved that the multi-dimensional feature selection layer that chose the most correlated meteorological variables and time range can make a positive effect in tropical cyclone track prediction.

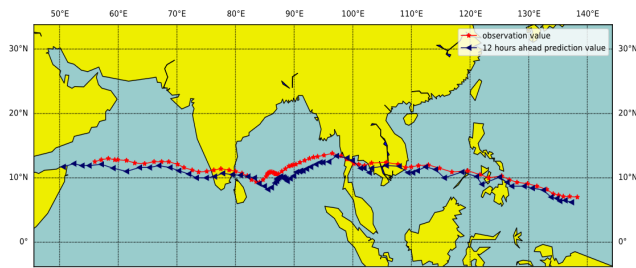


FIGURE 10. The track prediction of Typhoon Haiyan.

TABLE 7. The prediction results using evaluation metrics of RMSE, MAE, and MAPE for Typhoon Tip, Meranti, Dujuan, Haiyan.

Typhoon	Index	MAE	RMSE	MAPE
Tip	Latitude	0.00550	0.01041	0.02166
	Longitude	0.00261	0.00454	0.00335
	Avg	0.00405	0.00748	0.01251
Meranti	Latitude	0.00357	0.00452	0.01075
	Longitude	0.00182	0.00254	0.00251
	Avg	0.00270	0.00353	0.00663
Dujuan	Latitude	0.00332	0.00466	0.00934
	Longitude	0.00197	0.00283	0.00268
	Avg	0.00265	0.00375	0.00601
Haiyan	Latitude	0.00384	0.00523	0.02201
	Longitude	0.00926	0.01065	0.02020
	Avg	0.00655	0.00794	0.02111

Finally, in order to display the prediction effect of our proposed model better, we chose Typhoon Meranti, Haiyan, Tip, and Dujuan for visualization. The results of using previous 12 hours data to predict the tracks of Typhoon Tip, Meranti, Dujuan, and Haiyan were shown in Figure 7, 8, 9, and 10 respectively. The red color represented the real tropical cyclone track and the blue color represented the predicted track. The X-axis was longitude and the Y-axis was latitude. Table 7 showed the average results of MAE, RMSE, and MAPE between the normalized predicted value and observation value of the target typhoons. We can see that MAE and RMSE of the four typhoons were up to 0.001 level for both longitude and latitude. However, the model had a relatively larger error in predicting the longitude of Typhoon Haiyan, which was because that Typhoon Haiyan had a wider range of longitude motion in its life.

## V. CONCLUSION

In this paper, we proposed a novel deep learning model based on CNN and GRU models to predict the tropical cyclone’s tracks using historical tropical cyclones data and multiple meteorological variables. Moreover, for eliminating data redundancy and reducing the complexity of the prediction model, a multi-dimensional feature selection layer based on the Granger Causality test, ACF, and PACF was conducted. The Granger Causality test can analyze the causality between meteorological factors and tracks and get rid of the factors that have lower correlation with the track. The ACF and PACF can determine the most correlated time range for the sequential model. Then, the CNN layer was implemented to model and extract deep features from the selected meteorological variables and tropical cyclone locations. At last, the GRU layer was introduced to learn the nonlinearity and complexity of time-series information from the feature vectors, and then predict the center location of the target tropical cyclone at the next timestamp. Our proposed model can effectively handle the time series data and improve the generalization capability. The experimental results that using the JWTC dataset to predict the tropical cyclone tracks reveal that our proposed model can achieve a better result than traditional deep learning models, where the average MAE of longitude and latitude were only around 0.002 and 0.003, respectively.

In conclusion, the proposed model is suitable for processing tropical cyclone tracks with meteorological features. But there are still some limitations. First, in order to achieve the uniformity of dataset and experiment settings, some earlier prediction methods were used in comparison experiment. Second, some tropical cyclones happened on a close time range, they may influence each others moving track. Additionally, one tropical cyclone may split into multiple cyclones. And the tropical cyclones with different intensities may have different influences on predicting the center location of the next timestamp. In the future, we will compare our proposed method with more recent numeric and statistical methods and improve our model by considering multiple tropical cyclones and taking tropical cyclone intensities into consideration.

## REFERENCES

- [1] (Jun. 2019). *The Central People’s Government of the People’s Republic of China*. [Online]. Available: [http://www.gov.cn/xinwen/2018-02/01/content\\_5262947.htm](http://www.gov.cn/xinwen/2018-02/01/content_5262947.htm)
- [2] Y. Jinhua, T. Jiexiang, and D. Yuhan, “Analyses in errors and their causes of Chinese typhoon track operational forecasts,” (in Chinese), *Meteorol. Monthly*, vol. 38, no. 6, pp. 695–700, 2012.
- [3] J. S. Goerss, “Tropical cyclone track forecasts using an ensemble of dynamical models,” *Monthly Weather Rev.*, vol. 128, no. 4, pp. 1187–1193, Apr. 2000.
- [4] T. R. Knutson, J. L. McBride, J. Chan, K. Emanuel, G. Holland, C. Landsea, I. Held, J. P. Kossin, A. K. Srivastava, and M. Sugi, “Tropical cyclones and climate change,” *Nature Geosci.*, vol. 3, no. 3, pp. 157–163, 2010.
- [5] S. Wang, J. Cao, and P. S. Yu, “Deep learning for spatio-temporal data mining: A survey,” 2019, *arXiv:1906.04928*. [Online]. Available: <http://arxiv.org/abs/1906.04928>
- [6] T. M. Hall and S. Jewson, “Statistical modelling of North Atlantic tropical cyclone tracks,” *Tellus A, Dyn. Meteorol. Oceanogr.*, vol. 59, no. 4, pp. 486–498, Jan. 2007.

- [7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [8] R. Collobert and J. Weston, "A unified architecture for natural language processing: Deep neural networks with multitask learning," in *Proc. 25th Int. Conf. Mach. Learn. (ICML)*, 2008, pp. 160–167.
- [9] R. Dey and F. M. Salem, "Gate-variants of gated recurrent unit (GRU) neural networks," in *Proc. IEEE 60th Int. Midwest Symp. Circuits Syst. (MWSCAS)* Aug. 2017, pp. 1597–1600, doi: 10.1109/MWSCAS.2017.8053243.
- [10] K. Cho, B. van Merriënboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," 2014, *arXiv:1409.1259*. [Online]. Available: <http://arxiv.org/abs/1409.1259>
- [11] K. A. Emanuel, "Thermodynamic control of hurricane intensity," *Nature*, vol. 401, no. 6754, pp. 665–669, 1999.
- [12] H. C. Weber, "Hurricane track prediction using a statistical ensemble of numerical models," *Monthly Weather Rev.*, vol. 131, no. 5, pp. 749–770, May 2003.
- [13] M. DeMaria, M. Mainelli, L. K. Shay, J. A. Knaff, and J. Kaplan, "Further improvements to the statistical hurricane intensity prediction scheme (SHIPS)," *Weather Forecasting*, vol. 20, no. 4, pp. 531–543, Aug. 2005.
- [14] C. Davis, W. Wang, S. S. Chen, Y. Chen, K. Corbosiero, M. DeMaria, J. Dudhia, G. Holland, J. Klemp, J. Michalak, H. Reeves, R. Rotunno, C. Snyder, and Q. Xiao, "Prediction of landfalling hurricanes with the advanced hurricane WRF model," *Monthly Weather Rev.*, vol. 136, no. 6, pp. 1990–2005, Jun. 2008.
- [15] M. Grecu and W. F. Krajewski, "An efficient methodology for detection of anomalous propagation echoes in radar reflectivity data using neural networks," *J. Atmos. Ocean. Technol.*, vol. 17, no. 2, pp. 121–129, Feb. 2000.
- [16] S. Alemany, J. Beltran, A. Perez, and S. Ganzfried, "Predicting hurricane trajectories using a recurrent neural network," 2018, *arXiv:1802.02548*. [Online]. Available: <https://arxiv.org/abs/1802.02548>
- [17] M. Moradi Kordmahalleh, M. Gorji Sefidmazgi, and A. Homaifar, "A sparse recurrent neural network for trajectory prediction of Atlantic hurricanes," in *Proc. Genetic Evol. Comput. Conf. (GECCO)*, 2016, pp. 957–964.
- [18] J. Vohradsky, "Neural model of the genetic network," *J. Biol. Chem.*, vol. 276, no. 39, pp. 36168–36173, 2001.
- [19] J. Chen, L. Yang, Y. Zhang, M. Alber, and D. Z. Chen, "Combining fully convolutional and recurrent neural networks for 3D biomedical image segmentation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 3036–3044.
- [20] S. J. Lim, S. J. Jang, J. Y. Lim, and J. H. Ko, "Classification of snoring sound based on a recurrent neural network," *Expert Syst. Appl.*, vol. 123, pp. 237–245, Jun. 2019.
- [21] L. Chen, X. Pan, Y.-H. Zhang, M. Liu, T. Huang, and Y.-D. Cai, "Classification of widely and rarely expressed genes with recurrent neural network," *Comput. Struct. Biotechnol. J.*, vol. 17, pp. 49–60, Dec. 2018.
- [22] A. M. Rather, A. Agarwal, and V. N. Sastry, "Recurrent neural network and a hybrid model for prediction of stock returns," *Expert Syst. Appl.*, vol. 42, no. 6, pp. 3234–3241, Apr. 2015.
- [23] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, and Y. Bengio, "Show, attend and tell: Neural image caption generation with visual attention," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 2048–2057.
- [24] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [25] S. Gao, P. Zhao, B. Pan, Y. Li, M. Zhou, J. Xu, S. Zhong, and Z. Shi, "A nowcasting model for the prediction of typhoon tracks based on a long short term memory neural network," *Acta Oceanologica Sinica*, vol. 37, no. 5, pp. 8–12, May 2018.
- [26] S. Kim, H. Kim, J. Lee, S. Yoon, S. E. Kahou, K. Kashinath, and M. Prabhat, "Deep-hurricane-tracker: Tracking and forecasting extreme climate events," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2019, pp. 1761–1769.
- [27] D. L. Minh, A. Sadeghi-Niaraki, H. D. Huy, K. Min, and H. Moon, "Deep learning approach for short-term stock trends prediction based on two-stream gated recurrent unit network," *IEEE Access*, vol. 6, pp. 55392–55404, 2018.
- [28] R. Zhao, D. Wang, R. Yan, K. Mao, F. Shen, and J. Wang, "Machine health monitoring using local feature-based gated recurrent unit networks," *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1539–1548, Feb. 2018.
- [29] B. Hidasi, M. Quadrana, A. Karatzoglou, and D. Tikk, "Parallel recurrent neural network architectures for feature-rich session-based recommendations," in *Proc. 10th ACM Conf. Recommender Syst.*, Sep. 2016, pp. 241–248.
- [30] M. Ravanelli, P. Brakel, M. Omologo, and Y. Bengio, "Light gated recurrent units for speech recognition," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 2, no. 2, pp. 92–102, Apr. 2018.
- [31] X. Su, "Using deep learning model for meteorological satellite cloud image prediction," in *Proc. AGU Fall Meeting Abstracts*, vol. 2017, Dec. 2017, p. IN13B-0064
- [32] J. A. Carta, P. Cabrera, J. M. Matías, and F. Castellano, "Comparison of feature selection methods using ANNs in MCP-wind speed methods. A case study," *Appl. Energy*, vol. 158, pp. 490–507, Nov. 2015.
- [33] J.-H. Seo, Y. H. Lee, and Y.-H. Kim, "Feature selection for very short-term heavy rainfall prediction using evolutionary computation," *Adv. Meteorol.*, vol. 2014, pp. 1–15, Jan. 2014.
- [34] M. C. McGraw and E. A. Barnes, "Memory matters: A case for Granger causality in climate variability studies," *J. Climate*, vol. 31, no. 8, pp. 3289–3300, Apr. 2018.
- [35] C. Feng, M. Cui, B.-M. Hodge, and J. Zhang, "A data-driven multi-model methodology with deep feature selection for short-term wind forecasting," *Appl. Energy*, vol. 190, pp. 1245–1257, Mar. 2017.
- [36] J. Wang and G. Song, "A deep spatial-temporal ensemble model for air quality prediction," *Neurocomputing*, vol. 314, pp. 198–206, Nov. 2018.
- [37] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural Comput.*, vol. 1, no. 4, pp. 541–551, Dec. 1989.
- [38] K. Xu, Y. Feng, S. Huang, and D. Zhao, "Semantic relation classification via convolutional neural networks with simple negative sampling," 2015, *arXiv:1506.07650*. [Online]. Available: <http://arxiv.org/abs/1506.07650>
- [39] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 1097–1105.
- [40] Y. Kim, "Convolutional neural networks for sentence classification," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2014, pp. 1746–1751.
- [41] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. 28th Int. Conf. Neural Inf. Process. Syst. (NIPS)*, vol. 1. Cambridge, MA, USA: MIT Press, 2015, pp. 802–810.
- [42] E. Racah, C. Beckham, T. Maharaj, S. E. Kahou, M. Prabhat, and C. Pal, "Extremeweather: A large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events," Tech. Rep., Dec. 2017, pp. 3405–3416. [Online]. Available: <https://papers.nips.cc/paper/6932-extremeweather-a-large-scale-climate-dataset-for-semi-supervised-detection-localization-and-understanding-of-extreme-weather-events>
- [43] Z. Zhu, L. Zhuo, P. Qu, K. Zhou, and J. Zhang, "Extreme weather recognition using convolutional neural networks," in *Proc. IEEE Int. Symp. Multimedia (ISM)*, Dec. 2016, pp. 621–625.
- [44] M. Elhoseiny, S. Huang, and A. Elgammal, "Weather classification with deep convolutional neural networks," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 3349–3353.
- [45] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica, J. Econometric Soc.*, vol. 37, no. 3, pp. 424–438, Aug. 1969.
- [46] C.-J. Huang and P.-H. Kuo, "A deep CNN-LSTM model for particulate matter (PM<sub>2.5</sub>) forecasting in smart cities," *Sensors*, vol. 18, no. 7, p. 2220, Jul. 2018.
- [47] (Jan. 2019). *Western North Pacific Ocean Best Track Data*. [Online]. Available: <http://www.metoc.navy.mil/jtwc/jtwc.html?western-pacific>
- [48] L. A. Acosta, E. A. Eugenio, P. B. M. Macandog, D. B. Magcale-Macandog, E. K. Lin, E. R. Abucay A. L. Cura, M. G. Primavera, "Loss and damage from typhoon-induced floods and landslides in the Philippines: Community perceptions on climate impacts and adaptation options," *Int. J. Global Warming* vol. 9, no. 1, pp. 33–65, 2016.
- [49] J. Han, D. Zhang, G. Cheng, N. Liu, and D. Xu, "Advanced deep-learning techniques for salient and category-specific object detection: A survey," *IEEE Signal Process. Mag.*, vol. 35, no. 1, pp. 84–100, Jan. 2018.
- [50] M. Kim, M.-S. Park, J. Im, S. Park, and M.-I. Lee, "Machine learning approaches for detecting tropical cyclone formation using satellite data," *Remote Sens.*, vol. 11, no. 10, p. 1195, May 2019.

- [51] B. Feng, "A neural network regression model for tropical cyclone forecast," in *Proc. Int. Conf. Mach. Learn. Cybern.*, vol. 7, Aug. 2005, pp. 4122–4128.
- [52] J. Jin, M. Li, and L. Jin, "Data normalization to accelerate training for linear neural net to predict tropical cyclone tracks," *Math. Problems Eng.*, vol. 2015, pp. 1–8, Jul. 2015.
- [53] R. S. T. Lee and J. N. K. Liu, "Tropical cyclone identification and tracking system using integrated neural oscillatory elastic graph matching and hybrid RBF network track mining techniques," *IEEE Trans. Neural Netw.*, vol. 11, no. 3, pp. 680–689, May 2000.
- [54] R. Pradhan, R. S. Aygun, M. Maskey, R. Ramachandran, and D. J. Cecil, "Tropical cyclone intensity estimation using a deep convolutional neural network," *IEEE Trans. Image Process.*, vol. 27, no. 2, pp. 692–702, Feb. 2018.
- [55] X. Dong, J. Shen, D. Wu, K. Guo, X. Jin, and F. Porikli, "Quadruplet network with one-shot learning for fast visual object tracking," *IEEE Trans. Image Process.*, vol. 28, no. 7, pp. 3516–3527, Jul. 2019.
- [56] Z. Liang and J. Shen, "Local semantic siamese networks for fast tracking," *IEEE Trans. Image Process.*, vol. 29, no. 99, pp. 3351–3364, Dec. 2019.
- [57] J. Shen, Z. Liang, J. Liu, H. Sun, L. Shao, and D. Tao, "Multiobject tracking by submodular optimization," *IEEE Trans. Cybern.*, vol. 49, no. 6, pp. 1990–2001, Jun. 2019.
- [58] S. Giffard-Roisin et al., "Fused deep learning for hurricane track forecast from reanalysis data," Tech. Rep., 2018.
- [59] W. Wang and J. Shen, "Deep visual attention prediction," *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2368–2378, May 2018.
- [60] X. Dong and J. Shen, "Triplet loss in siamese network for object tracking," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 459–474.
- [61] X. Dong, J. Shen, W. Wang, L. Shao, H. Ling, and F. Porikli, "Dynamical hyperparameter optimization via deep reinforcement learning in tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Nov. 29, 2019, doi: 10.1109/TPAMI.2019.2956703.
- [62] J. Shen, X. Tang, X. Dong, and L. Shao, "Visual object tracking by hierarchical attention siamese network," *IEEE Trans. Cybern.*, early access, Sep. 12, 2019, doi: 10.1109/TCYB.2019.2936503.
- [63] C. Roy and R. Kovordányi, "Tropical cyclone track forecasting techniques—A review," *Atmos. Res.*, vol. 104, pp. 40–69, Feb. 2012.
- [64] R. A. Jeffries and R. J. Miller, "Tropical cyclone forecasters reference guide 3. Tropical cyclone formation," NRL, Monterey, CA, USA, 1993, Tech. Rep. NRL/PU/7515C93-0007.
- [65] M. Demaria, S. D. Aberson, K. V. Ooyama, and S. J. Lord, "A nested spectral model for hurricane track forecasting," *Monthly Weather Rev.*, vol. 120, no. 8, pp. 1628–1643, Aug. 1992.
- [66] J. S. Goerss, C. R. Sampson, and J. M. Gross, "A history of western north pacific tropical cyclone track forecast skill," *Weather Forecasting*, vol. 19, no. 3, pp. 633–638, Jun. 2004.
- [67] J. Shen, D. Yu, L. Deng, and X. Dong, "Fast online tracking with detection refinement," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 1, pp. 162–173, Jan. 2018.
- [68] Q. Lai, W. Wang, H. Sun, and J. Shen, "Video saliency prediction using spatiotemporal residual attentive networks," *IEEE Trans. Image Process.*, vol. 29, no. 8, pp. 1113–1126, Aug. 2020.
- [69] P. Dong, J. Lian, and Y. Zhang, "A novel data-driven approach for tropical cyclone tracks prediction based on Granger causality and GRU," in *Proc. IEEE Int. Conf. Service Oper. Logistics, Informat. (SOLI)*, Nov. 2019, pp. 70–75.

• • •