JGR | Machine Learning and Computation

RESEARCH ARTICLE
10.1029/2025JH000816

Detection of Eye Occurrence in Sequential Satellite Infrared Imagery and Its Application to Improve Deep Learning-Based Tropical Cyclone Intensity Estimation

Yi Liu[1,2], Jing-Yi Zhuo[3], Kekuan Chu[1,2], and Zhe-Min Tan[1,2]

[1]State Key Laboratory of Severe Weather Meteorological Science and Technology, Nanjing University, Nanjing, China, [2]Key Laboratory of Mesoscale Severe Weather/Ministry of Education and School of Atmospheric Sciences, Nanjing University, Nanjing, China, [3]High Meadows Environmental Institute, Princeton University, Princeton, NJ, USA

Key Points:

- A deep learning-based model is developed to detect tropical cyclone (TC) eye from infrared imagery with 96.24% accuracy
- The TC eye persistence index (EPI) provides additional physical insight that enhances deep learning-based TC intensity estimation
- Extending the EPI time windows can further reduce the TC intensity estimation errors, especially for intense TCs

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

K. Chu,
kkchu@nju.edu.cn

Citation:

Liu, Y., Zhuo, J.-Y., Chu, K., & Tan, Z.-M. (2026). Detection of eye occurrence in sequential satellite infrared imagery and its application to improve deep learning-based tropical cyclone intensity estimation. *Journal of Geophysical Research: Machine Learning and Computation*, *3*, e2025JH000816. https://doi.org/10.1029/2025JH000816

Received 11 JUN 2025
Accepted 12 JAN 2026

Abstract Tropical cyclone (TC) intensity estimation remains a critical yet challenging task, especially for intense storms that are often underestimated by both conventional and deep learning satellite-based methods. Among various structural features, eye formation is the most distinct transformation during TC development and is closely linked to rapid intensification and peak intensity. Despite its significance, TC eye occurrence information remains underutilized in existing deep learning-based intensity estimation models. In this study, we improve TC intensity estimation by incorporating TC eye occurrence information as a physical auxiliary input to deep learning models. We first develop a deep learning-based classifier (DeepTCEye) to detect TC eyes from infrared imagery, achieving classification accuracies of 96.54% for eye scenes and 95.93% for non-eye scenes. Using the output probabilities from DeepTCEye, we construct eye persistence index (EPI) of varying time windows and use them as auxiliary inputs to the intensity estimation model. Results show that incorporating EPI leads to substantial improvements on TC intensity estimation, especially for intense TCs (intensity >96 kt), reducing the root mean square error by up to 15.65%. Moreover, EPI remains effective when combined with other known physical auxiliary input such as TC fullness, a structural indicator. Our results demonstrate the value of bringing the TC eye occurrence information into focus, showing that temporal evolution of TC eye provides deep learning models with a simple but physically grounded signal that significantly improves intensity estimation, particularly for intense storms.

Plain Language Summary Tropical cyclones (TCs) are powerful storms that can cause severe damage. Accurately estimating their intensity is crucial but remains challenging, especially for intense storms. Although the eye of a TC is a key indicator of storm intensity, existing deep learning models have not fully used this information. This study develops a deep learning method to detect the presence of the eye using satellite infrared images. Based on these detections, we construct eye persistence index that captures how the eye comes and goes over time. Incorporating these sequences significantly improves deep-learning based intensity estimates, particularly for strong TCs. Moreover, the longer EPI time series lead to better performance. Importantly, EPI can work not only on its own but also together with other physical features to support more accurate intensity estimation. Our findings show that including detailed eye evolution data helps deep learning models better understand storm intensity, paving the way for more accurate forecasts of high-impact TCs.

1. Introduction

As one of the most iconic and important features of a tropical cyclone (TC), the TC eye plays a crucial role in TC internal dynamics (e.g., Hendricks, 2012; Willoughby, 1998). Eye formation is often strongly associated with TC intensification (Kimball & Mulekar, 2004; Knaff & DeMaria, 2017; Knapp et al., 2018; Vigh et al., 2012) and nearly all TCs that reach a minimum sea level pressure below 950 hPa exhibit an eye (Weatherford & Gray, 1988). Furthermore, TCs with large, symmetric eyes and minimal outer-band convections are called annular TCs, which can sustain peak intensity for extended periods, leading to more severe damage (Chu & Tan, 2014; Knaff et al., 2003, 2008). Conversely, the disintegration of the TC eye is possibly associated with the weakening process (Knaff & DeMaria, 2017; Wood & Ritchie, 2015). Considering the close connection between TC eye and intensity, TC eye has long been emphasized in operational satellite-based intensity estimation. For example, the widely used Dvorak technique relies on subjective interpretation of satellite imagery and analysts often need to break standard constraints to provide a more accurate estimate if the eye becomes increasingly well developed

(Dvorak, 1975, 1984; Velden et al., 2006). Objective methods developed on this basis also include dedicated modules to account for eye-related information (objective Dvorak technique, ODT, Velden et al., 2006; advanced Dvorak technique, ADT, Olander & Velden, 2019; Satellite Consensus, SATCON, Velden & Herndon, 2020).

Accurately estimating TC intensity is essential for disaster preparedness, yet operational methods still face limitations in uncertainty and often struggle in emerging eye scenes and extreme storms (Olander & Velden, 2019; Velden et al., 2017; Velden & Herndon, 2020). With the recent surge in deep learning, convolutional neural networks (CNNs) have been widely applied to TC intensity estimation, leveraging their strengths in image recognition to extract features from satellite imagery (B.-F. Chen et al., 2019; Griffin et al., 2022, 2024; Pradhan et al., 2018; A. Wimmers et al., 2019; Zhuo & Tan, 2021). Further improvements have been achieved by incorporating attention mechanisms, residual connections, spatiotemporal encoding, and heterogeneous data processing (B. Chen et al., 2021; Tan et al., 2022; Wang et al., 2022; Z. Zhang et al., 2022). Meanwhile, alternative architectures such as graph neural networks (Xu et al., 2023; Zhao et al., 2024), swin transformers (Choo et al., 2024; C.-J. Zhang, Wang, Ma, & Lu, 2024), and vision transformers (Y. Tian et al., 2025; Tong et al., 2023) have also been explored. Reported root mean square errors (RMSEs) in these studies typically fall in the range of 8–15 kt, depending on the basin, data set, and evaluation period. Yet, despite growing architectural complexity, these purely data-driven models still face many challenges due to high data requirements and a lack of general world knowledge.

Therefore, adopting a physical perspective is important for overcoming these limitations and improving model performance. An effective approach is to introduce physical constraints during model training to guide the learning process (Karniadakis et al., 2021; Meng et al., 2022). For example, Zhuo and Tan (2021) significantly improved estimation accuracy by incorporating TC fullness, a newly defined structural indicator of TC (Guo & Tan, 2017, 2022). Other studies have integrated environmental factors such as vertical wind shear, ocean heat content, and sea surface temperature to enhance performance (Griffin et al., 2024; W. Tian et al., 2023; Yang et al., 2024; C.-J. Zhang, Wang, Lu, & Sun, 2024). However, given the relatively low case number, existing deep learning-based models tend to systematically underestimate intense TCs that are of greater concern (Duong et al., 2023). Therefore, introducing physical information closely tied to intense TCs may provide an important pathway to alleviating underestimation and thus further improve model performance. In this context, incorporating TC eye information may offer a promising solution. Despite the operational focus on the eye, recent deep learning–based intensity estimation models have not paid sufficient attention to this critical feature. It should be acknowledged that deep learning models could inherently focus more on the eye region as indicated by feature attribution analyses (Griffin et al., 2024). And efforts such as applying fisheye distortion to prioritize eye information (Higa et al., 2021) and modifying loss functions to encode the eye's link to intensity indices (Li et al., 2025) have improved both interpretability and performance. Nonetheless, exploring more explicit and efficient ways to incorporate TC eye information into deep learning-based TC intensity estimation remains important to pursue better performance.

In this context, constructing consistent long-term eye information data sets is the first step. Some recent studies have focused on the wind distribution in the eye using Himawari-8 data (Bessho et al., 2016), but these data are limited by the satellite's relatively short temporal coverage and the sparse observations of the wind field (Tsujino et al., 2021; Tsukada et al., 2024; Tsukada & Horinouchi, 2020, 2023). The Automated Rotational Center Hurricane Eye Retrieval (ARCHER) algorithm, originally designed for microwave-based TC center detection, also provides eye-related parameters (Wimmers & Velden, 2010; Wimmers & Velden, 2016a, 2016b). However, despite its adaptation to some infrared (IR) data, available data sets have mostly been sparse microwave-based and only provide partial IR-based eye parameters after 2016. Additionally, these algorithms are often computationally demanding. Some attempts have used machine learning to detect TC eyes from IR imagery, but the classification accuracy in these studies could be further improved (DeMaria et al., 2015; L. Zhao et al., 2020; see more details in Section 3.2). Overall, developing a deep learning-based, high-accuracy eye detection model is important for both real-time TC monitoring and the construction of long-term eye occurrence data to enhance deep learning-based TC intensity estimation models.

This study aims to objectively detect the TC eye and explicitly incorporate eye occurrence information into deep learning models to improve performance in real-time TC intensity estimation. We make two key tasks: (a) the development of a deep learning-based eye detection model (DeepTCEye), which identifies the presence or absence of a TC eye from IR imagery; and (b) the proposal and application of a simple yet effective eye-related

physical auxiliary input (Eye Persistence Index, EPI), which indicates the temporal evolution of eye occurrence and cessation, to enhance the deep learning-based intensity estimation models. The structure of this paper is as follows: Section 2 provides a detailed description of the data sets and methodology used in this study. Section 3 presents the development and evaluation of DeepTCEye. Section 4 investigates the application of EPI to TC intensity estimation. Finally, Section 5 summarizes the findings and discusses potential future directions.

2. Data and Methods

2.1. Data

The databases used in this study are described in this section. These include the satellite imagery, the ground truth for algorithm development and evaluation, and basic TC information, such as intensity and size.

2.1.1. Satellite Imagery

All deep learning models in this study were trained using TC-centered IR images (near 11 μm window channel) from the Hurricane Satellite data set (HURSAT-B1; Knapp & Kossin, 2007a, 2007b), which provides global IR imagery on an equal-angle latitude–longitude grid with a spatial resolution of 8 km and a temporal resolution of 3 hr, covering the period from 1982 to 2016. For post-2016 data, we employed the GridSat-B1 archive (Knapp et al., 2011), which maintains the same 11 μm window channel, spatiotemporal resolution and map projection as HURSAT-B1, ensuring consistency across the entire data set. In this study, each IR image has been standardized by normalizing brightness temperature values to approximate a normal distribution, facilitating the training and applications of the deep learning model.

2.1.2. Definition and Labeling of TC Eyes in IR Imagery

To train the eye detection model (DeepTCEye), we used labels derived from ARCHER algorithm (Wimmers and Velden, 2016a, 2016b), which has been applied in the HURSAT-B1 data set and provides IR-based eye parameters, including eye probability, eye completeness (the percent of the eye that the eyewall encompasses), and eye radius. We adopt the thresholding approach validated by Knapp et al. (2018), who performed statistical analysis to align ARCHER-derived eye classifications with the Cyclone Center data set, which was shown to closely match expert consensus (Knapp et al., 2016), making it a reliable reference. Based on this validation, if the ARCHER-derived eye probability exceeds 5% and the eye completeness is greater than 95%, the sample is labeled as having an eye (eye scene, label 1); otherwise, it is labeled as without an eye (non-eye scene, label 0).

2.1.3. TC Data

The TC data used in this study were obtained from the Joint Typhoon Warning Center best track records in the IBTrACS database (Knapp et al., 2010a, 2010b). For training and evaluation of intensity estimation model, we use the 1-min mean maximum sustained surface wind (MSW). In addition, two TC size measures, the gale force wind radii (34 kt; 1 kt = 0.51 m/s; R34) and the radius of maximum wind (RMW), are used to calculate TC fullness. TC fullness is defined as 1- (RMW/R34), which is a novel concept for characterizing TC structure and has been shown to correlate strongly with intensity and intensity change (Guo & Tan, 2017, 2022). Following Zhuo and Tan (2021), TC fullness is also considered here as physical auxiliary input to enhance the performance of TC intensity estimation model.

2.1.4. TC Case Selection and Training/Test Data Set Preparation

In this study, only TC cases with a lifetime maximum intensity (LMI) exceeding 34 kt are included, while extratropical systems and tropical waves are excluded.

For the DeepTCEye model, 8,110 IR images labeled as eye scenes were extracted from HURSAT-B1, covering the Northern Hemisphere during 1982–2016 based on the ARCHER criteria. In parallel, 300 randomly selected images labeled as non-eye scenes were drawn from each year during 1982–2016, resulting in 10,005 eye-absent samples. For years with fewer than 300 available non-eye scenes (1982, 1983, 1986, and 1987), all available images were included. Each image was cropped to a 56 × 56 pixel patch (approximately 448 km × 448 km) centered on the TC. Overall, eye scenes accounted for 44.8% of the data set, ensuring a nearly balanced distribution for training. The data set was split chronologically into training (1982–2008 for eye scenes and 1982–2006

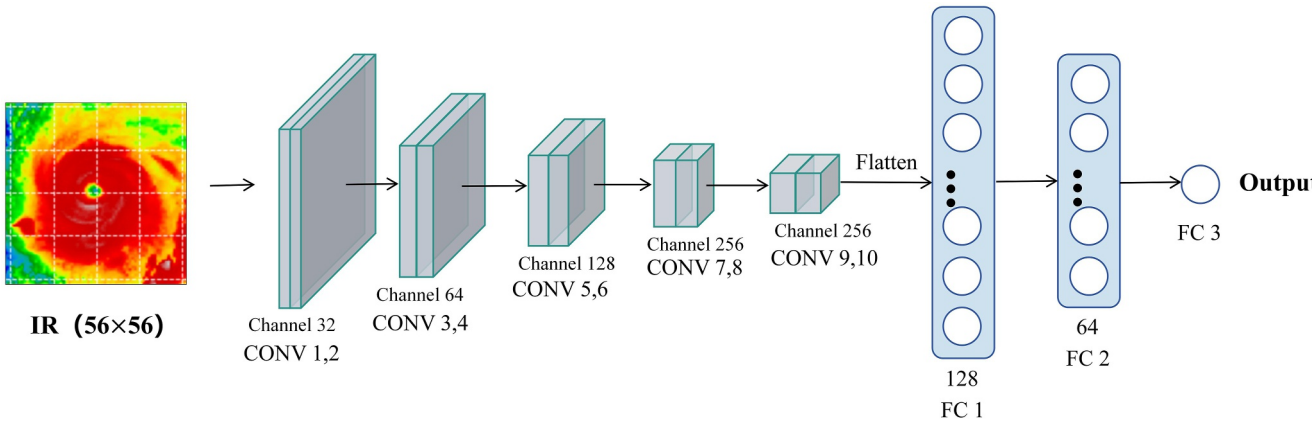

Figure 1. The configuration of DeepTCEye.

for non-eye scenes; 5,731 eye and 7,005 non-eye samples), validation (2009–2013 for eye scenes and 2007–2011 for non-eye scenes, 1,078 eye and 1,500 non-eye samples), and test (2014–2016 for eye scenes and 2012–2016 for non-eye scenes, 1,301 eye and 1,500 non-eye samples) sets.

For the TC intensity estimation model, only TCs in the Northwest Pacific were used. A total of 14,142 samples were selected and split chronologically into training (2000–2011; 9,831 samples), validation (2012–2014; 1,948 samples), and test (2015–2016; 2,363 samples) sets. Each image was cropped to a 58 × 58 pixel patch (approximately 464 km × 464 km) centered on the TC, the same as Zhuo and Tan (2021).

2.2. Deep Learning Algorithm

In this study, we first develop a deep learning model to accurately detect TC eyes from satellite imagery, and then explore how the derived eye information can be used to enhance TC intensity estimation. We employ CNNs (LeCun et al., 2015) for both TC eye detection and intensity estimation, given their effectiveness in capturing spatial hierarchies in visual data and extracting relevant cloud features from satellite IR imagery. Specifically, we adopt the visual geometry group network (VGGNet; Simonyan & Zisserman, 2015) as the backbone for all models, following Zhuo and Tan (2021), who demonstrated its superior performance in TC-related estimation tasks compared to other CNN variants.

The architecture of the eye detection model DeepTCEye is illustrated in Figure 1. It consists of 10 convolutional layers with channel sizes of 32, 32, 64, 64, 128, 128, 256, 256, 256, and 256, followed by 3 fully connected layers with sizes of 128, 64, and 1. A 3 × 3 max-pooling layer follows every two convolutional layers, and batch normalization is applied after each convolutional block to stabilize training and reduce overfitting. The ReLU function is used in the first two fully connected layers, while the final layer uses a sigmoid activation to output a probability between 0 and 1.

For the TC intensity estimation model, we adopt the single-task learning branch of DeepTCNet (DeepTCNet-I; Zhuo & Tan, 2021), keeping the network architecture and hyperparameters the same as the original design except for the auxiliary input layer. DeepTCNet-I is specifically designed for TC intensity estimation from IR imagery and improves accuracy by integrating physical auxiliary inputs such as TC fullness into the top layer of a VGGNet-based architecture. This makes it an appropriate reference model for our study. By varying the specific physical auxiliary inputs, we evaluate the impact of explicitly incorporating eye occurrence information on intensity estimation. Details of the experimental settings are provided in the following sections.

3. Development and Evaluation of DeepTCEye

Building on the framework introduced in Section 2.2, we trained the DeepTCEye model using binary cross-entropy loss over 100 epochs with the Adam optimizer (learning rate = 1e−3, batch size = 32). Given the relatively balanced distribution of eye and non-eye scenes (4:5), classification accuracy was used as the monitoring metric during training. To reduce randomness, training was repeated five times, and the final model was

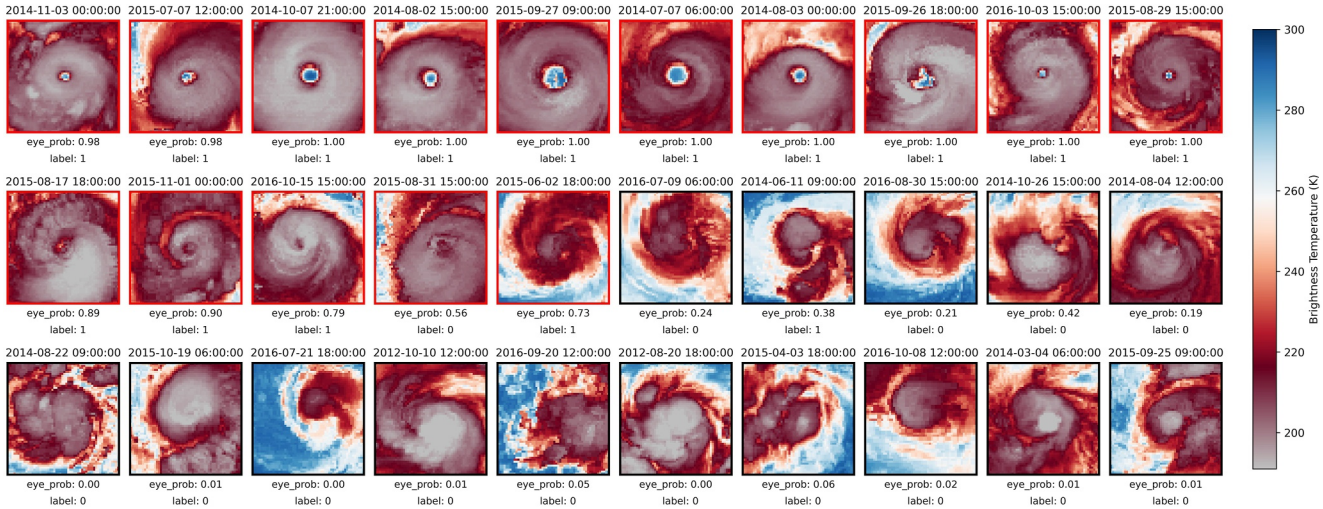

Figure 2. Examples of TC IR imagery under different predicted eye probabilities by DeepTCEye. Red borders indicate scenes with predicted eye probability greater than 0.5; black borders indicate less than or equal to 0.5.

selected based on the lowest validation loss (0.1372 at epoch 69), which corresponded to the highest validation accuracy of 0.9527. This section provides a detailed evaluation of DeepTCEye's performance. In addition to achieving high classification accuracy, the model produces interpretable probabilistic outputs, enabling uncertainty-aware labeling and supporting robust downstream applications.

3.1. Eye Probability Output and Threshold Selection

Although DeepTCEye was trained using binary labels (0 for non-eye scenes, 1 for eye scenes), the model outputs a continuous probability between 0 and 1 through a sigmoid activation. To examine whether these outputs align with TC physical structure, we analyzed the probability distribution on the test set. Results show that 41.5% of the samples had probabilities greater than 0.9, and 47.5% had values below 0.1. Only 11.1% fell between 0.1 and 0.9. We further visualized representative IR images from each group (Figure 2). Samples with high probabilities generally displayed a well-defined, symmetric eye, while those with low probabilities lacked organized features. Scenes in the middle range often showed partially formed or asymmetric eye-like structures, indicating either a developing or dissipating TC core. This pattern suggests that the output probability has meaningful correspondence with physical characteristics. To determine an appropriate classification threshold, we manually reviewed the ambiguous cases (0.1–0.9). This inspection showed that scenes with probabilities above 0.5 typically featured recognizable eye structures, while those below 0.5 did not. Therefore, we adopt 0.5 as the final cutoff, as it aligns well with both human interpretation and observed storm morphology.

3.2. Performance Evaluation and Model Comparison

DeepTCEye demonstrates high performance on the test set, achieving classification accuracies of 96.54% for eye scenes, 95.93% for non-eye scenes, and an overall accuracy of 96.24% across 2,801 samples. As shown in Table 1, we compare these results with existing approaches, including machine learning methods such as Quadratic Discriminant Analysis (QDA; DeMaria et al., 2015) and CNN-based models like L. Zhao et al. (2020). The accuracy rates of the QDA method in eye scenes and non-eye scenes are 78% and 72%, respectively. DeepTCEye slightly outperforms L. Zhao et al. (2020) in eye scenes, while it falls slightly below their results in non-eye scenes. It should be noted that, precise eye detection is more crucial in practical applications. Moreover, their data set was not partitioned by year or by individual TC, meaning that images from different times of the same TC could appear in both the training and testing sets. Given the temporal continuity of TC structure, if some samples of a given TC are included in the training set, the model may perform better on the remaining samples of the same TC that appear in the test set. Overall, DeepTCEye achieves comparable or superior performance to existing

Table 1
Comparison of DeepTCEye Accuracy Against Existing Methods

Model	Eye scenes	Non-eye scenes	Overall
DeMaria et al. (2015)	78%	72%	75%
L. Zhao et al. (2020)	94.22%	99.43%	96.83%
DeepTCEye	96.54%	95.93%	96.24%

methods and shows particular strength in detecting eye scenes, which are essential for reliable TC intensity estimation. Furthermore, unlike prior models trained on subjectively labeled data sets with limited temporal coverage, DeepTCEye is trained on a large, objectively labeled data set spanning multiple decades. This improves the consistency, reliability, and representativeness of the model and provides a solid foundation for using EPI as a physically meaningful predictor in downstream applications.

4. Application of TC Eye to Intensity Estimation

Since a well-defined TC eye is an important indicator of intensification, we explicitly introduce the Eye Presence Index (EPI), a series describing eye occurrence over a recent window, and evaluate its contribution to deep learning-based TC intensity estimation. The motivation is twofold: (a) eye formation and persistence are commonly associated with intense or intensifying TCs (Knaff & DeMaria, 2017), and (b) temporal information may help reduce uncertainty when the eye is obscured or too small to be identified in a single IR image. Specifically, we adopted the DeepTCNet-I framework described in Section 2.2 and conducted controlled experiments. The baseline model (CNTL) does not incorporate any physical auxiliary inputs, while the experimental models include EPI, TC fullness, or their combination to evaluate the independent contribution of EPI as well as its complementarity with other known physical auxiliary inputs. For each experiment, we traine 20 independent runs to reduce the randomness inherent in deep learning model training. However, when quantifying the evaluation of each sample, we used the average prediction from the three top-performing models (with the lowest RMSE) as the final estimate to improve efficiency for real-time applications. And all RMSE and bias values discussed in this study are the comparisons between the deep learning model results and IBTrACS ground truth on the test set.

4.1. Impact of EPI on TC Intensity Estimation

We defined EPI and investigated its utility in improving TC intensity estimation. Specifically, EPI is a sequence recording the presence (labeled 1) or absence (labeled 0) of a TC eye at each preceding 3-hr interval, with the last element representing the current time. For example, a 6h-EPI would be a 3-element label (e.g., [0, 0, 1]), indicating eye presence/absence at −6h, −3h, 0h (current time), respectively. If a TC sample lacked records for certain prior times due to missing data or early-stage formation, or a lifecycle shorter than the specified EPI duration, the corresponding values were set to 0. This ensured that all samples had standardized, temporally consistent EPI inputs for model training. Subsequently, we conducted a set of controlled experiments. The baseline model (CNTL) used only IR images as input, while experimental models incorporated 6h-EPI, 12h-EPI, 18h-EPI, and 24h-EPI to evaluate the effect of varying time windows for eye information in intensity estimation.

We analyzed the intensity estimation results across different TC intensity categories, classified following the Saffir-Simpson Hurricane Wind Scale. Under this classification, tropical depressions (TD, MSW < 34 kt), tropical storms (TS, MSW between 34 and 64 kt), Category 1 hurricanes (C1, 64–83 kt), Category 2 hurricanes (C2, 83–96 kt) and Category 3 and above (C3+, >96 kt) are distinguished. Notably, TCs with MSW exceeding 96 kt (i.e., C3+) are generally considered as intense TCs. As shown in Figure 3a, RMSE distributions for different TC intensity categories are compared across the experiments. For weaker TCs (TD and TS), incorporating EPI showed limited impact. However, for C1 and above, integrating EPI consistently reduced the RMSE in intensity estimation, with the improvements becoming more pronounced as both TC intensity and EPI time window increased. The bias distributions exhibit a consistent pattern (Figure 3b). Table 2 quantitatively summarizes the results. The model incorporating 24h-EPI achieved the best performance, reducing the overall RMSE from 12.37 kt to 11.50 kt, and the RMSE of intense TCs (C3+) from 15.07 kt to 12.71 kt. These findings are particularly noteworthy given that intense TCs have long posed a challenge for deep learning models, mainly due to their underrepresentation in training data sets. Even with dual-band microwave inputs that could capture 3D structural information, B.-F. Chen et al. (2019) reported RMSEs around 13–15 kt for Category 3+ storms, underscoring the inherent difficulty of such cases. The observed reduction in estimation error highlights the value of EPI as a simple yet effective tool for capturing structurally intense TCs that are otherwise difficult to capture.

Given the superior performance of the 24h-EPI configuration, we further quantitatively compared the EPI24 and CNTL models across different TC intensity categories. As shown in Figure 4a, predictions from the model incorporating 24h-EPI align more closely with the 1:1 reference line, indicating better agreement with the IBTrACS records and reduced estimation bias. Figure 4b further illustrates the variation of bias and RMSE with

10.1029/2025JH000816

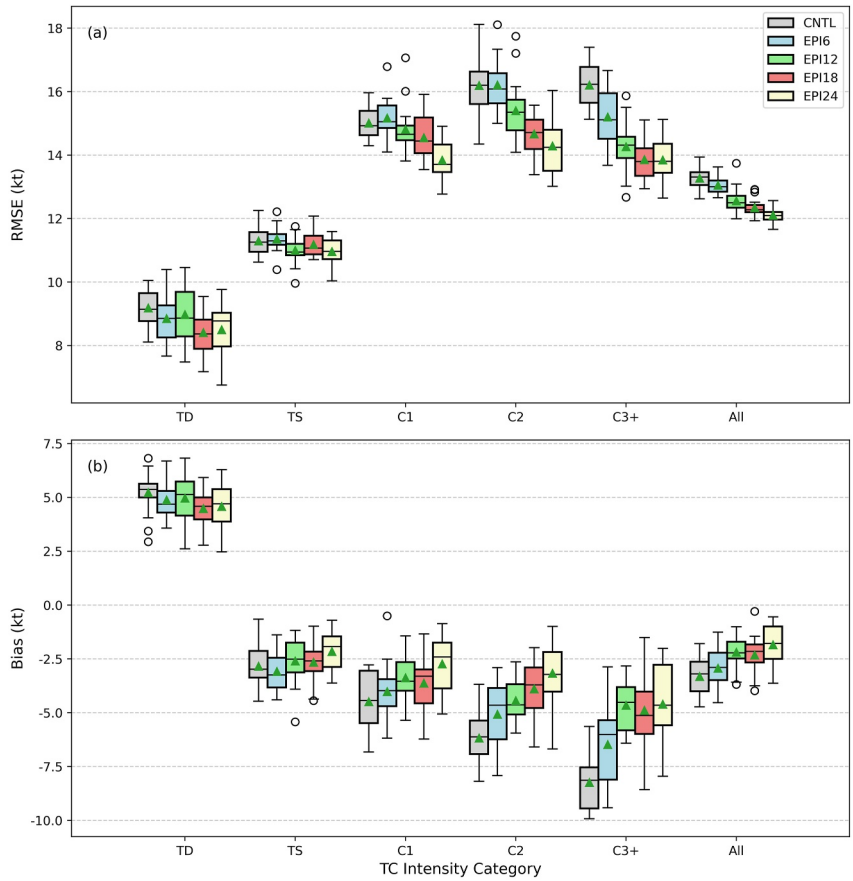

Figure 3. Root mean square error (a) and bias (b) distribution of TC intensity estimation across different intensity categories for models incorporating EPI of varying time windows (6, 12, 18, 24h). Categories follow the Saffir-Simpson Hurricane Wind Scale. "All" represents the full data set.

intensity, showing that 24h-EPI systematically reduces high-intensity underestimation while maintaining comparable performance at lower intensities. The RMSE of predictions is also notably smaller, especially for higher intensity ranges. Overall, the incorporation of 24h-EPI improves the estimation accuracy by 7.07% across all samples, and for intense TCs (C3+), the accuracy improvement reaches as high as 15.65%. Notably, Zhuo and Tan (2021) also attempted to enhance intensity estimation by incorporating historical IR images to reflect TC development processes. However, their results showed that including images beyond 18 hr often degraded performance, likely due to the introduction of excessive irrelevant information. Expanding input channels at the bottom of the model to include past IR images also significantly reduce training and inference efficiency. In contrast, our approach integrates a compact sequence at the model's top layer, which requires negligible additional computation. This further highlights the efficiency and effectiveness of using EPI to improve intensity estimation.

Table 2
Intensity Estimation RMSE and Bias Incorporating EPI of Varying Time Windows

		RMSE (kt)		Bias (kt)	
Experiment	Auxiliary input	Overall	C3+[a]	Overall	C3+[a]
CNTL	/	12.37	15.07	−2.32	−7.98
EPI6	6h-EPI[b]	12.36	13.84	−1.85	−5.54
EPI12	12h-EPI[c]	11.72	13.24	−2.03	−5.93
EPI18	18h-EPI[c]	11.57	13.18	−1.76	−5.78
EPI24	24h-EPI[c]	11.50	12.71	−0.79	−3.01

[a]$C3^+$ refers to TC scenes with MSW exceeding 96 kt, corresponding to Category 3 and above based on the Saffir–Simpson Hurricane Wind Scale. [b]"6h-EPI" refers to a 0–1 sequence indicating the presence or absence of a TC eye over the past 6 hr at 3-hr intervals (i.e., a 3-element sequence for 6h EPI). [c]"12h-EPI" "18h-EPI" and "24h-EPI" follow the same construction as "6h," with increasing historical time window.

4.2. The Interpretation of EPI's Impact on Intensity Estimation

The previous results demonstrated clear improvements in intensity estimation with the inclusion of EPI, especially for intense TCs. This improvement may stem from the fact that eye formation, persistence and disintegration, as captured by EPI, are physically linked to organized convection and intensification, thereby providing the structural cues that guide the model toward more accurate intensity predictions.

10.1029/2025JH000816

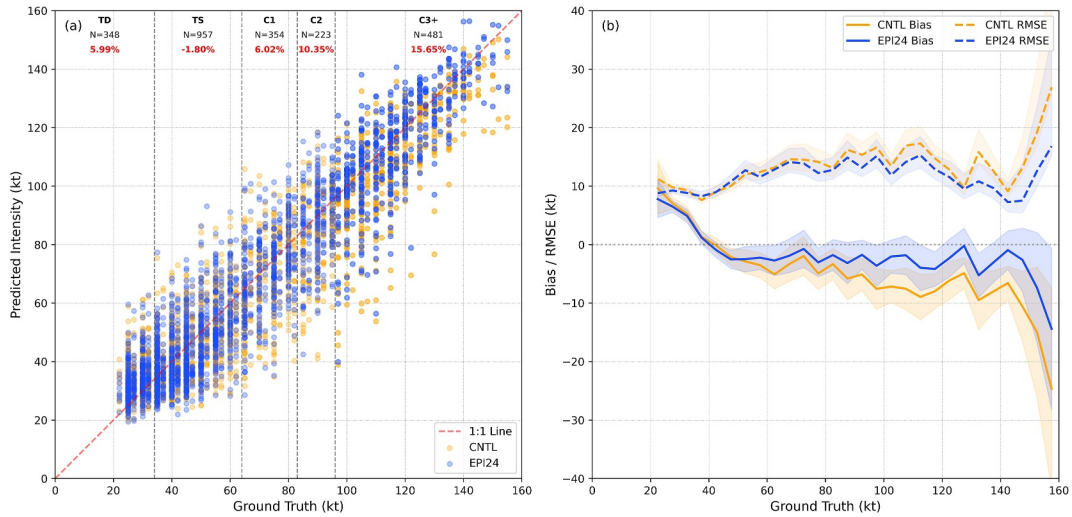

Figure 4. (a) Scatterplot comparing intensity predictions from CNTL and EPI24 against IBTrACS records across different intensity categories. Blue dots represent the model incorporating 24h EPI; orange dots represent the baseline model (CNTL). A 1:1 line is included for reference, and the gray dashed lines demarcate the boundaries between tropical depressions, tropical storms, C1, C2, and C3+ categories. Sample size (*N*) for each category is annotated, and root mean square error (RMSE) reduction (%) achieved by EPI24 is highlighted in bold red text. (b) Binned verification of model errors (using a 5 kt bin size). Shown are the mean Bias (solid lines) and RMSE (dashed lines) as a function of best-track intensity, with shading indicating the 95% confidence intervals based on *t*- and χ^2-based estimates. Orange denotes CNTL and blue denotes EPI24.

To further interpret how the model utilizes TC eye occurrence information, we apply permutation feature importance (Altmann et al., 2010), an explainable artificial intelligence technique (Gunning et al., 2019) that quantifies the importance of an input feature by measuring the change in model performance when the feature values are permuted. In this context, the EPI is treated as a physical input feature. To evaluate its impact, we compare model performance under three specific EPI configurations with a control model (CNTL) that does not use EPI: (a) the original EPI derived from observations, (b) all EPI values set to 0, representing the absence or disintegration of an eye, and (c) all values set to 1, indicating persistent eye presence. This design is based on the consideration that the position and number of "1"s in the EPI encode the occurrence and recent evolution of the eye—two factors that are strongly associated with TC intensity. By systematically altering this sequence, we aim to assess whether the model has learned to recognize and utilize such physical cues in its intensity estimation. A notable performance drop, particularly under permutation, would indicate that the model has learned to meaningfully leverage eye-related structural signals for intensity estimation. As shown in Figure 5, setting EPI to all 1 for weak TCs resulted in substantial overestimation, while setting them to all 0 for intense TCs led to underestimation. We also constructed additional EPI series to mimic the onset and cessation of the eye, showing that eye appearance leads to higher predicted intensities while disappearance has the opposite effect (Figure S1). These patterns suggest that the model has indeed learned to utilize EPI meaningfully, interpreting the presence, frequency, and position of eye occurrence information as important physical cues. Moreover, the CNTL model tends to underestimate intensity for TS and above. And this underestimation is notably alleviated when 24h-EPI is included, suggesting that 24h-EPI may play an especially important role in intense TC cases.

To further assess the validity of 24h-EPI for TC intensity estimation, we evaluate four super TCs (LMI > 113 kt) from 2018 to 2022: Mangkhut (2018), Goni (2020), Chanthu (2021), and Hinnamnor (2022). They were selected not only for their extremity but also for their varied eye characteristics. Mangkhut exhibited a large, well-defined eye during its mature phase; Goni and Chanthu had compact eyes that were less distinguishable in IR imagery and led to notable underestimation by the baseline model during early intensification; Hinnamnor featured a small but reasonably identifiable eye. These structural differences enable us to assess how EPI supports intensity estimation across varying degrees of eye clarity and development. None of the selected cases were included in the training set, ensuring an independent evaluation. As shown in Figure 6, incorporating 24h-EPI consistently improved the intensity estimation throughout the entire TC lifecycle. Across all samples, the RMSE decreased for every case, with reductions ranging from 14.11% (Mangkhut) to 36.30% (Chanthu). Notably, when focusing on intense

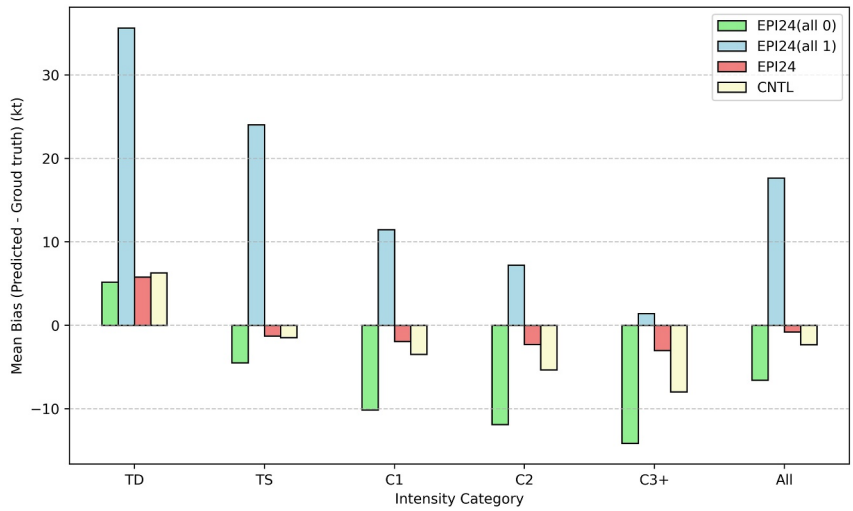

Figure 5. Intensity estimation bias across different intensity categories under varying 24h-EPI configurations. Four EPI settings are compared: normal EPI input (red), all-zeros EPI (green), all-ones EPI (blue), and control model without EPI (yellow).

periods (MSW > 96 kt), the improvements were even more substantial, with a 41.15% RMSE reduction for Chanthu and over 30% for both Goni and Hinnamnor. This pattern is physically intuitive: for Mangkhut, where eye signals were already clear, the EPI provided modest additional benefit. In contrast, for storms with small or visually ambiguous eyes, particularly Chanthu and Goni, EPI served as a key supplementary indicator, helping the model better capture the structural maturity of the system.

A closer look at Chanthu further illustrates this point. During its first episode of rapid intensification, which began at 03:00 UTC on 6 September and reached Category 5 intensity at 06:00 UTC on 8 September, Chanthu featured an exceptionally small eye. At certain times, the eye was likely too small to be reliably identified, resulting in poor visual continuity (Figure 6e). During the period marked by the yellow shaded region in Figure 6c, this lack of clear structural signals likely contributed to the underestimation of Chanthu's intensity by the baseline model. In contrast, the inclusion of EPI helped the model better infer a coherent core structure, improving estimation accuracy even when direct eye features were inconspicuous. This highlights the value of EPI not only in conveying temporal context but also in reinforcing critical structural cues that may be subtle or partially obscured in IR imagery. It is also worth noting that the IR imagery used in this study has a spatial resolution of 8 km, which inherently limits the ability of DeepTCEye to detect pinhole eyes with radii smaller than 8 km. Nevertheless, our results suggest that a DeepTCEye model finetuned on higher-resolution IR data, such as the 2 km resolution provided by the Himawari-8 satellite's Advanced Himawari Imager (Bessho et al., 2016), as well as upcoming missions like Himawari-10 (Bessho et al., 2025) and NOAA's GeoXO (Lindsey et al., 2024), could further improve TC eye detection capability. And the derived EPI may thus play a more significant role in enhancing TC intensity estimation.

4.3. Performance of EPI With Other Physical Auxiliary Input

To fully assess EPI's robustness and complementarity, we examine its performance alongside established auxiliary inputs such as TC fullness—a physically interpretable predictor strongly correlated with TC intensity (Guo & Tan, 2017) and the dominant auxiliary variable in DeepTCNet-I (Zhuo & Tan, 2021). Since TC fullness already captures key structural characteristics of TC, it often dominates the learning process and may obscure other auxiliary signals. We therefore investigate whether EPI can still contribute meaningfully when TC fullness is already included. To address this, we conducted two additional experiments: the first used IR imagery with TC fullness only (WTCF) as input, and the second combined IR imagery with both 24h-EPI and TC fullness (EPI24&TCF) as input.

The quantitative results are summarized in Table 3, each row reflects a specific model configuration aimed at evaluating the individual and combined effects of TC fullness and 24h-EPI on TC intensity estimation. Compared

10.1029/2025JH000816

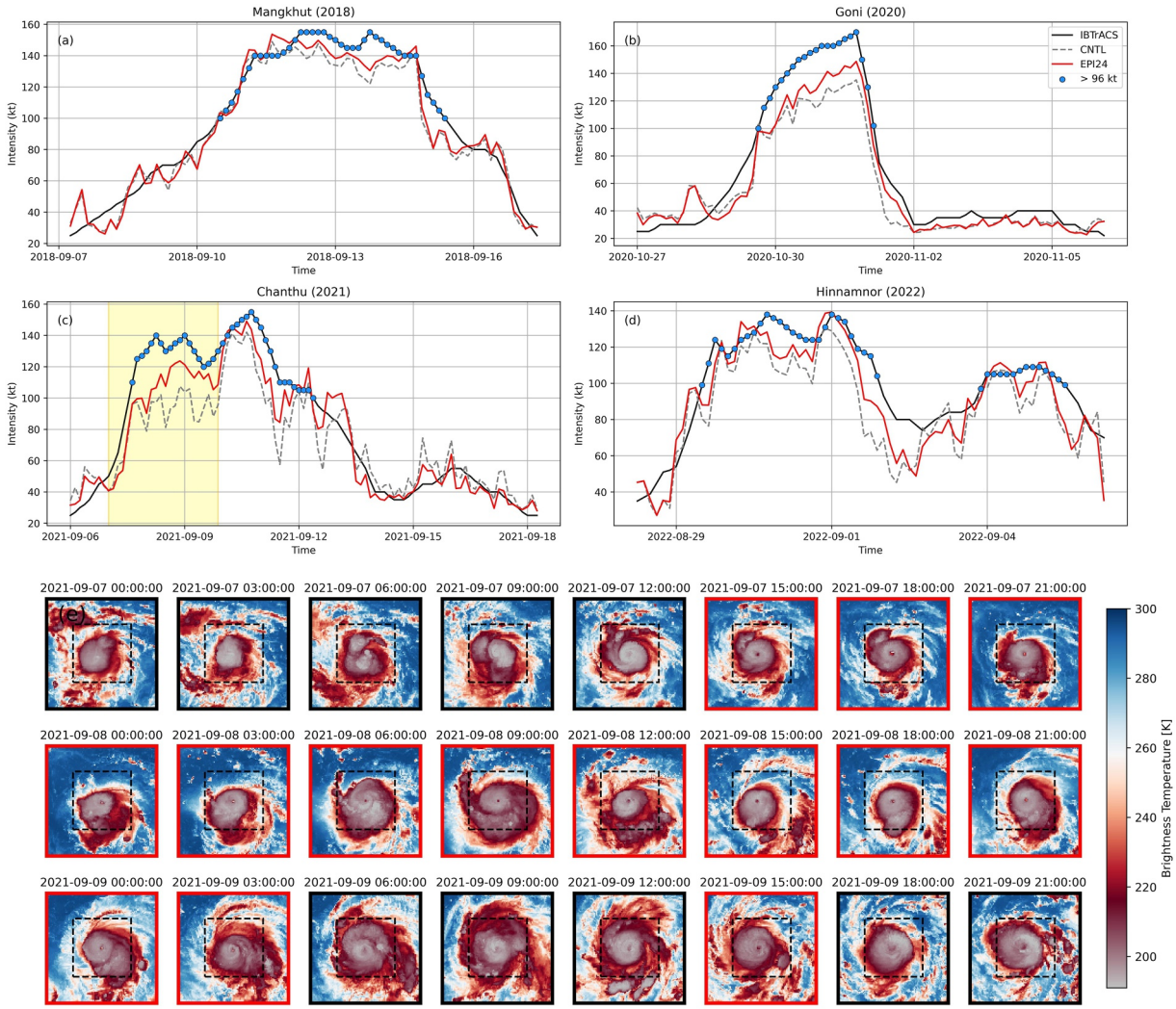

Figure 6. TC Intensity estimation results for real intense case. Panel (a)–(d) show the comparison of IBTrACS ground truths (black), model predictions incorporating 24h-EPI (red), and the control model without EPI (gray). Intensity estimates during intense phases (MSW > 96 kt) are highlighted with blue solid dots. Panel (e) shows IR images of Chanthu (2021) from 00:00 UTC on 7 September 2021 to 21:00 UTC on 9 September 2021, corresponding to the yellow shaded region in (c); red and black borders indicate scenes identified by DeepTCEye as eye and non-eye scenes, respectively, and the dashed black box marks the cropped region used as model input.

to the IR-only baseline (CNTL), incorporating TC fullness (WTCF) significantly reduces RMSEs across all intensity categories, with the most pronounced improvement observed for weaker TCs (e.g., 44.44% reduction for TD and 24.97% reduction for TS). This confirms TC fullness's strong role in capturing structural information relevant to general intensity estimation, especially in early developmental stages. In comparison, incorporating 24h-EPI (EPI24) shows limited improvement for weaker TCs but increasing gains with intensity, achieving a 15.65% RMSE reduction for intense TCs (C3+). These results reinforce the standalone value of EPI, particularly in capturing the dynamics of stronger systems. Crucially, when both TC fullness and EPI are used together (EPI24&TCF), the model achieves further gains over WTCF, particularly for intense scenes, with a 9.79% improvement in the C3+ category. This reinforces the complementary nature of the two inputs: while TC fullness captures quasi-static structure, EPI introduces temporal signals indicative of ongoing intensification.

Overall, EPI24&TCF delivered the best performance when considering all configurations tested. As the final model, we further benchmarked it against several existing operational and deep learning–based TC intensity estimation methods. It should be noted that due to differences in test sets and input information across methods, these results are not strictly comparable and should be interpreted as reference values only. As summarized in Table 4, our final model achieved an RMSE of 8.95 kt on the test set comprising 2,363 TC samples, outperforming

Table 3
Intensity Estimation RMSE (kt) of EPI and TC Fullness as Auxiliary Inputs Across Different Intensity Categories

Experiment	TD	TS	C1	C2	C3+	All
CNTL[a]	9.60	10.30	14.02	14.86	15.07	12.37
EPI24[b]	9.02	10.49	13.17	13.32	12.71	11.50
WTCF[c]	5.33	7.73	10.22	10.95	12.39	9.31
EPI24&TCF[d]	5.01	7.89	10.15	10.34	11.24	8.95

Note. All RMSE values are reported with two decimal places, while the improvement ratios reported in the main text are calculated from the original eight-decimal RMSE values and then rounded to two decimals. Therefore, slight discrepancies may exist if one computes the ratios directly from the tabled RMSE values. [a]"CNTL" represents the experiment using only IR imagery as input. [b]"EPI24" represents the experiment using IR imagery together with 24h-EPI as input. [c]"WTCF" represents the experiment using IR imagery together with TC fullness only as input. [d]"EPI24&TCF" represents the experiment using IR imagery together with both TC fullness and 24h-EPI as input.

widely used operational approaches such as ADT (12.24 kt) and SATCON (9.97 kt), and also demonstrating strong competitiveness among deep learning–based methods. These results highlight the value of incorporating temporal eye occurrence information like EPI into deep learning frameworks, particularly in basins such as the Northwest Pacific, where aircraft-based observations are sparse and intense storms are more frequent. Moreover, EPI is structurally simple and can be easily integrated into other deep learning models with negligible additional computational cost, while effectively guiding the model toward predictions that better align with the underlying TC dynamics. Therefore, the key issue of this study lies in demonstrating these practical value of EPI for real-time applications rather than pursuing minimal error in TC intensity estimation.

5. Summary and Discussion

This study proposes a physics-augmented deep learning approach to improve TC intensity estimation derived from satellite IR imagery by incorporating temporal eye occurrence information. Specifically, we introduce the novel physical auxiliary input EPI, which imposes temporal constraints that guide deep learning models toward more physically consistent predictions. The proposed approach is motivated by the strong relationship between eye structure and TC intensity and intensification, with particular importance for intense TCs. To obtain reliable TC eye occurrence information, we first developed a CNN-based TC eye detection model (DeepTCEye), which achieved an overall classification accuracy of 96.24% in detecting TC eyes from IR imagery, providing a robust foundation for constructing EPI.

Using predicted eye probabilities from DeepTCEye, we evaluated the impact of EPI with varying time windows (previous 6, 12, 18, and 24 hr) in the TC intensity estimation model. Our results demonstrated that: (a) EPI consistently improved intensity estimation performance, with longer sequences yielding greater benefits; and (b) the improvements were most pronounced for intense TCs (MSW > 96 kt). The longest-time window setting (EPI24) achieved the best performance with largest reduction in RMSE up to 15.65%. Furthermore, even when used in combination with another fundamental TC structural indicator, TC fullness, EPI continued to provide additional benefits, reducing RMSE by 9.79% for intense TCs and 3.77% overall. Additionally, this reinforcement through EPI is applicable to other deep learning models, offering clear practical value for real-time TC intensity estimation.

The current framework relies exclusively on IR imagery. Future work will investigate model fine-tuning with multi-source and higher-resolution satellite inputs to enhance generalizability across basins and observation systems. This is particularly relevant for challenging cases such as pinhole-eye TCs and episodes of rapid

Table 4
Comparison of Final Model Accuracy With Existing TC Intensity Estimation Methods

Method	TC years	TC basin	RMSE (kt)	Input	Reference
Ours[a]	2015–2016	WNP	8.95	IR, TC fullness, EPI	This study
ADT[a]	2015–2016	WNP	12.24	IR, visible, PMW imagery	Olander & Velden (2019)
SATCON[a]	2015–2016	WNP	9.97	IR, visible, PMW imagery	Velden & Herndon (2020)
CNN	1999–2014	Atlantic and Pacific	10.18	IR	TABLE IX, Pradhan et al., 2018
DeepMicroNet	2007, 2012	NA and EP	10.6	IR, WV, PMW rain rate	Table 5, Wimmers et al., 2019
CNN-TC	2015–2016	Atlantic and EP	8.79	IR, PMW	Table 3, B.-F. Chen et al., 2019
D-PRINT[b]	2019–2021	WNP	8.4	IR, 27 ESFs[c]	Table 9, Griffin et al., 2024
D-MINT[b]	2019–2021	WNP	7.2	IR, MW, 27 ESFs[c]	Table 9, Griffin et al., 2024

Note. WNP refers to the Western North Pacific, NA to the North Atlantic, EP to the Eastern North Pacific, PMW to passive microwave, MW to microwave, and WV to water vapor. [a]Methods have RMSE values calculated on the 2,363 test samples used in this study to enable direct comparison. [b]Methods also have RMSE values calculated from their homogeneous sample. RMSE values for the other methods are taken directly from the published literature without recalculation. [c]"27 ESFs" refers to the 27 environmental scalar features.

intensity change, where structural signals are subtle yet crucial. Moreover, the derived EPI could be extended beyond analysis to support real-time intensity forecasting, offering additional guidance for short-term prediction. With further development, DeepTCEye itself may also evolve into a predictive model for eye formation and evolution, contributing to earlier detection of intensification signals. Overall, this study demonstrates that explicitly incorporating eye information, which is closely tied to intense TCs, can substantially alleviate the underestimation caused by insufficient training data. This underscores the importance of introducing physically meaningful signals strongly associated with sparse cases to improve their representation. For extreme weather systems where in situ measurements are sparse, such physics-augmented artificial intelligence based on satellite data offers a promising and practical direction for improving operational forecasting and early warning capabilities.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

The HURSAT-B1 data set used for training and evaluation of deep learning models is available from Knapp and Kossin (2007a, 2007b), provided by the NOAA National Centers for Environmental Information (NCEI). ARCHER-based TC eye parameters used for labeling are included in the HURSAT-B1 data set and can be directly accessed therein, with the original ARCHER output data also available from Wimmers and Velden (2016a, 2016b), provided by the Cooperative Institute for Meteorological Satellite Studies (CIMSS), University of Wisconsin–Madison. Post-2016 IR imagery was obtained from the GridSat-B1 data set (Knapp et al., 2011), provided by NOAA NCEI. The IBTrACS-WMO v4 data set was provided by Knapp et al. (2010a, 2010b), archived at NOAA NCEI. The code and data set supporting this study have been openly published on Zenodo (Liu et al., 2026). As the implementation is adapted from Zhuo and Tan (2021), users are requested to cite Zhuo and Tan (2021) when using the code.

Acknowledgments

The authors acknowledge Chris Velden and the two anonymous reviewers for valuable feedback that greatly improved this study. This study is jointly supported by the National Natural Science Foundation of China under Grants 42192553, 42475004, and 423B1005, as well as the Nanjing University Integrated Research Platform of the Ministry of Education—Top Talents Program (2024300460). The development of the deep learning framework in this study is based on code and methodologies introduced by Zhuo and Tan (2021). The satellite data sets employed, including HURSAT, GridSat-B1, and ARCHER, are provided by the National Centers for Environmental Information (NCEI) and the Cooperative Institute for Meteorological Satellite Studies (CIMSS) at the University of Wisconsin–Madison. The authors express sincere gratitude to these institutions for making the data publicly accessible.

References

Altmann, A., Toloşi, L., Sander, O., & Lengauer, T. (2010). Permutation importance: A corrected feature importance measure. *Bioinformatics*, *26*(10), 1340–1347. https://doi.org/10.1093/bioinformatics/btq134

Bessho, K., Andou, A., Sumida, Y., & Abe, M. (2025). Himawari-10 project overview. *SOLA*, *21*, 251–257. https://doi.org/10.2151/sola.2025-031

Bessho, K., Date, K., Hayashi, M., Ikeda, A., Imai, T., Inoue, H., et al. (2016). An introduction to Himawari-8/9— Japan's new-generation geostationary meteorological satellites. *Journal of the Meteorological Society of Japan. Series II*, *94*(2), 151–183. https://doi.org/10.2151/jmsj.2016-009

Chen, B., Chen, B.-F., & Chen, Y.-N. (2021). Real-time tropical cyclone intensity estimation by handling temporally heterogeneous satellite data. *Proceedings of the AAAI Conference on Artificial Intelligence*, *35*(17), 14721–14728. https://doi.org/10.1609/aaai.v35i17.17729

Chen, B.-F., Chen, B., Lin, H.-T., & Elsberry, R. L. (2019). Estimating tropical cyclone intensity by satellite imagery utilizing convolutional neural networks. *Weather and Forecasting*, *34*(2), 447–465. https://doi.org/10.1175/WAF-D-18-0136.1

Choo, M., Kim, Y., Lee, J., Im, J., & Moon, I. J. (2024). Bridging satellite missions: Deep transfer learning for enhanced tropical cyclone intensity estimation. *GIScience and Remote Sensing*, *61*(1), 2325720. https://doi.org/10.1080/15481603.2024.2325720

Chu, K., & Tan, Z.-M. (2014). Annular typhoons in the eastern North Pacific. *Weather and Forecasting*, *29*(2), 241–251. https://doi.org/10.1175/WAF-D-13-00060.1

DeMaria, R. T., Chirokova, G., Knaff, J. A., & Dostalek, J. F. (2015). Machine learning algorithms for tropical cyclone center fixing and eye detection. *Paper presented at the 95th Annual Meeting of the American Meteorological Society, Phoenix, AZ, USA*.

Duong, Q.-P., Wimmers, A., Herndon, D., Tan, Z.-M., Zhuo, J.-Y., Knaff, J., et al. (2023). Objective satellite methods including AI algorithms reviewed for the tenth international workshop on tropical cyclones (IWTC-10). *Tropical Cyclone Research and Review*, *12*(4), 259–266. https://doi.org/10.1016/j.tcrr.2023.11.001

Dvorak, V. F. (1975). Tropical cyclone intensity analysis and forecasting from satellite imagery. *Monthly Weather Review*, *103*(5), 420–430. https://doi.org/10.1175/1520-0493(1975)103<0420:TCIAAF>2.0.CO;2

Dvorak, V. F. (1984). Tropical cyclone intensity analysis using satellite data. *NOAA Technical Report NESDIS*, *11*. https://repository.library.noaa.gov/view/noaa/19322

Griffin, S. M., Wimmers, A., & Velden, C. S. (2022). Predicting rapid intensification in North Atlantic and eastern North Pacific tropical cyclones using a convolutional neural network. *Weather and Forecasting*, *37*(8), 1333–1355. https://doi.org/10.1175/WAF-D-21-0194.1

Griffin, S. M., Wimmers, A., & Velden, C. S. (2024). Predicting short-term intensity change in tropical cyclones using a convolutional neural network. *Weather and Forecasting*, *39*(1), 177–202. https://doi.org/10.1175/WAF-D-23-0085.1

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G.-Z. (2019). XAI—Explainable artificial intelligence. *Science Robotics*, *4*(37), eaay7120. https://doi.org/10.1126/scirobotics.aay7120

Guo, X., & Tan, Z.-M. (2017). Tropical cyclone fullness: A new concept for interpreting storm intensity. *Geophysical Research Letters*, *44*(9), 4324–4331. https://doi.org/10.1002/2017GL073680

Guo, X., & Tan, Z.-M. (2022). Tropical cyclone intensification and fullness: The role of storm size configuration. *Geophysical Research Letters*, *49*(16), e2022GL098449. https://doi.org/10.1029/2022GL098449

Hendricks, E. A. (2012). Internal dynamical control on tropical cyclone intensity variability – A review. *Tropical Cyclone Research and Review*, *1*(1), 97–105. https://doi.org/10.6057/2012TCRR01.11

Higa, M., Tanahara, S., Adachi, Y., Ishiki, N., Nakama, S., Yamada, H., et al. (2021). Domain knowledge integration into deep learning for typhoon intensity classification. *Scientific Reports*, *11*(1), 12972. https://doi.org/10.1038/s41598-021-92286-w

Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. *Nature Reviews Physics*, *3*(6), 422–440. https://doi.org/10.1038/s42254-021-00314-5

Kimball, S. K., & Mulekar, M. S. (2004). A 15-year climatology of North Atlantic tropical cyclones. Part I: Size parameters. *Journal of Climate*, *17*(18), 3555–3575. https://doi.org/10.1175/1520-0442(2004)017<3555:AYCONA>2.0.CO;2

Knaff, J. A., Cram, T. A., Schumacher, A. B., Kossin, J. P., & DeMaria, M. (2008). Objective identification of annular hurricanes. *Weather and Forecasting*, *23*(1), 17–28. https://doi.org/10.1175/2007WAF2007031.1

Knaff, J. A., & DeMaria, R. T. (2017). Forecasting tropical cyclone eye formation and dissipation in infrared imagery. *Weather and Forecasting*, *32*(6), 2103–2116. https://doi.org/10.1175/WAF-D-17-0037.1

Knaff, J. A., Kossin, J. P., & DeMaria, M. (2003). Annular hurricanes. *Weather and Forecasting*, *18*(2), 204–223. https://doi.org/10.1175/1520-0434(2003)018<0204:AH>2.0.CO;2

Knapp, K. R., Ansari, S., Bain, C. L., Bourassa, M. A., Dickinson, M. J., Funk, C., et al. (2011). Globally gridded satellite observations for climate studies [Dataset]. *Bulletin of the American Meteorological Society*, *92*(7), 893–907. https://doi.org/10.1175/2011BAMS3039.1

Knapp, K. R., & Kossin, J. P. (2007a). Hurricane satellite (HURSAT-B1), version 6 [Dataset]. *NOAA National Centers for Environmental Information*. https://doi.org/10.7289/V55X274D

Knapp, K. R., & Kossin, J. P. (2007b). New global tropical cyclone data set from ISCCP B1 geostationary satellite observations. *Journal of Applied Remote Sensing*, *1*(1), 13505. https://doi.org/10.1117/1.2712816

Knapp, K. R., Kruk, M. C., Levinson, D. H., Diamond, H. J., & Neumann, C. J. (2010a). International best track archive for climate stewardship (IBTrACS), version 4r01 [Dataset]. *NOAA National Centers for Environmental Information*. https://doi.org/10.25921/82ty-9e16

Knapp, K. R., Kruk, M. C., Levinson, D. H., Diamond, H. J., & Neumann, C. J. (2010b). The international best track archive for climate stewardship (IBTrACS) unifying tropical cyclone data. *Bulletin of the American Meteorological Society*, *91*(3), 363–376. https://doi.org/10.1175/2009BAMS2755.1

Knapp, K. R., Matthews, J. L., Kossin, J. P., & Hennon, C. C. (2016). Identification of tropical cyclone storm types using crowdsourcing. *Monthly Weather Review*, *144*(10), 3783–3798. https://doi.org/10.1175/MWR-D-16-0022.1

Knapp, K. R., Velden, C. S., & Wimmers, A. J. (2018). A global climatology of tropical cyclone eyes. *Monthly Weather Review*, *146*(7), 2089–2101. https://doi.org/10.1175/MWR-D-17-0343.1

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Li, W., Li, Y., Zhou, Y., Song, D., Zhang, J., Wei, Z., & Liu, A. A. (2025). Knowledge-guided prompt learning for tropical cyclone intensity estimation. *IEEE Transactions on Geoscience and Remote Sensing*, *63*, 4102113. https://doi.org/10.1109/TGRS.2025.3539823

Lindsey, D. T., Heidinger, A. K., Sullivan, P. C., McCorkel, J., Schmit, T. J., Tomlinson, M., et al. (2024). GeoXO: NOAA's future geostationary satellite system. *Bulletin of the American Meteorological Society*, *105*(3), E660–E679. https://doi.org/10.1175/BAMS-D-23-0048.1

Liu, Y., Zhuo, J.-Y., Chu, K., & Tan, Z.-M. (2026). Dataset for detection of eye occurrence in sequential satellite infrared imagery and its application to improve deep learning-based tropical cyclone intensity estimation [Dataset]. *Zenodo*. https://doi.org/10.5281/zenodo.18169468

Meng, C., Seo, S., Cao, D., Griesemer, S., & Liu, Y. (2022). When physics meets machine learning: A survey of physics-informed machine learning. *arXiv*. https://doi.org/10.48550/arXiv.2203.16797

Olander, T. L., & Velden, C. S. (2019). The advanced dvorak technique (ADT) for estimating tropical cyclone intensity: Update and new capabilities. *Weather and Forecasting*, *34*(4), 905–922. https://doi.org/10.1175/WAF-D-19-0007.1

Pradhan, R., Aygun, R. S., Maskey, M., Ramachandran, R., & Cecil, D. J. (2018). Tropical cyclone intensity estimation using a deep convolutional neural network. *IEEE Transactions on Image Processing*, *27*(2), 692–702. https://doi.org/10.1109/TIP.2017.2766358

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *arXiv*. https://doi.org/10.48550/arXiv.1409.1556

Tan, J., Yang, Q., Hu, J., Huang, Q., & Chen, S. (2022). Tropical cyclone intensity estimation using Himawari-8 satellite cloud products and deep learning. *Remote Sensing*, *14*(4), 812. https://doi.org/10.3390/rs14040812

Tian, W., Lai, L., Niu, X., Zhou, X., Zhang, Y., & Kenny, L. K. S. T. C. (2023). Estimation of tropical cyclone intensity using multi-platform remote sensing and deep learning with environmental field information. *Remote Sensing*, *15*(8), 2085. https://doi.org/10.3390/rs15082085

Tian, Y., Zhou, W., Cheung, P. K. Y., Zhang, X., & Wang, Y. (2025). Vision transformer for extracting tropical cyclone intensity from satellite images. *Advances in Atmospheric Sciences*, *42*(1), 79–93. https://doi.org/10.1007/s00376-024-3191-1

Tong, B., Fu, J., Deng, Y., Huang, Y., Chan, P., & He, Y. (2023). Estimation of tropical cyclone intensity via deep learning techniques from satellite cloud images. *Remote Sensing*, *15*(17), 4188. https://doi.org/10.3390/rs15174188

Tsujino, S., Horinouchi, T., Tsukada, T., Kuo, H.-C., Yamada, H., & Tsuboki, K. (2021). Inner-core wind field in a concentric eyewall replacement of typhoon trami (2018): A quantitative analysis based on the Himawari-8 satellite. *Journal of Geophysical Research: Atmospheres*, *126*(7), e2020JD034434. https://doi.org/10.1029/2020JD034434

Tsukada, T., & Horinouchi, T. (2020). Estimation of the tangential winds and asymmetric structures in typhoon inner core region using Himawari-8. *Geophysical Research Letters*, *47*(11), e2020GL087637. https://doi.org/10.1029/2020GL087637

Tsukada, T., & Horinouchi, T. (2023). Strong relationship between eye radius and radius of maximum wind of tropical cyclones. *Monthly Weather Review*, *151*(2), 569–588. https://doi.org/10.1175/MWR-D-22-0106.1

Tsukada, T., Horinouchi, T., & Tsujino, S. (2024). Wind distribution in the eye of tropical cyclone revealed by a novel atmospheric motion vector derivation. *Journal of Geophysical Research: Atmospheres*, *129*(9), e2023JD040585. https://doi.org/10.1029/2023JD040585

Velden, C., Harper, B., Wells, F., Beven, J. L., Zehr, R., Olander, T., et al. (2006). The dvorak tropical cyclone intensity estimation technique: A satellite-based method that has endured for over 30 years. *Bulletin of the American Meteorological Society*, *87*(9), 1195–1210. https://doi.org/10.1175/BAMS-87-9-1195

Velden, C., Olander, T., Herndon, D., & Kossin, J. P. (2017). Reprocessing the most intense historical tropical cyclones in the satellite era using the advanced dvorak technique. *Monthly Weather Review*, *145*(3), 971–983. https://doi.org/10.1175/MWR-D-16-0312.1

Velden, C. S., & Herndon, D. (2020). A consensus approach for estimating tropical cyclone intensity from meteorological satellites: Satcon. *Weather and Forecasting*, *35*(4), 1645–1662. https://doi.org/10.1175/WAF-D-20-0015.1

Vigh, J. L., Knaff, J. A., & Schubert, W. H. (2012). A climatology of hurricane eye formation. *Monthly Weather Review*, *140*(5), 1405–1426. https://doi.org/10.1175/MWR-D-11-00108.1

Wang, C., Zheng, G., Li, X., Xu, Q., Liu, B., & Zhang, J. (2022). Tropical cyclone intensity estimation from geostationary satellite imagery using deep convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1–16. https://doi.org/10.1109/TGRS.2021.3066299

Weatherford, C. L., & Gray, W. M. (1988). Typhoon structure as revealed by aircraft reconnaissance. Part I: Data analysis and climatology. *Monthly Weather Review*, *116*(5), 1032–1043. https://doi.org/10.1175/1520-0493(1988)116<1032:TSARBA>2.0.CO;2

Willoughby, H. E. (1998). Tropical cyclone eye thermodynamics. *Monthly Weather Review*, *126*(12), 3053–3067. https://doi.org/10.1175/1520-0493(1998)126<3053:TCET>2.0.CO;2

Wimmers, A., Velden, C., & Cossuth, J. H. (2019). Using deep learning to estimate tropical cyclone intensity from satellite passive microwave imagery. *Monthly Weather Review*, *147*(6), 2261–2282. https://doi.org/10.1175/MWR-D-18-0391.1

Wimmers, A. J., & Velden, C. S. (2010). Objectively determining the rotational center of tropical cyclones in passive microwave satellite imagery. *Journal of Applied Meteorology and Climatology*, *49*(9), 2013–2034. https://doi.org/10.1175/2010JAMC2490.1

Wimmers, A. J., & Velden, C. S. (2016a). Advancements in objective multisatellite tropical cyclone center fixing. *Journal of Applied Meteorology and Climatology*, *55*(1), 197–212. https://doi.org/10.1175/JAMC-D-15-0098.1

Wimmers, A. J., & Velden, C. S. (2016b). Automated rotational center hurricane eye retrieval (ARCHER) [Dataset]. *Cooperative Institute for Meteorological Satellite Studies (CIMSS)*. http://tropic.ssec.wisc.edu/real-time/archerOnline/cyclones/

Wood, K. M., & Ritchie, E. A. (2015). A definition for rapid weakening of North Atlantic and eastern North Pacific tropical cyclones. *Geophysical Research Letters*, *42*(22), 10091–10097. https://doi.org/10.1002/2015GL066697

Xu, G., Li, Y., Ma, C., Li, X., Ye, Y., Lin, Q., et al. (2023). TFG-Net: Tropical cyclone intensity estimation from a fine-grained perspective with the graph convolution neural network. *Engineering Applications of Artificial Intelligence*, *118*, 105673. https://doi.org/10.1016/j.engappai.2022.105673

Yang, W., Fei, J., Huang, X., Ding, J., & Cheng, X. (2024). Enhancing tropical cyclone intensity estimation from satellite imagery through deep learning techniques. *Journal of Meteorological Research*, *38*(4), 652–663. https://doi.org/10.1007/s13351-024-3186-y

Zhang, C.-J., Wang, Y., Lu, X.-Q., & Sun, F.-Y. (2024). Rapid weakening tropical cyclone intensity estimation based on deep learning using infrared satellite images and reanalysis manuscript data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *17*, 17598–17611. https://doi.org/10.1109/JSTARS.2024.3465829

Zhang, C.-J., Wang, Y., Ma, L.-M., & Lu, X.-Q. (2024). TCIC-SC-RNet: A hybrid model for TC intensity estimation based on transfer learning. *Weather and Forecasting*, *39*(10), 1469–1485. https://doi.org/10.1175/WAF-D-23-0080.1

Zhang, Z., Yang, X., Wang, X., Wang, B., Wang, C., & Du, Z. (2022). A neural network with spatiotemporal encoding module for tropical cyclone intensity estimation from infrared satellite image. *Knowledge-Based Systems*, *258*, 110005. https://doi.org/10.1016/j.knosys.2022.110005

Zhao, L., Chen, Y., & Sheng, V. S. (2020). A real-time typhoon eye detection method based on deep learning for meteorological information forensics. *Journal of Real-Time Image Processing*, *17*(1), 95–102. https://doi.org/10.1007/s11554-019-00899-2

Zhao, Z., Zhang, Z., Tang, P., Wang, X., & Cui, L. (2024). MT-GN: Multi-task-learning-based graph residual network for tropical cyclone intensity estimation. *Remote Sensing*, *16*(2), 215. https://doi.org/10.3390/rs16020215

Zhuo, J.-Y., & Tan, Z.-M. (2021). Physics-augmented deep learning to improve tropical cyclone intensity and size estimation from satellite imagery. *Monthly Weather Review*, *149*(7), 2097–2113. https://doi.org/10.1175/MWR-D-20-0333.1