Typhoon Comprehensive Classification and Travel Route Intelligent Forecast Based on Convolutional Neural Networks

Tao Wen *, Xiangling Tian, Hongjuan Liu, Lifeng Wei, Chengbi Sheng
School of Mathematics and Big Data, Guizhou Education University,Guiyang, China, 550000
* Corresponding Author Email: wt16666661@163.com

Abstract. This study, based on typhoon data provided by the China Meteorological Administration and focusing on the critical issue of typhoon disaster warning and defense, thoroughly examines two core components: typhoon classification and intelligent prediction of typhoon paths. An innovative evaluation system that integrates time series clustering with hierarchical clustering is developed to standardize various indicators and accurately capture the characteristics of different typhoon categories. Furthermore, the combination of a three-dimensional convolutional neural network (CNN) and an improved long short-term memory network (LSTM) significantly enhances the accuracy and efficiency of path recognition. The experimental results demonstrate that the goodness of fit for wind speed as a category indicator and precipitation as a path indicator reaches 0.9973 and 0.9906, respectively, thereby validating the robust performance of the proposed model. This innovative outcome not only provides a new direction for typhoon path prediction but also substantially advances the field of typhoon disaster warning and defense.

Keywords: 3D-Cnn-Lstm, Time Sequence Cluster, Weight Coefficient, Pressure Gradient Force.

1. Introduction

In the context of global climate change, typhoons, as a kind of highly destructive tropical cyclone phenomenon, have exerted profound influences on human society and the natural environment. The disastrous weather conditions they trigger, such as heavy rainfall, fierce winds, and floods, often cause considerable casualties and property losses, becoming one of the significant natural disasters that seriously threaten human security and economic development. Although there is a certain degree of understanding of the formation mechanism of typhoons, which originate from disturbances in the tropical atmosphere and evolve into cyclones due to the large-scale evaporation of seawater in the high-temperature tropical ocean environment, the formation process is filled with uncertainties due to the interweaving of numerous factors, such as the Coriolis force on the Earth's surface and changes in the atmospheric circulation, which significantly increases the difficulty of related research.

Hou Hui et al. [1] proposed a CNN-LSTM-XGBoost typhoon rainstorm and power meteorology hybrid prediction model based on the attention mechanism, aiming to enhance the prediction accuracy of typhoon-related meteorological elements by integrating multiple technical means; Wang Kun et al. [2] utilized a CNN-LSTM model based on the attention mechanism for trajectory prediction, providing a new idea for typhoon path research; Chen Hongsheng et al. [3] adopted the long short-term memory neural network (LSTM) for the optimization and application of numerical simulation of storm surges, further deepening the understanding of secondary disasters caused by typhoons; Qian Meichao et al. [4] studied the similarity of typhoon movement trajectories based on the DTW algorithm, which is conducive to summarizing the movement rules of typhoons; Zheng Desheng et al. [5] conducted a review of the clustering algorithms for multivariate time series, providing theoretical references for the classification and processing of typhoon data; Meng Jinfeng et al. [6] employed the K-means clustering method and EOF analysis to study the characteristics of typhoon rainstorms in Guigang City, refining the research on the impacts of typhoons from a regional perspective; Qin Zhipeng et al. [7] carried out research on typhoon wave height forecasting methods based on the BO-LSTM neural network model, enriching the technical means for predicting typhoon marine disasters; Yu Taoran [8] expounded on the method of observing typhoon movements from satellite cloud images and radar images, providing an intuitive means for typhoon monitoring; Yao

Fei et al. [9] analyzed the encounter probability of typhoon precipitation and meiyu rainfall intensity during the meiyu period in the Jianghuai region based on the Copula function, revealing the correlations between different meteorological phenomena and typhoons; Zhou Ziying et al. [10] used the C-Vine Copula to study the joint probability distribution of the typhoon disaster chain “wind-rain-tide”, deepening the risk assessment of typhoons from the perspective of disaster chains.

Based on previous studies, this paper is dedicated to constructing an innovative comprehensive prediction model. By deeply integrating satellite remote sensing data and ground meteorological station observation data, and applying data mining and deep learning algorithms, it conducts all-round and refined simulations and predictions of typhoon generation, development, path changes, and the scope of disaster impacts. Through the introduction of new feature extraction techniques and optimized deep learning architectures, it aims to improve the model's ability to capture the complex characteristics of typhoons, thereby providing more accurate, timely, and reliable scientific basis for disaster warning, emergency response, and the formulation of disaster prevention and mitigation strategies.

2. The Fundamental Principles of Time Series Clustering and 3D CNN-LSTM

The data source of this study is: http://typhoon.nmc.cn/web.html

2.1. The Structure of Time Series Clustering

Time series clustering is an unsupervised data analysis approach, intended to partition a given set of time series data objects into multiple clusters, so that the time series within the same cluster are as similar as possible under some predefined similarity metric, while the time series among different clusters are as dissimilar as possible. The structure is depicted in Figure 1.

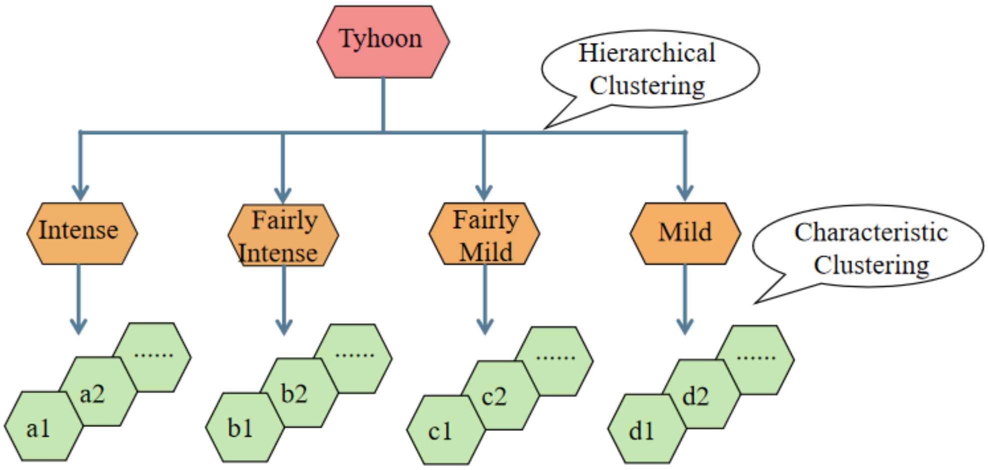

Figure 1: The Hierarchical and Feature Clustering Structure

2.2. The primary clustering of typhoons based on hierarchical characteristics

Employing the pandas and sklearn.model_selection libraries within Python, the features and target variables of typhoons are extracted and subsequently cleansed in accordance with the ensuing four steps:

Step 1: Address the missing values in “Typhoon Intensity” and carry out classified statistics.

Step 2: Classify “Wind Speed”, eliminate outliers, and ensure that “Wind Speed” is in alignment with “Typhoon Intensity”.

Step 3: Format “Typhoon Start Time” and “Typhoon End Time”. For example, if the start time of a typhoon is 1978-12-9, it is identified as abnormal data and automatically adjusted to the correct format of 1978-12-09.

Step 4: Conduct a duplicate check on the complete data and clean the duplicate sections.

In light of the outcomes of the preprocessing of typhoon characteristics, the evaluation scores of each typhoon grade are transformed into a percentile system based on their severity. As the evaluation scores of all grades have been converted into the percentile system, the whitening weight functions for the five grade categories are identical. The comprehensive situation is as follows:

$$f_j^i(x)=\begin{cases}0, x \notin [a_i, b_i] \\ \dfrac{d_i - x}{d_i - c_i}, x \in [c_i, d_i] \\ \dfrac{x - d_i}{e_i - d_i}, x \in [d_i, e_i] \\ i = 1,2,3,4,5\end{cases} \tag{1}$$

$$\begin{cases} f_j^1(x): a_1 = 0, b_1 = 65, c_1 = 40, d_1 = 55, e_1 = 65 \\ f_j^2(x): a_2 = 55, b_2 = 75, c_2 = 55, d_2 = 65, e_2 = 75 \\ f_j^3(x): a_3 = 65, b_3 = 85, c_3 = 65, d_3 = 75, e_3 = 85 \\ f_j^4(x): a_4 = 75, b_4 = 95, c_4 = 75, d_4 = 85, e_4 = 95 \\ f_j^5(x): a_5 = 85, b_5 = 100, c_5 = 85, d_5 = 95, e_5 = 100 \end{cases} \tag{2}$$

Where, the whitening weight functions of each grade indicator regarding the "tropical depression and tropical storm" grade category are lower limit measure whitening weight functions. The whitening weight functions of each indicator with respect to the "super typhoon" grade category are upper limit measure whitening weight functions. The whitening weight functions of each indicator concerning the "severe tropical storm", "typhoon", and "severe typhoon" grade categories are all triangular whitening weight functions, as depicted in Figure 2.

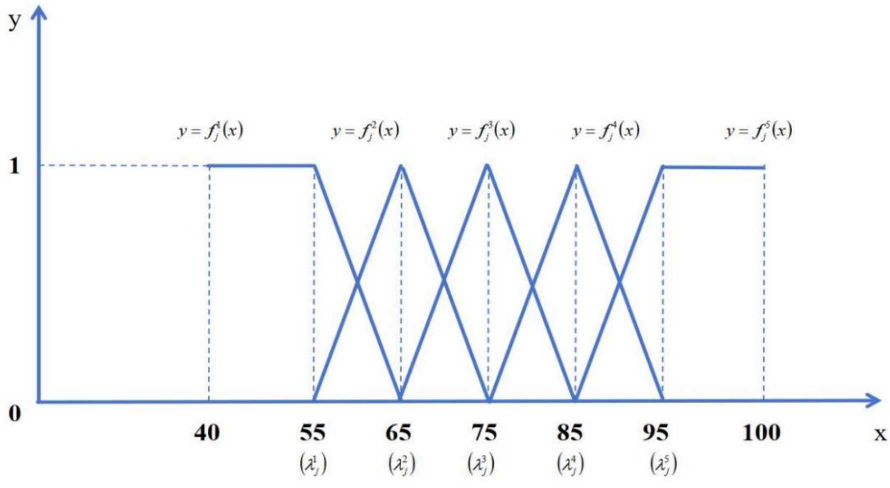

Figure 2: Whitening Weight Function for Grade Evaluation

Select the features that can reflect the characteristics of the typhoon to form the feature vector, which serves as the feature vector for the secondary clustering of the typhoon, that is, $Zn = \{z_1, z_2, ..., z_5\}$. The specific meanings of $z_1, z_2, ..., z_5$ are respectively: the rate of change of the typhoon center pressure, the rate of change of the maximum wind speed of the typhoon, the rate of change of the typhoon duration, the rate of change of the typhoon moving speed, and the rate of change of the typhoon moving direction. To determine the typhoon path, it is necessary to study the rate of change of the typhoon moving direction, and the objective function is:

$$R=\frac{1}{n-2}\sum_{i=2}^{n-1}\frac{|\theta_{i+1}-\theta_i|}{\Delta t} \tag{3}$$

$$s.t.\begin{cases}\vec{d}_i=(x_{i+1}-x_i,y_{i+1}-y_i)\\ |\vec{d}_i|=\sqrt{(x_{i+1}-x_i)^2+(y_{i+1}-y_i)^2}\\ \cos\theta_i=\dfrac{\vec{d}_{i-1}\cdot\vec{d}_i}{|\vec{d}_{i-1}||\vec{d}_i|}\\ \vec{d}_{i-1}\cdot\vec{d}_i=(x_i-x_{i-1})(x_{i+1}-x_i)+(y_i-y_{i-1})(y_{i+1}-y_i)\end{cases} \tag{4}$$

Based on the aforementioned functions, the grades, wind speeds, and latitude membership degrees of each typhoon category were derived, and their radar diagrams were respectively plotted, as depicted in Figure 3.

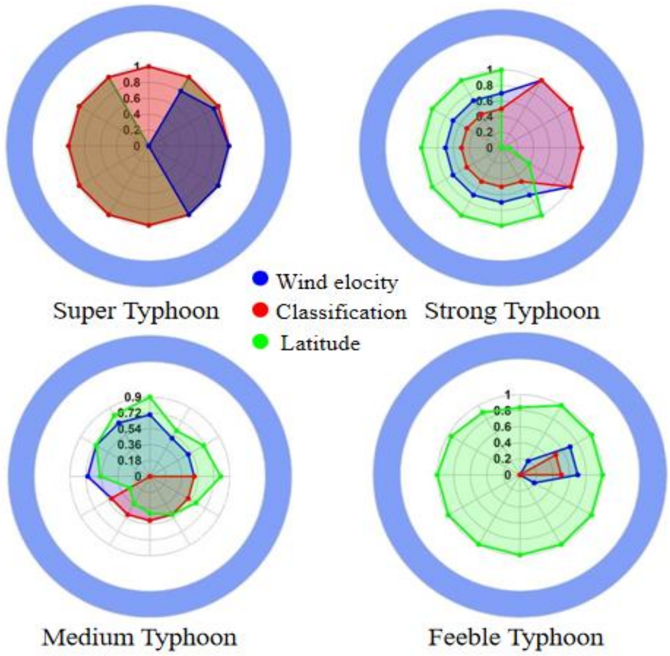

Figure 3 Radar Graph of the Membership Degrees of Each Typhoon

2.3. Construction of the Typhoon Movement Path Model

Firstly, make the assumption:

(1) Assume that the Earth is a standard sphere with a radius of R .

(2) The position of the typhoon is expressed in spherical coordinates (λ, φ, r), among which λ represents the longitude, φ represents the latitude, and $r=R+h$ (h is the height of the typhoon center relative to the Earth's surface, which is usually negligible and regarded as $r\approx R$).

(3) The time variable is t , and the velocity vector of the typhoon is $\vec{v}(t)=(v_\lambda(t),v_\varphi(t),v_r(t))$. Nevertheless, given that the vertical motion of the typhoon is relatively more complex than the horizontal motion and has a relatively minor influence on the path, herein, the focus is primarily on the horizontal velocity, namely $\vec{v}(t)=(v_\lambda(t),v_\varphi(t))$.

Based on the aforementioned assumptions and definitions, five significant factors influencing the typhoon path are introduced as follows:

Environmental impact air current:

$$u(\lambda,\varphi,t)=(u_\lambda(\lambda,\varphi,t),u_\varphi(\lambda,\varphi,t)) \tag{5}$$

It is a function of longitude, latitude and time.

Coriolis Force:

In spherical coordinates, the horizontal component of the Coriolis Force is:

$$\begin{cases} F_{C\lambda} = -2\omega v_{\varphi} \sin\varphi \\ F_{C\varphi} = 2\omega v_{\lambda} \sin\varphi, \end{cases} \tag{6}$$

Where, ω represents the angular velocity of the Earth's rotation.

Forces resulting from factors such as the asymmetric structure of the typhoon itself:

The resultant force of the internal forces in the horizontal direction can be expressed by the following vector:

$$\vec{F}_I(\lambda,\varphi,t) = \left(F_{I\lambda}(\lambda,\varphi,t), F_{I\varphi}(\lambda,\varphi,t)\right) \tag{7}$$

This is also a function of longitude, latitude, and time.

Based on the foregoing fundamental conditions, we can determine the approximate direction of the typhoon's path. Specifically, the dynamics equation of the typhoon path can be employed for calculation, namely:

$$\begin{cases} \vec{F} = m\vec{a} \\ \text{In terms of the longitude direction:} \\ \dfrac{dv_{\lambda}}{dt} = -\dfrac{1}{r\cos\varphi}\dfrac{\partial p}{\partial \lambda} + F_{C\lambda} + F_{I\lambda} + u_{\lambda} \\ \text{In respect of the latitude direction:} \\ \dfrac{dv_{\varphi}}{dt} = -\dfrac{1}{r}\dfrac{\partial p}{\partial \varphi} + F_{C\varphi} + F_{I\varphi} + u_{\varphi}, \end{cases} \tag{8}$$

Where, $\frac{1}{r\cos\varphi}\frac{\partial p}{\partial \lambda}$ denotes the component of the pressure gradient force in the longitude direction, and $\frac{1}{r}\frac{\partial p}{\partial \varphi}$ indicates the component of the pressure gradient force in the latitude direction.

2.4. The Establishment of 3D CNN-LSTM Typhoon Trajectory Prediction Model

Based on 3D CNN, it is conducive for us to extract the spatial characteristics of the atmospheric variable data. Atmospheric variables exhibit three-dimensional correlations in space. 3D CNN can, through convolution operations, learn these complex spatial relationships and capture the impact of variations in atmospheric characteristics at different altitudes, latitudes, and longitudes on the typhoon path. The construction of group convolution is depicted in Figure 4.

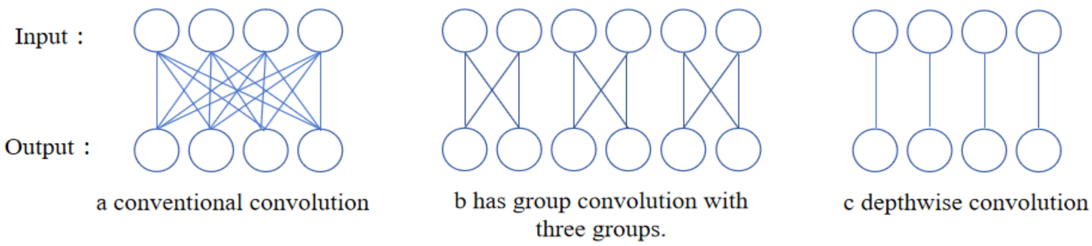

Figure 4 Structure of Group Convolution

Given that the real-time trajectory of a typhoon is controlled by its central position (latitude and longitude), and all are numerical data, therefore, in a mathematical form, the typhoon trajectory prediction model can be represented as follows:

Firstly, define the data dimension as: $N \times M \times P \times D$ (representing length, width, height, and the number of feature channels respectively). Secondly, the dimension of the convolution kernel is: $a \times b \times c \times D$ (the length, width, height, and the number of feature channels of the convolution kernel), the stride is s, and the padding is p. Then, the output dimension after convolution is as follows:

$$\begin{cases} \text{Length}: \left\lfloor \dfrac{N + 2\text{p} - \text{a}}{\text{s}} + 1 \right\rfloor \\ \text{breadth}: \left\lfloor \dfrac{\text{M} + 2\text{p} - \text{b}}{\text{s}} + 1 \right\rfloor \\ \text{height}: \left\lfloor \dfrac{P + 2\text{p} - \text{c}}{\text{s}} + 1 \right\rfloor \end{cases} \tag{9}$$

The value of D is contingent upon the quantity of convolution kernels. Hence, the convolution computation of a solitary convolution kernel in a certain spatial dimension can be represented as:

$$O(x,y,z) = \sum_{i=0}^{a-1} \sum_{j=0}^{b-1} \sum_{k=0}^{c-1} \sum_{l=0}^{D-1} W(i,j,k,l) \bullet I(x+i, y+j, z+k, l) \tag{10}$$

Where, $O(x,y,z)$ is the value of the output at position (x,y,z), $W(i,j,k,l)$ is the value of the convolution kernel at position (i,j,k,l), and $I(x+i,y+j,z+k,l)$ is the value of the output at position $(x+i,y+j,z+k,l)$.

Secondly, for three-dimensional max pooling, the data dimension is still defined as $N \times M \times P \times D$, the dimension of the pooling window is $a \times b \times c$, and the stride is s. Subsequently, the output dimension after pooling is:

$$\begin{cases} \text{Length}: \left\lfloor \dfrac{N - a}{s} + 1 \right\rfloor \\ \text{breadth}: \left\lfloor \dfrac{M - b}{s} + 1 \right\rfloor \\ \text{height}: \left\lfloor \dfrac{P - c}{s} + 1 \right\rfloor \end{cases} \tag{11}$$

The computation of the max pooling operation at a specific position is presented as follows:

$$O(x,y,z) = \max_{i=0}^{a-1} \max_{j=0}^{b-1} \max_{k=0}^{c-1} I(x \times s + i, y \times s + j, z \times s + k) \tag{12}$$

Furthermore, it is necessary to compute the Mean Squared Error (MSE) loss and the Mean Absolute Error (MAE) loss. The formulas are presented as follows:

$$\begin{cases} MSE = \dfrac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \\ MAE = \dfrac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|, \end{cases} \tag{13}$$

Where, y_i denotes the actual typhoon path value (that is, position coordinates), and $\hat{y}_i$ represents the typhoon path value predicted by the model. Subsequently, the weight update formula and bias update formula for the convolutional layer in the backpropagation process can be computed as follows:

$$\begin{cases} \Delta W = -\eta \dfrac{\partial L}{\partial W} \\ \Delta b = -\eta \dfrac{\partial L}{\partial b}, \end{cases} \tag{14}$$

Where, ΔW and Δb respectively represent the update quantities of the weight and the bias, Δb stands for the learning rate, L is the loss function, and $\frac{\partial L}{\partial W}$ and $\frac{\partial L}{\partial b}$ are the partial derivatives of the loss function with respect to the weight and the bias respectively.

3. Results

3.1. Solution to the Model of Typhoon Movement Path

Historical typhoon data, encompassing information such as typhoon positions, movement velocities, environmental airflow conditions, and pressure fields, etc., are imported as the input and validation basis for the model. Representative data samples of the output are presented in Table 1.

Table.1. Latitude and Longitude Information of Sample Typhoons

Typhoon number	longitude	latitude
1	131.6	19.9
18	125.3	23.5
...	...	...
45	120.8	26.9
81	122.9	37.7

The box-and-whisker plot depicting the relationship between the typhoon path and related factors is presented in Figure 5.

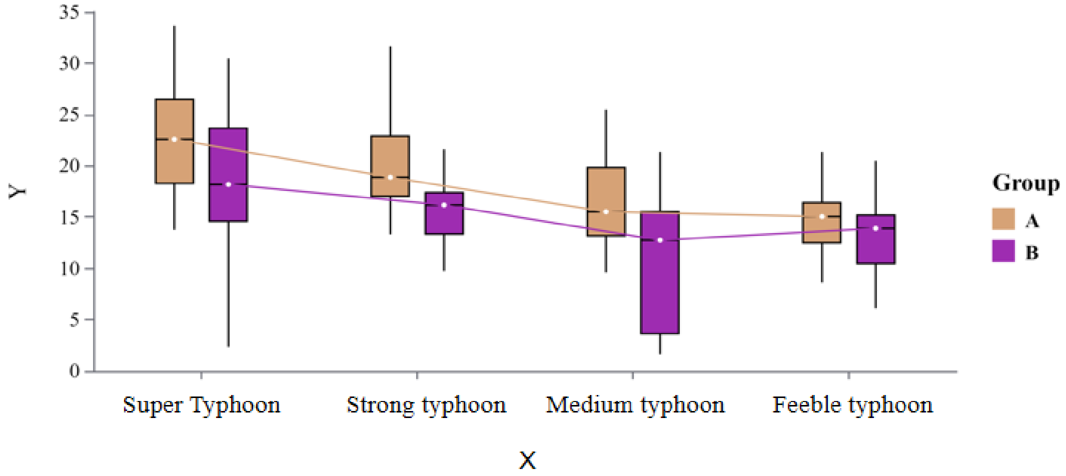

Figure 5 Box-and-Whisker Plot of Typhoon Paths

Based on Table 1 and Figure 5, the quantity of moderate typhoons predominates, followed by that of super typhoons. The majority of typhoons traverse the mid-low latitude regions. Consequently, the approximate path of the typhoon's movement direction can be ascertained. The map depicting the specific direction is presented in Figure 6.

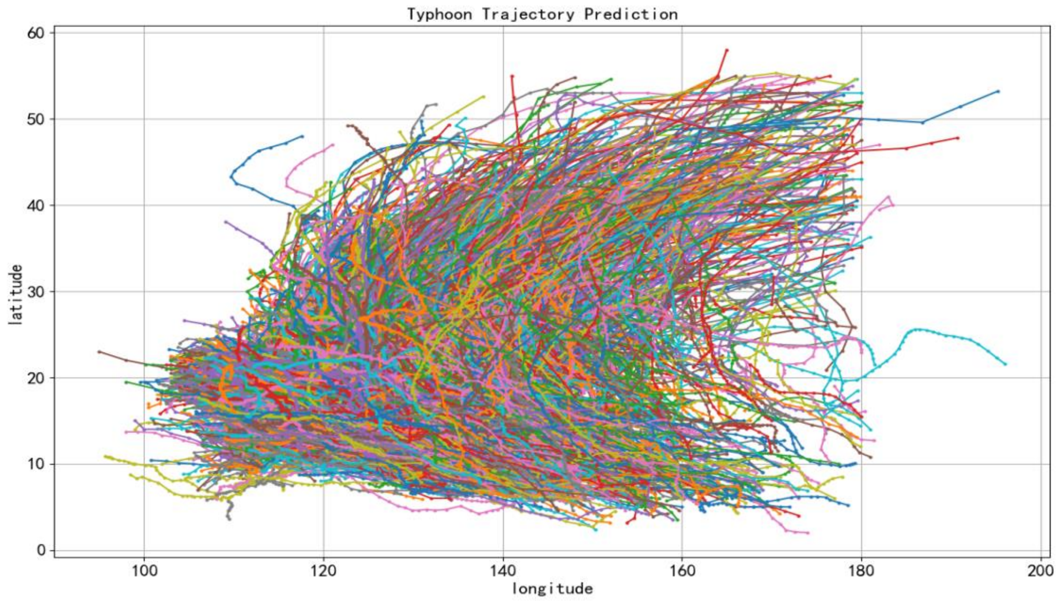

Figure 6 Typhoon Path Prediction

It is conspicuously observable that the activity tracks of the majority of typhoons will ultimately come to an end in the vicinity of the 50th degree of latitude and the 180th degree of longitude.

3.2. 3D CNN-LSTM Typhoon Trajectory Prediction Model Resolution

The accuracy and reliability of the typhoon trajectory prediction model can be verified through computing the error between the actual trajectory and the predicted trajectory by the DTW algorithm. The actual and predicted errors are extracted as presented in Table 2.

Table .2. Error Degree of Actual Path and Predicted Path

Actual longitude	Actual Latitude	Predicted Longitude	Predicted Latitude	Error Degree
124.8	23.8	124.8000001	23.80000004	7.17327E-08
121.4	26.3	121.4000003	26.29999987	1.27629E-07
118.9	29.9	118.9000006	29.90000007	6.38473E-07
122.9	37.7	122.9000005	37.69999973	2.73238E-07

The error comparison is presented as shown in Figure 7.

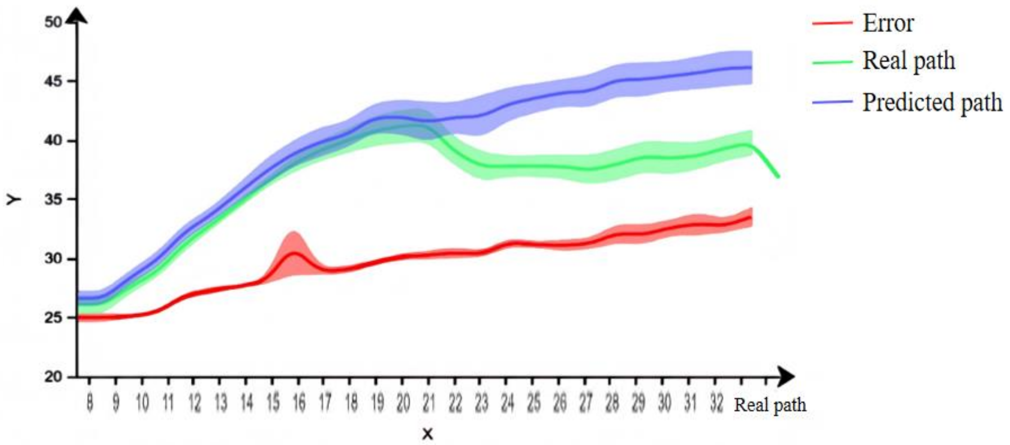

Figure 7 Analysis of Path Prediction Error

Based on the above data, the remote sensing map of the typhoon path obtained through the 3D CNN-LSTM model is shown in Figure 8.

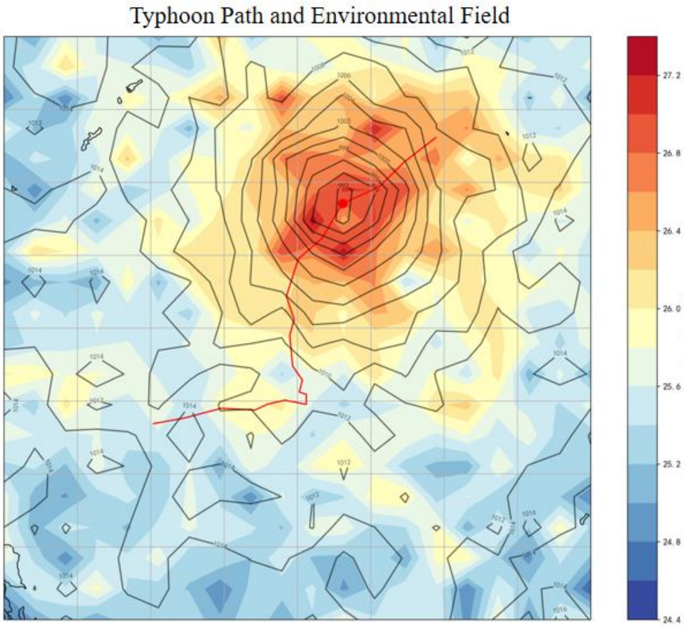

Figure 8 Typhoon Path and Environmental Field

During the course of its movement, a typhoon tends to gradually shift from areas of low temperature and high pressure to those of high temperature and low pressure. In this process, the typhoon may undergo a period of slow movement with relatively small positional changes. However, after the stagnation period concludes, the typhoon's trajectory will undergo a significant alteration. Subsequently, the typhoon will resume its movement towards another area of low temperature and high pressure.

4. Conclusions

Regarding the classification of typhoons, in this paper, time series clustering, primary clustering, and secondary clustering are comprehensively integrated to establish an evaluation system. A time adjustment factor and an iterative optimization strategy are introduced to determine the weight coefficients, thereby achieving a refined classification of typhoons. Weights are allocated to the severe category, relatively severe category, relatively mild category, and mild category to ensure higher homogeneity within each category. Through continuous adjustments, the optimal classification threshold is identified to enhance the accuracy and adaptability of the classification. Concerning the prediction of typhoon paths, in this paper, CNN and LSTM are ingeniously combined to construct a novel mathematical prediction model, extract the spatial features of typhoons and explore the temporal features, and an attention mechanism is introduced to enhance the recognition of long-term dependencies.

In the classification and evaluation of typhoons, this paper constructs a model that addresses the issues of single-feature dependence and inaccurate classification in traditional approaches, making the classification results more refined and reliable and offering novel ideas for subsequent studies. Regarding typhoon path prediction, the constructed CNN-LSTM model effectively improves the processing capability of complex factors and prediction accuracy and overcomes the deficiencies of traditional models in handling the long-term dependency relationships in spatio-temporal feature capture.

The relevant data information output by the model can serve as a reference for disaster prevention and reduction efforts, minimizing the losses caused by typhoon disasters to the greatest extent possible. Future research directions can be further expanded by integrating multi-source data fusion techniques to make more precise predictions of the impact scope and intensity variations of typhoon disasters. For instance, combining satellite remote sensing data, geographic information data, and model output data to conduct in-depth analyses of the spatiotemporal evolution patterns of typhoon disasters and provide more comprehensive information for disaster early warnings. Additionally, in practice, it facilitates relevant practitioners to gain a deeper understanding of the characteristics and changing patterns of typhoons, holding significant reference value for fields such as meteorological research and climate prediction, and promoting the advancement of related domains. Future studies can enhance interdisciplinary collaborations by organically integrating meteorology with disciplines such as oceanography and geography to offer more scientific and effective solutions for dealing with typhoon disasters.

References

[1] Hou Hui, Wu Wenjie, Wei Ruizeng, et al. Attention Mechanism Based CNN-LSTM-XGBoost Electric Power Meteorological Hybrid Forecasting Model of Typhoon Rainstorm[J]. Smart Power, 2024, 52(10): 96-102.

[2] Wang Kun, Zhou Zhichong, Qu Kai, et al. Real-Time Track Prediction of CNN-LSTM Model Based on Attention Mechanism[J]. Journal of Air Force Engineering University, 2023, 24(06): 50-57.

[3] Chen Hongsheng, Lin Xiaogang, Lin Xiaozhen, et al. Application of Long Short-Term Memory neural network for optimization of numerical simulation results of storm surge[J]. Marine Forecasts, 2024, 41(04): 1-10.

[4] Qian Meichao, Huang Yan, Zhang Fang, et al. Research on the Similarity of Typhoon Movement Trajectories Based on DTW Algorithm [J]. Electronic Technology and Software Engineering, 2022, (19): 230-233.

[5] Zheng Desheng, Sun Hanming, Wang Liyuan, et al. A Review of Multivariate Time Series Clustering Algorithms[J]. Journal of Frontiers of Computer Science and Technology, 2024: 1-21.

[6] Meng Jinfeng, Long Junyu, Liang Hong, et al. Analysis the Characteristics of typhoon rainstorm in Guigang based on K-Means Clustering Method and EOF Analysis[J]. Journal of Meteorological Research and Application, 2024, 45(03): 44-49.

[7] Qin Zhipeng, Chen Yongping, Pan Yi, et al. Research on typhoon wave height prediction method based on BO-LSTM neural network model[J]. Haiyang Xuebao, 2024: 1-10.

[8] Yu Taoran. How to Track Typhoon Movements from Satellite Cloud Images and Radar Maps [N]. Jiefang Daily. 2024-9-18.

[9] Yao Fei, Duan Yaobin, Qin Yong, et al. Analysis of the Encounter Probability of Typhoon Precipitation and Meiyu Rain Intensity during the Meiyu Period in the Yangtze-Huaihe River Basin Based on Copula Function [J]. Water Resources and Power, 2024, (11): 41-45+40.

[10] Zhou Ziying, Yang Saini, Liu Xiaoyan, et al. Research on the Joint Probability Distribution of Typhoon Disaster Chain “Wind-Rain-Tide” Based on C-Vine Copula Function [J]. Tropical Geography, 2024, (06): 1036-1046.