Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A geophysically constrained Transformer model for tropical cyclone track and intensity simulation over the western North Pacific and its contribution analysis

Ziqi Lin
Beijing Normal University
Wenxuan Li
Beijing Normal University
Shuo Wang
Beijing Normal University
Jing Xu
Monash University
Siqi Zhang
Beijing Normal University
Heng Ma
Ministry of Emergency Management of the People's Republic of China
Peijun Shi
Beijing Normal University
Gangfeng Zhang
gangfeng@bnu.edu.cn

Beijing Normal University

Article

Keywords: Tropical cyclone prediction, geophysically constrained Transformer model, Multi-modal data, contribution analysis

Posted Date: October 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11123951/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A geophysically constrained Transformer model for tropical cyclone track and intensity simulation over the western North Pacific and its contribution analysis

Ziqi Lin[1,2,3], Wenxuan Li[1,2], Shuo Wang[4], Jing Xu[5], Siqi Zhang[1,2], Heng Ma[6], Peijun Shi[1,2,3*], Gangfeng Zhang[1,2,3*]

[1]*State Key Laboratory of Earth Surface Processes and Disaster Risk Reduction, Beijing Normal University, Beijing 100875, China;*

[2]*Academy of Disaster Reduction and Emergency Management, Ministry of Emergency Management and Ministry of Education, Beijing Normal University, Beijing 100875, China;*

[3]*Key Laboratory of Environmental Change and Natural Disaster of Ministry of Education, Beijing Normal University, Beijing, 100875, China;*

[4] *School of Systems Science, Beijing Normal University, Beijing 100875, China*

[5]*Department of Data Science and AI, Monash University, Melbourne 3800, Australia*

[6] *National Institute of Natural Hazards, Ministry of Emergency Management of China, Beijing 100085, China*

Manuscript submitted to npj Climate and Atmospheric Science

Research Article

31 August 2026

* Corresponding author: Peijun Shi

E-mail: spj@bnu.edu.cn

Gangfeng Zhang

E-mail: gangfeng@bnu.edu.cn

Abstract: Tropical cyclone represents one of the most devastating natural hazards in global coastal zones, inflicting massive economic damage and loss of life annually. Precise multi-step tropical cyclone forecasting is still confronted with great challenges, primarily attributed to intricate ocean-atmosphere interactions governing tropical cyclone movement. Here, we propose a geophysically constrained Transformer model to jointly predict tropical cyclone tracks and intensity using historical best-track records over the western North Pacific and reanalysis data, benchmarked against conventional XGBoost and LSTM models. Experimental results show that the geophysically constrained Transformer model realizes stable tropical cyclone forecasting, with track RMSEs of 200.72 km (6 h) and 357.02 km (36 h), and intensity RMSEs of 2.59 m/s (6 h) and 7.54 m/s (36 h). When compared with LSTM and XGBoost baselines, it reduces 6-hour track RMSE by 30% and 55%, and intensity RMSE by 23% and 36%, respectively, maintaining better accuracy over most forecast periods for both tasks. SHAP analysis further reveals that surface air pressure (contributing 32%) and wind speed (contributing 25%) are the dominant predictive variables for tropical cyclone track and intensity forecasting. By contrast, large-scale circulation and sea surface temperature function as secondary influencing factors, which further validates the physical interpretability and rationality of the developed model.. Overall, our findings advance the integration of deep learning techniques with meteorological and geophysical datasets for real-time operational tropical cyclone prediction, yielding critical implications for tropical cyclone monitoring and disaster risk governance.
Key word: Tropical cyclone prediction; geophysically constrained Transformer model; Multi-modal data; contribution analysis

Tropical cyclones are intense vortical systems that form over tropical and subtropical oceans, they are also referred to as typhoon hurricanes, or tropical storms depending on the region (Knutson et al., 2020). They originate from warm, moist air that rises from the ocean surface and develops into a warm-core low-pressure vortex under favorable dynamic and thermodynamic conditions(Emanuel, 1986, 2003; Knutson et al., 2020; Kossin et al., 2020). In a warming climate, both the intensity and number of extreme tropical cyclones have increased, making them among the most damaging natural hazards in global coastal regions and surroundings(Kossin et al., 2020; Knutson et al., 2020). Strong tropical cyclones produce extreme winds and storm surges

that damage buildings, infrastructure, forests and crops, while the concurrent heavy rainfall often triggers secondary disasters such as coastal flooding and landslides, resulting in severe economic losses and casualties(Knutson et al., 2020). The western North Pacific (WNP) is the world's most active ocean basin for tropical cyclone activity, generating over one-third of the global total number of TCs annually, most of which track westward and northwestward, frequently affecting coastal countries in East Asia. Situated on the western side of the WNP, China is among the countries most affected by tropical cyclones, incurring annual direct economic losses of roughly 10 billion yuan, about 15% of global tropical cyclone-related losses. Enhancing tropical cyclone monitoring, early warning, and risk prevention has therefore become a national priority.

Accurate and timely tropical cyclone track and intensity prediction is notoriously challenging, given that cyclone dynamics are governed by an array of complex variables such as sea surface temperature and atmospheric circulation(Bauer et al., 2021). Over the past several decades, physics-driven numerical weather prediction (NWP) models have remained the cornerstone of operational track forecasting(Bauer et al., 2021). These models solve nonlinear equations governing atmospheric dynamics and thermodynamics, supplemented by parameterizations of cloud microphysics, radiation, and related processes, to simulate large-scale circulation and tropical cyclone movement on supercomputers(Bauer et al., 2021). Continuous advances in spatial resolution, physical process representation, and data assimilation have substantially improved the skill of global models such as ECMWF and NCEP, making them essential to meteorological operations and disaster-risk management(DeMaria et al., 2005).

However, growing climate complexity and increasing demands for real-time hazard response have highlighted key limitations of traditional physically-based NWP(Bauer et al., 2021). Since high-resolution simulations of those models require processing massive atmospheric datasets and thus depend heavily on computational resources, leading to long forecast turnaround times that hinder rapid updates during fast-evolving tropical cyclone events(Bi et al., 2023). Moreover, most NWP models provide deterministic forecasts, making it difficult to efficiently generate multi-scenario track ensembles needed for risk assessment, emergency decision-making, and insurance loss estimation(Wadler et al. 2023). Under constrained compute capacity, further improvements in resolution or physical schemes alone are unlikely to yield simultaneous gains in both accuracy and timeliness. For instance, during Tropical

cyclone Lekima (2019), although major models captured its turning behavior in advance, substantial errors remained in predicting landfall intensity and rainfall distribution(Fowler et al. 2021), while each forecast cycle required 4 – 6 hours, insufficient for operational rolling updates.

To address these challenges, existing studies have increasingly turned to data-driven methods. By learning statistical regularities of cyclone evolution from large archives of observations and reanalysis datasets, data-driven models can reduce computation time from hours to minutes while maintaining competitive accuracy, and can efficiently produce probabilistic forecasts and scenario ensembles (Pathak et al., 2022). This emerging integration of physically-based and data-driven paradigms is viewed as a promising direction for next-generation forecasting systems (Reichstein et al., 2019). Moreover, recent advances in deep learning have further accelerated this trend. Architectures such as convolutional neural networks (CNNs), recurrent networks (RNNs/LSTMs), spatiotemporal Transformers, and graph neural networks (GNNs) have been applied to tropical cyclone track prediction, intensity estimation, and generative simulation, demonstrating performance broadly comparable to NWP in early studies (Giffard-Roisin et al., 2020; Wang et al., 2023; Zhao et al., 2022; Huang et al., 2025). Nonetheless, most research remains focused on feasibility testing and accuracy comparison, without fully exploiting deep learning's advantages in rapid inference, parallelism, and model reusability. The inability to fully capture the complex spatiotemporal interactions between large-scale atmospheric circulation and the autonomous evolution of tropical cyclones leads to significant error accumulation during medium-range forecasting. Furthermore, while physics-informed machine learning has shown promise in reconstructing tropical cyclone wind fields and enforcing dynamical consistency (Hu and Li, 2024), existing deep learning-based tropical cyclone prediction frameworks largely ignore geophysical environmental effects and typically suffer from poor model interpretability, leaving the contributions of various meteorological factors unquantified.

Given the unsatisfactory prediction performance of tropical cyclone forecasting and the disconnection between prevailing machine learning methods and physical meteorological processes, this study proposed a deep learning model for combined track and intensity prediction of tropical cyclones. Meanwhile, SHAP interpretability analysis is introduced to quantitatively evaluate the contribution of multi-source

environmental and tropical cyclone endogenous features and further clarifying the physical driving logic of tropical cyclone evolution.

Result

Performance of the proposed GCTF model

Given the pronounced nonlinear relationships governing tropical cyclone evolution and related dynamic-thermal mechanisms, the GCTF model adopts multi-source meteorological factors to realize tropical cyclone prediction. Overall, track prediction errors of GCTM rise nonlinearly with extended forecast lead time, as longer forecasts entail greater atmospheric uncertainty and resultant positional deviations (Table 1). The model achieves favorable accuracy in short-term forecasts, with the 6-hour forecast RMSE reaching 200.72 km, and still keeps acceptable prediction capability at the 36-hour forecast with RMSE limited to 357.02 km. MAE is consistently smaller than RMSE at all forecast moments, manifesting that the model generates stable outputs without obvious extreme deviations. Moreover, the intensity RMSE grows from 2.59 m/s at 6 h to 7.54 m/s at 36 h, while MAE increases from 1.93 m/s to 5.70 m/s, mirroring the nonlinear error accumulation pattern observed in track prediction. The consistent pattern of MAE being smaller than RMSE across all steps further indicates stable intensity estimation without catastrophic outlier errors. Combined with the foregoing typical tropical cyclone case analysis, the statistical results further verify that the constructed model possesses reliable prediction capacity and satisfactory stability in short and medium-range tropical cyclone track forecasting.

Table.1 Performance of the proposed GCTF model. Quantitative evaluation metrics for different time steps are presented in the three rightmost columns.

steps	Time	Path forecasting error			Intensity forecasting error		
		MSE	RMSE	MAE	MSE	RMSE	MAE
1	6 h	40287.924	200.719	177.290	6.719	2.592	1.928
2	12 h	47020.495	216.842	192.390	11.290	3.360	2.501
4	24 h	74713.825	273.338	240.621	35.025	5.918	4.407
6	36 h	127463.213	357.020	311.239	56.891	7.543	5.698

Scatter pairs of observed and simulated maximum sustained wind speeds congregate closely about the 1:1 reference line, confirming the model reliably reproduces the magnitude and temporal variability of wind speed. (Figure. 1). Note that prediction uncertainty varies systematically with wind speed. Prediction errors are small for low wind speeds (< 40 m/s), where predictions closely match observations. By contrast, prediction spread increases with wind speed rises, with the largest deviations occurring for intense tropical cyclone cases exceeding 60 m/s, reflecting reduced model reliability under extreme wind speed conditions. The model predictions maintain comparatively good consistency with observations across the intermediate wind-speed band (40–60 m/s), a feature partially attributable to the well-balanced sample size within this range.

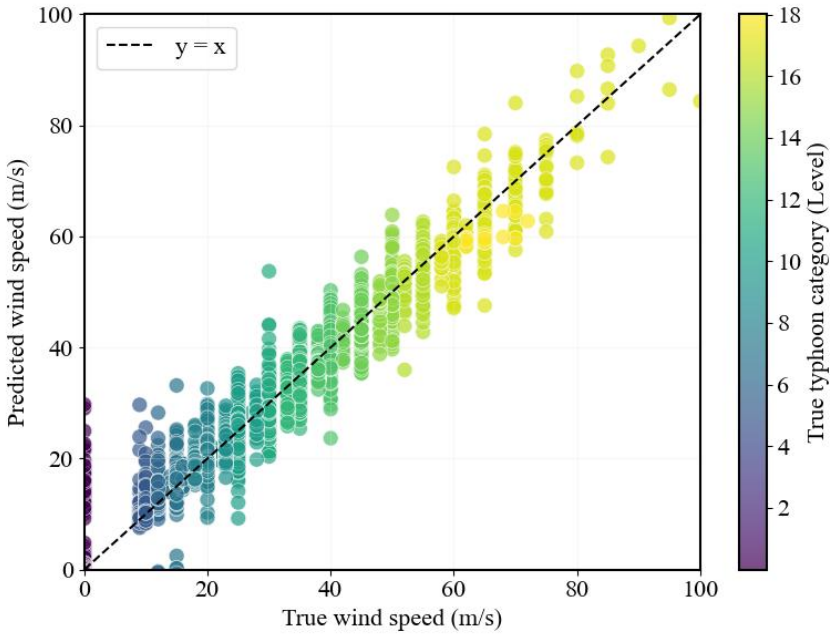


Fig. 1 Predicted versus observed maximum wind speeds for the test set. The dashed line indicates y = x line. Colors denote observed tropical cyclone intensity, with warmer colors corresponding to stronger storms.

Comparison with conventional deep learning models

The multi-step forecast performance of the proposed GCTF model was further compared with LSTM, and XGBoost models in terms of RMSE, MAE, and MSE metrics. Across all forecast steps, the proposed GCTF model yields the lowest prediction errors, with the most pronounced advantage at short lead times (Figure 2). At the initial step (6 h), the GCTF model achieves an RMSE of ~200 km, substantially

lower than LSTM (~287 km) and XGBoost (~480 km), corresponding to reductions of approximately 30% and 55%, respectively. Note that the GCTF model maintains RMSE values within 200-244 km over the first three steps, indicating more accurate short-term trajectory prediction. As the forecast lead times extend to steps 4-8, the GCTF model continues to perform competitively, with RMSE ranging from 273 to 480 km. While the multi-metric gaps between GCTF and LSTM narrows and LSTM achieves slight advantages at certain longer horizons ($\geqslant$ 7 steps), the GCTF model outperforms XGBoost by more than 50%. These outcomes suggest that the proposed GCTF model effectively captures spatiotemporal dependencies, leading to robust performance across varying forecast horizons when compare to conventional deep learning models.

For intensity, GCTF again outperforms both baselines through the first six steps (Fig. 2b, d, f). At 6 h, its RMSE is ~2.7 m/s versus ~3.5 m/s for LSTM and ~4.2 m/s for XGBoost (23% and 36% reductions), and MAE shows the same ordering (~1.9 vs ~2.4 vs ~2.6 m/s). GCTF maintains the lowest RMSE, MAE, and MSE up to step 6. At steps 7-8, XGBoost has marginally lower RMSE/MSE and LSTM slightly lower MAE—different architectures appear to accumulate error differently in extreme intensity regimes. GCTF nonetheless provides the most consistent intensity predictions in the operationally important 6–36 h range.

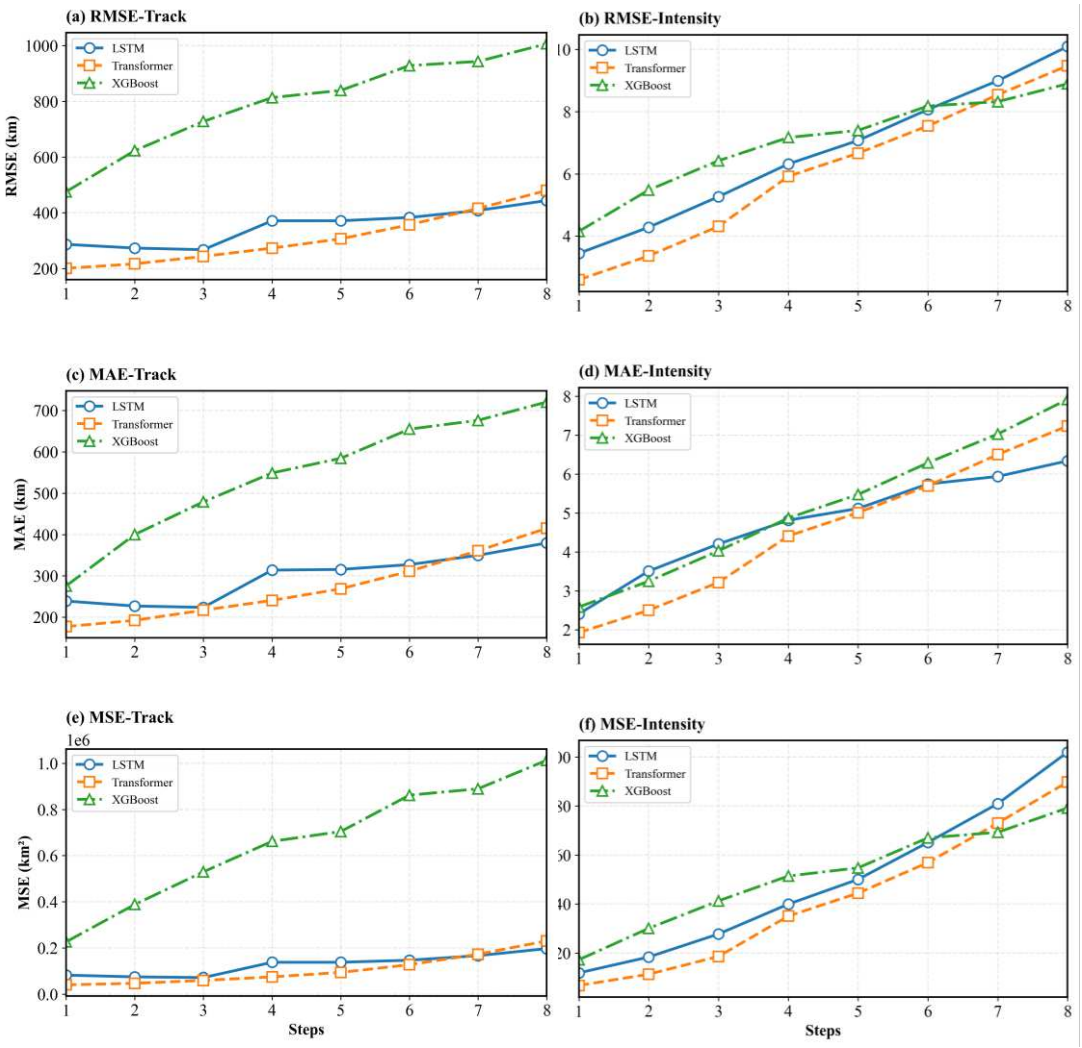


Fig 2. Comparison of prediction errors for tropical cyclone tracks and intensity multi-step forecasting using LSTM, Transformer, and XGBoost models. The variability of (a) RMSE, (c) MAE, and (e) MSE with increasing prediction steps are presented.

The performance of the GCTF model in real cases

To further verify the proposed GCTF model performance in individual tropical cyclone cases, two typical tropical cyclones are selected for detailed case evaluation, including northwestward-moving open-ocean Tropical cyclone Megi and rapidly intensifying offshore Tropical cyclone Talim. In terms of track simulation (Fig. 3a, b), the GCTF model basically reproduced the actual movement paths of the two tropical cyclones. Tropical cyclone Megi achieved good consistency between simulated and observed tracks in the early stage, and its west-northward movement characteristics were well reflected. However, a slight westward deviation occurred in the middle phase, and track prediction errors gradually increased during northward progression. For

Tropical cyclone Talim, the model effectively captured its westward advancing trend towards South China, with limited positional deviation appearing near landfall. In intensity simulation (Fig. 3c, d), the GCTF model reasonably reproduced the overall variation of tropical cyclone strengthening and weakening.

The GCTM model simulated maximum sustained wind speed of Tropical cyclone Megi was close to observed values, with peak intensity deviation below 3 m/s at approximately 40 m/s. The rapid intensification process of Tropical cyclone Talim could be basically identified, yet its maximum sustained wind speed was slightly underestimated, with simulated value of 58 m/s against the observed 68 m/s. The simulations remained stable throughout the whole forecast period without unreasonable numerical fluctuation. The above case evaluations indicate that the established GCTF model yields stable short-to-medium term prediction of tropical cyclone track and intensity under diverse environmental conditions, highlighting its practical implications in operational tropical cyclone forecasting.

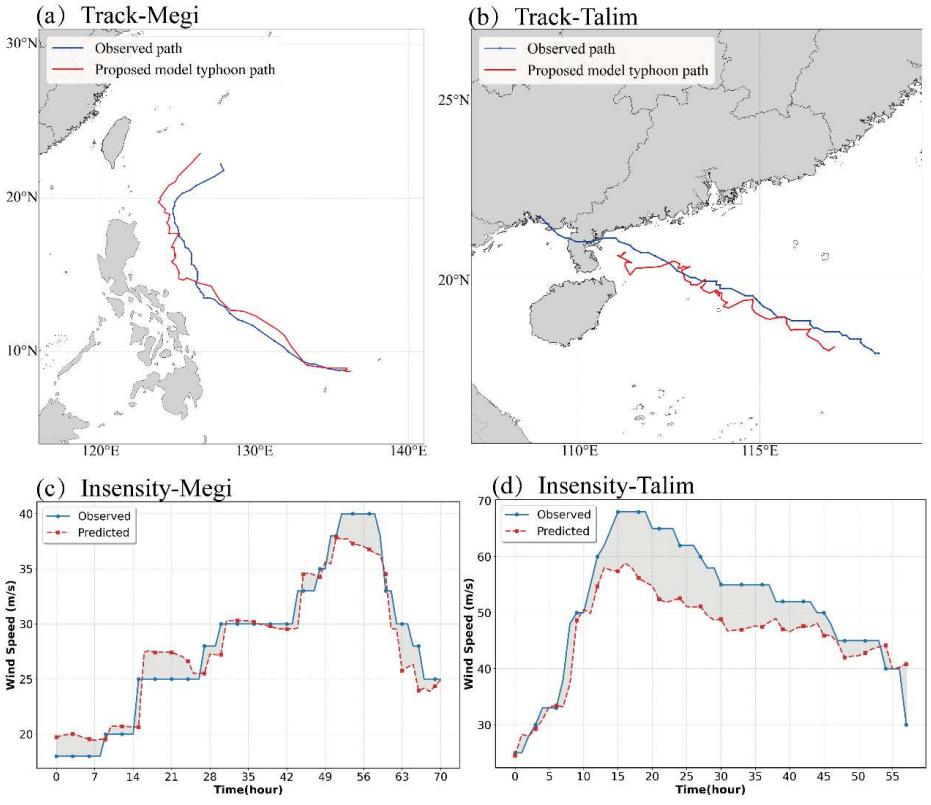


Fig 3. Comparison of observed and predicted tropical cyclone track and intensity. (a)-(b) 72-hour track forecast performance for Tropical cyclone Megi and Tropical cyclone Talim, respectively. (c)-(d) 72-hour intensity forecast performance for Tropical cyclone Megi and Tropical cyclone Talim, respectively. Blue lines represent observed data, and

red lines represent predictions from the proposed model. The shaded area indicates the prediction error range. Tropical cyclone intensity is quantified using maximum sustained wind speed.

Contribution of various features in GCTF model prediction

SHAP-based feature importance analysis was employed to elucidate the contributions of various features on tropical cyclone prediction. Overall, surface air pressure emerged as the dominant predictor, accounting for 32% of the total contribution, with lower central surface pressure consistently correlating with stronger tropical cyclone intensity, in line with classic meteorological theories (Figure 4). With a contribution rate of 25%, maximum wind speed(vamx) ranked second key features, for which initial wind field conditions effectively constrain subsequent tropical cyclone intensity changes. Moreover, mid-tropospheric wind fields and geopotential height exerted moderate contribution (18%) to GCTF model simulations, reflecting the regulatory role of large-scale atmospheric circulation on tropical cyclone movement.

For the remaining factors, sea surface temperature supplies fundamental environmental conditions for tropical cyclone propagation yet contributes merely 8% to GCTF model simulations, imposing limited influence on short-term intensity variabilities. In contrast, 100 m wind speed and direction had the weakest effect (combined contribution < 5%) on GCTM model simulations, as they are primarily generated by internal tropical cyclone circulation rather than core driving forces. Overall, GSTM model predictions were dominated by real-time tropical cyclone inherent states, with external environmental factors playing an auxiliary role.

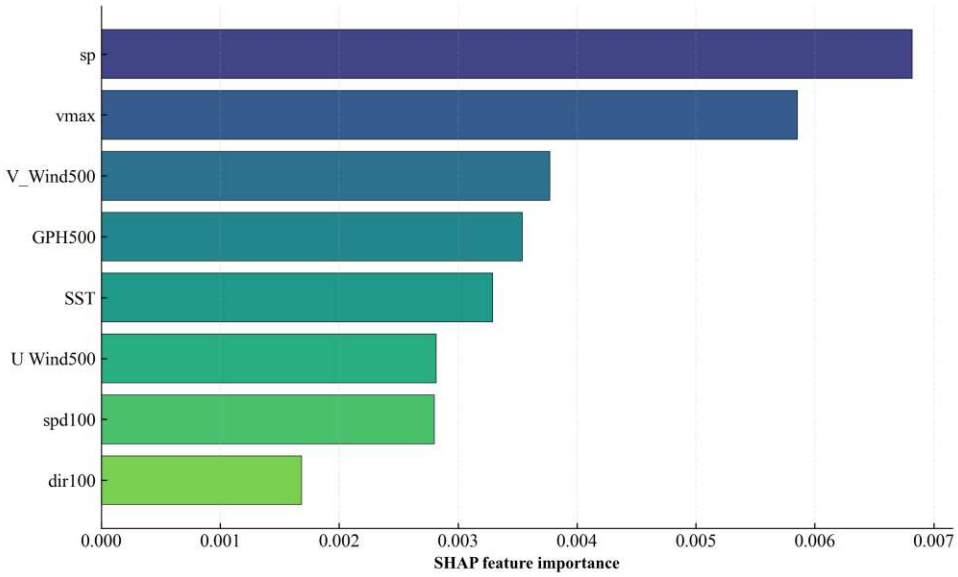


Fig. 4. SHAP-based feature importance for tropical cyclone track and intensity prediction. Bars show mean absolute SHAP values, with surface pressure (sp) and wind speed identified as the dominant predictors, while 100 m wind speed and direction (spd100, dir100) contribute the least.

Discussion

To tackle the defects in spatiotemporal modeling and the lack of meteorological physical constraints in existing deep learning-based tropical cyclone forecasting, we proposed a geophysically constrained Transformer model. Combining ERA5 reanalysis and CMA observations as multi-modal inputs, the model performs joint prediction of tropical cyclone tracks and intensity across the western North Pacific within a 72-hour forecast window. Overall, the proposed GCTF model generates reliable forecasts of tropical cyclone track and intensity. While prediction errors increase nonlinearly with lead time, they remain acceptable for short-to-medium range projections. Intensity estimates agree well with observations under low-to-moderate wind conditions, yet deteriorate for intense tropical cyclones, owing to insufficient representation of inner-core structures and rapid intensification processes (Qu et al., 2025; Wang et al., 2024). The model outperforms LSTM and XGBoost counterparts, particularly in short-term forecasts, by better capturing spatiotemporal dependencies in meteorological fields (Bi et al., 2023), although such advantages diminish in longer lead times. Accordingly, the model is suitable for real-time short-term tropical cyclone monitoring and early warning with satisfactory stability.

Nevertheless, track deviations in real-case simulations can be attributed to three common physical factors: biases in large-scale steering flows (Yang et al., 2024; Ma et al., 2025), amplified multi-scale interactions between inner-core convection and environmental circulation (Wang et al., 2024; Chen et al., 2023), and unresolved land–atmosphere interactions near coastlines (Hu et al., 2024). The interactions among these physical mechanisms cannot be well represented by either inner-core data or environmental field data alone. Despite these biases, the model reasonably reproduces the overall evolution of tropical cyclone track and intensity without spurious fluctuations, supporting its stability and effectiveness. Furthermore, case validation further confirms that the model achieves dependable performance for short-to-medium range forecasts under varied atmospheric conditions.

In addition, SHAP-based feature importance analysis reveals a hierarchical structure of feature contributions. Short-term intensity variability is dominated by surface air pressure and wind speed, indicating strong dependence on the initial storm state (Qu et al., 2025). Mid-tropospheric winds and geopotential height provide moderate contributions, reflecting the regulatory role of large-scale circulation systems (Ma et al., 2025; Yang et al., 2024). Sea surface temperature and 100 mwind parameters exert weak influences due to slow temporal variation and secondary characteristics of internal tropical cyclone circulation (Guan et al., 2025). This hierarchical contribution pattern confirms that the model relies on physically consistent mechanisms rather than purely data-driven fitting, consistent with recent interpretable deep learning applications for tropical cyclone intensity (Lee et al., 2024; Tang et al., 2025). Thus, the proposed model can serve as a transferable baseline for deep learning-based operational tropical cyclone prediction (Bi et al., 2023; Lam et al., 2023) and support interdisciplinary progress toward more accurate and timely tropical cyclone early warning (Bauer et al., 2021).

In this study, we establish a practical multi-modal deep learning framework that connects artificial intelligence and atmospheric science, yet several limitations remain. The generalizability of the GCTF model is restricted by limited case coverage and sample imbalance, particularly for extreme intensification events (Ascenso et al., 2024; Yan et al., 2025). Feature interpretation only considers individual factors rather than coupled interactions (Lundberg et al., 2020). Universal model parameters lack region-

and intensity-specific calibration, and distribution shifts across basins and intensity regimes remain challenging (Choo and Kim, 2024; Yan et al., 2025). Validation of the GCTF model largely depends on indirect best-track and reanalysis datasets without sufficient high-resolution in-situ observations. These limitations are associated with data scarcity, computational resources, and the inherent complexity of tropical cyclone systems. Future work will expand the sample library to include extreme tropical cyclones, quantify interactive effects among predictive features (Lundberg et al., 2020), implement region- and intensity-specific parameter calibration via transfer learning (Choo and Kim, 2024), and incorporate high-quality satellite and remote-sensing observations for more rigorous validation (Zhang et al., 2026).

Conclusion

This paper proposes a novel multi-modal deep learning method for typhoon track and intensity prediction, providing a feasible scheme for the application of advanced deep learning techniques in tropical cyclone forecasting. 1) Quantitative results show that the GCTF model achieves reliable short-to-medium-range forecasting performance. For track prediction, RMSE values are 200.72 km and 357.02 km for 6 h and 36 h forecasts, respectively, and all MAE values are consistently lower than RMSE, indicating stable model outputs. For intensity prediction, the GCTF model achieves an RMSE of 2.59 m/s at 6 h and 7.54 m/s at 36 h, demonstrating strong capacity in capturing maximum sustained wind speed variability at short lead times. 2) Compared with LSTM and XGBoost baseline models, the proposed model significantly reduces short-term forecast errors. At the 6-hour lead time, it reduces track RMSE by 30% against LSTM and 55% against XGBoost, and likewise lowers intensity RMSE by approximately 23% and 36% relative to LSTM and XGBoost, respectively, maintaining optimal prediction accuracy across most forecast durations for both track and intensity. 3) Real-case tests on Typhoon Megi and Typhoon Talim further verify that the GCTF model can stably reproduce typhoon track and intensity evolution, with only minor deviations for strong typhoon intensity simulation. 4) SHAP interpretation demonstrates that surface air pressure (32%) and wind speed (25%) serve as the dominant predictive factors, while large-scale circulation and sea surface temperature provide auxiliary contributions, confirming the model's physically consistent prediction mechanism.

In short, this multi-modal model exhibits prominent advantages in short-term

typhoon forecasting and can provide valuable implications for future tropical cyclone prediction combining deep learning techniques and meteorological big data.

Methods

Dataset

This subsection presents the two datasets used in thisstudy, for 2D tropical cyclones and 3D tropical cyclones, respectively. Moreover, the preprocessing methods for both datasets are described.

The tropical cyclone best track dataset over the western North Pacific, compiled by the Shanghai Typhoon Institute of the China Meteorological Administration (CMA, https://tcdata.typhoon.org.cn/zjljsjj.html), covers tropical cyclones in the western North Pacific and South China Sea from 1949 to 2024. It provides 6-hourly records of tropical cyclone center position, intensity (pressure and wind speed), classification, and wind radii, serving as an authoritative source for tropical cyclone characteristics, climate impacts, and disaster risk assessment in previous studies. In addition, the fifth-generation ECMWF atmospheric reanalysis (ERA5) dataset (Hersbach et al., 2020) provides 6-hourly multi-level atmospheric and surface variables, including sea surface temperature, geopotential height, relative humidity, and horizontal wind components, which serve as environmental predictors characterizing the large-scale circulation surrounding each tropical cyclone. More details of both datasets are summarized in Table 2.

Table 2. Summary of the tropical cyclone best-track dataset over the western North Pacific and ERA5 reanalysis datasets used in this work. Each column lists the name, content, unit, temporal resolution and data source of individual variables.

Dataset	Variable	Content	Unit	Temporal resolution	Data source
Tropical cyclone Track	lat	Latitude of Tropical cyclone center	°N	6 h	CMA Best Track (https://tcdata.typhoon.
	lon	Longitude of Tropical cyclone center	°E	6 h	
	pmin	Minimum central	hPa	6 h	

		pressure			org.cn/zjljsjj.html)
	vmax	Maximum wind speed	$m \cdot s^{-1}$	6 h	
Meteorology	SST	Sea surface temperature	K	6 h	ERA5（https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels?)
	sp	Surface air pressure	pa	6 h	
	GPH500	Geopotential height at 500 hPa	m	6 h	
	Spd100	Wind direction at 100 m height	$m \cdot s^{-1}$	6 h	
	Dir100	Wind speed at 100 m height	°	6 h	
	U_Wind 500	Zonal wind at 500 hPa	$m \cdot s^{-1}$	6 h	
	V_Wind 500	Meridional wind at 500 hPa	$m \cdot s^{-1}$	6 h	

To prepare the datasets for subsequent analysis and model training, a series of preprocessing procedures were applied. The raw tropical cyclone best-track dataset was first quality-controlled to remove incomplete or anomalous records. Missing or placeholder values were excluded to ensure data reliability. All variables were temporally aligned to a uniform time interval to maintain consistency among multi-source inputs. The predictors were then normalized to reduce scale differences and improve model convergence. Finally, the processed data were organized into time-ordered sequences suitable for further model training and prediction. Representative tropical cyclone tracks derived from the dataset are illustrated in Fig.5.

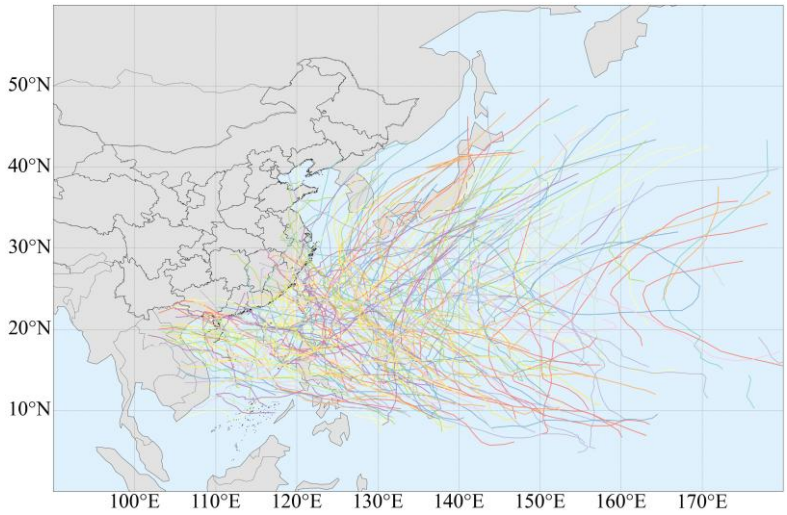


Fig 5. Historical tropical cyclone tracks in the western North Pacific. Each colored line depicts the full path of one tropical cyclone, showing typical genesis regions and propagation routes affecting East Asia.

Transformer model architecture

To address two key limitations in tropical cyclone (TC) forecasting—insufficient prediction precision and inadequate integration of deep learning with domain physical knowledge—we propose a geophysically constrained Transformer (GCTF) model **(Fig 6)**. GCTF forecasts future TC center positions and intensity levels from historical track sequences and multi-source environmental fields, employing an encoder-decoder architecture in which the encoder extracts spatio-temporal representations and the decoder generates future trajectories via cross-attention.

Formally, given a historical track sequence of length L and a prediction horizon of T steps (T = 6, corresponding to 36 hours at 6-hourly intervals), the model learns a sequence-to-sequence mapping:

$$\hat{Y} = f_\theta(X, E, C) \qquad (1)$$

where $X \in R^{L \times F_x}$ denotes the historical TC track features at 6-hourly intervals, including latitude, longitude, central sea-level pressure, maximum sustained wind speed, and translational speed and direction (F_x = 6). $E \in R^{L \times F_e}$ denotes multi-source environmental field features — including sea surface temperature, 500-hPa geopotential height, 850-hPa relative vorticity, and vertical wind shear—extracted at the corresponding TC center locations from reanalysis data (F_e is the number of environmental variables). $C \in \{1, \ldots, K\}$ is the trajectory cluster assignment obtained

from offline trajectory clustering (Section X), serving as a discrete prior on the TC motion regime. The output $\hat{Y} \in R^{T\times 3}$ represents the predicted TC positions (longitude and latitude) and intensity level for the next T time steps.

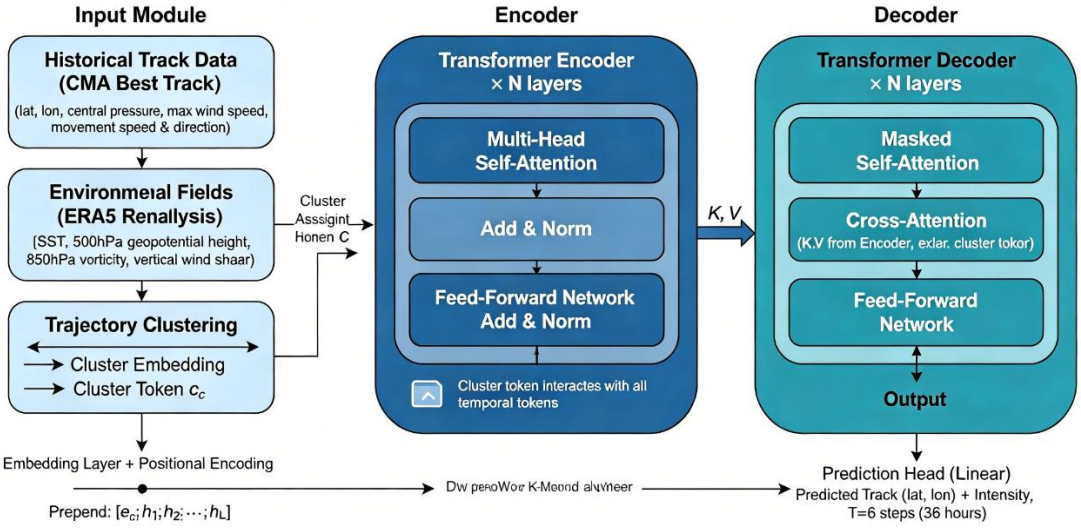


Fig 6. Architecture of the geophysically constrained Transformer model for tropical cyclone intensity and trajectory prediction. The model consists of an encoder-decoder structure. The encoder (left) takes multi-source data as input, and extracts spatio-temporal features using a multi-head self-attention mechanism. The decoder (right) leverages cross-attention to model the correlation between historical features and future states, generating predictions for the tropical cyclone's 6-hour trajectory and intensity level.

Attention mechanism and cluster-guided embedding

To capture both the temporal evolution of TC motion and the influence of large-scale environmental fields, GCTF employs a multi-head self-attention mechanism in the encoder. At each encoder layer, the input representation $\mathbf{H} \in \mathbb{R}^{L\times d}$ is linearly projected into query (**Q**), key (**K**), and value (**V**) matrices:

$$\mathbf{Q} = \mathbf{H}\mathbf{W}_\mathrm{Q}, \quad \mathbf{K} = \mathbf{H}\mathbf{W}_\mathrm{K}, \quad \mathbf{V} = \mathbf{H}\mathbf{W}_\mathrm{V} \tag{2}$$

where $\mathbf{W}_\mathrm{Q}$, $\mathbf{W}_\mathrm{K}$, $\mathbf{W}_\mathrm{V} \in \mathbb{R}^{d\times d_k}$ are learnable projection matrices and d_k is the dimension of each attention head. The scaled dot-product attention is computed as:

$$\mathrm{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \mathrm{softmax}(\mathbf{Q}\mathbf{K}^T / \sqrt{d_k})\,\mathbf{V} \tag{3}$$

Multi-head attention concatenates the outputs of h parallel attention heads and projects the result through a learnable matrix $\mathbf{W}_O \in \mathbb{R}^{h \cdot d_k \times d}$, enabling the model to jointly attend to information from different representation subspaces.

The cluster prior is incorporated through a cluster-guided token embedding. First, each TC trajectory is assigned to one of K motion regimes via offline trajectory clustering (Section X), yielding a discrete cluster label:

$$C = \text{argmin}_\{k \in \{1, \ldots, K\}\} \text{DTW}(\mathbf{T}, \mathbf{c}_k) \quad (4)$$

where $\mathbf{T}$ is the historical track sequence, $\mathbf{c}_k$ is the centroid of the k-th cluster, and DTW is the dynamic time warping distance. A learnable embedding matrix $\mathbf{W}_\text{clu} \in \mathbb{R}\^\{K \times d_\text{model}\}$ maps the discrete label to a dense vector:

$$\mathbf{e}_c = \mathbf{W}_\text{clu}[C], \quad \mathbf{e}_c \in \mathbb{R}^d \quad (5)$$

This cluster embedding is then prepended to the input sequence as a special token, allowing the self-attention mechanism to dynamically modulate the weights of temporal dependencies based on the trajectory type:

$$\mathbf{H}_\text{in} = [\mathbf{e}_c; \mathbf{h}_1^{(0)}; \mathbf{h}_2^{(0)}; \ldots; \mathbf{h}_L^{(0)}] \quad (6)$$

where $\mathbf{h}_t^{(0)} \in \mathbb{R}\^\{d_\text{model}\}$ is the embedded representation of the t-th time step (after track/environmental feature projection and positional encoding), and $[\cdot;\cdot]$ denotes concatenation along the sequence dimension. The resulting sequence has length $L + 1$ and is fed into the Transformer encoder. Within the encoder, the cluster token interacts with all temporal tokens via self-attention, enabling the model to attend to different environmental patterns for recurving TCs versus straight-moving TCs.

In the decoder, cross-attention layers take the encoder output (excluding the cluster token) as key and value, while decoder self-attention operates on previously predicted positions (teacher forcing during training). This allows the decoder to align each future prediction step with the most relevant historical and environmental features.

Benchmark models and evaluation metrics

To evaluate the performance of the GCTM model, two benchmark approaches were also implemented: (i) LSTM (Long Short-Term Memory, Hochreiter and

Schmidhuber, 1997), a recurrent neural network widely used for sequence prediction and previously applied to real-time hurricane evolution forecasting (Bose et al., 2022); (ii) XGBoost (Extreme Gradient Boosting, Chen and Guestrin, 2016), a highly efficient gradient boosting decision tree algorithm renowned for outstanding performance in tabular data prediction and regression tasks.

The detailed configurations of all benchmark models are provided in Table S1. All deep learning models were trained on the NVIDIA Tesla V100 (32G) GPU with a batch size of 64, learning rate of 1e-4, and early stopping with patience of 10 epochs based on validation loss.

In this study, the prediction performance of the models is evaluated using the Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Square Error (MSE)and coefficient of determination (R^2). The corresponding formulas are defined as follows:

$$MAE = \left(\frac{1}{N}\right)\sum\left|Y_i - \hat{Y}_i\right| \quad (7)$$

$$RMSE = \sqrt{\left(\frac{1}{N}\right)\sum_{\{i=1\}}^{\{N\}}\left(Y_i - \hat{Y}_i\right)^2} \quad (8)$$

$$MSE = \left(\frac{1}{N}\right)\sum_{\{i=1\}}^{\{N\}}\left(Y_i - \hat{Y}_i\right)^2 \quad (9)$$

$$R^2 = 1 - \frac{\left(\sum(Y_i - \hat{Y}_i)^2\right)}{\left(\sum(Y_i - \bar{Y})^2\right)} \quad (10)$$

where N denotes the total number of samples; represents the actual observed values; $Y_i$refers to the TC values predicted by the models; and $\bar{Y}$ is the mean of the actual values. MAE quantifies the average absolute deviation between predicted and observed values, while RMSE provides a measure of the standard deviation of the prediction errors.

SHapley Additive exPlanations (SHAP)

To quantitatively disentangle the predictive contribution of each input variable to short-term tropical cyclone intensity forecasts, SHapley Additive exPlanations (SHAP) framework was adopted to interpret the trained spatiotemporal Transformer model.

SHAP values calculate the marginal predictive impact of each feature by simulating all possible feature combinations, which avoids bias caused by feature correlation and delivers physically interpretable global feature importance ranking(Lundberg and Lee 2017). For the constructed dataset containing TC core observational features and ERA5 variables, we computed the mean absolute SHAP

value of each variable across all test samples as the global feature importance metric. The averaged SHAP magnitudes were sorted to quantify the relative weight of surface vortex characteristics, mid-tropospheric circulation fields, low-level wind parameters and sea surface thermal conditions in regulating TC intensity variability.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (42330502, 42101027), and Independent Research Project of the National Key Laboratory for Earth Surface Processes and Disaster Risk Reduction (2025-ZY-09). This work was also supported by the DOWNBURST (Ref. CIPROM/2023/38, Generalitat Valenciana -GVA- for Excellence Research Groups); Ministry for the Ecological Transition and the Demographic Challenge (MITECO) and European Commission – NextGenerationEU (Regulation EU 2020/2094), through CSIC's Interdisciplinary Thematic Platform Clima (PTI Clima); and the RED-CLIMA 2 project (Ref. LINCG24042, CSIC-LINCGLOBAL).

References

Ascenso, G., Giuliani, M., Dainelli, F., et al. (2024). A systematic framework for data augmentation for tropical cyclone intensity estimation using deep learning. Journal of Geophysical Research: Machine Learning and Computation, 1(3), e2024JH000206. https://doi.org/10.1029/2024JH000206

Bauer, P., Dueben, P.D., Hoefler, T., et al. (2021). Artificial intelligence for weather and climate. Nature, 596(7870), 472–478. https://doi.org/10.1038/s41586-021-03854-8

Bi, K., Xie, L., Zhang, H., Chen, X., He, B., Wang, F., et al. (2023). Accurate medium-range global weather forecasting with 3D neural networks. Nature, 619(7970), 533–538. https://doi.org/10.1038/s41586-023-06185-3

Bose, R., Pintar, A., & Simiu, E. (2022). A real time prediction methodology for hurricane evolution using LSTM recurrent neural networks. Neural Computing and Applications, 34(20), 17491–17505. https://doi.org/10.1007/s00521-022-07384-1

Chen, B.-F., Kuo, Y.-T., & Huang, T.-S. (2023). A deep learning ensemble approach for predicting tropical cyclone rapid intensification. Atmospheric Science Letters, 24(4), e1151. https://doi.org/10.1002/asl.1151

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785

Choo, M., & Kim, Y. (2024). Bridging satellite missions: deep transfer learning for enhanced tropical cyclone intensity estimation. GIScience & Remote Sensing, 61(1), 2325720. https://doi.org/10.1080/15481603.2024.2325720

DeMaria, M., Knaff, J.L., & Sampson, C.R. (2005). The SHIPS forecast system: Update and

verification. Weather and Forecasting, 20(4), 531–543. https://doi.org/10.1175/WAF876.1

Emanuel, K.A. (1986). An air-sea interaction theory for tropical cyclones. Part I: Steady-state maintenance. Journal of the Atmospheric Sciences, 43(6), 585–604. https://doi.org/10.1175/1520-0469(1986)043<0585:AASITF>2.0.CO;2

Emanuel, K.A. (2003). Tropical cyclones. Annual Review of Earth and Planetary Sciences, 31, 75–104. https://doi.org/10.1146/annurev.earth.31.100901.135348

Giffard-Roisin, S., Yang, M., Charpiat, G., Kumler-Bonfanti, C., Kégl, B., & Monteleoni, C. (2020). Tropical cyclone track forecasting using fused deep learning from aligned reanalysis data. Frontiers in Big Data, 3, 8. https://doi.org/10.3389/fdata.2020.00008

Guan, S., Wang, C., Wang, Y., et al. (2025). Weak self-induced cooling of tropical cyclones amid fast sea surface warming. Nature Geoscience. https://doi.org/10.1038/s41561-025-01879-x

Hersbach, H., Bell, B., Berrisford, P., et al. (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hu, C., Tam, C.-Y., Li, Z., Chen, J., Chow, T.N., Li, Y., Cheung, K.K.W., Chu, J.-E., & Yang, Z.-L. (2024). Impacts of urban heat and surface roughness on landfalling tropical cyclone intensity: A case study based on TC Victor (1997) in coastal South China. Journal of Geophysical Research: Atmospheres, 129(23), e2024JD040915. https://doi.org/10.1029/2024JD040915

Hu, F., & Li, Q. (2024). Reconstruction of tropical cyclone boundary layer wind field using physics-informed machine learning. Physics of Fluids, 36(11), 116608. https://doi.org/10.1063/5.0234728

Huang, H., Deng, D., Hu, L., Chen, Y., & Sun, N. (2025). Beyond the backbone: A quantitative review of deep-learning architectures for tropical cyclone track forecasting. Remote Sensing, 17(15), 2675. https://doi.org/10.3390/rs17152675

Knutson, T.R., Camargo, S.J., Chan, J.C.L., et al. (2020). Tropical cyclones and climate change assessment: Part II: Projected response to anthropogenic warming. Bulletin of the American Meteorological Society, 101(3), E303–E322. https://doi.org/10.1175/BAMS-D-18-0194.1

Kossin, J.P., Hall, T., Knutson, T.R., et al. (2020). Global increase in major tropical cyclone exceedance probability over the past four decades. Proceedings of the National Academy of Sciences, 117(22), 11975–11980. https://doi.org/10.1073/pnas.1922589117

Lam, R., Sanchez-Gonzalez, M., Willson, M., Das, A., Mercier, A., Oh, C., et al. (2023). Learning skillful medium-range global weather forecasting. Science, 382(6677), 1416–1421. https://doi.org/10.1126/science.adi2336

Lee, J., Im, J., & Shin, Y. (2024). Enhancing tropical cyclone intensity forecasting with explainable deep learning integrating satellite observations and numerical model outputs. iScience, 27(6), 109768. https://doi.org/10.1016/j.isci.2024.109768

Lundberg, S.M., & Lee, S.I. (2017). A unified approach to interpreting model predictions.

Advances in Neural Information Processing Systems, 30.

Lundberg, S.M., Erion, G., Chen, H., et al. (2020). From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence, 2(1), 56–67. https://doi.org/10.1038/s42256-019-0138-9

Ma, Y., Qiu, H., Feng, B., Qi, H., Zhang, J., Gong, W., & Liu, J. (2025). Impacts of environment on the sudden track change of super typhoon GAEMI (2024). Frontiers in Earth Science, 13, 1601840. https://doi.org/10.3389/feart.2025.1601840

Pathak, J., Subramanian, S., Harrington, P., Raja, S., Chattopadhyay, A., Mardani, M., et al. (2022). FourCastNet: A global data-driven high-resolution weather model using adaptive Fourier neural operators. arXiv:2202.11214.

Qu, H., Xu, H., Dong, L., Xiang, C., & Nie, G. (2025). Accurate tropical cyclone intensity forecasts using a non-iterative spatiotemporal transformer model. npj Climate and Atmospheric Science, 9, 9. https://doi.org/10.1038/s41612-025-01279-3

Reichstein, M., Camps-Valls, G., Stevens, B., et al. (2019). Deep learning and process understanding for data-driven Earth system science. Nature, 566(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1

Tang, Y., Zhang, Y., Li, X., Tu, S., Xu, J., Zeng, S., Shao, D., Hu, J., & Tan, Q. (2025). Advancing tropical cyclone intensity forecasts over the South China Sea with a meteorology-aware and attention-enhanced deep learning model. Journal of Geophysical Research: Machine Learning and Computation, 3(3), e2025JH001102. https://doi.org/10.1029/2025JH001102

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.

Wang, L., Wan, B., Zhou, S., Sun, H., & Gao, Z. (2023). Forecasting tropical cyclone tracks in the northwestern Pacific based on a deep-learning model. Geoscientific Model Development, 16, 2167–2179. https://doi.org/10.5194/gmd-16-2167-2023

Wang, Y., Li, X., & Zhang, Q. (2024). A review of typhoon inner core characteristics and their relationship with intensity changes. Atmosphere, 15(12), 1522. https://doi.org/10.3390/atmos15121522

Yan, H., Mu, P., Zhang, S., Zhu, Y., Zhang, J., & Bai, C. (2025). IDOL: Meeting diverse distribution shifts with prior physics for tropical cyclone multi-task estimation. ICLR 2025 Conference Proceedings.

Yang, X., Liu, D., Wang, Y., et al. (2024). A hybrid machine learning/physics-based modeling framework for 2-week extended prediction of tropical cyclones. Journal of Geophysical Research: Oceans, 129(7), e2024JC000207. https://doi.org/10.1029/2024JH000207

Zhang, J., Tan, Z.-M., Guo, S., & Zhu, K. (2026). Integrating diurnal pulsing signatures for AI-driven tropical cyclone intensity prediction. Geophysical Research Letters, 53, e2025GL119496. https://doi.org/10.1029/2025GL119496

Zhao, J., Huang, H., & Wang, X. (2022). A review on the application of machine learning methods in tropical cyclone forecasting. Frontiers in Earth Science, 10, 902596. https://doi.org/10.3389/feart.2022.902596