atmosphere

MDPI

Article

DGF-Net: A Novel Approach for Tropical Cyclone Path Prediction Using Multimodal Meteorological Data

Yuxue Wang *, Sheng Li and Baoqin Chen

College of Mathematics and Computer, Guangdong Ocean University, Zhanjiang 524094, China; lish_ls@gdou.edu.cn (S.L.); cbqchen@gdou.edu.cn (B.C.)

* Correspondence: 18320384097@stu.gdou.edu.cn

Abstract

Tropical cyclones are among the most destructive meteorological systems on Earth. Accurate track forecasting of tropical cyclones remains a core challenge in atmospheric science, and it is of great significance for disaster prevention and mitigation. This study targets the critical limitations of existing tropical cyclone track forecasting models: the insufficient ability to extract non-linear spatiotemporal features from 3D atmospheric circulation fields and the long-standing bottlenecks in multi-source heterogeneous meteorological data fusion. To address these issues, we propose a Dual-Stream Gated Fusion Network (DGF-Net), a high-precision track forecasting method tailored to the Northwest Pacific basin. The proposed framework takes the Best Track dataset and ERA5 Reanalysis Dataset as primary inputs: a Bidirectional Gated Recurrent Unit (Bi-GRU) is adopted to capture the temporal evolution characteristics of 2D tropical cyclone trajectory sequences, and a SpatioTemporal Convolutional Gated Recurrent Unit (STConvGRU) is used to extract complex non-linear features from 3D atmospheric environmental fields. Then, a multimodal fusion module integrating gating and attention mechanism is constructed to achieve deep fusion of cross-dimensional features, which effectively mines the intrinsic physical correlations between tropical cyclone track evolution and environmental driving factors. Comparative experiments based on historical observational datasets of the Northwest Pacific show that DGF-Net achieves superior forecasting performance, with the 6 h, 12 h, and 24 h Great Circle Distance (GCD) errors of 35.62 km, 43.53 km, and 135.49 km, respectively. The results significantly outperform mainstream baseline models, which validates the effectiveness of DGF-Net in feature extraction and multimodal fusion and provides solid technical support for tropical cyclone disaster prevention and operational decision-making.

Keywords: tropical cyclones; tracking forecast; fused deep learning; gated mechanism

Check for updates

Academic Editor: Qiusheng Li

Received: 25 January 2026
Revised: 28 February 2026
Accepted: 4 March 2026
Published: 6 March 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Tropical cyclones (TCs) [1] are intense warm-core low-pressure systems originating over tropical and subtropical oceanic surfaces, characterized by rapid cyclonic rotation around a central low-pressure eye region. Based on their intensity, TCs are categorized into hierarchical levels, ranging from tropical depressions to typhoons. As one of the most destructive natural hazards globally [2], TCs frequently trigger extreme disaster chains—including torrential rain, gale-force winds, storm surges, and massive waves—leading to large-scale infrastructural damage in coastal regions. The Northwest Pacific basin is particularly notable as the most active and intense region for TC activity, averaging approximately 30 TCs annually, which accounts for over 30% of the global total.

Nearly 70% of these systems directly impact coastal nations such as China and Japan, resulting in catastrophic economic losses and casualties [3]. Consequently, developing high-precision TC track forecasting techniques for this basin is of paramount scientific value and practical urgency for disaster prevention and mitigation.

The methodology for TC track forecasting has evolved through several generations [4]. Early approaches primarily relied on Numerical Weather Prediction (NWP) models based on non-hydrostatic primitive equations, such as the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecasting System (IFS) [5] and the Weather Research and Forecasting (WRF) model [6]. While these models approximate multi-scale interactions via parameterization schemes, they are constrained by high computational costs and extreme sensitivity to initial conditions. Another class of early statistical models constructed forecasting frameworks based on historical correlation analysis and covariance matrices of climatological variables. However, their accuracy was limited by the completeness of hand-crafted features and linear assumptions. For instance, the classic Climatology and Persistence (CLIPER) model [7] fails to adequately account for dynamic environmental factors such as Sea Surface Temperature (SST) anomalies and Vertical Wind Shear (VWS), leading to significant forecasting biases for complex cases involving extreme intensities or anomalous tracks.

Recently, the rapid advancement of artificial intelligence (AI), specifically deep learning, has established it as a pivotal methodology for TC forecasting due to its capacity to process massive multimodal meteorological datasets and extract complex non-linear features. The evolution of DL in TC prediction has progressed from simple architectures to sophisticated systems. Initial efforts focused on Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). Moradi et al. [8] proposed a sparse RNN with a flexible topology for Atlantic hurricane track forecasting, utilizing a customized genetic algorithm to optimize network topology and connection weights. Subsequently, more complex architectures such as Long Short-Term Memory (LSTM) [9] and Gated Recurrent Units (GRUs) [10] were introduced. For example, Alemany et al. [11] employed LSTMs to capture the spatiotemporal correlations between wind shear and SST anomalies, thereby improving directional forecasting accuracy.

As research deepened, it became evident that pure temporal or spatial modeling could not fully characterize the three-dimensional (3D) spatial distribution of atmospheric environmental fields. Consequently, researchers began exploring hybrid models. Kim et al. [12] applied Convolutional LSTM (ConvLSTM) [13] to ERA-Interim reanalysis data, demonstrating that spatiotemporal feature extraction significantly outperforms traditional 2D-CNNs. Subsequent studies integrated physical mechanisms with hybrid networks [14] to enhance predictive capabilities. Despite these technical foundations, existing methods still struggle to fully extract 3D spatial features from multi-layer atmospheric reanalysis data, overlook vertical correlations between different pressure levels, and fail to exploit the inherent physical dependencies between meteorological variables, which limits feature representation and further accuracy gains.

With the continuous upgrading of observation systems such as the Fengyun (FY) satellite series and Global Navigation Satellite System Radio Occultation (GNSS-RO), meteorological data has seen exponential growth, exhibiting high dimensionality, heterogeneity, and strong coupling. This has driven the transition from single-modality to multimodal feature learning. Xu et al. [15] proposed a multimodal TC track prediction study using CMA Best Track data to construct 2D representations based on CLIPER characteristics for a Generalized Linear Model (GLM). Simultaneously, they utilized a Multi-ConvGRU structure with a Recurrent Channel Attention (ReCA) mechanism to process 3D atmospheric reanalysis data, fusing these features via a Wide and Deep framework [16]. Building on this,

Fang et al. [17] further optimized the extraction and fusion strategies using an attention-based Multi-ConvGRU to capture 3D spatial and dynamic temporal features from the ERA5 dataset, adaptively focusing on key regions and physical quantities while suppressing redundant noise. Similarly, Li et al. [18] proposed a fusion model using U-Net for 3D spatial feature extraction from ERA5 data combined with an LSTM encoder–decoder for Best Track data. Other representative works include LSTM-CNN hybrids [19] and Three-Stream Networks based on Multi-Layer Perceptrons (MLPs) [20], designed to achieve multi-source fusion of 1D historical data, 2D surface variables, and 3D pressure-level data. Despite these advances, current multimodal TC forecasting still relies heavily on simple MLP-based concatenation or fixed-weight fusion, lacking sophisticated crossmodal interaction mechanisms. This often results in a lack of semantic consistency between heterogeneous features and the loss of critical physical correlation information during fusion.

To address these challenges, this study proposes the Dual-Stream Gated Fusion Network (DGF-Net), a novel forecasting framework that integrates ERA5 reanalysis data and the CMA Best Track dataset for short-term (6–24 h) TC track prediction in the NWP. To ensure high-precision forecasting, the model utilizes the CMA dataset, which covers central coordinates and intensity for all TCs in the region since 1949, and high-resolution 3D atmospheric fields from ERA5. Architecturally, DGF-Net consists of two core modules: (1) a Multi-scale Spatiotemporal Feature Extraction Module, which employs Bidirectional GRUs (Bi-GRU) for 2D TC track sequence data and a SpatioTemporal Convolutional Gated Recurrent Unit (STConvGRU) for 3D features, incorporating a Squeeze-and-Excitation Network (SENet) to adaptively calibrate key channel-wise features of the 3D atmospheric representations; (2) a Gated Multimodal Fusion Module (GMFM), which combines gating and attention mechanisms to dynamically model interactions between modalities, ensuring semantic consistency between latent representations and explicit physical variables. Validated on the Northwest Pacific samples from 2005 to 2023, DGF-Net achieved a 6 h mean Great Circle Distance (GCD) error of 35.62 km and a 24 h error of 135.49 km, demonstrating its utility for TC early-warning decision support. However, limitations remain: the model is currently optimized for the Northwest Pacific basin, and its generalization to other basins or its stability in long-term forecasting scenarios requires further validation.

The main contributions of this paper are summarized as follows:

1. We propose the STConvGRU module, which employs a dual spatiotemporal convolution strategy to overcome the inherent limitations of traditional temporal convolutions in capturing multi-scale spatial features and multi-layer atmospheric dependencies.
2. We propose the GMFM, which integrates gating and attention mechanisms to capture complex dependencies within multimodal data, thereby enhancing the interpretability of feature interactions and significantly improving overall forecasting performance.

2. Dataset

This study utilizes the CMA TC Best Track dataset [21,22] and the fifth-generation global atmospheric reanalysis dataset (ERA5) released by ECNWF [23] as data foundations for TC track-prediction research. Specifically, the CMA dataset is used to construct a 2D TC feature dataset, which systematically records all TCs formed in the Northwest Pacific basin since 1949 according to this dataset. It provides the position and intensity of each TC at 6 h intervals. The Best Track dataset, regarded as one of the most accurate records of TC tracks, is derived through the integration and correction of multiple data sources. TCs in this dataset are presented in Figure 1. Each line in a different color represents the track of a TC that occurred in the Northwest Pacific between 2005 and 2023. Table 1 provides a detailed description of the variables included in the dataset along with their explanations. This study utilizes 18 years of TC data from 2005 to 2023, comprising 518 TCs

and a total of 15,074 records. To maintain consistency in the time series, the data are uniformly sampled at 6 h intervals, specifically at 00:00, 06:00, 12:00, and 18:00. To better capture the movement trends of TCs, the forecasting factors based on the CMA dataset are expanded by incorporating variables such as acceleration and movement direction.

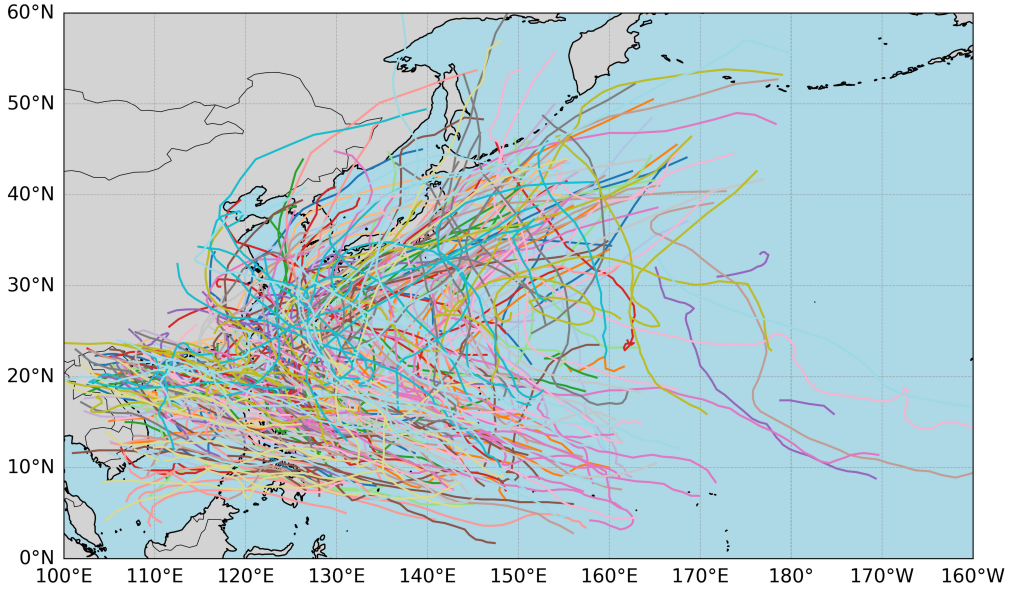


Figure 1. Visualization of TC tracks in the CMA dataset.

Table 1. Explanation of variables in CMA dataset.

Variable Name	Variable Description
YYYY-MM-DD-HH	YYYY year, MM month, DD day, HH hour
I	Intensity marker
LAT	Latitude (0.1° N)
LONG	Longitude (0.1° E)
PRES	Central minimum pressure (hPa)
WND	2-min average maximum near-center wind speed (m/s)
OWD	2-min average wind speed (m/s)

Table 2 presents these extended TC features. LAT_t and $LONG_t$ represent the latitude and longitude of the TC at time t, respectively; WND_t and PRE_t denote the wind speed and atmospheric pressure at time t. $Angle(LAT_t, LON_t)$ represents the angle of the path at time t, and its calculation formula can be expressed as

$$Angle(X_t, Y_t) = \arccos\left(\frac{(X_{t+1} - X_t)(X_t - X_{t-1}) + (Y_{t+1} - Y_t)(Y_t - Y_{t-1})}{\sqrt{(X_{t+1} - X_t)^2 + (Y_{t+1} - Y_t)^2} \cdot \sqrt{(X_t - X_{t-1})^2 + (Y_t - Y_{t-1})^2}}\right). \quad (1)$$

EERA5 is the successor to ERA-Interim, generated by a four-dimensional variational (4D-Var) data assimilation system that fuses global meteorological observations, serving as the current benchmark for high-resolution global atmospheric reanalysis. This study selects ERA5 as the source for 3D environmental features of TCs. ERA5 offers several advantages: temporally, it provides hourly data from 1940 to the present; spatially, it features a horizontal resolution of 0.25° × 0.25° and includes 37 isobaric levels from 1000 hPa to 1 hPa in the vertical direction, forming a fine-grained 3D grid-based atmospheric dataset. The ERA5 dataset contains 16 fundamental atmospheric variables, enabling a comprehensive description of the dynamical, thermodynamical, and moisture characteristics of the

atmosphere, such as zonal wind (u), meridional wind (v), vertical velocity (w), temperature (T), and specific humidity (q).

Table 2. 2D features of tropical cyclone tracks.

ID	Name	Description
X_1	LAT_t	Longitude
X_2	$LONG_t$	Latitude
X_3	$Month$	Month
X_4	$LAT_t - LAT_{t-1}$	First-order difference of longitude
X_5	$LONG_t - LONG_{t-1}$	First-order difference of latitude
X_6	WND_t	Wind speed
X_7	PRE_t	Atmospheric pressure
X_8	$\sqrt{LAT_t}$	Sum of longitude squares
X_9	$\sqrt{LONG_t}$	Sum of latitude squares
X_{10}	$\sqrt{(LAT_t - LAT_{t-1})^2 + (LONG_t - LONG_{t-1})^2}$	Displacement

To construct the 3D TC data, this study adopts the method by Xu [15] to segment the ERA5 reanalysis data. This approach employs a 31° × 31° grid at each time step, using zonal wind (u), meridional wind (v), and geopotential height (z) as inputs. A 31° × 31° grid extracts spatial information covering a diameter of approximately 3445 km. However, according to WMO statistics [24], the average diameter of a TC is 200–500 km; thus, a 31° × 31° grid contains significant redundant background information, increasing noise interference during model training. To address this, Sattar et al. [25] optimized the cropping range by extracting a 3° × 3° local grid at 1° resolution centered on the TC. While this precisely focuses on the core dynamical circulation (covering 222 km), the spatial coverage is too limited to capture the large-scale environmental field features that influence TC movement and development. Synthesizing these two approaches, this study extends 10° in all directions from the TC center, resulting in a 20° × 20° region that corresponds to an 81 × 81 grid matrix. This spatial configuration covers both the TC's core structure and its key peripheral environment while reducing redundant data. Vertically, we select five standard isobaric levels that play a critical role in modulating TC track evolution: 100 hPa, 300 hPa, 500 hPa, 700 hPa, and 900 hPa. Regarding variable selection, Giffard et al. [26] experimentally demonstrated that u, v, and z are highly correlated with TC motion. However, relying solely on these three variables limits the performance ceiling of predictive models. Tan et al. [27] utilized 33 predictors across various categories, including surface thermodynamical conditions and mid-to-low-level circulation fields, proving that multi-physical variable fusion improves forecasting. Combining these insights, we utilize zonal wind, meridional wind, vertical velocity, geopotential height, potential vorticity (PV), and SST as the six key variables for 3D environmental inputs. This scheme avoids a single-variable bias and supplements variables with high correlation to cyclone evolution. The resulting multi-layer 3D environmental time series structure is shown in Figure 2. Based on this preprocessing framework, an integrated time series fusing 2D TC features and 3D environmental data was generated. This framework dynamically records positional evolution throughout the life cycle and 3D atmospheric structural changes, providing high-resolution inputs for capturing complex vortex–environment interactions.

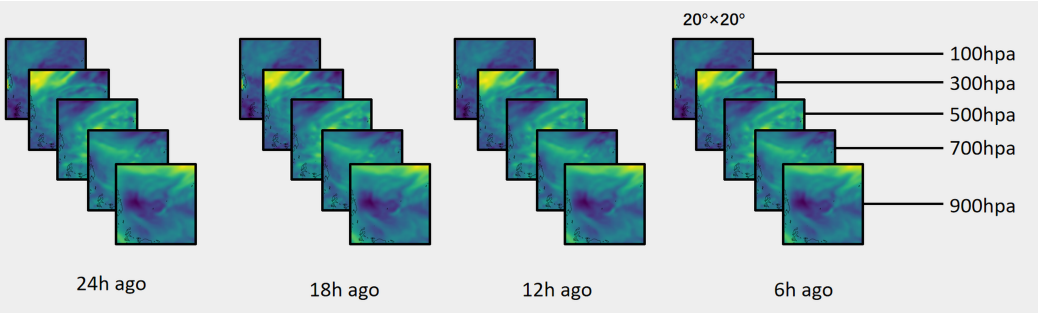


Figure 2. Structure of the TC environmental field time series.

3. Model Introduction

3.1. Model Structure

This study proposes the DGF-Net model to address the challenges of multi-source data in TC track prediction. As depicted in Figure 3, the architecture employs a dual-component design: the Meteorological Data Feature Extraction Layer learns TC characteristics through an RNN-BLOCK module and a CNN-BLOCK module, while the subsequent Multimodal Decoding Layer captures interactive information among different modal data sources and predicts TCs. The RNN-BLOCK processes 2D TC trajectories through Bi-GRU networks for temporal dependency modeling. The CNN-BLOCK handles 3D environmental fields by integrating STConvGRU for spatiotemporal correlations, SENet for channel-wise feature enhancement, and spatial pooling for dimensionality reduction. The subsequent Multimodal Decoding Layer combines the GMFM with an MLP to achieve spatiotemporal alignment and non-linear decoding, where the GMFM's dual mechanisms—a Gated Fusion Module (GFM) for dynamic crossmodal weighting and a Gated Attention Module (GAM) for establishing semantic–physical correlations—synergistically capture TC–environment interactions, eliminate feature redundancy, and mitigate abrupt path prediction errors. This end-to-end "feature extraction–fusion prediction" framework ultimately enables multi-dimensional high-precision forecasting by exploiting complementary information from heterogeneous data sources.

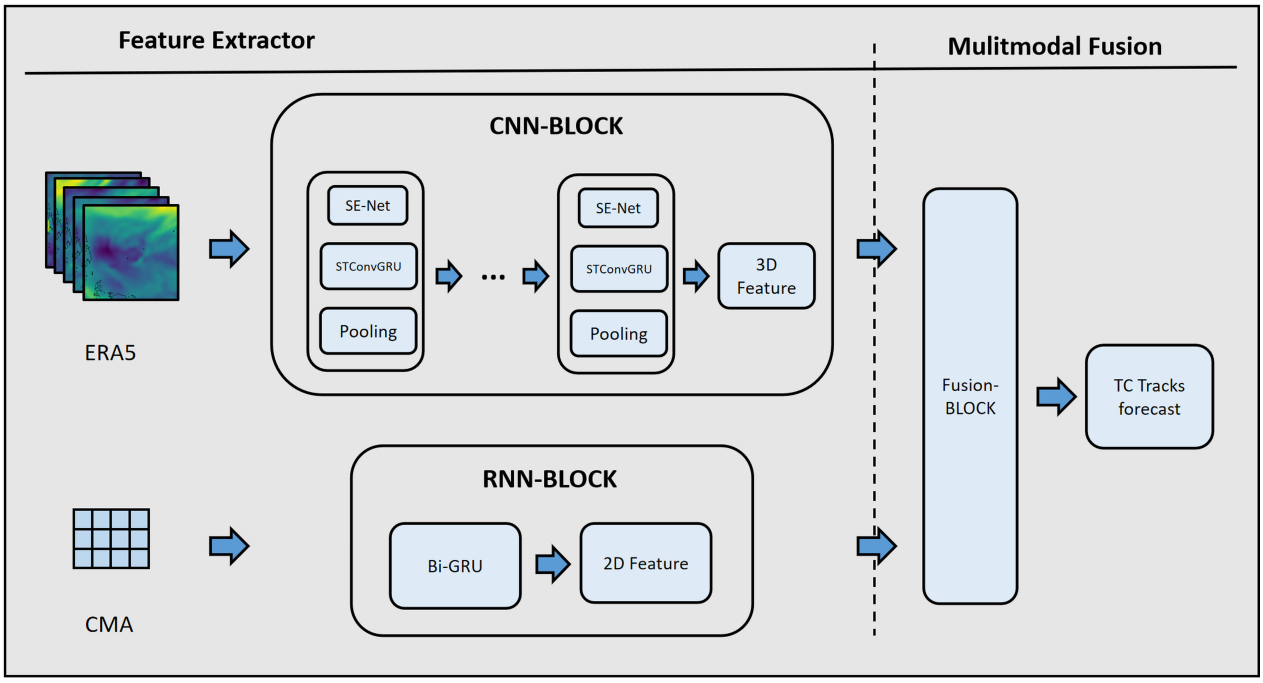


Figure 3. Basic architecture of the DGF-Net model.

3.2. 2D Feature Extraction Module

The primary objective of the 2D Feature Extraction Module is to characterize the inherent non-linear evolutionary patterns and temporal dependencies within TC trajectories, thereby providing robust temporal feature support for subsequent multimodal fusion.

In this study, the Bidirectional GRU (Bi-GRU) is adopted as the sequential modeling backbone to achieve an optimal trade-off between computational efficiency and feature representation capability. Compared with vanilla RNNs, the GRU architecture effectively mitigates the vanishing gradient problem through its streamlined gated mechanism. Although LSTM networks offer similar advantages, the parameter count of a GRU is approximately two-thirds that of an LSTM, making it less prone to overfitting on TC trajectory datasets where sample sizes are relatively limited. To clarify these structural differences, Figure 4 provides a topological comparison among RNN, LSTM, and GRU units, highlighting how varying gating complexities regulate internal signal propagation. The mathematical formulation of the GRU is defined as follows:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z), \tag{2}$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r), \tag{3}$$

$$\tilde{h}_t = \tanh(W_h \cdot x_t + r_t \odot h_{t-1} + b_h), \tag{4}$$

$$h_t = (1 - z_t) \odot \tilde{h}_t + z_t \odot h_{t-1}, \tag{5}$$

where $\cdot$ denotes matrix multiplication, which is used to realize linear projection and dimension transformation of features, and $\odot$ denotes element-wise product, which is used to perform bit-wise weighting on feature channels. z_t and r_t represent the update gate and reset gate, respectively, which are the core components of the GRU's adaptive information flow control mechanism: z_t is used to perform linear interpolation between the previous hidden state and the current candidate state to control the retention degree of long-term temporal context information; r_t acts on the hidden state h_{t-1} at the previous time step, which can dynamically adjust the integration ratio of historical information to characterize the variation trend at the current time step and effectively suppress the interference of redundant noise on the candidate state. W and b denote the weight matrix and bias term, respectively, which are learnable parameters in the temporal operation of the model.

To further exploit the bidirectional temporal correlations within historical TC tracks, this research processes sequences using a Bi-GRU framework. The Bi-GRU extends the conventional unidirectional GRU cell into forward and backward branches: the forward branch extracts temporal dependencies in chronological order, while the backward branch captures inverse associations from the historical sequence in reverse-chronological order [28]. In the context of TC track forecasting, this dual-perspective modeling integrates complementary information across the entire historical sequence, significantly enhancing the model's capacity to capture local trajectory anomalies and sudden path deflections—features that are often overlooked by unidirectional architectures. Research by Song et al. [29] on TC track forecasting in the Northwest Pacific further confirms that the bidirectional temporal mining capability of Bi-GRU effectively reduces forecasting errors, which is highly consistent with the methodology and research scenarios of this study. The complete formulation of the Bi-GRU is as follows:

$$\overrightarrow{h}_t = \mathrm{GRU}(x_t, \overrightarrow{h}_{t-1}), \tag{6}$$

$$\overleftarrow{h}_t = \mathrm{GRU}(x_t, \overleftarrow{h}_{t+1}), \tag{7}$$

$$h_t = [\overrightarrow{h}_t, \overleftarrow{h}_t], \tag{8}$$

where $\overrightarrow{h}_t$ and $\overleftarrow{h}_t$ denote the hidden states of the forward and backward GRUs, respectively, and h_t represents the final concatenated output.

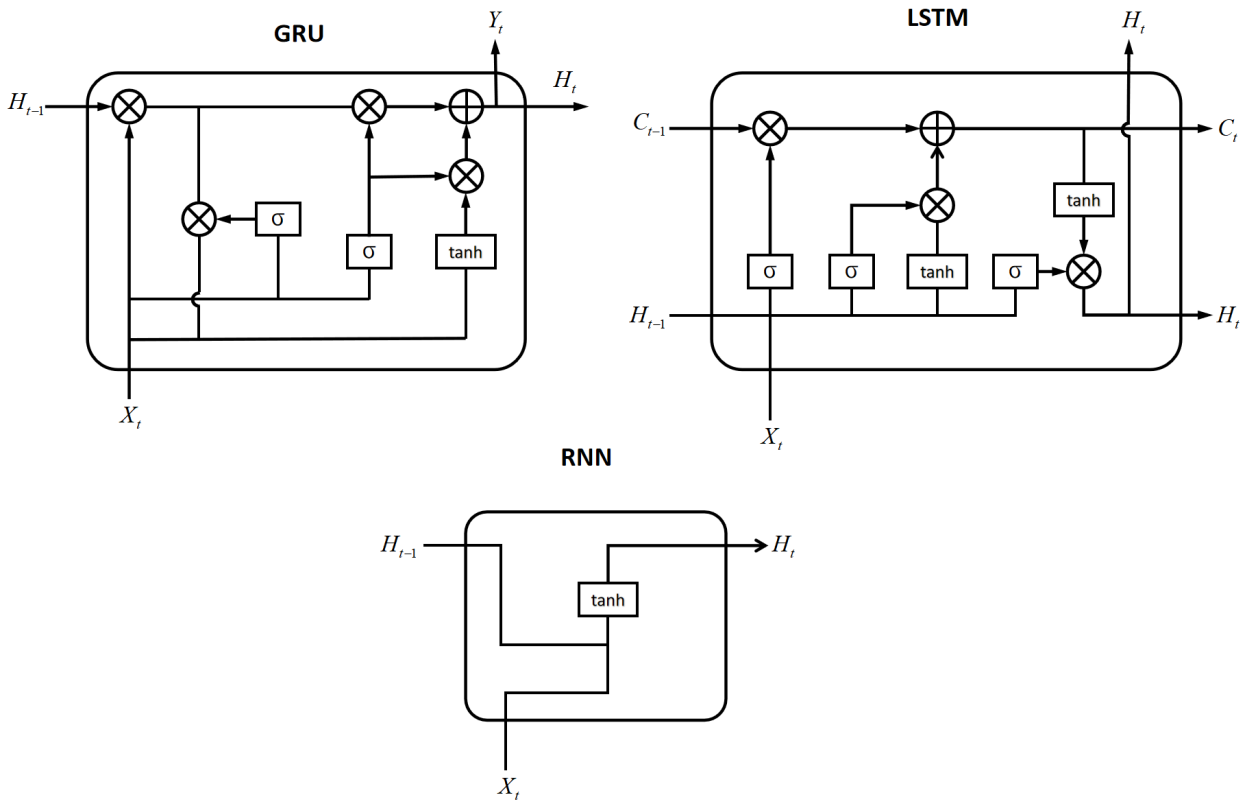


Figure 4. Basic architecture of RNN, LSTM and GRU.

The salient advantage of Bi-GRU lies in its ability to synthesize historical motion inertia with future trend anticipation. This dual-perspective modeling significantly improves the detection of local trajectory anomalies and sudden path deflections. By integrating complementary information from the entire sequence, the Bi-GRU provides more robust and spatially aware feature representations for the subsequent multimodal fusion layer.

3.3. 3D Feature Extraction Module

This study develops a 3D feature extraction backbone comprising four cascaded CNN-Blocks to distill hierarchical spatiotemporal representations from ERA5-derived TC environmental fields. At the core of this module is the STConvGRU, which integrates convolutional operators within the gated recurrent framework to facilitate joint spatiotemporal feature learning. By performing systematic traversal over the atmospheric data manifolds, the STConvGRU effectively captures multi-scale temporal dynamics and non-linear spatial correlations inherent in 3D atmospheric systems. To further refine the feature space, an SENet is integrated to implement adaptive channel-wise feature recalibration. This mechanism dynamically amplifies the responses of track-critical environmental drivers while suppressing redundant noise, thereby enhancing the overall signal-to-noise ratio and representational precision. The module concludes with a spatial pooling operation to perform dimensionality reduction. This operation not only mitigates computational overhead and prevents overfitting by reducing the parameter volume but also enforces scale invariance and local translation invariance. Consequently, the model can extract more robust and transferable environmental patterns, providing a discriminative feature set for high-precision TC track forecasting.

3.3.1. STConvGRU

The Spatiotemporal LSTM (ST-LSTM) network [30] is an architectural extension of the conventional LSTM, originally proposed for skeleton-based joint detection. As validated by Lian et al. [31], ST-LSTM has demonstrated significant advantages in atmospheric data modeling, owing to its ability to effectively capture complex spatial dependencies and temporal evolution patterns within meteorological fields. In this study, the proposed STConvGRU is an improved architecture derived from the ST-LSTM framework. The core innovation involves replacing the LSTM gating modules within the ST-LSTM with a streamlined GRU gating structure. This modification preserves the superior spatiotemporal correlation modeling capabilities of the original framework while substantially reducing the parameter count and mitigating the risk of overfitting. Architecturally, STConvGRU utilizes a dual memory flow structure that decouples the 3D atmospheric field into two independent, parallel convolutional branches. Specifically, the system comprises two distinct memory flows: (1) a temporal memory flow, dedicated to modeling the continuous evolutionary trends of tropical cyclone (TC) movement along the temporal axis, and (2) a spatial memory flow, which employs a specialized gating mechanism to capture cross-scale dependencies across the horizontal dimensions (longitude and latitude) and the vertical dimension (isobaric levels) of the 3D atmospheric environment. To optimize information interaction across these dimensions, a dynamic weight assignment mechanism is introduced to adaptively adjust the fusion ratio of the two memory flows, facilitating a progressive integration of deep spatiotemporal features. Ultimately, this architecture enables a more precise characterization of the complex vortex–environment coupling, generating robust feature representations that serve as high-quality inputs for the precision forecasting of TC tracks in the Northwest Pacific.

Figure 5 shows the structure of the proposed STConvGRU model. This model breaks a key limitation of traditional recurrent units. Traditional units only maintain a single temporal sequence. STConvGRU adopts a dual-track mechanism. It simultaneously maintains two hidden states: temporal state H_t and spatial state H_s. This design collaboratively captures spatiotemporal information. The module builds on traditional GRU temporal propagation. It adds a cross-layer spatial propagation mechanism. This strengthens vertical hierarchical feature interactions in meteorological fields. Consequently, each STConvGRU unit takes two essential inputs: the temporal hidden state $H^l_{T,t-1}$ from the previous time step and the spatially fused hidden state $H^{l-1}_{S,t}$ from upper layer transmission. Additionally, a composite gated mechanism dynamically adjusts cross-layer spatiotemporal fusion weights. The mathematical formulations of STConvGRU are given below:

$$Z_t = \sigma(W_z * [H^{l-1}_{S,t}, H^l_{T,t-1}, X_t] + b_z), \tag{9}$$

$$R_S = \sigma(W_{rs} * [H^{l-1}_{S,t}, X_t] + b_{rs}), \tag{10}$$

$$R_T = \sigma(W_{rt} * [H^l_{T,t-1}, X_t] + b_{rt}), \tag{11}$$

$$C_{ST} = \sigma(W_c * [H^{l-1}_{S,t}, H^l_{T,t-1}] + b_c), \tag{12}$$

$$\tilde{H}t = \tanh(W_h * X_t + R_S \odot (U_s * H^{l-1}_{S,t}) + R_T \odot (U_t * H^l_{T,t-1}) + b_h), \tag{13}$$

$$H^l_t = (1 - Z_t) \odot \tilde{H}_t + Z_t \odot ((1 - C_{ST}) \odot H^{l-1}_{S,t} + C_{ST} \odot H^l_{T,t-1}), \tag{14}$$

where X_t represents the input data at the current time step; $H^{l-1}_{S,t}$ is the spatial hidden state from the previous layer at the same time step; $H^l_{T,t-1}$ is the temporal hidden state (containing long-term memory information) from the previous time step in the same layer; and H_t is the output of the cell. The functionalities of the gating units are as follows:

the Update Gate Z_t controls the fusion weight between the hidden state and the candidate state; the Spatiotemporal Reset Gate R_T regulates the contribution of the temporal hidden state $H^l_{T,t-1}$ to the current state; the Spatial Reset Gate R_S dynamically filters the relevance between the spatial information $H^{l-1}_{S,t}$ and the current input X_t; and the Spatiotemporal Collaboration Gate C_{ST} modulates the association between the two types of hidden states, determining their fusion ratio by adjusting the weights of $H^l_{T,t-1}$ and $H^{l-1}_{S,t}$; $*$ denotes the convolutional operator, which enables the extraction of local spatial patterns from the 3D atmospheric grid. The operator $\odot$ represents the Hadamard product, facilitating the gating mechanism by pixel-wise modulation of the information flow. The STConvGRU dynamically adjusts information weights through the aforementioned gates to achieve adaptive fusion of temporal continuity and spatial hierarchy, thereby strengthening global spatial modeling capability and enhancing spatiotemporal sequence prediction performance. The spatial states are propagated hierarchically across layers to reconstruct the atmospheric structure, while the temporal hidden states capture the evolutionary patterns of the cyclone trajectory through sequential recurrence. The two are then integrated via the Spatiotemporal Collaboration Gate to complete the synergistic extraction of spatiotemporal features, significantly improving path forecasting accuracy.

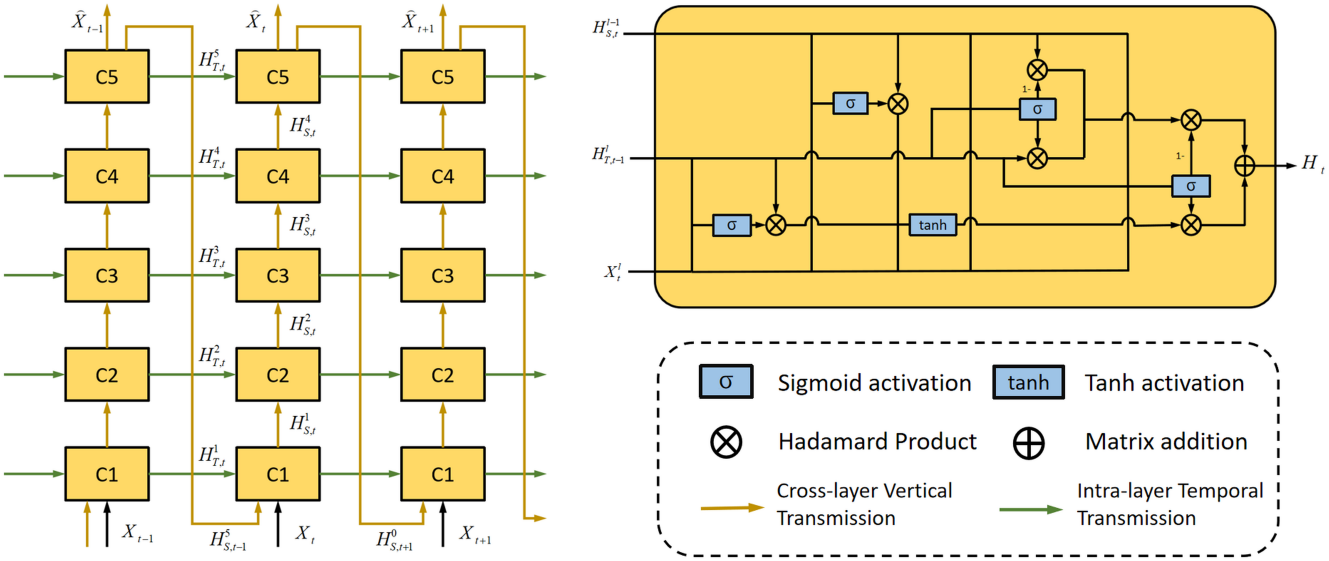


Figure 5. Model structure of STConvGRU.

3.3.2. SENet

SENet was proposed by the authors of [32] to address the limitation in traditional convolutional operations where a single convolutional kernel can only capture information within a local receptive field. In this study, to prevent critical channels from being overwhelmed by redundant information during the learning process, SENet is introduced. This network enhances the sensitivity to information features through a Squeeze-and-Excitation mechanism. Specifically, the squeeze operation compresses the spatiotemporal information of each channel into a single scalar via average pooling, thereby reducing spatial redundancy. The excitation stage employs fully connected layers to learn non-linear weights across channels, recalibrating the importance of each channel. By integrating SENet into the 3D feature extraction module, its channel attention mechanism adaptively focuses on key environmental field variables while reducing the weight of abnormal noisy channels, suppressing spurious information along land–sea boundaries. Furthermore, its lightweight design introduces only a small number of additional parameters to the original network, improving prediction performance with minimal computational overhead.

Figure 6 shows the structure of SENet. The specific workflow is as follows: the original information is compressed into channel descriptors through global average pooling, followed by a gating mechanism combined with a sigmoid activation function to capture dependencies between channels. Finally, the generated weights are reapplied to the input feature maps to enhance important features. The formulation can be described as follows:

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i,j), \tag{15}$$

$$s = F_{ex}(u_c) = \sigma(W_2 \cdot \mathrm{ReLU}([W_1 \cdot z_c])), \tag{16}$$

$$\hat{u}_c = s_c \cdot u_c, \tag{17}$$

where z_c represents the global feature statistics of the c-th channel; $F_{sq}(\cdot)$ denotes the operation for extracting global feature statistics; $u_c(i,j)$ refers to the feature at the i-th row and j-th column of the c-th channel; H is the height of the input feature map; W is the width of the input feature map, s denotes the gated weights generated by excitation; $F_{ex}(\cdot)$ indicates channel feature activation operation; and W_1 and W_2 are weight matrices of fully connected layers, where W_1 compresses dimensionality while W_2 expands dimensionality.

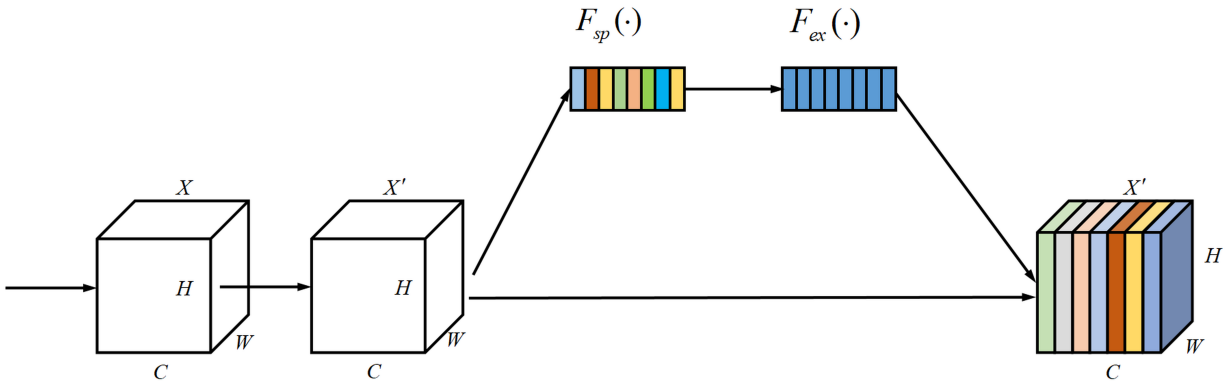


Figure 6. Architecture diagram of SENet.

In the 3D feature extraction module proposed in this study, the incorporation of the SENet mechanism enables dual enhancement of spatiotemporal modeling capabilities. On one hand, the global pooling operation expands the model's receptive field of multi-scale structures of TCs, compensating for the limitation of traditional convolutional kernels with local receptive fields in capturing cross-regional correlation information. On the other hand, the channel attention mechanism significantly improves the model's generalization performance in complex meteorological scenarios by reinforcing coupled features across influencing factors.

3.4. Multimodal Fusion Module

In meteorological forecasting tasks, unimodal features often fail to comprehensively capture the complex processes of TC genesis and movement. To address this, this study proposes the GMFM, which serves as a crossmodal progressive decoder tailored for 2D TC tracks and 3D atmospheric fields. As illustrated in Figure 7, it clearly presents the core components of the GMFM, the connection relationships between each modules, and the complete processing flow of multimodal features from input to output, intuitively demonstrating the collaborative working mechanisms of the dual-branch GRU module, GFM module and GAM module. Such GRU-based multimodal encoder–decoder architectures have been validated for their superior spatiotemporal interpretation capabilities in

multi-source data feature extraction and sequence decoding tasks [33]. At each time step, the 2D temporal features and 3D spatiotemporal features first undergo temporal alignment and feature refinement via a dual-branch GRU module. Subsequently, the GFM module facilitates adaptive fusion and complementary enhancement of the decoded multimodal features through a two-stage gating mechanism. Finally, the GAM module performs global refinement of the fused features by leveraging the crossmodal interaction between the original inputs X_I and X_S, effectively mitigating the redundant information and noise introduced during the multimodal integration process.

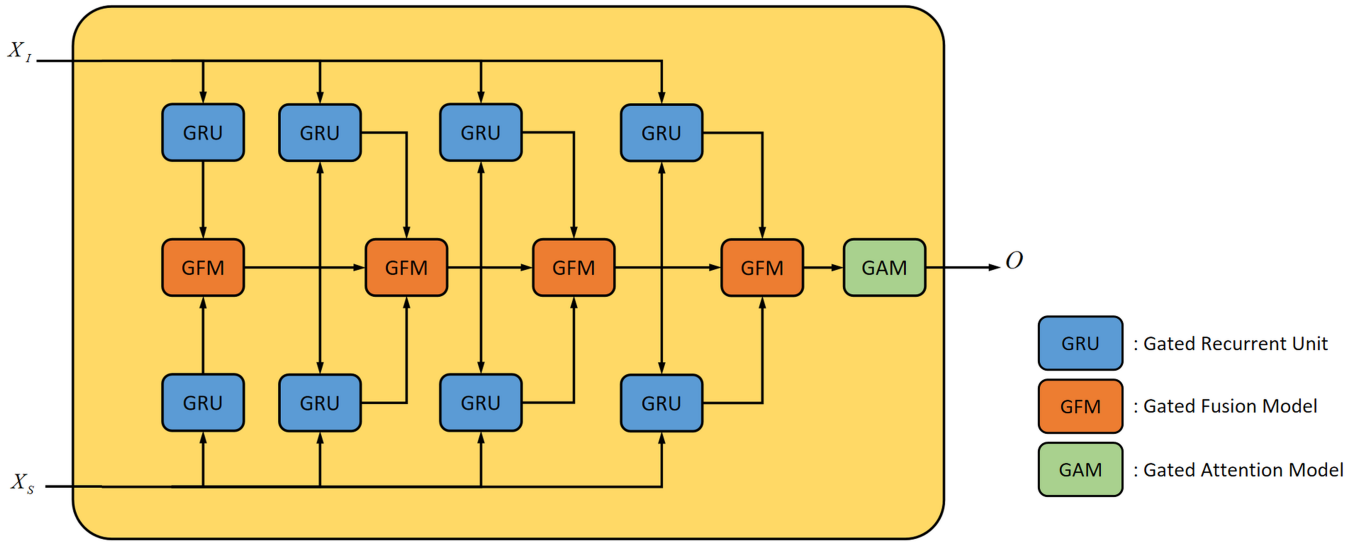


Figure 7. Structure of the GMFM.

3.4.1. GFM and GRU Interaction

To facilitate progressive feature integration across heterogeneous sequential data, the GFM establishes a cyclic dependency with the backbone feature extraction layers, constructing a closed-loop "process-then-fuse" feedback system. As depicted in the interaction linkage diagram (Figure 8), the GRU and GFM modules are interleaved along the temporal axis to achieve fine-grained spatiotemporal alignment. Among them, the I-GRU takes the 3D TC atmospheric environmental features as input, while the S-GRU takes the 2D tropical cyclone track features. Unlike traditional late-fusion strategies that aggregate modalities only at the decision layer, this recurrent interleaving mechanism enables the model to perform crossmodal calibration at each hidden state transition. Specifically, the GRU at time step t ingests the global multimodal context O'_{t-1}—refined and rectified by the preceding GFM unit—as its initial hidden state. Its updated latent representations are subsequently projected into the current GFM to undergo multi-source synergistic modeling. To support this recursive flow, the GFM dynamically adapts its internal state transitions based on the temporal coordinates, ensuring that the synthesized representations maintain long-term semantic consistency while effectively capturing the instantaneous vortex–environment interactions.

The internal topology of the GFM, illustrated in Figure 9, is governed by a dual-stage gating mechanism designed to process the distinct 3D spatiotemporal environmental stream x_I^t and the 2D trajectory stream x_S^t in parallel. Initially, the inputs undergo Correlation-based Filtering (G_1), which functions as a soft-attention mask to model the non-linear interdependencies between modalities.

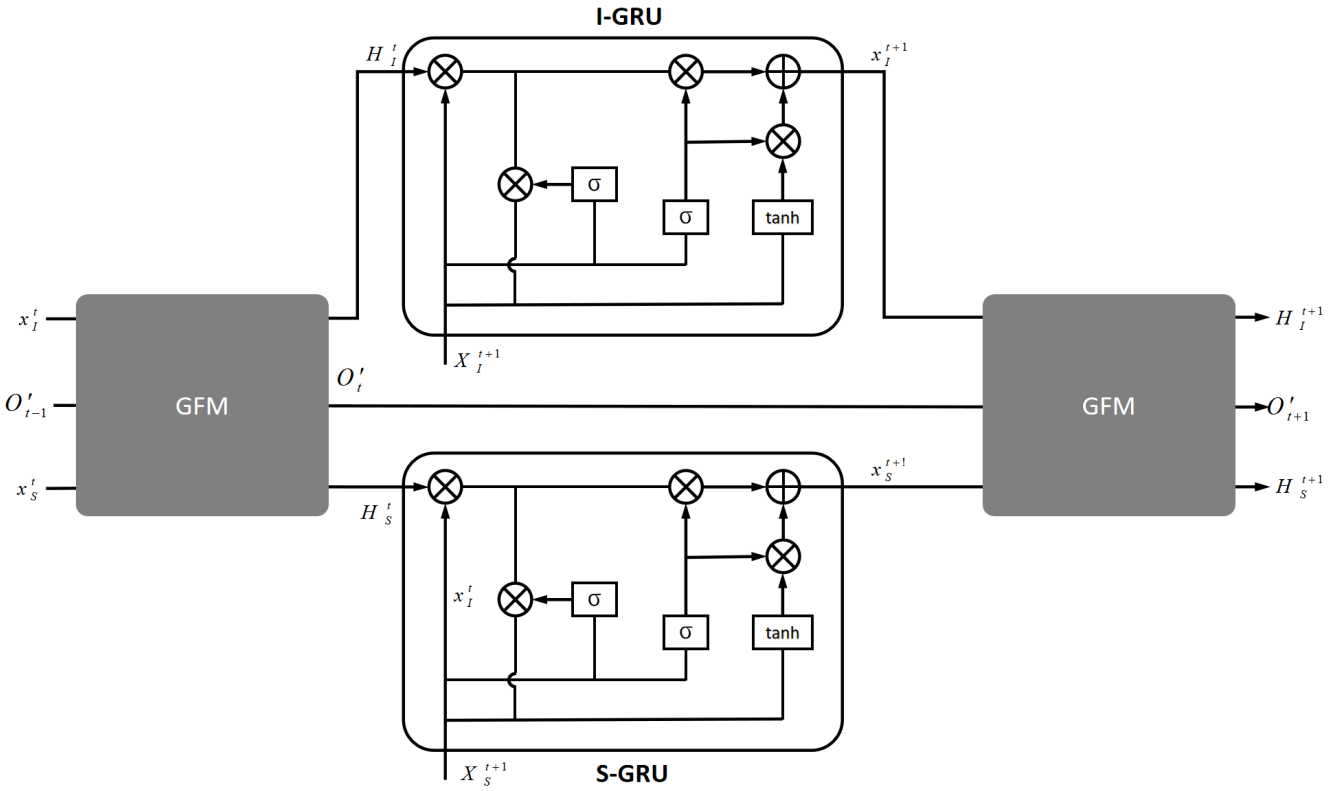


Figure 8. A schematic of the recurrent interaction logic between the GRU and GFM.

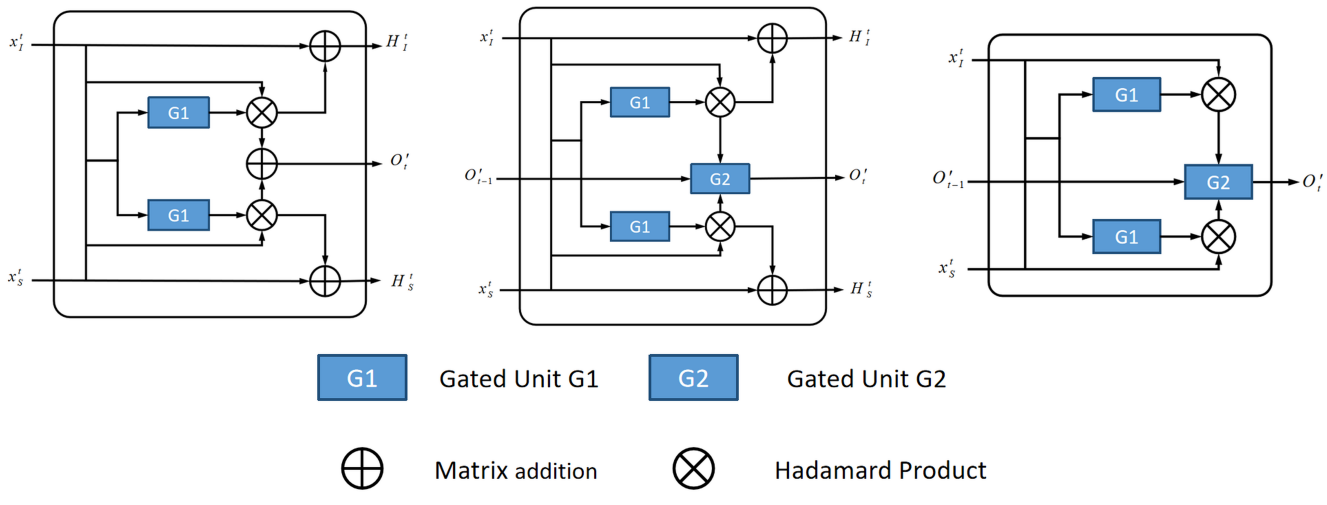


Figure 9. Detailed architectural topology of the GFM model.

In complex meteorological scenarios, 3D atmospheric grids often contain significant background redundancy that may overshadow the sparse 2D trajectory signals. G_1 effectively mitigates this by computing crossmodal relevance scores, thereby adaptively suppressing low-relevance noise and spurious correlations along the land–sea boundaries. Formally, the filtered features are aggregated to form a candidate hybrid representation O_t'', which serves as the foundational feature map for subsequent temporal alignment:

$$G_I = x_I^t \odot \sigma(W_{fI} \cdot [x_I^t, x_S^t]), \tag{18}$$

$$G_S = x_S^t \odot \sigma(W_{fS} \cdot [x_I^t, x_S^t]), \tag{19}$$

$$O_t'' = G_I + G_S, \tag{20}$$

where σ represents the sigmoid function and $\odot$ denotes element-wise multiplication; G_I denotes the latent spatiotemporal representation derived from the 3D environmental features after being processed by the GFM unit; and G_S represents the refined 1D temporal features obtained from the trajectory data following gated modulation.

Subsequently, to preserve temporal continuity, a Temporal Alignment Update mechanism in unit $G2$ dynamically regulates the fusion ratio between the historical context O'_{t-1} and the current candidate O''_t. This step is critical to mitigate fusion biases caused by abrupt signal changes and is governed by the following formulations:

$$z_t = \sigma(W_{zt} \cdot [O''_t, O'_{t-1}]), \tag{21}$$

$$O'_t = (1 - z_t) \cdot O''_t + z_t \cdot O'_{t-1}, \tag{22}$$

Finally, residual connections are incorporated to bypass the fusion gate and generate modality-specific hidden states for next-step GRU processing, ensuring efficient gradient flow through these definitions:

$$H_I^t = G_I + x_I^t, \tag{23}$$

$$H_S^t = G_S + x_S^t, \tag{24}$$

Through this multi-stage architecture, the GFM achieves a synergistic augmentation of 2D temporal patterns and 3D spatial structures by effectively balancing historical inertia with current observation updates.

3.4.2. GAM

The GAM dynamically regulates interaction weights across modalities, with primary emphasis on suppressing low-correlation noise in multimodal fusion features. As illustrated in Figure 10, the architecture of GAM integrates multi-head attention with gated filtering. Operationally, it processes the raw 3D and 2D modal features through integrated transformations to form the key matrices (k) and value matrices (v), while utilizing processed fusion features as the query matrix (q). These computed q, k, and v are then fed into the multi-head attention mechanism to quantify associations between original modalities and fused representations, dynamically adjusting their weights. Finally, the gated mechanism regulates fusion-induced noise and prevents deviations from source modalities, thereby enhancing predictive performance.

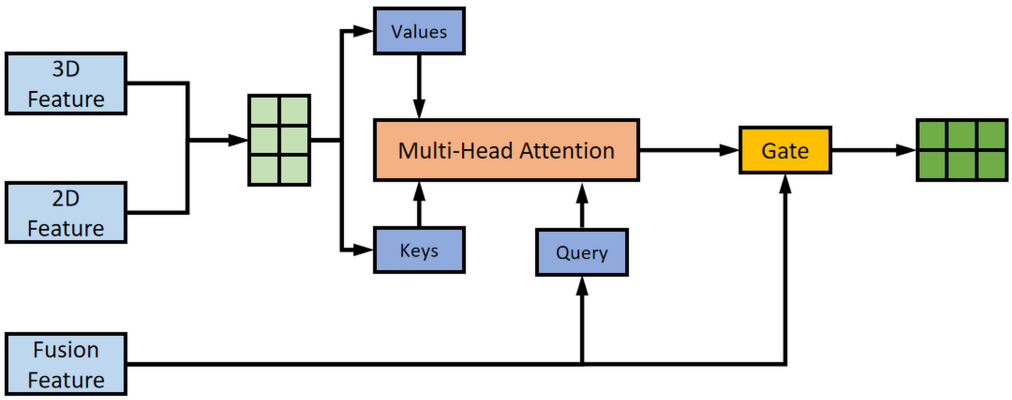


Figure 10. Architecture of the GAM module.

The attention mechanism is one of the most important tools in deep learning, inspired by the cognitive function of biological attention [34]. Multi-head attention, proposed by Vaswani et al. [35] as the core component of the Transformer architecture, fundamentally captures diversified input features from multiple subspaces by parallelizing multiple

attention heads. For given q, k, and v matrices, the attention mechanism computation follows the formula

$$Attention(q, k, v) = softmax(\frac{qk^T}{\sqrt{d_k}})v, \tag{25}$$

where d_k denotes the dimension of the key vector. The multi-head attention mechanism extends the standard attention mechanism by projecting input information into multiple subspaces, computing attention independently in each subspace, and integrating multi-dimensional attention outcomes. The mechanism captures rich contextual information, thereby enhancing the model's capacity to model complex data patterns. It can be represented by the following formula:

$$head_j = Attention(q_j, k_j, v_j), \tag{26}$$

$$MultiheadAttention(q, k, v) = concat(head_1, head_2, ..., head_n)W_o, \tag{27}$$

where q_j denotes the query matrix of the j-th attention head; k_j represents the key matrix of the j-th attention head; v_j corresponds to the value matrix of the j-th attention head; and W_o is the output linear transformation matrix, designed to integrate multi-head attention outputs into the original input dimensionality. This study proposes GAM to filter noise and redundancy in multimodal fusion features.

Based on the aforementioned mechanisms, GAM employs multi-head attention to capture discrepancies between modal features and fused representations, with a gated mechanism to suppress noise in fusion data. The formal representation of GAM is expressed as

$$v^i = W_v \cdot [x_I^i, x_S^i], \tag{28}$$

$$k^i = W_k \cdot [x_I^i, x_S^i], \tag{29}$$

$$q^i = W_q \cdot O'^i, \tag{30}$$

$$A = MultiheadAttention(q, k, v), \tag{31}$$

$$O_t = W_{of}\dot{O}_t' \odot \sigma(W_{oa} \cdot [A, O_t']), \tag{32}$$

where O_t' denotes the input of the GAM module; q^i, k^i and v^i are the inputs of each head in the multi-head attention mechanism; and O_t represents the final output after GAM processing. Thus, the GMFM incorporating GAM not only excavates interactive features across modalities but also filters noise and redundancy within these interactions, significantly enhancing the accuracy and robustness of TC predictions.

4. Experiment

To validate the scientific rigor and superiority of the proposed DGF-Net, this study establishes a four-stage experimental hierarchy: (1) verification of the feature extraction capabilities of Bi-GRU and STConvGRU; (2) evaluation of the multimodal fusion framework; (3) ablation studies of core modules; and (4) assessment of comprehensive model performance. These stages are designed to be progressively incremental, providing a holistic validation of the model's architectural value.

4.1. Experimental Preparation

The CMA and ERA5 datasets are initially aligned across their respective temporal and spatial domains, followed by a partitioning process where the data are divided into training, validation, and testing sets. Individual TCs are used as the fundamental units for this partitioning, covering the period from 2005 to 2023. Specifically, 70% of the TCs are randomly allocated to the training set, while the validation and testing sets are

each assigned 15% of the total TC samples to ensure an unbiased evaluation of model generalization. During the preprocessing phase, TC sequences containing fewer than eight records—representing a lifecycle of less than 48 h—are systematically excluded to ensure sufficient temporal dependency for recursive modeling, as these truncated trajectories are insufficient to support robust 24 h lead-time forecasting. Furthermore, to ensure numerical stability and facilitate model convergence during the training process, both 2D and 3D TC input features are normalized using the Min-Max scaling method, thereby mapping all input variables into a standardized range for deep learning optimization.

To ensure experimental parity and reproducibility, the hyperparameters are kept consistent across all tests. Specifically, the batch size is set to 128, and the models are trained for 128 epochs. The Adam optimizer is employed to implement adaptive learning rate adjustments, thereby accelerating model convergence. The detailed distribution of the dataset is summarized in Table 3.

Table 3. Distribution of the experimental dataset.

Category	Quantity
Temporal dimension	24 h
Number of TC datasets	518 TCs
Number of TCs in training sets	363 TCs
Number of TCs in validation sets	78 TCs
Number of TCs in test sets	77 TCs
Number of training sets	8425 tuples
Number of validation sets	1551 tuples
Number of test sets	1476 tuples

The experiments were conducted on a high-performance computing system equipped with a 12-vCPU Intel® Xeon® Platinum 8336C CPU (Intel Corporation, Santa Clara, CA, USA) and dual NVIDIA RTX 2080 Ti GPUs (NVIDIA Corporation, Santa Clara, CA, USA). Each GPU is equipped with 11 GB of GDDR6 memory, resulting in a total of 22 GB for the dual-card configuration. The model architecture was developed and trained using Python (v3.8) running on Ubuntu (v20.04), alongside the CUDA (v11.8) toolkit and the PyTorch (v2.0.0) deep learning framework.

To evaluate the prediction results more accurately, this study employs the Great Circle Distance (GCD). This is a tool commonly used to measure latitude and longitude, which can quantify the error between the model's predictions and the actual ground positions, with its calculation method presented as follows:

$$D = \frac{1}{n}\sum_{n}^{i=1}(2R\arcsin(\sqrt{\sin^2(\frac{\Delta y_i}{2}) + \cos(y_i)\cos(\widehat{y}_i)\sin^2(\frac{\Delta x_i}{2})})), \quad (33)$$

where R denotes the radius of the Earth and $R = 6371$ km; Δy represents the difference between the actual latitude and the predicted latitude; and Δx denotes the difference between the actual longitude and the predicted longitude.

4.2. Experimental Design

4.2.1. Experiment I: Feature Extraction Module Performance Validation

This experiment aims to verify the adaptability of the Bi-GRU and STConvGRU architectures for extracting features from 2D and 3D TC data through two comparative sets. For 2D feature extraction, Bi-GRU is benchmarked against XGBoost [36], Random Forest (RF) [37], GBDT, LSTM, GRU, and vanilla RNN. The three traditional machine learning methods incorporate the feature construction logic of the CLIPER method [15], while the deep learning models autonomously extract temporal features from the input sequences.

Regarding 3D feature extraction, STConvGRU is compared with 3D-RNN, CNN-LSTM, ConvLSTM [14], and STConvLSTM [31]. While the baseline models require dimensionality reduction of horizontal planes and meteorological parameters, the proposed STConvGRU and STConvLSTM process the five standard isobaric levels of ERA5 reanalysis data without such constraints. All models are evaluated using GCD error at 6 h, 12 h, and 24 h lead times.

4.2.2. Experiment II: Multimodal Fusion Framework Performance Evaluation

This experiment quantifies the efficacy of the proposed GMFM by benchmarking it against an ablated baseline, DGF-Base. While DGF-Base shares the same backbone (STConvGRU and Bi-GRU) as our final model, it employs a naive feature-level concatenation for multimodal integration, serving as a lower bound for fusion performance. To evaluate the architectural superiority of the GMFM, we substitute the fusion unit with five heterogeneous alternatives—including the Self-Attention-based Transformer [35], GRU, LSTM, XGBoost, and RF. Furthermore, to explore the sensitivity of the model to vertical atmospheric structures, four isobaric level configurations are rigorously tested: 1 layer (500 hPa), 3 layers (300–900 hPa), 5 layers (200–1000 hPa), and 7 layers (200–800 hPa). By analyzing the 24 h GCD error across these variants, we aim to disentangle the interplay between multimodal fusion depth and vertical feature granularity, providing insights into how the GMFM mitigates information redundancy as the input dimensionality increases.

4.2.3. Experiment III: Core Module Ablation Study

Using a controlled variable approach, this experiment quantitatively analyzes the performance contribution and necessity of each core component within the DGF-Net architecture. Three sets of ablation variants are constructed by systematically removing SENet, GFM, or GAM, while keeping all other components and hyperparameters constant. The evaluation is conducted across two dimensions: prediction accuracy (6 h, 12 h, and 24 h GCD errors) and computational complexity, specifically measured by Floating Point Operations (FLOPs) and parameter counts (Params), thereby validating the performance gain versus the computational overhead for each specific module.

4.2.4. Experiment IV: Comprehensive Performance Benchmark

This experiment evaluates the comprehensive performance and practical utility of DGF-Net against state-of-the-art (SOTA) TC forecasting models and operational standards. The CLIPER model [7] is selected as the statistical baseline, while the Central Meteorological Observatory (CMO) official operational mode [38] serves as the NWP benchmark [39]. Additionally, mainstream deep learning models, including AM-ConvGRU [15], TAM-CL [14], and GBRNN [11], are included for comparison. GCD errors at 6, 12, and 24 h are utilized as evaluation metrics to clarify the performance positioning and application value of DGF-Net.

4.2.5. Statistical Significance Test

To rigorously verify the reliability of the performance differences between models, this study employs statistical significance testing as a quantitative analytical tool [40]. The core objective is to determine whether the observed differences stem from the inherent superiority of the algorithm or merely from random fluctuations. Given the multiple comparisons involved, Analysis of Variance (ANOVA) [41] is utilized to systematically evaluate global differences between models. The evaluation is based on the calculated test statistics and the p-value, which quantifies the probability that the observed differences resulted from random factors. Using a significance level of $\alpha = 0.05$ as the threshold, a p-value smaller than α indicates statistical significance, implying that the differences arise from the inherent distinctions in the methods.

4.2.6. Visualization Experiments

To intuitively validate the forecasting robustness of DGF-Net at critical trajectory nodes, this section selects representative typhoon cases for visualization and error characteristic analysis. Super Typhoon Doksuri (2023) and Typhoon Kajiki (2019) were designated as typical validation cases, as they encompass two core forecasting challenges in the Northwest Pacific, thereby facilitating a comprehensive evaluation of the model's integrative performance. Specifically, Doksuri represents a typical landfalling TC characterized by coastal rapid intensification and drastic adjustments in its warm-core structure and vertical circulation. Its movement involved two significant directional deflections, with its near-shore trajectory being heavily influenced by the complex interactions between coastal topography and multi-scale circulations, leading to extreme path uncertainty. This case serves as a prototypical sample for capturing the features of the rapid intensification process. Conversely, Kajiki serves as a prototypical example of an extremely weak and poorly organized TC in the South China Sea. Maintaining only tropical storm intensity throughout its life cycle, it featured an exceptionally loose circulation structure lacking a clear, stable closed center or strong, coherent cyclonic characteristic signals. Its erratic and lingering movement path provides a rigorous test for the model's ability to process weak structural signals and low signal-to-noise ratio features.

5. Experimental Results and Discussion

This chapter presents an objective evaluation of the aforementioned experimental results. Following the progressive logic of the experimental design, we sequentially disclose the core findings regarding the feature extraction modules, multimodal fusion framework, ablation studies, comprehensive performance benchmarks, statistical significance tests, and visualization experiments. Each result is subjected to an in-depth analysis integrated with current research trends and operational requirements in the field of TC forecasting.

5.1. Results of Experiment I

The forecasting performance comparison of the 2D feature extraction modules is presented in Table 4. The results indicate that decision-tree-based machine learning algorithms, including XGBoost, Random Forest, and GBDT, exhibit suboptimal overall performance. The fundamental reason is that linear patterns and non-linear evolutionary characteristics are highly coupled within 2D TC trajectory data; these algorithms struggle to construct complex feature mapping relationships, leading to limited model generalization. Conversely, the forecasting performance of deep learning models generally surpasses that of traditional algorithms, validating the superiority of deep learning in modeling non-linear temporal features. Among them, Bi-GRU demonstrates the most prominent performance, achieving the minimum error across all forecasting lead times. This success is attributed to the bidirectional propagation structure of Bi-GRU, which simultaneously captures the historical evolutionary patterns and future trend dependencies of TC movement. Compared to the unidirectional structures of GRU and LSTM, Bi-GRU extracts critical information from 2D temporal data more comprehensively, highlighting its feature extraction adaptability in this specific research context.

The performance comparison for the 3D feature extraction modules is summarized in Table 5. The results show that STConvGRU significantly outperforms 3D-RNN, CNN-LSTM, and ConvLSTM. This performance disparity stems from the loss of inter-layer coupling information during the feature compression process in the baseline models. The spatial convolutional module of STConvGRU directly extracts local spatial features from each isobaric level, while the GRU units capture the temporal correlations between layers, thereby achieving spatiotemporal synergistic modeling. Compared with STConvL-

STM, STConvGRU exhibits distinct advantages in 6 h and 24 h forecasting scenarios. This is because STConvGRU maintains higher update efficiency of spatiotemporal states when processing multi-layer dynamic data, which mitigates the risk of overfitting and enables a better capture of long-term TC trajectory trends.

Table 4. Comparison of 2D prediction performance of feature extraction modules.

Models	6 h (km)	12 h (km)	24 h (km)
XGBoost	46.82	117.71	227.11
Random Forest	47.83	117.08	237.54
GBDT	42.95	104.01	216.46
LSTM	43.92	104.45	217.46
GRU	43.19	106.46	216.72
Simple RNN	46.05	115.77	240.24
Bi-GRU	41.75	98.19	207.56

Table 5. Comparison of prediction performance of 3D feature extraction modules.

Models	6 h (km)	12 h (km)	24 h (km)
3D-RNN	41.32	102.10	195.65
CNN-LSTM	98.92	117.08	198.52
ConvLSTM	40.28	95.63	183.37
STConvLSTM	38.53	84.21	171.53
STConvGRU	38.82	85.83	170.48

The comparative validation results of the aforementioned 2D and 3D feature extraction modules fully confirm the architectural rationality and performance superiority of the proposed dual-branch feature extraction framework at the data level, achieving efficient and precise feature extraction from both 2D temporal trajectory data and 3D spatiotemporal environmental fields. For the 2D temporal data of historical TC tracks, Bi-GRU leverages its bidirectional gated propagation structure to compensate for the deficiencies of unidirectional Recurrent Neural Networks in capturing temporal trend correlations, realizing the synergistic modeling of historical evolution and future trends. This modeling effect is entirely consistent with the research conclusions regarding Northwest Pacific TC trajectory forecasting in the literature [29], providing external validation for the scientific validity of our Bi-GRU branch design. Regarding the 3D spatiotemporal data of multi-layer atmospheric environmental fields, STConvGRU achieves synergistic modeling of spatial atmospheric circulation features and vertical temporal correlations through its dual-stream spatiotemporal interval design. Furthermore, compared to the ST-LSTM architecture, it possesses a more streamlined gating mechanism and enhanced spatiotemporal trend-capturing capabilities, fully corroborating the unique advantages of this framework in processing multi-dimensional atmospheric reanalysis data.

5.2. Results of Experiment II

The forecasting performance comparison of the multimodal fusion framework is detailed in Table 6. This study systematically verifies the performance advantages and structural rationality of the GMFM across two dimensions: vertical isobaric level configurations and multimodal fusion methodologies. The results indicate that the DGF-Net model achieves peak performance under a five-layer isobaric configuration, yielding a forecasting error of 135.49 km, which represents a 15.8% reduction compared to the DGF-Base baseline. Generally, the forecasting error follows a trend of initial decline followed by a slight rebound as the number of input pressure levels increases. This trend suggests that while comprehensive

spatial features are critical for prediction, excessive dimensionality introduces informational redundancy and noise interference. Consequently, the five standard isobaric levels within the 200–1000 hPa range are identified as the optimal configuration for this study.

Table 6. Comparison of prediction performance of the multimodal fusion framework.

Model/Levels	1 Level (km)	3 Levels (km)	5 Levels (km)	7 Levels (km)
DGF-Base	176.77	169.39	160.93	161.16
DGF	147.79	141.59	135.49	137.19
DGF-Transformer	149.13	139.10	137.89	144.93
DGF-GRU	153.23	149.66	142.20	141.94
DGF-LSTM	154.67	148.88	140.32	143.31
DGF-XGBoost	161.31	160.47	155.89	156.42
DGF-Random Forest	158.84	156.43	153.419	157.1

In the horizontal comparison of various multimodal fusion approaches, the results show that traditional machine learning fusion layers, such as XGBoost and Random Forest, outperform the baseline but lag significantly behind deep learning fusion units. This disparity arises because traditional methods struggle to construct complex mapping relationships for highly intertwined linear and non-linear multimodal features. While deep learning methods like GRU, LSTM, and Transformer can capture temporal dependencies, they exhibit modal adaptation deficiencies. Specifically, GRU and LSTM focus on single-sequence dependencies and fail to effectively distinguish the semantic differences between 2D temporal features and 3D spatiotemporal features. Although the self-attention mechanism in Transformers enables global modeling, it remains susceptible to redundant information in multimodal scenarios. In contrast, the proposed GMFM utilizes a dual mechanism of attention and gating to precisely capture matching features between 2D TC intensity and 3D environmental fields. By completing fusion across both temporal and semantic dimensions, it generates highly discriminative features, reducing the 24 h forecasting error by 1.7% compared to the best-performing baseline, DGF-Transformer.

In summary, these results demonstrate the effectiveness and sophistication of the proposed GMFM. The five-layer isobaric configuration effectively covers critical circulation systems, ensuring that essential driving features are captured while avoiding the interference of redundant data. The GMFM module, through its dual-stage fusion structure consisting of time-step integration and attention-based interaction, effectively manages the fusion of heterogeneous data in multimodal meteorological forecasting. This progressive fusion approach aligns the intrinsic spatiotemporal relationships between 2D cyclone motion and 3D environmental features, effectively resolving common issues such as modal semantic misalignment and the fragmentation of temporal and spatial characteristics [33]. Combined with the dual-branch feature extraction architecture, this module forms a complete multimodal model that lays the core foundation for the overall performance enhancement of DGF-Net, providing a robust reference for multimodal fusion in Northwest Pacific tropical cyclone track forecasting.

5.3. Results of Experiment III

The results of the core module ablation study are summarized in Table 7. In this experiment, we utilized the controlled variable method to systematically remove the SENet, GFM, and GAM modules, constructing several ablation variants to quantitatively evaluate their functional contributions and architectural necessity across two dimensions: computational complexity and forecasting performance. The results indicate that the additional parameters introduced by SENet are negligible, with the floating-point operations (FLOPs)

increasing by only 0.02 GFLOPs. Consequently, the integration of SENet enhances model performance without significantly increasing the overall complexity of the system. In the configuration where GFM was omitted—utilizing GRU for the fusion of 2D and 3D features while employing GAM to filter fusion-induced noise—the variant's parameter count decreased by 6.63 M, while the FLOPs increased by only 0.24 GFLOPs, representing a negligible increment compared to the original model. In the variant without GAM, where only GFM was used to fuse features followed by an MLP for trajectory prediction, the parameter count and FLOPs were reduced by only 0.65 M and 0.06 GFLOPs, respectively. These findings demonstrate that the inclusion of GAM significantly elevates the overall performance of the model while incurring minimal computational overhead.

Table 7. Comparison of computational complexity in ablation experiments.

Model	FLOPs	Params
-SENet	701.49 G	15.80 M
-GFM	701.27 G	9.17 M
-GAM	701.45 G	15.15 M
DGF-Net	701.51 G	15.80 M

The synthesized results of these ablation experiments confirm the indispensability of each core module within the DGF-Net architecture and the rationality of the overall design. As shown in Table 8, the full DGF-Net consistently achieves the optimal performance among all ablation variants, demonstrating the synergistic advantage of the integrated modules. Specifically, the additional computational costs introduced by the multimodal fusion module and the 3D feature enhancement module are negligible, as the primary computational burden remains concentrated in the feature extraction stage. Among the components, the SENet channel attention module enhances 3D atmospheric feature extraction at a near-zero cost; however, its individual gain is relatively limited, as its removal led to only a 1.67 km increase in the 24 h forecasting error, suggesting that there is still room for improvement in enhancing 3D feature extraction through single-channel attention mechanisms alone. The GFM module serves as the pivotal unit for multimodal feature fusion, and its removal resulted in a significant attenuation of model performance, underscoring the strong adaptability of the gated fusion structure for multi-source heterogeneous meteorological data. The GAM module performs global refinement of the fused features and focuses on core driving features, thereby filtering redundant noise and mitigating modal semantic misalignment to generate high-discriminative features for track forecasting. Through the synergistic action of these three components, the model achieves a significant improvement in forecasting accuracy across all lead times with minimal computational overhead, providing a lightweight and high-performance technical solution for the intelligent short-term forecasting of Northwest Pacific tropical cyclone paths.

Table 8. Comparison of prediction performance in ablation experiments.

Models	6 h	12 h	24 h
-SENet	36.03	65.33	137.16
-GFM	37.15	69.05	145.96
-GAM	39.73	67.32	147.91
DGF-Net	35.62	63.53	135.49

5.4. Results of Experiment IV

The comprehensive performance comparison between DGF-Net and mainstream TC track forecasting models is summarized in Table 9. The results demonstrate that

DGF-Net consistently outperforms the statistical baseline model, CLIPER, with the 24 h track forecasting error decreasing from 162.54 km to 135.49 km, representing a significant 16.6% reduction. Among all evaluated deep learning models, DGF-Net also achieves superior forecasting performance; specifically, its 24 h forecasting error is 13.7% lower than that of the GBRNN model and 3.0% lower than the state-of-the-art AM-ConvGRU model. Comparisons with the Central Meteorological Observatory (CMO) operational mode reveal that while the NWP-based operational mode maintains a leading edge in 12 h and 24 h short-term forecasting windows, DGF-Net exhibits a unique performance advantage in 6 h nowcasting. Its Great Circle Distance (GCD) error reaches as low as 35.62 km, a 4.0% reduction compared to the CMO operational mode.

Table 9. Comparison of prediction performance of deep learning algorithms.

Models	6 h (km)	12 h (km)	24 h (km)
AM-ConvGRU	36.35	67.82	139.65
TAM-CL	38.75	74.53	143.56
GBRNN	39.21	73.39	156.95
CLIPER	36.89	72.44	162.54
DGF-Net	35.62	63.53	135.49
CMO	37.12	52.17	75.53

The error distribution characteristics of the DGF-Net for 24 h track forecasting are illustrated in Figure 11, where statistical results reveal a mean forecasting error of 135.49 km and a median of 128.05 km, exhibiting a slight right-skewed distribution. This pattern suggests that the overall mean is primarily elevated by a small number of extreme outliers rather than a systemic predictive bias, thereby confirming the model's stability and accuracy across the majority of cases. Quantile statistics further indicate that over 50% of the sample errors do not exceed 128.05 km, with 75% controlled within 166.31 km and 25% falling below 94.31 km, which substantiates the model's high precision and reliability in conventional TC forecasting scenarios. In terms of dispersion metrics, the standard deviation of 55.44 km and the range of 357.89 km reflect that while isolated complex cases may exhibit extreme deviations, the dispersion of the core predictive results remains within a controllable threshold. Within the scope of operational TC forecasting, these extreme deviations are largely attributable to complex meteorological processes such as rapid track deflections and explosive intensification, or to input data errors stemming from representative biases in observations and dynamical discontinuities within reanalysis meteorological fields, both of which remain persistent bottlenecks in the field. Collectively, these distribution characteristics comprehensively demonstrate the superior forecasting robustness of DGF-Net and its capability to maintain low-error performance in core operational scenarios.

These comprehensive results confirm the advancement and robustness of the proposed DGF-Net in short-term TC track forecasting. The performance gains over statistical and mainstream deep learning models primarily stem from the dual-branch architecture's precision in extracting 2D trajectories and 3D atmospheric data, alongside the deep integration of multi-source meteorological data facilitated by the GMFM. This effectively addresses the limitations of existing models regarding insufficient feature representation and poor multimodal fusion adaptability. Compared to the CMO operational numerical models, the advantage of DGF-Net in 6 h nowcasting indicates that it can serve as a potent supplement to traditional NWP modes, providing high-value references for operational forecasting and disaster mitigation.

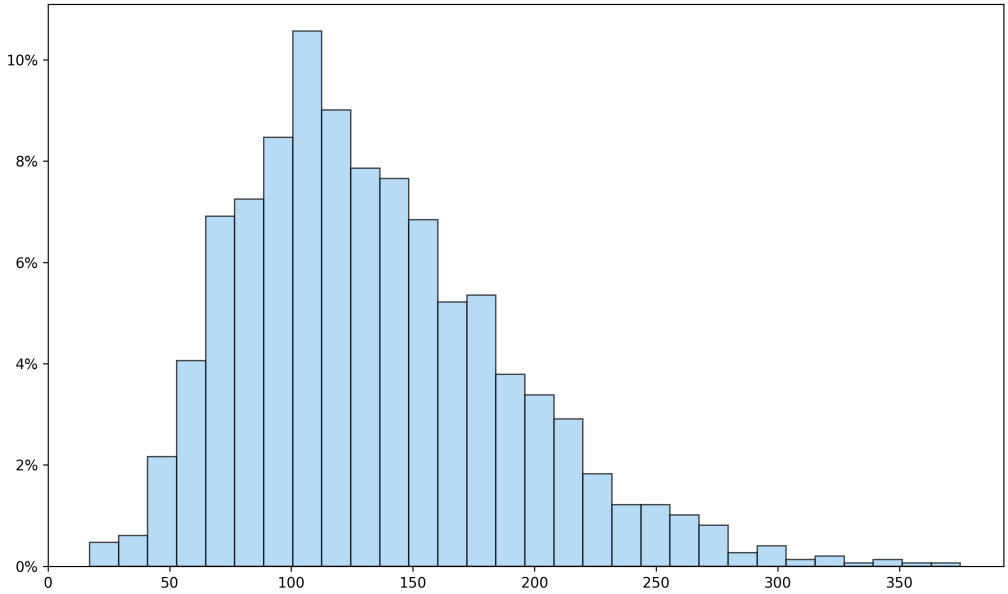


Figure 11. Error distribution of DGF-Net.

5.5. Results of Statistical Significance Test

The statistical analysis results of the Analysis of Variance (ANOVA) for all experiments are summarized in Table 10. The results indicate that the p-values for the experiments corresponding to Tables 5, 6 and 9 are all below the significance level $\alpha = 0.05$. This demonstrates that the performance differences between the models in these scenarios possess clear statistical significance. Only the 2D feature extraction experiment corresponding to Table 4 yielded a p-value of 0.1658921, which is greater than 0.05 and thus failed to pass the significance test.

Table 10. Statistical testing results in Tables 4–6 and 9.

Table	Statistic	p-Value
Table 4	1.523466	0.1658921
Table 5	2.370041	0.0499995
Table 6	7.167994	0.0000005
Table 7	3.362113	0.0179833
Table 9	10.594941	0.0000000

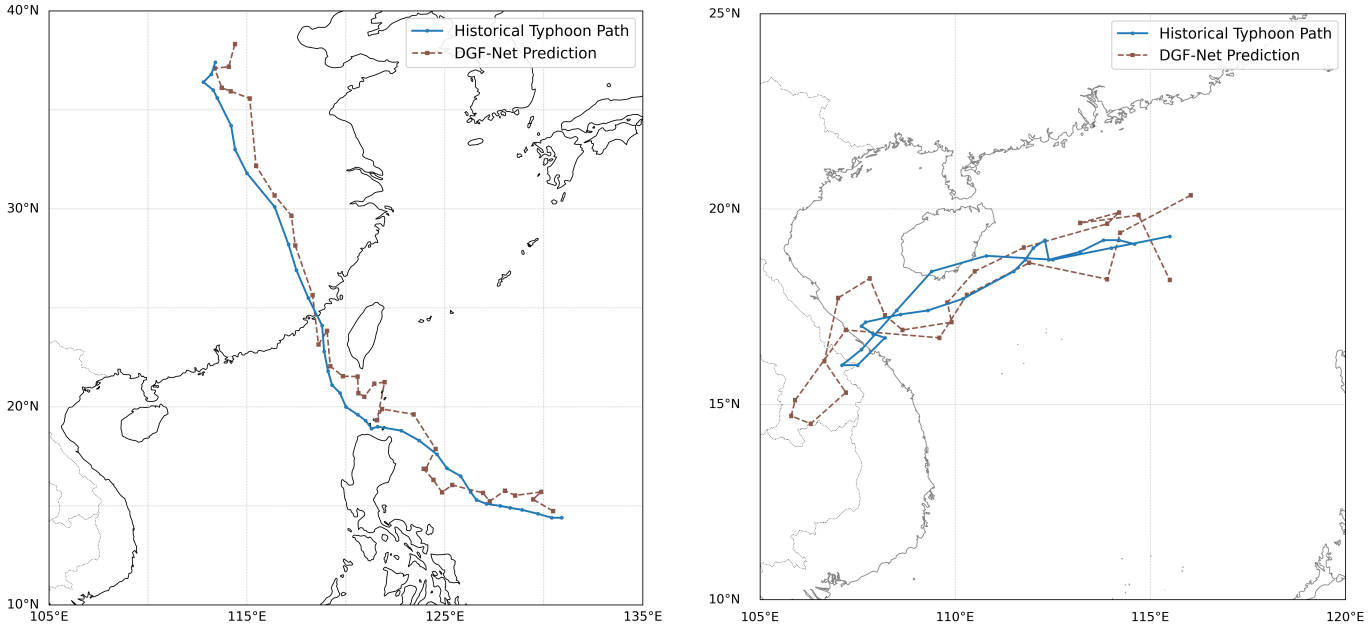
These statistical significance findings provide rigorous mathematical support for the core conclusions of this study. As shown in the consolidated ANOVA results in Table 10, the 3D feature extraction experiments (Table 5), the multimodal fusion framework experiments (Table 6), and the ablation studies (Table 9) all achieved p-values lower than the significance threshold $\alpha = 0.05$. This statistical significance indicates that the performance improvements brought about by the proposed architectural refinements are inherent advantages of the model rather than stochastic fluctuations. Such results confirm that the experimental conclusions are reliable from a statistical perspective, remaining stable and reproducible rather than being driven by accidental factors during the training process.

Conversely, the 2D feature extraction experiment (Table 4) reported a p-value of 0.1658921, exceeding the 0.05 threshold. This outcome suggests that although Bi-GRU holds a marginal lead in mean forecasting performance, the performance gains across different RNN variants are far less pronounced than those achieved through 3D spatial dimension optimization, primarily due to the inherent complexity and noise within 2D

trajectory data. This discrepancy is the fundamental underlying reason why the 2D feature extraction experiment did not pass the statistical significance test.

5.6. Results of Visualization Experiments

The visualization results of the 24 h lead-time forecasting tracks for typical typhoon cases are illustrated in Figure 12.



(**a**) Twenty-four hour forecast track of TC "DOKSURI" (**b**) Twenty-four hour forecast track of TC "Kajiki"

Figure 12. Twenty-four hour forecast track of TC "DOKSURI" and TC "Kajiki".

Super Typhoon Doksuri (2023) initially underwent a rapid intensification process, escalating to super typhoon intensity. Its subsequent trajectory evolution was significantly modulated by the adjustment of the mid-latitude Western Pacific Subtropical High and the topographic friction effects of Taiwan Island, eventually making landfall along the coast of Fujian [42]. The 24 h Great Circle Distance (GCD) error of DGF-Net for this typhoon was 115.9 km. As shown in Figure 12a, the model effectively mitigated the forecasting biases induced by the interaction between coastal topography and multi-scale circulations, precisely capturing the directional deflection of the tropical cyclone (TC) and significantly enhancing forecasting accuracy. This result corroborates the efficacy of the DGF-Net 3D feature extraction mechanism in capturing the dynamic variations of the vertical structure during the rapid intensification phase.

In contrast, Typhoon Kajiki (2019) represents a typical case of an extremely weak and poorly organized TC in the South China Sea. It was characterized by an exceptionally loose circulation structure, the absence of a clear and stable closed circulation center, and a lack of coherent cyclonic characteristic signals. Furthermore, its trajectory exhibited an extreme anomaly with a near-180° "V-shaped" erratic recurvature. DGF-Net yielded a mean 24 h GCD error of 148.2 km for this TC, with a maximum error reaching 318.19 km. As illustrated in Figure 12b, while DGF-Net demonstrated a certain capability to capture the overall movement trend, it exhibited significant positional deviations during the critical stage of anomalous recurvature, which corresponded to the peak error. The overall forecasting precision for this case was markedly lower than that for strong typhoon cases with well-defined structures and clear steering flows. This phenomenon of larger overall errors and extreme anomalous deviations exposes the model's forecasting bottleneck in complex scenarios such as weak systems, weak steering fields, and drastic trajectory changes. It also

reflects that such extreme errors may be closely related to sudden shifts in the environmental field, the uncertainty of data assimilation in weak systems, and insufficient feature learning for anomalous paths.

In summary, DGF-Net demonstrates excellent forecasting robustness at critical nodes such as landfall and track recurvature, effectively meeting the demands for path prediction in complex scenarios involving rapid intensity changes. However, for extremely weak, poorly organized TCs with highly anomalous paths, there remains further room for optimization in the model's predictive performance.

6. Limitations and Future Directions

6.1. Data Constraints and Extension Strategies

Although the ERA5 reanalysis dataset provides a consistent 3D atmospheric environmental field for the model, several constraints limit its direct operational deployment. Since reanalysis products are grid-based data generated through data assimilation and numerical model hindcasts, they suffer from a fixed release latency that hinders real-time forecasting. Since reanalysis products are generated via hindcast assimilation, models trained exclusively on ERA5 may suffer from data distribution shifts (covariate shifts) when deployed in real-time operational pipelines filled with observation noise. Furthermore, this study only incorporates six physical variables across five isobaric levels, omitting other critical quantities that modulate TC evolution. To facilitate operationalization, future research will transition toward training with real-time observations or high-performance dynamical models such as HWRF. These products offer real-time synchronization and mature physical schemes while providing grid products with higher spatial resolution and more comprehensive physical elements. Additionally, the integration of multi-source point-and-grid data can mitigate single-source assimilation errors and feature noise, thereby enhancing the completeness and robustness of the input feature set.

6.2. Expansion of Forecasting Scenarios

The current research focuses on 24 h track forecasting in the Northwest Pacific, which remains relatively narrow compared to the full-spectrum demands of operational forecasting. Given that official CMA guidance provides up to 5-day predictions, the proposed model can be extended to longer lead times by expanding the historical input time window to capture long-term temporal dependencies. Moreover, operational requirements prioritize multi-element forecasting, including TC intensity and Maximum Sustained Wind. By leveraging the existing multimodal fusion framework, auxiliary forecasting branches can be introduced to implement multi-task learning, bringing the model closer to the multi-scenario needs of real-world applications.

6.3. Structural Refinement and Interpretability

Bulk testing and case studies indicate that DGF-Net still faces challenges regarding physical interpretability and feature extraction for weak or poorly organized systems. Future iterations can incorporate Physics-Informed Neural Networks by introducing physical constraint loss functions or interpretable modules that embed atmospheric prior knowledge into the architecture. To improve the capture of weak cyclonic signals, the attention mechanisms will be optimized to focus on core motion features. Furthermore, a transfer learning strategy could be employed, wherein the model is pre-trained on high-intensity TC samples and subsequently fine-tuned on weak TC data. This approach allows the model to adapt to features with a low signal-to-noise ratio, thereby elevating the forecasting performance for systems with unstable circulation structures.

7. Conclusions

This study introduces DGF-Net, an intelligent tropical cyclone track forecasting model based on gated and attention mechanisms. By integrating historical TC trajectory sequences with ERA5 reanalysis data, we constructed a dual-branch architecture for deep feature extraction and multimodal fusion. The STConvGRU unit effectively captures the spatiotemporal features of multi-layer atmospheric data, while the GMFM implements dynamic integration of heterogeneous data, resolving the issue of modal semantic misalignment and providing a novel technical solution for precise TC track prediction. Experimental results in the Northwest Pacific demonstrate that DGF-Net significantly outperforms traditional deep learning methods by reducing track errors, particularly in nowcasting (6 h) scenarios, where it shows unique advantages over operational numerical models. The study confirms the utility of blending deep learning with meteorological domain logic and provides a feasible reference for the application of AI in meteorological forecasting.

Despite the progress achieved, the model's predictive capability remains subject to further optimization across multiple dimensions. Future work will involve utilizing real-time HWRF products to improve dataset latency and resolution, expanding the temporal window for long-term forecasting and constructing multi-task models for intensity and wind speed prediction. Architecturally, we aim to integrate physical constraints and interpretable modules while utilizing transfer learning to bolster the feature representation of weak TCs. In summary, by building a multimodal fusion framework, this research offers a new paradigm for intelligent TC forecasting. Moving forward, the focus will remain on refining the model's accuracy and adaptability to promote its operational deployment, ultimately providing more comprehensive support for disaster prevention and mitigation and contributing to the intelligent evolution of meteorological forecasting.

Author Contributions: Conceptualization, B.C., Y.W. and S.L.; methodology, Y.W.; software, Y.W.; validation, Y.W.; formal analysis, Y.W.; investigation, Y.W.; resources, Y.W.; data curation, Y.W.; writing—original draft preparation, Y.W.; writing—review and editing, S.L. and B.C.; visualization, Y.W.; supervision, S.L.; project administration, B.C. and S.L.; funding acquisition, B.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the National Science Fund of China (No. 12101138).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The CMA Tropical Cyclone Best Track dataset is available at https://tcdata.typhoon.org.cn/zjljsjj.html (accessed on 18 July 2025); ERA-5 data are available at https://cds.climate.copernicus.eu/datasets/reanalysis-era5-pressure-levels?tab=overview (accessed on 18 July 2025); and the historical forecasting accuracy data are obtained from the Typhoon Network of the Central Meteorological Observatory (CMO) via https://typhoon.nmc.cn/web.html (accessed on 1 July 2025).

Acknowledgments: During the preparation of this manuscript, the authors used the GPT-5.2 artificial intelligence tool for translation and language editing. The authors have reviewed and edited the content of this publication and take full responsibility for its accuracy and integrity.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Wu, L.; Zhao, H.; Wang, C.; Cao, J.; Liang, J. Understanding of the Effect of Climate Change on Tropical Cyclone Intensity: A Review. *Adv. Atmos. Sci.* **2022**, *39*, 205–221. [CrossRef]
2. Zhang, K.; Huang, Q.; Li, J. Hazard Analysis of Tropical Cyclone in Coastal Areas of China. *Geospat. Inf.* **2024**, *22*, 92–96.

3. Wang, H.; Zhao, P.; Wang, Y. Regional Characteristics of Tropical Cyclones Intensity and Frequency in Western North Pacific and an Analysis of Local Environmental Factors. *Meteorol. Sci. Technol.* **2025**, *53*, 520–534.
4. Chen, L.; Zhang, W.; Wang, X. Machine Learning in Tropical Cyclone Forecast Modeling: A Review. *Atmosphere* **2020**, *11*, 676. [CrossRef]
5. Duan, M.; Ma, J.; Wang, P. Preliminary comparison of the CMA, ECMWF, NCEP, and JMA ensemble prediction systems. *Acta Meteorol. Sin.* **2012**, *26*, 26–40. [CrossRef]
6. Skamarock, W.C.; Klemp, J.B.; Dudhia, J.; Gill, D.O.; Barker, D.M.; Duda, M.G.; Huang, X.-Y.; Wang, W.; Powers, J.G. A Description of the Advanced Research WRF Version 3. In *NCAR Technical Note*; National Center for Atmospheric Research (NCAR): Boulder, CO, USA, 2008; pp. 113–239.
7. Neumann, C.J. An Alternate to the HURRAN (Hurricane Analog) Tropical Cyclone Forecast System. In *NOAA Technical Memorandum NWS SR*; National Oceanic and Atmospheric Administration (NOAA): Silver Spring, MD, USA, 1972; pp. 1–23.
8. Moradi Kordmahalleh, M.; Gorji Sefidmazgi, M.; Homaifar, A. A Sparse Recurrent Neural Network for Trajectory Prediction of Atlantic Hurricanes. In Proceedings of the Genetic and Evolutionary Computation Conference 2016, Denver, Co, USA, 20–24 July 2016; pp. 957–964.
9. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
10. Cho, K.; Van Merrienboer, B.; Culcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, 25–29 October 2014; pp. 1724–1734.
11. Alemany, S.; Beltran, J.; Perez, A.; Ganzfried, S. Predicting hurricane trajectories using a recurrent neural network. In Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; pp. 468–675.
12. Kim, S.; Kim, H.; Lee, J.; Yoon, S.; Kahou, S.E.; Kashinath, K.; Prabhat, M. Deep-Hurricane-Tracker: Tracking and Forecasting Extreme Climate Events. In Proceedings of the 2019 IEEE Winter Conference on Applications of Computer Vision, Los Alamitos, CA, USA, 7–11 January 2019; pp. 1761–1769.
13. Shi, X.; Chen, Z.; Wang, H.; Yeung, D.Y.; Wong, W.K.; Woo, W.C. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 802–810.
14. Li, T.; Lai, M.; Nie, S.; Liu, H.; Liang, Z.; Lv, W. Tropical Cyclone Trajectory Based on Satellite Remote Sensing Prediction and Time Attention Mechanism ConvLSTM Model. *Big Data Res.* **2024**, *36*, 100439. [CrossRef]
15. Xu, G.; Xian, D.; Fournier-Viger, P.; Li, X.; Ye, Y.; Hu, X. AM-ConvGRU: A spatio-temporal model for typhoon path prediction. *Neural Comput. Appl.* **2022**, *34*, 5905–5921. [CrossRef]
16. Cheng, H.T.; Koc, L.; Harmsen, J.; Shaked, T.; Chandra, T.; Aradhye, H.; Anderson, G.; Corrado, G.; Chai, W.; Ispir, M.; et al. Wide & Deep Learning for Recommender Systems. In Proceedings of the 1st Workshop on Deep Learning for Recommender Systems, Boston, MA, USA, 15 September 2016; pp. 7–10.
17. Fang, W.; Lu, W.; Li, J.; Zou, L. A Novel Tropical Cyclone Track Forecast Model Based on Attention Mechanism. *Atmosphere* **2022**, *13*, 1607. [CrossRef]
18. Li, T.; Che, K.; Lu, J.; Zeng, Y.; Lv, W.; Liang, Z. Tropical Cyclone Ensemble Forecast Framework Based on Spatiotemporal Model. *Earth Sci. Inform.* **2024**, *17*, 4791–4807. [CrossRef]
19. Chen, R.; Wang, X.; Zhang, W.; Zhu, X.; Li, A.; Yang, C. A hybrid CNN-LSTM model for typhoon formation forecasting. *GeoInformatica* **2019**, *23*, 375–396. [CrossRef]
20. Ma, D.; Wang, L.; Fang, S.; Lin, J. Tropical cyclone intensity prediction by inter-and intra-pattern fusion based on multi-source data. *Environ. Res. Lett.* **2023**, *18*, 014020. [CrossRef]
21. Ying, M.; Zhang, W.; Yu, H.; Lu, X.; Feng, J.; Fan, Y.; Zhu, Y.; Chen, D. An Overview of the China Meteorological Administration Tropical Cyclone Database. *J. Atmos. Ocean. Technol.* **2014**, *31*, 287–301. [CrossRef]
22. Lu, X.; Yu, H.; Ying, M.; Zhao, B.; Zhang, S.; Lin, L.; Bai, L.; Wan, R. Western North Pacific Tropical Cyclone Database Created by the China Meteorological Administration. *Adv. Atmos. Sci.* **2021**, *38*, 690–699. [CrossRef]
23. Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horányi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; et al. The ERA5 Global Reanalysis. *Q. J. R. Meteorol. Soc.* **2020**, *146*, 1999–2049. [CrossRef]
24. World Meteorological Organization. *Manual on Tropical Cyclone Forecasting (WMO-No. 259)*; World Meteorological Organization: Geneva, Switzerland, 2009.
25. Sattar, K.; Muhammad Saad Missen, M.; Saher, N.; Bashir, R.N.; Zahra, S.Z.; Faheem, M.; Khan, A.R. Tropical Cyclone Intensity Prediction Using Spatio-Temporal Data Fusion. *IEEE Access* **2025**, *13*, 70095–70104. [CrossRef]
26. Giffard-Roisin, S.; Yang, M.; Charpiat, G.; Kumler Bonfanti, C.; Kégl, B.; Monteleoni, C. Tropical Cyclone Track Forecasting Using Fused Deep Learning From Aligned Reanalysis Data. *Front. Big Data* **2020**, *3*, 1. [CrossRef] [PubMed]

27. Tan, J.; Chen, S.; Wang, J. Western North Pacific Tropical Cyclone Track Forecasts by a Machine Learning Model. *Stoch. Environ. Res. Risk Assess.* **2021**, *35*, 1113–1126. [CrossRef]
28. Bahdanau, D.; Cho, K.; Bengio, Y. Neural machine translation by jointly learning to align and translate. In Proceedings of the Third International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015; pp. 1–15.
29. Song, T.; Li, Y.; Meng, F.; Xie, P.; Xu, D. A Novel Deep Learning Model by BiGRU with Attention Mechanism for Tropical Cyclone Track Prediction in the Northwest Pacific. *J. Appl. Meteorol. Climatol.* **2022**, *61*, 3–12. [CrossRef]
30. Song, S.; Lan, C.; Xing, J.; Zeng, W.; Liu, J. Spatio-Temporal Attention-Based LSTM Networks for 3D Action Recognition and Detection. *IEEE Trans. Image Process.* **2018**, *27*, 3459–3471. [CrossRef]
31. Lian, J.; Chen, R. A Sequence-to-Sequence Based Multi-Scale Deep Learning Model for Satellite Cloud Image Prediction. *Earth Sci. Inform.* **2023**, *16*, 1207–1225. [CrossRef]
32. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E. Squeeze-and-Excitation Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 2011–2023. [CrossRef]
33. Boussioux, L.; Zeng, C.; Guénais, T.; Bertsimas, D. Hurricane Forecasting: A Novel Multimodal Machine Learning Framework. *Weather Forecast.* **2022**, *37*, 817–831. [CrossRef]
34. Niu, Z.; Zhong, G.; Yu, H. A review on the attention mechanism of deep learning. *Neurocomputing* **2021**, *452*, 48–62. [CrossRef]
35. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł; Polosukhin, I. Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 6000–6010.
36. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
37. Huang, M.F.; Wang, Q.; Jing, R.Z.; Lou, W.J.; Hong, Y.; Wang, L.Z. Tropical Cyclone Full Track Simulation in the Western North Pacific Based on Random Forests. *J. Wind Eng. Ind. Aerodyn.* **2022**, *228*, 105119. [CrossRef]
38. Central Meteorological Observatory (CMO). Typhoon Network. Available online: https://typhoon.nmc.cn/web.html (accessed on 1 July 2025).
39. Li, Z.C.; Zhang, L.; Qian, Q.F.; Ma, S.; Xu, J.; Dai, K.; Chen, Y.; Wang, Y. The Development and Consideration of Typhoon Forecast Operation of Central Meteorological Center. *Trans. Atmos. Sci.* **2020**, *43*, 10–19. [CrossRef]
40. Cox, D.R. Statistical Significance. *Annu. Rev. Stat. Its Appl.* **2020**, *7*, 1–10. [CrossRef]
41. St, L.; Wold, S. Analysis of Variance (ANOVA). *Chemom. Intell. Lab. Syst.* **1989**, *6*, 259–272. [CrossRef]
42. Liu, D.; Xiang, C.; Zhang, L.; Xu, Y.; Luo, Q. Analysis on main characteristics of Typhoon Doksuri (2305) and difficulties in its track and intensity forecast. *J. Mar. Meteorol.* **2023**, *43*, 1–10.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.