CAAI Transactions on Intelligence Technology

WILEY

The Institution of Engineering and Technology

ORIGINAL RESEARCH OPEN ACCESS

Cascade Deep Learning With Physics Guidance for Typhoon Intensity Prediction

Zhengya Sun[1,2] | Bojie Fan[3] | Bo Yin[3,4] | Jie Nie[3] | Anan Liu[5]

[1]University of Chinese Academy of Sciences, Beijing, China | [2]Institute of Automation, Chinese Academy of Sciences, Beijing, China | [3]Faculty of Information Science and Engineering, Ocean University of China, QingDao, China | [4]Qing Dao Jari Industry Control Technology Co. Ltd, QingDao, China | [5]School of Electrical and Information Engineering, Tianjin University, Tianjin, China

Correspondence: Bo Yin (ybfirst@ouc.edu.cn)

Received: 16 December 2025 | **Revised:** 10 June 2026 | **Accepted:** 24 June 2026

Keywords: physics-guided neural networks | spatio-temporal sequence | typhoon intensity prediction

ABSTRACT

Typhoons pose a significant threat to both human safety and economic stability. As a crucial metric for assessing their destructive potential, typhoon intensity (TI) prediction has become an important research focus, with numerous methods developed. However, effectively combining two-dimensional typhoon structure domain-expert knowledge (2D-TSDK) with three-dimensional typhoon structure data-driven knowledge (3D-TSDK) for accurate TI prediction remains a challenge. In this paper, we propose a novel cascaded deep learning framework that combines domain expertise with data-driven refinement for typhoon intensity prediction. The two-stage architecture first generates initial intensity estimates by fusing 2D-TSDK features with established physical wind-pressure relationships via a specialised loss function. These estimates are subsequently refined by a vision transformer that extract fine-grained spatial patterns from 3D wind field reanalysis data. Both stages incorporate temporal attention mechanisms with dual (absolute and relative) position encoding, providing comprehensive modelling of typhoon evolution dynamics. Extensive experiments on CMA–BST and ERA-interim datasets demonstrate consistent performance improvements over baseline methods, confirming the framework's effectiveness for TI prediction.

1 | Introduction

Tropical cyclones (TCs), among the most destructive natural disasters, originate over warm tropical oceans. They frequently cause substantial economic losses and human casualties through intense winds, heavy rainfall and storm surges during and after landfall, particularly in coastal areas vulnerable to TCs [1]. As global warming intensifies, rising sea surface temperatures further exacerbate typhoon activity by increasing both their frequency and intensity. Typhoon intensity (TI), a crucial metric for disaster risk assessment, is primarily determined by the maximum sustained surface wind speed near the storm's centre. Higher intensity levels are directly linked to greater economic impacts and more severe environmental damage. Therefore, accurate and timely forecasts of TI are essential for effective disaster preparedness and mitigation.

Traditional typhoon intensity prediction methods primarily fall into two categories: numerical modelling and statistical modelling. Numerical models solve a system of fluid dynamics equations through discretisation and numerical computation [2], providing a physically grounded framework for typhoon forecasting. However, their performance is constrained by the accuracy of initial conditions and inherent simplifications in the governing equations. Statistical models, on the other hand, rely

Zhengya Sun and Bojie Fan contributed equally to this work.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *CAAI Transactions on Intelligence Technology* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology and Chongqing University of Technology.

on historical typhoon data to construct regression-based predictive models [3]. However, their predictive capability is highly dependent on data quality and volume, and they may not adequately capture the complex nonlinear dynamics of TC intensity evolution [4].

Recent advances in deep learning have spurred growing interest in typhoon intensity prediction. Unlike traditional methods that rely on manually engineered predictors, deep learning can automatically extract relevant features from raw or minimally processed data. This capability stems from its remarkable ability to capture the complex nonlinear physical interactions within typhoon systems, which are often oversimplified in parametric or statistical schemes. Furthermore, deep learning provides a flexible framework for integrating heterogeneous data sources (e.g., satellite imagery and gridded atmospheric data) into a unified predictive model. Currently, the field primarily utilises two distinct data representations: two-dimensional typhoon structure domain knowledge (2D-TSDK) and three-dimensional typhoon structure data-driven knowledge (3D-TSDK). The 2D-TSDK method focuses on essential typhoon characteristics, including geographic coordinates (latitude/longitude), central pressure and other key parameters mainly derived from infrared satellite imagery [5–9]. Although computationally efficient, this method often fails to capture crucial contextual information about typhoon evolution and environmental interactions. In contrast, the 3D-TSDK method utilises comprehensive atmospheric profiles through gridded representations that incorporate longitude, latitude and altitude dimensions, typically obtained from advanced assimilation systems [10, 11]. Despite its richer spatial representation, this method is hindered by assimilation biases, error propagation from source models and computational complexity when processing the high-dimensional data fields. These limitations highlight the importance of effectively integrating both representations to leverage their complementary strengths while mitigating their respective weaknesses.

Current attempts to integrate 2D-TSDK and 3D-TSDK primarily follow one of two strategies. The first strategy involves encoding the two data types separately and concatenating their features [12–15], whereas the second treats them as complementary multimodal inputs, using adversarial fusion for joint representation learning [16]. Although both strategies rely on feature-level fusion, their single-stage prediction frameworks often yield suboptimal results. This challenge is further compounded in typhoon prediction, which operates in a complex hypothesis space with nonstationary variable relationships. These limitations highlight the need for multistage frameworks that integrate physical knowledge to enhance prediction accuracy.

In this paper, we propose CPC (cascade physics-coupled network), a novel deep learning framework that integrates 2D and 3D data representations with physics-based constraints for robust typhoon prediction. The framework adopts a two-stage sequential architecture, beginning with a 2D-TSDK model that generates initial predictions, followed by a 3D-TSDK corrector that refines residual errors to improve accuracy. Specifically, the 2D-TSDK model processes multivariate time series inputs while incorporating a physics-informed loss function derived from the established thermodynamic relationship between tropical cyclone central pressure and maximum wind speeds. This design ensures both empirical error minimisation and adherence to meteorological principles. The 3D-TSDK corrector then refines the predictions by analysing 3D wind field reanalysis data through a vision transformer, which captures fine-grained spatial patterns often missed in 2D representations. To effectively model the temporal dynamics of typhoon evolution, both stages utilise a temporal attention mechanism with dual positional encoding, combining absolute and relative position information. Extensive experiments on CMA–BST and ERA-interim datasets demonstrate consistent performance improvements over baseline methods, confirming the frameworks effectiveness for TI prediction.

The main contributions of this paper are summarised as follows:

1. A novel two-stage deep learning framework that synergistically combines a 2D-TSDK feature extractor with a 3D-TSDK spatiotemporal corrector to model planar and volumetric dynamics sequentially.
2. A physics-guided training paradigm that incorporates the fundamental pressure–wind relationship as a learning objective, enforcing consistency with fundamental meteorological principles.
3. Comprehensive experimental validation showing the model's superior accuracy over both conventional and state-of-the-art deep learning baselines, with residue magnitudes progressively decreasing across successive stages.

The remainder of this paper is organised as follows. Section 2 introduces the related work on traditional approaches and deep learning-based techniques for TI prediction. Section 3 presents the problem definition and our two-stage deep learning framework with physics guidance for typhoon intensity estimation, and the experimental results are reported in Section 4. Section 5 summarises the conclusion of this work.

2 | Related Work

Accurate prediction of typhoon intensity remains a critical yet challenging task. Existing methods can be broadly classified into two categories: traditional approaches (including numerical and statistical modelling) and deep learning-based techniques.

2.1 | Traditional Approaches

Numerical models simulate typhoon intensity changes by solving fluid dynamics equations, incorporating key factors such as typhoon structure, environmental conditions and large-scale meteorological fields [17, 18]. The development of these models traces back to Vilhelm Bjerknes' pioneering work on numerical weather prediction in the early 20th century [19]. Today, they have evolved into three main categories: global models (e.g., ECMWF [20] and GFS [21]), regional models (e.g., WRF [22] and MM5 [23]) and ensemble forecasting systems (e.g., GEFS [24]). However, these models require extensive atmospheric condition parameters as input and depend on large-scale computing resources [25]. Despite these demands, their performance in TI prediction remains limited.

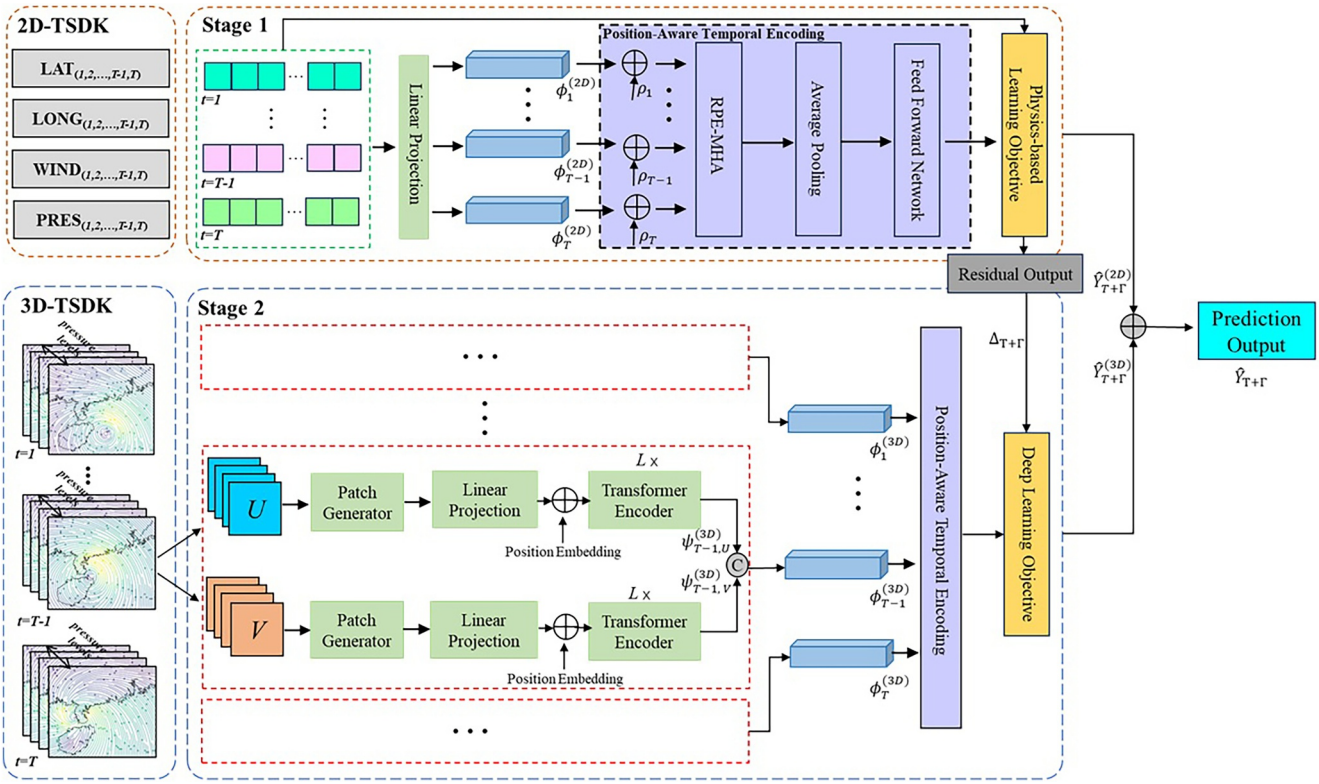


FIGURE 1 | Diagram of the overall framework of CPC. The model consists of two stages. The first stage is a physics-guided 2D-TSDK model that uses 2D-TSDK from CMA–BST dataset to generate initial predictions. The second stage is a 3D-TSDK corrector, which uses 3D-TSDK from ERA-Interim dataset to refine the residual errors. The final prediction is the sum of both stages' outputs.

Statistical models predict typhoon intensity by analysing historical records of key parameters, such as wind speed, central pressure and other relevant factors, to establish predictive frameworks. A typical example is the SHIFOR method, which employs backward stepwise regression to integrate climatological and persistence factors into a statistical model [26]. Another widely used method is the Statistical Hurricane Intensity Prediction Scheme (SHIPS), which applies multiple linear regression (MLR) to relate TI to climatology, persistence, and environmental factors [27]. Subsequent improvements to SHIPS involved dynamically optimising predictors by incorporating atmospheric and thermodynamic variables whilst continuously updating the dataset to remove redundancy, thereby enhancing forecast accuracy [28]. Beyond linear regression, Lee et al. [29] proposed an autoregressive model combing a deterministic MLR with a stochastic process to simulate the climatological distribution of global tropical cyclone (TC) intensity. Further advancing statistical techniques, Jing et al. [30] used a hidden Markov model (HMM) to simulate TC intensity evolution by treating transitions between intensity states as a Markov chain dependent on the large-scale environment. This method demonstrated significant improvements over previous statistical models.

2.2 | Deep Learning-Based Techniques

With the significant advances of deep learning in spatiotemporal prediction in recent years, these techniques have been increasingly adopted for typhoon intensity forecasting. The models primarily utilise two types of input data: two-dimensional typhoon structure domain-expert knowledge (2D-TSDK) and three-dimensional typhoon structure data-driven knowledge (3D-TSDK).

Early studies employed satellite imagery with CNN-based methods for TI prediction, leveraging their capability to capture 2D spatial features through multispectral data and typhoon motion patterns. Initial work by Pradhan et al. [31] established a baseline CNN model comprising five convolutional and three fully connected layers. Subsequently, Chen et al. [32] proposed an improved AlexNet architecture, where pooling and dropout layers were removed to preserve critical spatial details. To further exploit multispectral correlations, Lee et al. [33] introduced a hybrid 2D-CNN/3D-CNN framework. Later, Chen et al. [34] developed a GAN–CNN fusion framework to address temporal data heterogeneity in satellite observations. To better model spatiotemporal dependencies, Tong et al. [35] implemented a convolutional LSTM (ConvLSTM) network specifically designed for short-term TI predictions in the Northwest Pacific basin. Several optimised CNN architectures, including the visual geometry group network (VGGNet) and you only look once-neural architecture search (YOLO-NAS), have also been explored [36, 37]. Beyond convolutional models, recent studies have introduced vision transformers (ViTs), which employ a self-attention mechanism to extract comprehensive global features and contextual information from entire images [38, 39]. Building on these developments, Katragadda [40] proposed an optimised vision transformer integrated with a dynamic graph convolutional network to enhance the precision of TI prediction.

To improve typhoon system modelling, researchers have increasingly incorporated 3D-TSDK data as a complement to the 2D-TSDK data. Most existing approaches employ separate encoders for 2D and 3D data before concatenating the features [11–13]. For example, Chen et al. [12] proposed a hybrid CNN–LSTM model, where a 2D-CNN extracts spatial features from sea surface variables whereas a 3D-CNN captures atmospheric variable patterns, followed by an LSTM to model temporal dependencies in typhoon evolution. Similarly, Jiang et al. [11] introduced a spatial attention module combining 2D and 3D convolutions to extract multimodal features, which were then fed into a ConvLSTM-based spatio-temporal forecasting model. Xu et al. [13] further advanced this direction by developing a three-branch CNN for 3D-TSDK analysis and integrating it with 2D-TSDK using a wide and deep framework for typhoon intensity prediction. More recently, Li et al. [16] proposed a multimodal adversarial fusion method, where feature-encoded data from different modalities are fused adversarially to align cross-modal feature distributions. Lian et al. [41] further developed a sequence-to-sequence model that integrates multimodal data and captures correlational features between different data sources based on the wide and deep learning paradigm, leading to improved prediction accuracy.

2.3 | Physics-Guided Neural Networks (PGNN) Framework

Physics-guided data science (PGDS) is an emerging paradigm that integrates established physical principles with data-driven models to enhance both interpretability and predictive performance [42]. At its core, this methodology relies on physics-informed loss functions, where physical laws, such as differential equations, conservation relations and inequality constraints, are formally embedded into the model's training objective. For practical implementation, the ReLU activation function (or its variants) can either directly approximate these constraints or smoothly relax them into differentiable penalty terms.

Recent advances in physics-guided data science (PGDS) have demonstrated the effectiveness of physics-informed loss functions across diverse applications. McGowan et al. [43] developed a hybrid model for additive manufacturing (AM) that combines empirical and simulated laser metal deposition (LMD) data, showing how physics-based loss functions enhance porosity prediction. Similarly, Li et al. [44] proposed a physically guided neural network (PGNN) with a novel loss function that integrates physical constraints to improve sensor fault diagnosis under dynamic conditions, addressing the limitations of purely data-driven approaches. Kapusuzoglu et al. [45] further extended this paradigm to fused filament fabrication (FFF), embedding physics constraints into deep neural networks (DNNs) through a multistage approach. In environmental modelling, Daw et al. [46] designed a PGNN for lake temperature prediction, where a custom loss function enforced depth-dependent density relationships using ReLU-based penalties to ensure physical consistency. Meanwhile, Gu et al. [47] applied similar principles to quadcopter dynamics, embedding momentum conservation laws as soft constraints in a DNN to model highly nonlinear high-dimensional systems with improved efficiency. Collectively, these studies highlight how integrating domain-specific physics into loss functions can significantly improve model performance whilst maintaining interpretability.

3 | Methodology

3.1 | Problem Definition

The TI prediction can be formulated as a spatialtemporal regression problem. Given a sequence of historical typhoon features $\{X_1, X_2, ..., X_T\}$ observed up to time step T, we aim to learn a mapping function that predicts the TI values $Y_{T+\Gamma}$ at lead time step Γ. This can be formally expressed as follows:

$$X_1, X_2, ..., X_T \xrightarrow{f} Y_{T+\Gamma}, \tag{1}$$

where $X_t = \left(X_t^{(2D)}, X_t^{(3D)}\right)$, $X_t^{(2D)}$ denotes the 2D-TSDK and $X_t^{(3D)}$ denotes the 3D-TSDK. Here, the time interval between adjacent time steps is 6 h. With $T = 4$ (corresponding to a 24-h lead time), the model utilises 18 h typhoon features as input to predict TI 24 h in advance.

The network architecture of the proposed model is illustrated in Figure 1, which consists of two stages, that is, a physics-guided 2D-TSDK model that generates initial predictions, and a 3D-TSDK corrector that refines the residual errors. The final prediction is the sum of both stages' outputs.

3.2 | Pressure-Wind Modelling

The intensity of a typhoon is governed by multiple factors, with sea surface temperature (SST) playing a pivotal role. When SST exceeds a certain threshold, it drives intense ocean evaporation, supplying abundant moisture that fuels typhoon development. As the warm moist air rises, water vapour condenses at higher altitudes, releasing latent heat. This heating of the upper atmosphere induces a pressure drop through hydrostatic adjustment, strengthening the pressure gradient and ultimately increasing the typhoon's maximum sustained wind speed (MSW) [48].

Consistent with this physical understanding, minimum sea level pressure (MSLP) near the typhoon core is a key intensity indicator. It can be directly measured via reconnaissance aircraft or dropsondes, whereas satellite and radar retrievals provide alternative estimates in data-sparse regions. In contrast, measuring MSW is far more challenging due to the transient nature of wind observations and limited spatial coverage [49]. Given these constraints, satellite-derived MSLP has become the primary benchmark for typhoon intensity assessment, with MSW often estimated through empirical wind-pressure relationships (WPRs). Thus, incorporating central pressure data and robust WPRs is critical for both physical consistency and forecast reliability.

Numerous empirical relationships have been developed to estimate the maximum surface wind speed of tropical cyclones from their minimum central pressure. Despite decades of research, a single universally stable pressure–wind relationship suitable for operational use across all ocean basins remains elusive. These models typically depend on several key parameters (e.g., environmental pressure, air density and the radius of maximum wind) which can be set to default values or regionally

calibrated when some training data are available. The underlying rationale for such calibration is to account for the inherent scatter in the pressure-wind relationship. This is typically achieved by using pressure as the dependent variable in regression models or by allowing key parameters to vary, thereby capturing a realistic range of maximum wind speeds for a given central pressure [50]. However, this introduces a fundamental trade-off: improving model accuracy through local calibration often limits its applicability to other ocean basins. A prominent example of this limitation is the Dvorak pressure–wind technique, which, although widely used, tends to overestimate maximum winds, particularly for major hurricanes. Furthermore, its application in the North Atlantic has been shown to underestimate the modulating effect of storm translation speed on intensity. This highlights the need for new approaches that can better balance accuracy and generalisability.

To address this challenge, this paper investigates an alternative modelling framework based on PGNN. In this approach, the outputs of a generic untuned physics-based model are combined with observational data to estimate typhoon intensity (e.g., TI values). The PGNN framework leverages the physical structure of traditional models while using data-driven learning to improve predictive performance. This integration is physically well-founded as WPRs reflect fundamental thermodynamic constraints in typhoon dynamics. If neural networks are trained using purely data-driven approaches without embedding these constraints, they risk producing physically implausible results (e.g., unrealistic pairings of low pressure with low wind speed or high pressure with high wind speed). The PGNN framework mitigates this risk by incorporating physical principles directly into the learning process.

3.3 | 2D-TSDK Temporal Encoding With Physics-Informed Hybrid Modelling

The 2D typhoon descriptors consist of four key parameters: the typhoon centre latitude (LAT, unit:·), longitude (LONG, unit:·), centre pressure (PRES, unit: hPa) and near-centre maximum wind speed (WIND, unit: m/s). These variables form the basis for deriving the 2D-TSDK features using the CLIPER method [51, 52]. Temporal variations are incorporated by appending a subscript i (e.g., LAT_6) to denote the variable's value i hours prior. The features are organised by time step, where the differences between consecutive steps assigned to the current time step as additional features.

For each time step, the 2D-TSDK features are categorised into four groups: (1) typhoon persistence factors (Features 1–4), representing the basic state variables at time t; (2) climatology factors (Features 5–8), which characterise the typhoon's variability; (3) domain-knowledge-guided SHIP features (Features 9–10) that provide helpful predictive information and (4) detailed wind-speed and pressure dynamics (Features 11–20). Additionally, Feature 21 is designed to be a seasonal factor. A comprehensive summary of these features is presented in Table 1. The proposed framework processes 2D typhoon feature sequences through a neural architecture that explicitly couples physics-based constraints with data-driven representations.

3.3.1 | *Position-Aware Temporal Encoding*

To model temporal evolution, we employ a self-attention based encoding scheme that preserves absolute positional information whilst adaptively learning relative temporal dependencies (as proposed in ref. [53]).

First, the input $X^{(2D)} \in \mathbb{R}^{T\times d_0}$ is first projected into a latent embedding space $\Phi^{(2D)} \in \mathbb{R}^{T\times d_1}$ via an initial linear transformation layer, d_0 the input feature dimension and d_1 the embedding dimension. This forms the foundational representation for subsequent temporal modelling.

To ensure temporal coherence, absolute positional encodings are added to the feature embeddings for each time step. This additive operation ensures the model retains sequence order information, with the encoding defined as follows:

$$\phi_t^{(2D)} = \phi_t^{(2D)} + \rho_t, \tag{2}$$

where the positional embedding of the tth time step is computed using sine and cosine functions as given by the following equation:

$$\rho_t(2k) = \sin t\omega_k, \rho_t(2k+1) = \cos t\omega_k, \tag{3}$$

where

$$\omega_k = \frac{10000^{-2k/d_1} \times d_1}{L}, \tag{4}$$

where k is an integer within the interval $\left[0, \frac{d_1}{2}\right]$.

Building on this position-aware representation, we employ a self-attention mechanism with relative positional biases to model temporal dynamics. This mechanism dynamically captures temporal dependencies across all time steps by computing pairwise attention scores, enabling context aware feature refinement. By adaptively adjusting each steps features based on interstep relationships, it prioritise meteorologically significant patterns (e.g., rapid pressure drops) whilst suppressing noise and irrelevant fluctuations.

TABLE 1 | Each time step's 2D typhoon structure domain-expert knowledge features table.

ID	Features
1–4	LAT_t, LON_t, WIND_t, PRES_t
5–6	$\mathrm{LAT}_t - \mathrm{LAT}_{t+6}$, $\mathrm{LON}_t - \mathrm{LON}_{t+6}$
7–8	$\mathrm{WIND}_t - \mathrm{WIND}_{t+6}$, $\mathrm{PRES}_t - \mathrm{PRES}_{t+6}$
9–10	$\cos(\mathrm{LAT}_t)$, $\cos(\mathrm{LAT}_t) - \cos(\mathrm{LAT}_{t+6})$
11–12	$\log(\mathrm{WIND}_t)$, $\log(\mathrm{PRES}_t)$
13–14	$(\mathrm{WIND}_t)^2$, $(\mathrm{WIND}_t)^3$
15–16	$\mathrm{PRES}_t/\mathrm{WIND}_t$, $\mathrm{WIND}_t/\mathrm{PRES}_t$
17–18	$\mathrm{PRES}_t/(\mathrm{WIND}_t)^2$, $(\mathrm{WIND}_t)^2/\mathrm{PRES}_t$
19–20	$\mathrm{PRES}_t/(\mathrm{WIND}_t)^3$, $(\mathrm{WIND}_t)^3/\mathrm{PRES}_t$
21	Month

The self-attention weight between positions j and q, incorporating relative positional embedding (RPE) is computed as follows:

$$\beta_{jq} = \frac{\exp(e_{j,q})}{\sum_{q=1}^{T} \exp(e_{j,q})} + \theta_{j-q}, \tag{5}$$

where θ_{j-q} is a learnable scalar representing the relative position weight between position j and q, $e_{j,q}$ is an attention weight from positions q to j, calculated as follows:

$$e_{jq} = \frac{\left((\phi^{(2D)} W^{Q})(\phi^{(2D)} W^{K})^{T}\right)_{jq}}{\sqrt{d_2}}, \tag{6}$$

where W^Q and W^K are the query and key projection matrices that map the input to their respective representation spaces.

To capture different aspects of temporal dynamics, the multi-head self-attention (MSA) mechanism is employed, enabling more robust feature extraction. H individual features are concatenated to obtain the following output feature representation:

$$\hat{X}_t^{(2D)} = \left(\|_{h=1}^{H} B_h\left(\Phi^{(2D)} W_h^{V}\right)\right) W^{O}[t, :], \tag{7}$$

where $\|$ is the concatenation operation, W^V is the value projection matrix that transforms the input to the value representation space and W^O is a trainable parameter matrix used to project the concatenated vector back to the d_1 dimension.

With the temporal relationships encoded, the features are aggregated via average pooling, followed by dimensionality reduction to distill the most salient patterns. The resulting condensed representation is then fed into a regression head for typhoon intensity estimation. The process can be summarised as follows:

$$\hat{Y}_{T+\Gamma}^{(2D)} = \mathrm{FFN}\left(\mathrm{AvgPool1D}\left(\hat{X}_1^{(2D)}, \hat{X}_2^{(2D)}, \ldots, \hat{X}_T^{(2D)}\right)\right). \tag{8}$$

3.3.2 | *Physics-Based Learning Objective*

To enhance the physical consistency of typhoon intensity predictions, we incorporate a well-established empirical relationship between MSLP and MSW as a physics-based constraint during model training. This ensures our physics-guided neural network (PGNN) not only minimises prediction errors but also adheres to fundamental atmospheric dynamics governing tropical cyclones.

Most pressure–wind models follow the general form [54]:

$$v = a\left(p_e - p_c\right)^{m}, \tag{9}$$

where v is the maximum wind, p_e is a defined external pressure, p_c is the minimum central pressure and a and m are empirically determined constants. This equation describes a nonlinear coupling between MSLP and MSW, where maximum wind speed decreases monotonically with increasing central pressure (as illustrated in Figure 2).

To enforce physical consistency, we require that for any two time points with pressure values p_1 and p_2, the corresponding wind speeds v_1 and v_2 satisfy:

$$p_1 < p_2 \Rightarrow v_1 > v_2. \tag{10}$$

This condition serves as a quantifiable measure of physical consistency, ensuring the model's predictions align with known atmospheric behaviour. Without such constraints, purely data-driven models may learn spurious correlations that minimise training error but violate physical principles, leading to unreliable predictions in real-world scenarios.

To reduce computational costs, our approach focuses on calculating pairwise wind speed differences between current and predicted time points. Specifically, we sort pressure values in ascending order and compute the corresponding wind speed differences:

$$\Lambda_i = \begin{cases} v_{i,\,T} - v_{i,\,T+\Gamma}, & \text{if } p_{i,\,T+\Gamma} < p_{i,\,T} \\ v_{i,\,T+\Gamma} - v_{i,\,T}, & \text{if } p_{i,\,T} < p_{i,\,T+\Gamma} \end{cases}, \tag{11}$$

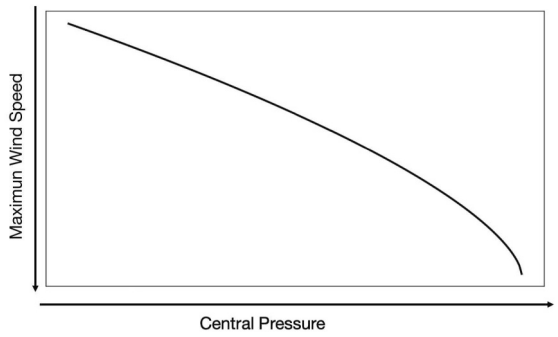


FIGURE 2 | Illustration of the pressure-wind relationship.

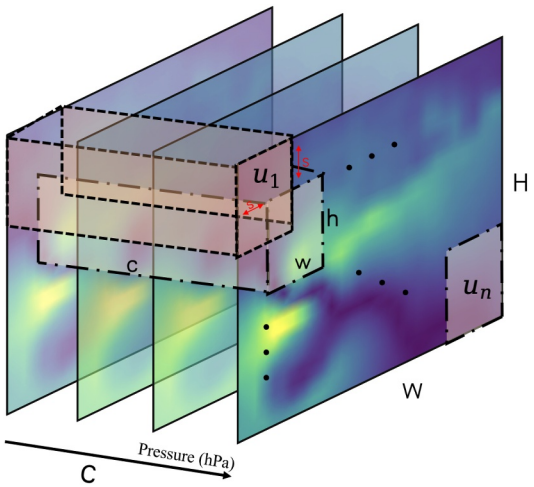


FIGURE 3 | Illustration of partitioning the 3D input into n overlapping patches.

where $v_{i,\ T+\Gamma}$ denotes the predicted wind speed $\hat{y}^{(2D)}_{i,T+\Gamma}$ for the ith instance at the time step $T + \Gamma$. A positive value of Λ_i indicates a violation of physical consistency. A physics-based loss term penalises such inconsistencies, guiding the network towards physically plausible predictions. The mean of these violations across all relative wind speed pairs defines the physics-based loss functions:

$$\mathcal{L}_{\text{phy}} = \frac{1}{N}\sum_{i=1}^{N} \text{RELU}(\Lambda_i), \tag{12}$$

where N is the number of training instances.

The complete learning objective of PGNN combines this physics-based loss with the standard prediction error:

$$\mathcal{L}_1 = \|\widehat{Y}^{(2D)}_{T+\Gamma} - Y_{T+\Gamma}\|_1 + \xi_1 L_{\text{phy}} + \xi_2 \|\mathbf{W}^{(2D)}\|_2, \tag{13}$$

where $\mathbf{W}^{(2D)}$ denotes the set of weights for the 2D-TSDK model, ξ_1 and ξ_2 are the hyperparameter of the model corresponding to physics-based loss function and regularisation term.

As the integrated loss function is differentiable almost everywhere, we compute gradients at both the output and hidden layers via backpropagation. Unlike conventional learning objectives, physics-based constraints improve generalisation, particularly for limited or noisy training data. Owning to the universally valid nature of physical principles, the proposed approach ensures robustness even beyond the training distribution. The resulting predicted residuals are then propagated to subsequent stages for iterative refinement.

3.4 | 3D-TSDK Spatial-Temporal Encoding for Residual Fitting

The residual reduction strategy is twofold: (1) employing a 3D transformer to capture atmospheric dynamics and (2) implementing position-aware temporal encoding scheme (as detailed in Section 3.3) to mitigate forecast errors.

This framework builds upon the 3D typhoon representation developed by ref. [13], which processes assimilation data on a latitude-longitude grid system. The data representation $X^{(3D)}$ extracts U and V wind components from a 31° × 31° domain (at 1° × 1° resolution) centred on the typhoon eye across four isobaric levels (1000 hPa, 750 hPa, 500 hPa and 250 hPa). The temporal dimension includes four time steps: the current observation plus three prior states at 6-h intervals ($T - 6h$, $T - 12h$ and $T - 18h$), creating a comprehensive spatialtemporal representation.

We process the U- and V-wind components independently across four pressure levels at each time step t using a vision transformer variant. For each 3D input, we partition the data into n overlapping patches of size $h \times w \times c$ with stride s (where $s < \min(h, w)$), where the stride controls patch overlap whereas zero-padding maintains spatial integrity. The details of the partition process are shown in Figure 3. This design ensures the learnt patch embeddings encode physically meaningful spatial relationships, effectively capturing the strong correlations inherent in neighbouring atmospheric conditions.

The spatial encoding pipeline is illustrated below using the U-wind component (with identical processing applied to the V-wind component). The 2D convolutional patch embedding layer first projects the input patches into a d_3-dimensional latent space, followed by flattening into a 1D token sequence. The resulting tokenised representation is as shown below:

$$\psi^{(3D)}_{t,U} = \left[u_{t,cls}, u_{t,1}, u_{t,2}, ..., u_{t,n}\right] + \sigma_{t,U}, \tag{14}$$

where the token sequence consists of $u_{t,cls}, u_{t,1}, ..., u_{t,n} \in \mathbb{R}^{d_3}$, with the *cls* token prepended as a classification token. Additionally, $\sigma_{t,U}$ with the embedding dimension $n \times d_3$ denotes a learnt spatial positional embedding.

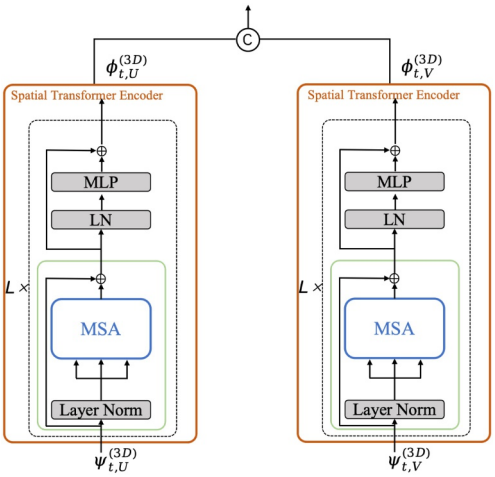


FIGURE 4 | Overall structure of the spatial transformer encoder.

TABLE 2 | The hyperparameter symbols, descriptions and values used in this study.

Symbol	Description	Value
ξ_1	Weight for physics-based loss	8×10^{-5}
image_size	Image size (px)	31
patch_size	Patch size (px)	4
Stride	Patch stride	4
in_channels	Input channels	4
out_dim	Output dimension	128
embed_dim	Embedding dimension	512
num_layers	Transformer layers	2
num_heads	Attention heads	2
hidden_dim	FFN hidden dimension	1024
atten_drop	Attention dropout rate	0.2
proj_drop	Projection dropout rate	0.2

TABLE 3 | Training details and inference performance.

Model size	Training time	Inference latency (avg)
5.11 M	1.62 h	0.23 s

The sequence is then encoded through L consecutive transformer layers, where each layer consists of three core components: multihead self-attention (MSA), layer normalisation (LN) and an MLP block, with the detailed computation specified in Figure 4.

We extract the cls token from the final layer's output for downstream processing. Following the same pipeline applied to the U-wind component, we derive the corresponding $\psi_{t,V}^{(3D)}$ for the V-wind component. These representations are then integrated through convolutional fusion and augmented with positional encoding ρ_t to generate the tth time step output $\phi_t^{(3D)}$, which can be expressed as follows:

$$\phi_t^{(3D)} = \mathrm{Conv}\left(\mathrm{Concat}\left(\psi_{t,U}^{(3D)}, \psi_{t,V}^{(3D)}\right)\right) + \rho_t. \tag{15}$$

The temporal encoding pipeline processes $\Phi_t^{(3D)} \in \mathbb{R}^{T \times d_3}$ through the position-aware temporal attention mechanism (detailed in Section 3.3.1) to generate $\widehat{X}_t^{(3D)}$.

With temporal dependencies encoded across T time steps, the model applies average pooling to aggregate sequential features, then perform dimensionality reduction to distill the most discriminative temporal patterns. The resulting condensed representation is then fed into a regression head for typhoon intensity estimation. The process can be summarised as follows:

$$\widehat{Y}_{T+\Gamma}^{(3D)} = \mathrm{FFN}\left(\mathrm{AvgPool1D}\left(\widehat{X}_1^{(3D)}, \widehat{X}_2^{(3D)}, ..., \widehat{X}_T^{(3D)}\right)\right). \tag{16}$$

We also adopt the mean of absolute errors (MAE) on the 3D-TSDK dataset as the loss function to evaluate and optimise the networks training performance.

$$\mathcal{L}_2 = \|\widehat{Y}_{T+\Gamma}^{(3D)} - \Delta_{T+\Gamma}\|_1 + \xi_3\|\mathbf{W}^{(3D)}\|_2, \tag{17}$$

where $\mathbf{W}^{(3D)}$ denotes the set of weights for the 3D-TSDK model, $\Delta_{T+\Gamma}$ is the residual propagated from the previous stage $\Delta_{T+\Gamma} = Y_{T+\Gamma} - \widehat{Y}_{T+\Gamma}^{(2D)}$ and ξ_3 is the hyperparameter of the model corresponding to the regularisation term.

4 | Experiments

In this section, we present the experimental setup and results of our proposed model CPC for typhoon intensity prediction. We first introduce the datasets used in our experiments, namely the CMA–BST dataset and the ERA-interim dataset. Then, we describe the experimental results, including the evaluation metrics and performance comparisons.

4.1 | Datasets

Two available datasets are used in our experiment, including the CMA–BST dataset [55] and the ERA-interim dataset [56].

The CMA–BST (China Meteorological Administration Best Track) dataset is provided by the China Meteorological Administration, which contains historical best-track records of tropical cyclones in the Northwestern Pacific and South China Sea. The key variables include: Tropical cyclone identifier, centre position (latitude and longitude in degrees), maximum sustained wind speed (unit: m/s) and minimum central pressure (unit: hPa). The CMA–BST dataset is used to generate the 2D-TSDK features noted in Section 3.3. The temporal resolution of this dataset is 6 h. Under this temporal resolution, tropical cyclones with lifespans shorter than 3 days (fewer than 12 observations) are discarded during pre-processing. This criterion is applied because our model requires 4 historical time steps to predict the typhoon intensity 24 h ahead. Cyclones with lifespans below this threshold lack sufficient observations to construct the required input sequence for prediction.

TABLE 4 | The MAE of the TI prediction at the lead times of 24 h (unit: m/s).

Method	2015	2016	2017	2018	Avg
CMA	4.26	5.10	5.63	4.00	4.75
JMA	5.08	5.30	5.05	3.90	4.83
JTWC	4.88	5.40	5.23	4.70	5.05
ECMWF-IFS	5.44	9.22	6.40	6.90	6.99
NCEP-GFS	6.42	6.42	5.92	4.80	5.89
CLIPER	5.20	6.17	4.85	5.45	5.42
CLIPER + EnvAvg	5.16	6.12	4.82	5.32	5.36
XGBoost	4.78	5.43	4.11	4.28	4.65
XGBoost + EnvAvg	4.69	5.39	4.13	4.12	4.58
CLIPER + ConvLSTM	4.77	5.34	4.03	4.25	4.60
CNN-based method	6.79	7.30	6.14	6.95	6.80
SAF-Net	4.54	4.78	3.95	3.94	4.30
MAF	4.34	**4.91**	3.82	**3.71**	4.19
CPC	**4.10**	4.92	**3.75**	3.72	**4.12**

The ERA-interim dataset is produced by ECMWF, which is a global atmospheric reanalysis dataset providing historical meteorological fields. It covers the period from January 1979 to December 2019. The ERA-interim dataset is employed to model the 3D spatial structure of tropical cyclones. We selected the U- and V-wind components at different isobaric planes from the dataset. For each isobaric plane, a 31° × 31° grid centred on the tropical cyclone was extracted, with a spatial resolution of 1° × 1°. The temporal resolution was synchronised with the CMA-BST dataset at 6-hourly intervals.

In order to maintain consistency with previous research methods [13, 16], we selected data from 2000 to 2014 for training and data from 2015 to 2019 for testing. The training and validation cohort spans 2000–2014, comprising 365 typhoons with 11,315 temporal records, where 10% (1131 records) are randomly allocated for validation to monitor potential overfitting. The independent test set contains 111 typhoons from 2015 to 2018, representing 2747 records for final model evaluation.

4.2 | Evaluation Criteria and Implementation Details

The mean absolute error (MAE) is employed as the primary metric to quantify prediction accuracy across all test instances. Calculated as follows:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|\widehat{Y}_{i,T+\Gamma} - Y_{i,T+\Gamma}|, \tag{18}$$

where N denotes the total yearly records, $\widehat{Y}_{i,T+\Gamma} = \widehat{Y}^{(2D)}_{i,T+\Gamma} + \widehat{Y}^{(3D)}_{i,T+\Gamma}$ represents the model's predicted intensity for the ith record and Y_i is the observed intensity value. Additionally, a consolidated evaluation metric the 4-year averaged MAE is derived by averaging annual MAE results across the entire test period (2015–2018).

4.3 | Experimental Setup

We train our model using Adam optimiser with an initial learning rate of 7e-3 and perform grid search over the hyper-parameters. Detailed configurations, including parameter symbols, descriptions and values, are given in Table 2. Our experiments were conducted with PyTorch 1.7.0 on a computer equipped with a 2.9 GHz Intel Xeon processor, 440 GB of RAM, 24 MB of L3 cache and 7 NVIDIA A40 (48 GB each) GPUs. Table 3 summarises the training time and inference latency for the model.

4.4 | Comparison Methods

To evaluate the effectiveness of our proposed method, we conduct comparative experiments against both up-to-date TI verification reports and machine learning/deep learning-based typhoon intensity prediction methods.

The methods in the up-to-date TI verification reports are from three authoritative meteorological agencies the China Meteorological Administration (CMA), Japan Meteorological Agency (JMA), Joint Typhoon Warning Center (JTWC) and two global numerical models the ECMWF-IFS integrated forecasting system and NCEP-GFS (Global Forecast System).

The methods in the second type are consist of CLIPER [57], XGBoost [58] methods and their different settings CLIPER + EnvAvg, XGBoost + EnvAvg. CLIPER is the classic statistical forecasting method, and CLIPER + EnvAvg enhances the baseline CLIPER method by incorporating spatially averaged key environmental variables. The XGBoost method leverages decision tree ensembles to perform nonlinear fitting of the input CLIPER features, whereas the XGBoost + EnvAvg enhances this framework by incorporating spatially averaged environmental field features to address the limitation of inadequate environmental modelling inherent in standard XGBoost method. The CNN-based method [59] primarily leverages convolutional neural networks (CNNs) to extract spatial temporal features from three-dimensional reanalysis data for tropical cyclone intensity prediction.

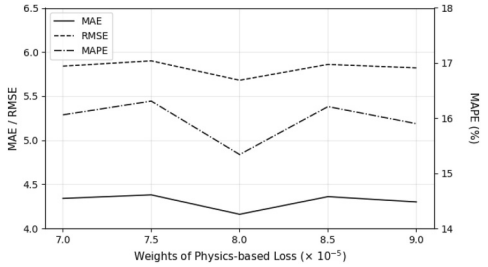


FIGURE 5 | Sensitivity analysis of CPC to different weights of physics-based loss.

Additionally, we adopted CLIPER + ConvLSTM, SAF-Net [13] and MAF [16] as benchmark models for comparative evaluation, all of which use the same dataset. The CLIPER+ConvLSTM method subsequently employs convolutional LSTM (ConvLSTM) networks to encode spatiotemporal dynamics after generating CLIPER features. SAF-Net uses a three-branch CNN for 3D-TSDK analysis and integrating it with 2D-TSDK using a wide and deep framework for typhoon intensity prediction. MAF uses block attention to extract observational data features and use CNNS to extract reanalysis data features, then uses adversarial fusion function to fuse these two types of data.

4.5 | Performance Comparison

A systematic experimental analysis demonstrates that the proposed model CPC exhibits comprehensive competitive advantages across multiyear evaluations. The MAE results for different methods are presented in Table 4, with the minimum prediction error for each year and the overall average indicated in bold.

As presented in Table 4, on the one hand, CPC maintains consistent superiority over operational systems from three major meteorological agencies during 2015–2018. Compared to the CMA, CPC reduce the average forecasting error from 4.75 to 4.12. When evaluated against the JMA, CPC exhibiting improvements of 14.7%. Compared to the 4-year average MAE of 5.05 from the JTWC, CPC demonstrates a further 18.4% accuracy enhancement with an overall error of 4.12. In numerical forecasting comparisons, the proposed model CPC achieves a 41.1% error reduction compared to the average MAE of 6.99 from the ECMWF-IFS system, with the most significant advantage in 2016 about 46.6% reduction. Compared to the NCEP-GFS system, CPC achieves a 30.1% improvement from 5.89 to 4.12 in 4-year average MAE.

On the other hand, comparative evaluation against machine learning and deep learning methods reveals that the proposed model consistently achieves lower MAE values both in the 4-

TABLE 5 | Ablation study of different strategies in terms of CPC for TI forecasting at the lead times of 24 h (unit: m/s).

Method	2015	2016	2017	2018	Avg
2D-TSDK	4.63	5.31	3.85	4.30	4.52
2D-TSDK + PATE	4.51	4.95	3.62	4.19	4.32
2D-TSDK + PATE + PILF	4.42	4.93	**3.50**	4.12	4.24
2D-TSDK + PATE + PILF + 3D-TSDK	**4.10**	**4.92**	3.75	**3.72**	**4.12**

TABLE 6 | The MAE of the TI prediction with one isobaric plane at the lead times of 24 h (unit: m/s).

Method	2015	2016	2017	2018	Avg
U	4.45	4.94	3.56	4.05	4.25
V	4.44	4.97	3.55	4.08	4.26
UV + add	4.44	4.89	3.58	3.99	4.23
UV + Conv	4.41	4.95	**3.51**	3.99	4.22
UV + MAX	**4.13**	5.12	3.83	3.91	4.24
UV + add + PATE	4.42	4.86	3.63	**3.90**	4.20
UV + Conv + PATE	4.41	**4.84**	3.54	3.98	**4.19**
UV + MAX + PATE	4.35	4.77	3.81	3.92	4.21

TABLE 7 | The MAE of the TI prediction with four isobaric planes at the lead times of 24 h (unit: m/s).

Method	2015	2016	2017	2018	Avg
U	4.47	4.96	**3.54**	4.04	4.25
V	4.45	4.99	3.57	4.07	4.27
UV + add	4.54	4.85	3.63	3.88	4.22
UV + Conv	4.38	4.91	3.45	3.92	4.17
UV + MAX	4.43	4.95	3.56	4.06	4.25
UV + add + PATE	4.44	**4.80**	3.65	3.84	4.18
UV + Conv + PATE	**4.10**	4.92	3.75	**3.72**	**4.12**
UV + MAX + PATE	4.25	4.89	3.62	4.05	4.19

TABLE 8 | The MAE of the TI prediction with eight isobaric planes at the lead times of 24 h (unit: m/s).

Method	2015	2016	2017	2018	Avg
U	4.50	**4.78**	3.68	3.91	4.22
V	4.45	4.99	3.53	4.05	4.26
UV + add	4.48	5.03	3.52	4.03	4.27
UV + Conv	4.39	5.01	3.57	3.88	4.21
UV + MAX	4.34	4.98	**3.50**	3.99	4.20
UV + add + PATE	4.48	4.85	3.59	3.83	4.19
UV + Conv + PATE	4.32	4.97	3.61	**3.72**	**4.16**
UV + MAX + PATE	**4.15**	4.96	3.83	3.75	4.17

year average and for each individual year, demonstrating superior predictive accuracy across all years. The proposed model reduces errors by 23.1% compared to the 4-year average MAE of 5.36 from the CLIPER + EnvAvg hybrid model, with a 35.9% improvement in 2016. Compared to the XGBoost + EnvAvg model, CPC achieves a 10.0% reduction in 4-year average error.

Within deep learning benchmarks, CPC demonstrates a 1.7% accuracy improvement over the MAF model's average error of 4.19. Compared to the SAF-Net model, our approach reduces the 2017 errors by 4.2%. These result highlights that these machine learning and deep learning methods have not effectively captured the temporal dependencies among meteorological elements in 2D-TSDK and the spatial characteristics in 3D-TSDK. In contrast, our proposed method CPC makes predictions that align with real-world physical laws by comprehensively considering both the temporal dependencies among meteorological elements and the spatial relationships among environmental variables.

4.6 | Sensitivity Analysis

We investigate the performance fluctuations of the CPC model caused by varying the weight assigned to the physics-based loss. The baseline range of this parameter is set to $[7 \times 10^{-5}, 9 \times 10^{-5}]$. As illustrated in Figure 5 (which reports RMSE and MAPE alongside MAE), our findings indicate that the resulting disturbances are relatively minor, although different parameter settings do introduce perturbations to the model.

4.7 | Ablation Study

To validate the effectiveness of our proposed method, we conducted comprehensive ablation studies encompassing the two-stage training strategy, the physics-based learning objective, the position-aware temporal encoding, the fusion scheme and the wind components.

The results for the first three aspects (the two-stage training strategy, the physics-based learning objective, and the position-aware temporal encoding) are shown in Table 5, with the minimum prediction error for each year and the overall average indicated in bold. Starting from a baseline 4-year average MAE of 4.52 using only 2D-TSDK, the incorporation of the position-

TABLE 9 | The MAE of the TI prediction with 16 isobaric planes at the lead times of 24 h (unit: m/s).

Method	2015	2016	2017	2018	Avg
U	4.51	**4.77**	3.72	4.07	4.27
V	4.37	5.07	3.53	4.07	4.26
UV + add	4.47	4.99	3.59	4.02	4.27
UV + Conv	4.47	5.04	**3.51**	4.08	4.28
UV + MAX	**4.28**	5.15	3.80	**3.65**	4.22
UV + add + PATE	4.46	4.85	3.64	3.90	**4.21**
UV + Conv + PATE	4.51	4.93	3.54	4.06	4.27
UV + MAX + PATE	4.44	4.97	3.55	4.03	4.25

aware temporal encoding (PATE) reduced the average MAE to 4.32. Subsequent integration of the physics-informed loss function (PILF) enhanced the physical consistency between the model's predictive results and real-world patterns, further compressing the average MAE to 4.24. Notably, in 2017, this configuration achieved a performance of 3.50, representing a 9.1% improvement over the 2D-TSDK-only baseline. The full model (2D-TSDK + PATE + PILF + 3D-TSDK) ultimately achieved a breakthrough average MAE of 4.12 through the strategic inclusion of ERA-Interim reanalysis data, demonstrating an 8.8% optimisation over the 2D-TSDK-only baseline. A dramatic error reduction of 13.5% was observed in 2018, underscoring the critical role of different modules.

For the remaining aspects (the fusion scheme and the wind components), Tables 6–9 present the 3D-TSDK ablation results, with the minimum prediction error for each year and the overall average indicated in bold. Specifically, for the three-dimensional spatial structure of the typhoon, we select isobaric planes in four distinct ways: one isobaric plane (500 *hPa*), four isobaric planes (250–1000 *hPa*, interval 250 *hPa*), eight isobaric planes (300–1000 *hPa*, interval 100 *hPa*) and 16 isobaric planes (250–1000 *hPa*, interval 50 *hPa*). With respect to wind components, we conduct experiments using the *U*-component alone, the *V*-component alone and both *U*- and *V*-components simultaneously. To evaluate the fusion schemes, we employ max fusion (Max), additive fusion (Add) and convolutional fusion (Conv).

TABLE 10 | 3D-TSDK encoding analysis of the TI forecasting at the lead times of 24 h (unit: m/s).

Method	2015	2016	2017	2018	Avg
Layers = 1	4.52	**4.66**	3.65	4.02	4.19
Layers = 2	**4.10**	4.92	3.75	**3.72**	**4.12**
Layers = 3	4.34	4.99	**3.55**	3.77	4.16
Layers = 4	4.31	4.89	3.87	3.81	4.22

The minimal prediction error is achieved when four isobaric planes are used. Employing a single plane yields worse results than using four. Furthermore, when the number of isobaric planes are increased to eight or 16, the error continues to rise. These observations indicate that a single isobaric plane is insufficient to model the spatial structure of a typhoon effectively, whereas an excessive number of planes introduce overly complex structures that the model cannot capture, thereby degrading predictive performance. Four distinct isobaric planes appear sufficient for effective modelling of the typhoon's three-dimensional structure.

For wind variables, simultaneous use of both *U*- and *V*-components achieves a lower prediction error. Because wind is a vector quantity, combining *U*- and *V*-components enriches the extracted features compared with using either component alone, enabling the network to better represent the typhoon's three-dimensional spatial structure.

Among the fusion schemes, convolution-based fusion outperforms additive fusion and max fusion. This advantage arises because convolution is a learnable process whose kernel parameters are updated during training, allowing smoother integration of the different wind components.

4.8 | Discussion

To further investigate the influence of the number of spatial attention layers within the ViT used in the 3D-TSDK encoder on the final prediction error, we conducted detailed experiments by

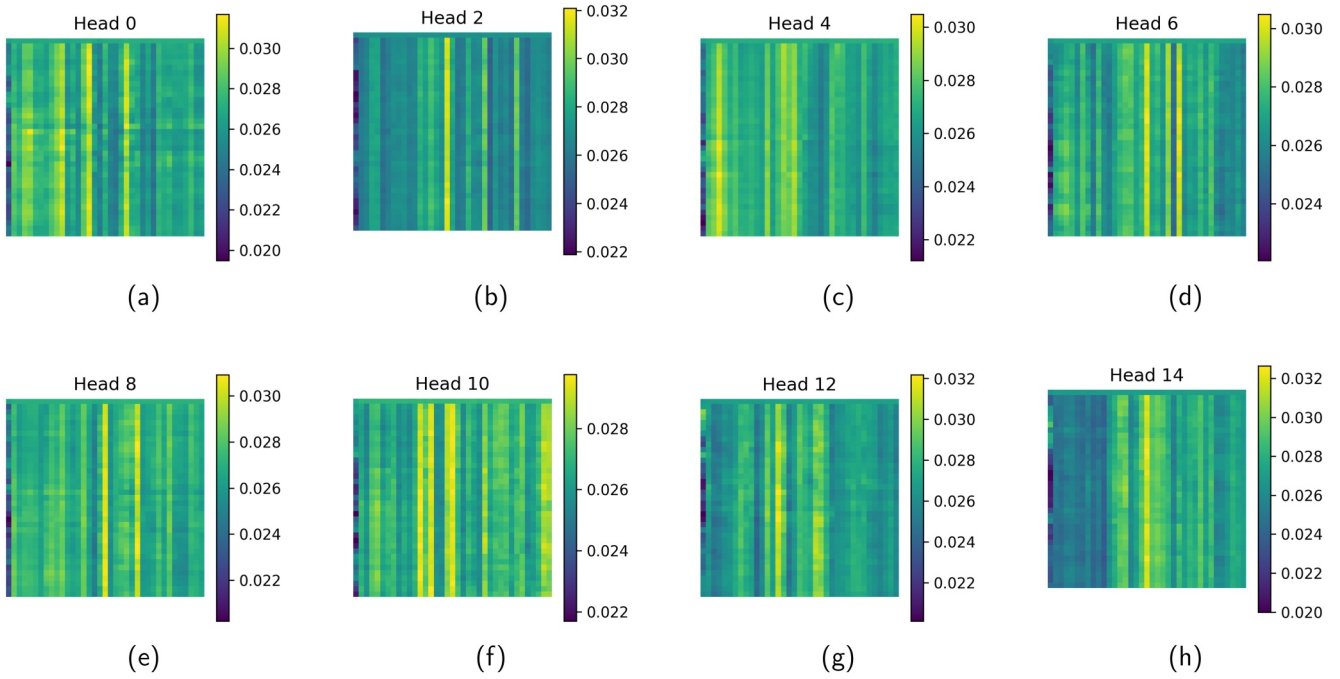
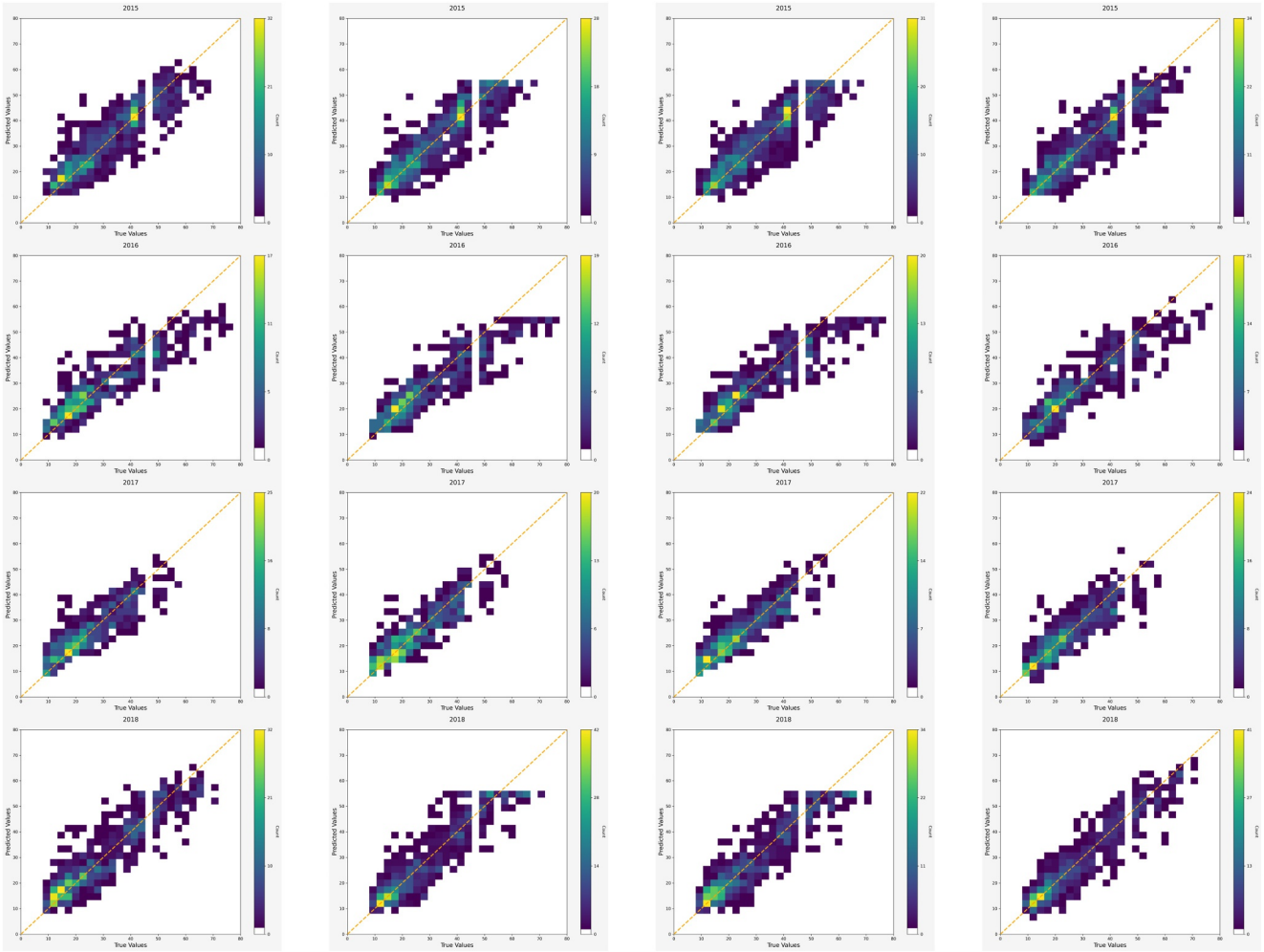


FIGURE 6 | Illustration of attention maps from Stage 2 of the vision transformer. Panels (a–d) show a subset of heads for the U-component, and panels (e–h) show a subset of heads for the V-component.



(a) 2D-TSDK (b) 2D-TSDK+PATE (c) 2D-TSDK+PATE +PILF (d) 2D-TSDK+PATE +PILF+3D-TSDK

FIGURE 7 | 2-D histogram of observed versus predicted values of maximum sustained wind near the typhoon centre. Colours indicate the number of corresponding data points. Our prediction distribution is closer to the midline and has better prediction results.

varying this number. The results are reported in Table 10, with the minimum prediction error for each year and the overall average indicated in bold. They indicate that the optimal prediction accuracy is obtained when the number of attention layers is set to 2. Fewer computations prevent the network from capturing the dependencies among different spatial locations in the typhoon's three-dimensional space, whereas an excessive number tends to cause overfitting.

As shown in Figure 6, the attention maps reveal that the second-stage ViT focuses on local regions rather than relying solely on global shape cues. This visualisation demonstrates that the model is capable of discriminating subtle spatial differences, which is crucial for fine-grained recognition tasks.

2-D histograms demonstrate the predictive performance of the maximum sustained wind speed near the typhoon centre from 2015 to 2018 by comparing the density distribution of predicted values with observed values. The colour bar represents the density of data points, and the diagonal represents the ideal prediction line. Figure 7 show a progressive improvement, with high-density regions gradually converging towards the diagonal, indicating a significant increase in prediction accuracy with the increase of model complexity.

Although this study offers valuable insights, several limitations should be acknowledged, which in turn open avenues for future research.

1. Data Source and Baseline Limitations: This study utilises the ERA-interim dataset for 3D atmospheric modelling and primarily compares its performance against global models (ECMWF-IFS and NCEP-GFS). Although this ensures a consistent baseline for evaluating our core contributions, it presents two limitations. First, ERA5, which benefits from a decade of advancements in model physics and data assimilation, could potentially provide more accurate inputs for TC intensity prediction. Second, high-resolution regional models, such as HWRF and COAMPS-TC, represent the operational state-of-the-art and generally outperform global models in intensity forecasting. Therefore, retraining and validating our model using ERA5, and establishing a systematic benchmark against these advanced regional models, constitutes an important direction for future research.

2. Handing of Extreme Events: This study was not specifically designed to address rapid intensification (RI) events, which remain one of the foremost challenges in modern tropical cyclone forecasting. We acknowledge that evaluating model performance solely based on overall MAE may obscure deficiencies in predicting these extreme yet high-impact cases. A dedicated analysis of RI prediction, incorporating techniques, such as specialised loss functions, resampling strategies or event-specific post-processing, will be a critical focus of our future work to enhance model robustness under extreme conditions.
3. Static Nature of Physical Constraints: The physical constraint employed in this study, based on an empirical wind–pressure relationship, remains relatively static in its current form. To achieve deeper physical coupling with the learning process, future work should consider incorporating dynamic constraints, such as vertical wind shear and sea surface temperature, directly into the model. These physical principles can be translated into learnable constraints and seamlessly integrated into the spatiotemporal modelling framework to improve both predictive accuracy and physical consistency. Furthermore, exploring additional physical constraints and extending validation across diverse meteorological scenarios would represent a significant step towards bridging physics-based modelling and AI for scientific discovery.

5 | Conclusion

This paper presents a cascade physics-coupled network for TI prediction that combines physics-guided constraints with complementary 2D and 3D feature learning. The proposed two-stage architecture first generates initial predictions using a 2D-TSDK model with a physics-informed loss function, then refines errors via a 3D-TSDK corrector based on a vision transformer. Temporal dynamics are captured by a dual positional encoding mechanism across both stages. Experiments show consistent improvements over baselines, demonstrating effective fusion of meteorological principles with data-driven learning.

Acknowledgements

This research work was supported by the Joint Funds of the National Natural Science Foundation of China (Grants U22A2068 and U23A20320), Key Research and Development Project of Shandong Province, China (Grant 2023ZLGX06) and Taishan Industrial Leading Talent Project Blue Talent Project.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Data will be made available on request.

References

1. D. Swain, "Tropical Cyclones and Coastal Vulnerability: Assessment and Mitigation," in *Geospatial Technologies for Land and Water Resources Management* (2022), 587–621.

2. M. A. Bender, I. Ginis, and Y. Kurihara, "Numerical Simulations of Tropical cyclone-ocean Interaction With a High-Resolution Coupled Model," *Journal of Geophysical Research: Atmospheres* 98, no. D12 (1993): 23245–23263, https://doi.org/10.1029/93jd02370.

3. M. E. Kozar, M. E. Mann, S. J. Camargo, J. P. Kossin, and J. L. Evans, "Stratified Statistical Models of North Atlantic Basin-Wide and Regional Tropical Cyclone Counts," *Journal of Geophysical Research: Atmospheres* 117, no. D18 (2012), https://doi.org/10.1029/2011jd017170.

4. L. M. Ma, "Research Progress on China Typhoon Numerical Prediction Models and Associated Major Techniques," *Progress in Geophysics* 29 (2014): 1013–1022, https://doi.org/10.6038/pg20140305.

5. B. Pan, X. Xu, and Z. Shi, "Tropical Cyclone Intensity Prediction Based on Recurrent Neural Networks," *Electronics Letters* 55, no. 7 (2019): 413–415, https://doi.org/10.1049/el.2018.8178.

6. S. Yuan, C. Wang, B. Mu, F. Zhou, and W. Duan, "Typhoon Intensity Forecasting Based on Lstm Using the Rolling Forecast Method," *Algorithms* 14, no. 3 (2021): 83, https://doi.org/10.3390/a14030083.

7. S. Jiang and L. Tao, "Classification and Estimation of Typhoon Intensity From Geostationary Meteorological Satellite Images Based on Deep Learning," *Atmosphere* 13, no. 7 (2022): 1113, https://doi.org/10.3390/atmos13071113.

8. M. Higa, S. Tanahara, Y. Adachi, et al., "Domain Knowledge Integration Into Deep Learning for Typhoon Intensity Classification," *Scientific Reports* 11, no. 1 (2021): 12972, https://doi.org/10.1038/s41598-021-92286-w.

9. P. Yin, H. Chen, H. Huang, H. Su, Q. Wu, and Q. Wan, "A Region-Based Convolutional Fusion Network for Typhoon Intensity Estimation in Satellite Images," *Engineering Applications of Artificial Intelligence* 134 (2024): 108671, https://doi.org/10.1016/j.engappai.2024.108671.

10. Z. Zhang, X. Yang, L. Shi, et al., "A Neural Network Framework for Fine-Grained Tropical Cyclone Intensity Prediction," *Knowledge-Based Systems* 241 (2022): 108195, https://doi.org/10.1016/j.knosys.2022.108195.

11. S. Jiang, H. Fan, and C. Wang, "Improvement of Typhoon Intensity Forecasting by Using a Novel Spatio-Temporal Deep Learning Model," *Remote Sensing* 14, no. 20 (2022): 5205, https://doi.org/10.3390/rs14205205.

12. R. Chen, X. Wang, W. Zhang, X. Zhu, A. Li, and C. Yang, "A Hybrid CNN-LSTM Model for Typhoon Formation Forecasting," *GeoInformatica* 23, no. 3 (2019): 375–396, https://doi.org/10.1007/s10707-019-00355-0.

13. G. Xu, K. Lin, X. Li, and Y. Ye, "Saf-Net: A Spatio-Temporal Deep Learning Method for Typhoon Intensity Prediction," *Pattern Recognition Letters* 155 (2022): 121–127, https://doi.org/10.1016/j.patrec.2021.11.012.

14. C. Wang, X. Li, and G. Zheng, "Tropical Cyclone Intensity Forecasting Using Model Knowledge Guided Deep Learning Model," *Environmental Research Letters* 19, no. 2 (2024): 024006, https://doi.org/10.1088/1748-9326/ad1bde.

15. D. Ma, L. Wang, S. Fang, and J. Lin, "Tropical Cyclone Intensity Prediction by Inter-and Intra-Pattern Fusion Based on Multi-Source Data," *Environmental Research Letters* 18, no. 1 (2023): 014020, https://doi.org/10.1088/1748-9326/aca9e2.

16. W. Li, Y. Li, Y. Yang, D. Song, Z. Wei, and A. A. Liu, "Multimodal Adversarial Fusion for Typhoon Intensity Forecasting," *IEEE Transactions on Geoscience and Remote Sensing* 62 (2024): 1–12, https://doi.org/10.1109/tgrs.2024.3367962.

17. M. DeMaria, C. R. Sampson, J. A. Knaff, and K. D. Musgrave, "Is Tropical Cyclone Intensity Guidance Improving?," *Bulletin of the American Meteorological Society* 95, no. 3 (2014): 387–398, https://doi.org/10.1175/bams-d-12-00240.1.

18. K. Emanuel and F. Zhang, "On the Predictability and Error Sources of Tropical Cyclone Intensity Forecasts," *Journal of the Atmospheric Sciences* 73, no. 9 (2016): 3739–3747, https://doi.org/10.1175/jas-d-16-0100.1.

19. P. Hewage, M. Trovati, E. Pereira, and A. Behera, "Deep Learning-Based Effective Fine-Grained Weather Forecasting Model," *Pattern Analysis & Applications* 24, no. 1 (2021): 343–366, https://doi.org/10.1007/s10044-020-00898-1.

20. F. Molteni, R. Buizza, T. N. Palmer, and T. Petroliagis, "The Ecmwf Ensemble Prediction System: Methodology and Validation," *Quarterly Journal of the Royal Meteorological Society* 122, no. 529 (1996): 73–119, https://doi.org/10.1002/qj.49712252905.

21. A. Latto, Y. J. Kim, M. B. Natoli, and E. M. Schaefer, "Preliminary Reporting on the Case Evaluation of Forecasters' Needs for the Global Forecast System (Gfs)," in *104th Annual AMS Meeting 2024* (2024), 427911.

22. J. G. Powers, J. B. Klemp, W. C. Skamarock, et al., "The Weather Research and Forecasting Model: Overview, System Efforts, and Future Directions," *Bulletin of the American Meteorological Society* 98, no. 8 (2017): 1717–1737, https://doi.org/10.1175/bams-d-15-00308.1.

23. R. S. Reddy, R. Karim, L. White, and A. Thorp, "Mesoscale Modeling Investigation Using Penn State/NCAR MM5 Model for Weather Simulation and Prediction," *Journal of the Mississippi Academy of Sciences* 55 (2010): 171–180.

24. N. M. Leonardo and B. A. Colle, "Verification of Multimodel Ensemble Forecasts of North Atlantic Tropical Cyclones," *Weather and Forecasting* 32, no. 6 (2017): 2083–2101, https://doi.org/10.1175/waf-d-17-0058.1.

25. D. J. Halperin, H. E. Fuelberg, R. E. Hart, J. H. Cossuth, P. Sura, and R. J. Pasch, "An Evaluation of Tropical Cyclone Genesis Forecasts From Global Numerical Models," *Weather and Forecasting* 28, no. 6 (2013): 1423–1445, https://doi.org/10.1175/waf-d-13-00008.1.

26. B. R. Jarvinen, and C. J. Neumann, "Statistical Forecasts of Tropical Cyclone Intensity for the North Atlantic Basin," NOAA Technical Memorandum NWS NHC-10 (National Hurricane Center, National Weather Service, 1979).

27. M. DeMaria and J. Kaplan, "A Statistical Hurricane Intensity Prediction Scheme (Ships) for the Atlantic Basin," *Weather and Forecasting* 9, no. 2 (1994): 209–220, https://doi.org/10.1175/1520-0434(1994)009<0209:aships>2.0.co;2.

28. M. DeMaria, M. Mainelli, L. K. Shay, J. A. Knaff, and J. Kaplan, "Further Improvements to the Statistical Hurricane Intensity Prediction Scheme (Ships)," *Weather and Forecasting* 20, no. 4 (2005): 531–543, https://doi.org/10.1175/waf862.1.

29. C. Y. Lee, M. K. Tippett, A. H. Sobel, and S. J. Camargo, "Autoregressive Modeling for Tropical Cyclone Intensity Climatology," *Journal of Climate* 29, no. 21 (2016): 7815–7830, https://doi.org/10.1175/jcli-d-15-0909.1.

30. R. Jing and N. Lin, "Tropical Cyclone Intensity Evolution Modeled as a Dependent Hidden Markov Process," *Journal of Climate* 32, no. 22 (2019): 7837–7855, https://doi.org/10.1175/jcli-d-19-0027.1.

31. R. Pradhan, R. S. Aygun, M. Maskey, R. Ramachandran, and D. J. Cecil, "Tropical Cyclone Intensity Estimation Using a Deep Convolutional Neural Network," *IEEE Transactions on Image Processing* 27, no. 2 (2017): 692–702, https://doi.org/10.1109/tip.2017.2766358.

32. B. F. Chen, B. Chen, H. T. Lin, and R. L. Elsberry, "Estimating Tropical Cyclone Intensity by Satellite Imagery Utilizing Convolutional Neural Networks," *Weather and Forecasting* 34, no. 2 (2019): 447–465, https://doi.org/10.1175/waf-d-18-0136.1.

33. J. Lee, J. Im, D. H. Cha, H. Park, and S. Sim, "Tropical Cyclone Intensity Estimation Using Multi-Dimensional Convolutional Neural Networks From Geostationary Satellite Data," *Remote Sensing* 12, no. 1 (2019): 108, https://doi.org/10.3390/rs12010108.

34. B. Chen, B. F. Chen, and Y. N. Chen, "Real-Time Tropical Cyclone Intensity Estimation by Handling Temporally Heterogeneous Satellite Data," in *Proceedings of the AAAI Conference on Artificial Intelligence* (2021), 14721–14728.

35. B. Tong, X. Wang, J. Fu, P. Chan, and Y. He, "Short-Term Prediction of the Intensity and Track of Tropical Cyclone via Convlstm Model," *Journal of Wind Engineering and Industrial Aerodynamics* 226 (2022): 105026, https://doi.org/10.1016/j.jweia.2022.105026.

36. W. Yang, J. Fei, X. Huang, J. Ding, and X. Cheng, "Enhancing Tropical Cyclone Intensity Estimation From Satellite Imagery Through Deep Learning Techniques," *Journal of Meteorological Research* 38, no. 4 (2024): 652–663, https://doi.org/10.1007/s13351-024-3186-y.

37. P. Nandal, P. Mann, N. Bohra, G. Aldehim, A. A. H. Elnour, and R. Allafi, "Tropical Cyclone Intensity Estimation Based on YOLO-NAS Using Satellite Images in Real Time," *Alexandria Engineering Journal* 113 (2025): 227–241, https://doi.org/10.1016/j.aej.2024.10.072.

38. H. Chen, P. Yin, H. Huang, Q. Wu, R. Liu, and X. Zhu, "Typhoon Intensity Prediction With Vision Transformer," preprint, arXiv: 2311.16450 (2023), https://doi.org/10.48550/arXiv.2311.16450.

39. Y. Tian, W. Zhou, P. K. Y. Cheung, and Z. Liu, "Vision Transformer for Extracting Tropical Cyclone Intensity From Satellite Images," *Advances in Atmospheric Sciences* 42, no. 1 (2025): 79–93, https://doi.org/10.1007/s00376-024-3191-1.

40. S. R. Katragadda, "Tropical Cyclone Intensity Prediction Using Optimized Vision Transformer Dynamic Graph Convolutional Network," *Knowledge-Based Systems* 330 (2025): 114603, https://doi.org/10.1016/j.knosys.2025.114603.

41. J. Lian, J. Shao, H. Yu, et al., "Multimodal Fusion Learning for Predicting Tropical Cyclone Intensity Over Western North Pacific," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 18 (2025): 7048–7063, https://doi.org/10.1109/jstars.2025.3544865.

42. A. Karpatne, G. Atluri, J. H. Faghmous, et al., "Theory-Guided Data Science: A New Paradigm for Scientific Discovery From Data," *IEEE Transactions on Knowledge and Data Engineering* 29, no. 10 (2017): 2318–2331, https://doi.org/10.1109/tkde.2017.2720168.

43. E. McGowan, V. Gawade, and W. G. Guo, "A Physics-Informed Convolutional Neural Network With Custom Loss Functions for Porosity Prediction in Laser Metal Deposition," *Sensors* 22, no. 2 (2022): 494, https://doi.org/10.3390/s22020494.

44. H. Li, L. Gou, H. Li, and Z. Liu, "Physics-Guided Neural Network Model for Aeroengine Control System Sensor Fault Diagnosis Under Dynamic Conditions," *Aerospace* 10, no. 7 (2023): 644, https://doi.org/10.3390/aerospace10070644.

45. B. Kapusuzoglu and S. Mahadevan, "Physics-Informed and Hybrid Machine Learning in Additive Manufacturing: Application to Fused Filament Fabrication," *JOM* 72, no. 12 (2020): 4695–4705, https://doi.org/10.1007/s11837-020-04438-4.

46. A. Daw, A. Karpatne, W. D. Watkins, J. S. Read, and V. Kumar, "Physics-Guided Neural Networks (PGNN): An Application in Lake Temperature Modeling," in *Knowledge Guided Machine Learning* (Chapman and Hall/CRC, 2022), 353–372.

47. W. Gu, S. Primatesta, and A. Rizzo, "Physics-Informed Neural Network for Quadrotor Dynamical Modeling," *Robotics and Autonomous Systems* 171 (2024): 104569, https://doi.org/10.1016/j.robot.2023.104569.

48. G. J. Holland, "The Maximum Potential Intensity of Tropical Cyclones," *Journal of the Atmospheric Sciences* 54, no. 21 (1997):

2519–2541, https://doi.org/10.1175/1520-0469(1997)054<2519:tmpiot>2.0.co;2.

49. J. A. Knaff and R. M. Zehr, "Reexamination of Tropical Cyclone Wind–Pressure Relationships," *Weather and Forecasting* 22, no. 1 (2007): 71–88, https://doi.org/10.1175/waf965.1.

50. G. Holland, "A Revised Hurricane Pressure-Wind Model," *Monthly Weather Review* 136, no. 9 (2007): 3432–3445, https://doi.org/10.1175/2008mwr2395.1.

51. J. A. Knaff, C. R. Sampson, and M. DeMaria, "An Operational Statistical Typhoon Intensity Prediction Scheme for the Western North Pacific," *Weather and Forecasting* 20, no. 4 (2005): 688–699, https://doi.org/10.1175/waf863.1.

52. Q. Jin, X. Fan, J. Liu, Z. Xue, and H. Jian, "Using Extreme Gradient Boosting to Predict Changes in Tropical Cyclone Intensity over the Western North Pacific," *Atmosphere* 10, no. 6 (2019): 341, https://doi.org/10.3390/atmos10060341.

53. N. M. Foumani, C. W. Tan, G. I. Webb, and M. Salehi, "Improving Position Encoding of Transformers for Multivariate Time Series Classification," *Data Mining and Knowledge Discovery* 38, no. 1 (2024): 22–48, https://doi.org/10.1007/s10618-023-00948-2.

54. B. Harper, *Tropical Cyclone Parameter Estimation in the Australian Region*, Vol. 83 (Systems Engineering Australia Pty Ltd for Woodside Energy Ltd, 2002).

55. X. Lu, H. Yu, M. Ying, et al., "Western North Pacific Tropical Cyclone Database Created by the China Meteorological Administration," *Advances in Atmospheric Sciences* 38, no. 4 (2021): 690–699, https://doi.org/10.1007/s00376-020-0211-7.

56. D. P. Dee, A. J. Uppala, P. Berrisford, et al., "The Era-Interim Reanalysis: Configuration and Performance of the Data Assimilation System," *Quarterly Journal of the Royal Meteorological Society* 137, no. 656 (2011): 553–597, https://doi.org/10.1002/qj.828.

57. J. H. Chu, "A Regression Model for the Western North Pacific Tropical Cyclone Intensity Forecast," Technical Report (1994).

58. T. Chen and C. Guestrin, "Xgboost: A Scalable Tree Boosting System," in *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining* (2016), 785–794.

59. C. Wang, G. Zheng, X. Li, Q. Xu, B. Liu, and J. Zhang, "Tropical Cyclone Intensity Estimation From Geostationary Satellite Imagery Using Deep Convolutional Neural Networks," *IEEE Transactions on Geoscience and Remote Sensing* 60 (2021): 1–16, https://doi.org/10.1109/tgrs.2021.3066299.