https://doi.org/10.5194/egusphere-2025-4333
Preprint. Discussion started: 9 December 2025

From Reanalysis to Climatology: Deep Learning Reconstruction of Tropical Cyclogenesis in the Western North Pacific

Duc-Trong Le[1], Tran-Binh Dang[1], Anh-Duc Hoang Gia[1], Duc-Hai Nguyen[1], Minh-Hoa Tien[1], Xuan-Truong Ngo[1], Quang-Trung Luu[3], Quang-Lap Luu[4], Tai-Hung Nguyen[4], Thanh T. N. Nguyen[1], and Chanh Kieu[2]

[1]Faculty of Information Technology, VNU University of Engineering and Technology, Hanoi, Vietnam
[2]Department of Earth and Atmospheric Sciences, Indiana University, Bloomington, IN 47405, USA
[3]Université Paris-Saclay - CNRS - CentraleSupélec - L2S, Gif-sur-Yvette, 91192, France
[4]School of Electrical and Electronic Engineering, Hanoi University of Science and Technology, Vietnam

Correspondence: Duc-Trong Le (trongld@vnu.edu.vn) and Chanh Kieu (ckieu@iu.edu)

Abstract. Tropical cyclogenesis (TCG) climatology is the key to understanding regional weather extremes and long-term cyclone risk, yet their large-scale environmental drivers remain difficult to characterize from observations or traditional physical-based modelling. In this study, we develop a deep learning (DL) framework, TCG-Net, based on an 18-layer residual convolutional neural network (ResNet-18) to reconstruct TCG climatology in the western North Pacific (WNP) basin from NASA's Modern-Era Retrospective analysis for Research and Applications Version 2 (MERRA-2). The framework addresses two tasks: the Past Domain (PD) task that predicts when TCG occurs in the WNP within the next 48 hours, and the Dynamic Domain (DD) task that predicts the spatial distribution of TCG at a given date and time. For each task, tailored labeling strategies define different negative samples to distinguish TCG from non-TCG conditions. To enhance the model's capability in handling the rarity of TCG data, temporal feature enrichment is used to incorporate environmental information from preceding 6-hour time steps, which helps improve the representation of each training task. In addition, random under-sampling (RUS) is applied with class weighting to address the severe imbalance caused by large numbers of negative TCG samples under these labeling strategies. With a training dataset from 1980-2016 and an independent set from 2017–2022, TCG-Net achieves an overall F1-score of 0.39 for the PD task and 0.33 for the DD task. In the PD task, feature selection experiments reveal that only a subset of environmental variables including vertical wind shear, low- to mid-level humidity, and mid-level vertical motion is required for robust performance, consistent with prior physical studies. In contrast, for the DD task, full-feature models perform better, likely due to their ability to exploit unknown or latent feature interactions. Both tasks reproduce key characteristics of the observed seasonality and spatial TCG distribution when evaluated against the best-track dataset. These results demonstrate that DL-based reconstructions, when coupled with task-specific labeling, temporal enrichment, and imbalance-aware training, can complement physics-based simulations and vortex-tracking algorithms and provide an efficient pathway for downscaling or projecting TCG climatology from coarse-resolution climate model outputs.

https://doi.org/10.5194/egusphere-2025-4333
Preprint. Discussion started: 9 December 2025

1 Introduction

The western North Pacific (WNP) basin has been well-documented to be the most active area of tropical cyclone (TC) activities (Kossin et al., 2016; Peduzzi et al., 2012; Tran-Quang et al., 2020). With favorable conditions for TC formation (also known as tropical cyclogenesis, or TCG) such as warmer sea surface temperature (SST), active monsoon trough formation, or frequent convectively coupled equatorial waves, WNP produces about 25-30 TCs annually with a quarter of that affecting Vietnam coastal regions (Defforge and Merlis, 2017; Trinh et al., 2021; Thanh et al., 2020). From the climate perspective, any change in TC main characteristics such as frequency, genesis locations, or intensity is often considered to be a manifestation of climate change. Thus, developing effective methods to construct TC climatology from different climate datasets is of significant importance for studying future projections of TC climatology from climate model outputs or outlooks from global forecasting systems (e.g., Knutson et al., 1998; Walsh et al., 2007; Bengtsson et al., 2007; Lee et al., 2020; Cha et al., 2020; Camargo et al., 2023; Kieu et al., 2023).

Traditionally, creating a TC climatology from gridded climate output involves using a vortex tracking algorithm to detect TC centers. This algorithm relies on a set of TC characteristics, such as absolute vorticity, surface maximum wind, minimum central pressure, warm core, or lifetime, which are checked at each model grid point (Walsh et al., 2015; Zhao et al., 2009; Strachan et al., 2013; Camargo and Zebiak, 2002; Horn et al., 2014; Zarzycki and Ullrich, 2017; Ullrich and Zarzycki, 2017; Vu et al., 2021). While being effective for high-resolution model outputs where TC characteristics are well-captured and thus suitable for weather forecast models or well-developed TCs, vortex tracking methods face challenges with coarse-resolution climate models (>0.5°). For these coarse-resolution models or datasets, TC characteristics, especially during early formation, are often unclear (Zarzycki and Ullrich, 2017; Tien et al., 2020). Consequently, directly tracking a vortex from such outputs can lead to uncertainties in the timing and location of early TCG. This issue becomes more serious when studying climate change aspects like shifts in TCG location or timing across different climate datasets (e.g., Tan et al., 2015; Tien et al., 2020).

The rapid advancement of machine learning (ML) techniques has opened new avenues for atmospheric research as well as operational forecasting. Given the vast amount of observational and model-generated data, weather and climate systems naturally provide a "big data" platform that are well-suited for training ML models not only for short-term weather prediction but also for capturing long-term climate variability (e.g., Schultz et al., 2021; Pathak et al., 2022; Bi et al., 2023; Lam et al., 2023; Nguyen and Kieu, 2024). In fact, several private and governmental organizations have recently developed deep learning architectures that outperform traditional physics-based models in weather forecasting, as demonstrated by Pathak et al. (2022) and Lam et al. (2023). Among recent applications of ML to TC research, most efforts have been limited to short-term contexts such as weather forecasting, satellite retrieval, or diagnostic studies. For examples, Miller et al. (2017a); Wimmers et al. (2019) developed a deep learning (DL) model with satellite data to train a convolutional neural network, which can categorize TC intensity based on different cloud patterns and satellite channels. This line of approach has been further advanced to help improve TC forecasts by integrating the tracking information and/or other reanalysis data, with some modest performance for nowcasting and diagnoses (e.g., Gao et al., 2018a; Kim et al., 2019b; Chen et al., 2020; Giffard-Roisin et al., 2020).

https://doi.org/10.5194/egusphere-2025-4333
Preprint. Discussion started: 9 December 2025

Focusing specifically on TCG from climate data, Nguyen and Kieu (2024, hereinafter NK2024) recently evaluated several DL architectures and identified promising potential for early warning of TCG events in the WNP basin. By treating climate reanalysis data as multi-channel input, they formulated TCG detection as a classification problem and demonstrated reasonable forecast skill, even at a coarse spatial resolution of $1^\circ \times 1^\circ$. While the model's predictive skill declines with increasing forecast lead time, their approach highlights an important capability of detecting TCG directly from climate model outputs. In particular, certain DL architectures can be used to identify both the location and timing of TCG events within a given domain, which is valuable for broader applications in TCG climatology or future projections. It is important to emphasize here the distinction between predicting and detecting TCG. While long-lead TCG prediction suffers from rapidly decreasing skill due to the inherent limits of predictability in tropical dynamics, detecting TCG from coarse-resolution climate data (also commonly known as downscaling) proves to be much more skillful and feasible. This is because detection relies primarily on identifying favorable environmental conditions at the time of genesis, making it a problem of climate downscaling rather than one constrained by the chaotic dynamics of the atmosphere.

Although existing ML applications for TC research show some promises, they have focused so far mostly on short-term prediction or nowcasting, rather than on TC climatology. Specifically, the use of ML for constructing TC climatology remains relatively preliminary,(e.g., Scher and Messori, 2019; Wang et al., 2022; Chen and Yuan, 2024). Key challenges include how to leverage ML to diagnose climate model outputs, identify and correct model biases, or construct robust statistics of large-scale climate features. Thus, the development of ML-based techniques for downscaling of TC climatology or distributions of any other extreme events from reanalysis data is largely unexplored, yet represents an important open direction for future applications of DL for climate research. Given the rapid advancements in DL techniques, the main objective of this study is to introduce a framework for reconstructing TCG climatology from climate reanalysis datasets. Specifically, we will extract key TCG characteristics such as frequency or spatial distribution from the NASA Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2), instead of forecasting TCG as in NK2024. Our DL-derived TCG climatology can serve as an independent validation and complement to those obtained from traditional vortex tracking methods. Furthermore, any TC climatology obtained from MERRA-2 data can also be used as a reference for examining the change of TC climatology in future projections, which justifies our DL approach herein.

The rest of this work is organized as follows. In Section 2, details of data pre-processing and our CNN algorithms are presented. An approach to generate and label TCG binary dataset for each application will also be discussed. Section 3 presents the detailed design of our DL pipeline in this study, along with DL experimental designs. Section 4 provides results and related discussions. Finally, a summary and concluding remarks are given in Section 5.

2 Methodology and Data

2.1 Input data

To make our work directly applicable to research in TC climate downscaling or projection, the MERRA-2 dataset (Gelaro et al., 2017) was used in this study. This dataset is an atmospheric reanalysis based on the Goddard Earth Observing System Model

https://doi.org/10.5194/egusphere-2025-4333
Preprint. Discussion started: 9 December 2025

Variable	MERRA-2	IBTrACS
LAT MIN	-90	-90
LAT MAX	90	90
LAT STEP	0.5	0.01
LON MIN	-180	0
LON MAX	180	360
LON STEP	0.625	0.01
TIME MIN	01/01/1980	01/01/1890
TIME MAX	12/31/2022	12/31/2022
TIME STEP	3 hours	6 or 3 hours
File format	NetCDF4	Csv
File split	By day	*No splitting*
File size	2.2–2.3 GB	306 MB

Table 1. A list of variables and their corresponding ranges in the MERRA-2 and IBTrACS datasets used for developing DL models.

(GEOS-5, Version 5) data assimilation system (Gelaro et al., 2017). Unlike the original MERRA, MERRA-2 employed a newer version of GEOS-5 that assimilated newer microwave sounders and infrared radiance, as well as other data types. In particular, all data collections from MERRA-2 are provided on the same horizontal grid, which has 576×361 points in the longitudinal and latitudinal direction, respectively (a resolution of $0.625 \times 0.5^\circ$ longitude-by-latitude grid), and interpolated to 42 standard pressure vertical levels. While the output collections of MERRA-2 are on the regular $0.625 \times 0.5^\circ$ that are relatively coarse for TC inner-core region, our focus in this study is on the climatology of TCG, which requires mostly environmental conditions at the meso to synoptic-scale. As such, 0.5-degree resolution data is sufficient for our purposes as discussed in NK2024.

Although several reanalysis datasets are available, MERRA-2 was selected for this study primarily due to two reasons: (i) its spatial and temporal resolution is suitable for detecting TCG, and (ii) its data format is convenient for ML model development. Unlike some TC metrics such as intensity or accumulated energy that require detailed inner-core structure, TCG is a process that is largely governed by environmental conditions. With a resolution of 0.5° MERRA-2 can therefore capture these environmental conditions effectively for DL purposes. Note that the MERRA-2 dataset provides gridded atmospheric data that includes 11 key meteorological variables at standard pressure levels, spanning the global domain from 90° S to 90° N latitude (at 0.5° resolution) and from 180° W to 180° E longitude (at 0.625° resolution). The dataset is available from January 1, 1980, to December 31, 2022, with data sampled every 3 hours. Each daily file contains 8 time slices and is stored in NETCDF4 format, with a file size of approximately 2.2–2.3 gigabytes. In total, the full dataset amounts to roughly 25 terabytes, offering a sufficiently large and detailed input for training DL models. One limitation of the MERRA-2 dataset as compared to other reanalysis datasets is that this dataset contains a single resolution of $0.5 \times 0.625^\circ$, while other reanalysis datasets such as ERA5 provide higher resolution up to 0.25° at an hourly interval. Using such higher resolution datasets is certainly an advantage, as it can help optimize ML models. However, most current global climate projection outputs are given on 0.5° to 1°

https://doi.org/10.5194/egusphere-2025-4333
Preprint. Discussion started: 9 December 2025

resolutions. Thus, using 0.5° data can help better demonstrate the usefulness and facilitate the applications of our DL models in reconstructing TC climatology from global climate outputs as designed. Yet, ERA5 is considered to be among the best for climate reanalysis and DL model development, whether ERA5 is better than MERRA-2 in terms of TC climatology at the 0.5° resolution has not been demonstrated. In this regard, our choice of MERRA-2 can be considered as a pre-learning step, which can be refined with ERA5 or any other climate datasets in the future if needed. For the purpose of implementing and evaluating our DL model, the MERRA-2 data is therefore sufficient.

Along with the use of MERRA-2dataset for training, we also employed the International Best Track Archive for Climate Stewardship (IBTrACS) (Knapp et al., 2010) to label all TCG events and locations, which contains global TC records. In the WNP basin, IBTrACS contains data from January 1, 1890 through present day, sampled every 3 hours and archived in a single CSV file with a total size of approximately 300 megabytes. The fact that this IBTrACS dataset is structured with the same synoptic times as the MERRA-2 dataset is of importance, because it allows us to pair these two datasets for supervised DL. The details of these datasets as well as their corresponding variables are summarized in Table 1.

2.2 Data pre-processing

As Table 1, MERRA-2 and IBTrACS datasets have their own format, structure, and parameters, it is necessary to first synchronize these datasets before any DL model development can be carried out. For this data pre-process step, we convert the longitude axis of MERRA-2 from $[-180; +180]$ to $[0; 360]$ to match with the coordinate definitions in IBTrACS, which is needed so that the `latitude` and `longitude` values of IBTrACS can be located properly on the MERRA-2 coordinate grids for data extraction. In our study, a domain of a size $[35^\circ S - 70^\circ N] \times [60^\circ - 220^\circ]$ in the Pacific Ocean is then extracted from the global MERRA-2 dataset. For each MERRA-2 file, a timestamp is output at an interval of 6 hours to match with the best track data (note that the original MERRA-2 dataset consists of daily data files at an interval of 3 hours). All details of these data outputs after pre-processing MERRA-2 and IBTrACS are provided in Table 2. The corresponding pre-process workflow is thoroughly described in our Zenodo repository (Le et al., 2025).

2.3 Supervised dataset design

Given the pre-processed data described above, the next step is to generate a labeled dataset for supervised ML models. Specifically for the study of TCG, we require a binary dataset that indicates whether or not a TCG event occurred in the WNP basin needed for the reconstruction of TCG climatology. The process of creating such a binary dataset is critical, as it must include not only positive TCG cases but also a well-designed set of negative samples. Such a requirement of both well-designed positive and negative samples is essential for ML models to effectively learn the distinct features between TCG and non-TCG conditions. On one hand, a strong contrast between positive and negative samples increases the likelihood that ML models will identify key patterns, thereby improving performance. Nevertheless, the selection of positive and negative samples must also align with practical applications and purposes in climate research. In fact, the criteria for labeling positive/negative TCG events vary depending on the specific type of TCG climatology or forecast objective under consideration as will be shown in the Section 4.

https://doi.org/10.5194/egusphere-2025-4333
Preprint. Discussion started: 9 December 2025

	MERRA-2	IBTrACS
LAT MIN	-35	0
LAT MAX	70	60
LAT STEP	0.5	0.01
LON MIN	60	100
LON MAX	220	180
LON STEP	0.625	0.01
TIME MIN	01.01.1980	01.01.1890
TIME MAX	31.12.2022	31.12.2022
TIME STEP	6 hours	6 hours
File format	NetCDF4	CSV
Sample size	By forecasts	1024 records
File size	97MB	4.0KB

Table 2. Pre-processed data format and structures obtained from the MERRA-2 and IBTrACS datasets for ML training data requirement.

In this study, we follow NK2024 and define a positive TCG event as the first time a storm was recorded in the best track data. One could also define a TCG as the first moment that a tropical depression stage is recorded to make sure TCG characteristics are well-defined for DL training (Kieu and Nguyen, 2024). However, our choice of the first time that a TC was recorded in the best track herein has the benefit of training a DL model that can detect TCG earlier, and so we will use this definition to label a positive TCG event.

With this definition of *positive TCG labels*, we can scan through all TC track records and take the first recorded location of each storm to create a target output for a TCG event. Given the typical scale of TCs, the positive TCG domain is chosen to be a square box of size $18 \times 18^\circ$ centered on the first recorded TCG location, which is equivalent to roughly 33×32 grid points with the MERRA-2 0.5° resolution. Finally, all relevant information related to a TCG event including its longitudes, latitudes, date, and time was stored in a csv database to facilitate our data sharing and input to the DL interface. Depending on the context and applications, one can have several different ways to define a *negative TCG event*, as discussed in Kieu and Nguyen (2024). From the practical perspective, we propose in this study two following sampling strategies.

The first strategy for generating negative-labeled TCG samples, referred to as the Past Domain (PD) strategy, uses temporal context to distinguish between positive and negative labels. Specifically, for each positive TCG event occurring at time t, samples within a past window from $t-k$ to t are labeled as positive, capturing the precursors leading up to the cyclone formation. In contrast, all earlier samples, i.e., those from $t-k-1$ and further back in time, are labeled as negative. This approach aims to answer the question of why a TC forms at a specific time but not earlier, as discussed in Kieu and Nguyen (2024). A key advantage of this strategy is that it preserves the geographical location between positive and negative samples while introducing temporal separation. However, depending on the chosen value of k, there may still be some overlap in

https://doi.org/10.5194/egusphere-2025-4333
Preprint. Discussion started: 9 December 2025

NW north-west	N north	NE north-east
W west	P **positive TCG**	E east
SW south-west	S south	SE south-east

Figure 1. Illustration of a TCG data labeling strategy based on the dynamical domain approach, for which a positive TCG label at one location is surrounded by 8 negative TCG labels for the sampling strategy.

favorable environmental conditions between the positive and negative samples. Moreover, as more historical data are labeled as negative, the dataset can become increasingly imbalanced.

The second approach for generating negative TCG data is referred to as the Dynamic Domain (DD) strategy. In this approach, for each positive TCG event at time t, we consider the surrounding spatial domains as negative samples. Specifically, one considers the eight adjacent regions to the positive TCG location, including the north-west, north, north-east, east, south-east, south, south-west, and west directions (see Figure 1). In this strategy, the eight surrounding domains are also shifted back in time, similar to the PD strategy. That is, negative samples are selected from the eight spatially adjacent regions but at earlier times, such as $t-n$ with $n \geq k+1$, to ensure a temporal gap with the positive window. This type of domain selection for negative TCG data helps answer a question of why TCG occurs in one place but not in other places at the same day/time. One could randomly choose one of the eight negative domains to construct a more balanced binary dataset, as in Kieu and Nguyen (2024), or choose all eight domains to increase the sample size. Unlike the PD approach, we note that this DM strategy will have a small chance of including a co-existing TC in the negative samples. This issue can be addressed by simply checking if there is any co-existing TC nearby within the negative TCG domain, and remove this domain or filter a TC out as discussed in Nguyen (2023). In this study, we use a simple approach of discarding all negative TCG domains that have a co-existing TC to avoid complications with changing environmental conditions due to vortex removal processes. This affects $< 1\%$ of all data points as we have in the WNP basin, as most TCG in this basin rarely overlap within a domain of an $18^\circ \times 18^\circ$ area.

As part of constructing the binary TCG dataset described above, it is important to note that TCs can form at any time of day, whereas the IBTrACS dataset provides records at fixed 6-hour intervals. Consequently, the actual genesis location of a TC may differ slightly from the position recorded in the best-track data. This spatial discrepancy typically depends on storm motion and the algorithms used for detecting vortex centers at operational centers. In this study, we assume that the TCG location does not vary significantly within a 6-hour window, which provides a reasonable basis for maintaining consistency in our binary TCG dataset design. This assumption also allows us to use the first recorded time in the IBTrACS dataset to define positively labeled TCG samples, which by convention are assigned a label of 1. All negative samples are labeled as 0.

Regarding undefined (NaN) values in the MERRA-2 dataset, we noticed a common issue of missing data at higher atmospheric levels. These missing values pose some challenges for both data interpolation and ML training. To address this, we

https://doi.org/10.5194/egusphere-2025-4333
Preprint. Discussion started: 9 December 2025

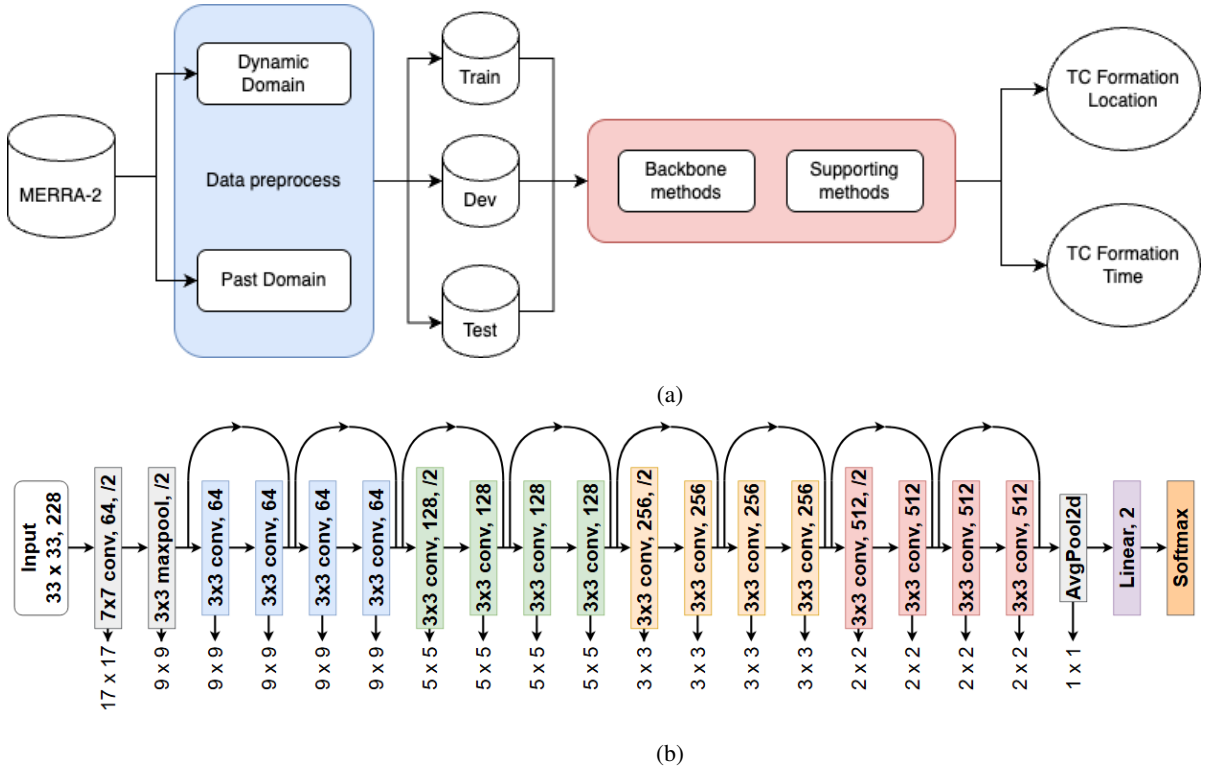

Figure 2. a) The pipeline of the DL design for reconstructing TCG climatology from the MERRA-2 dataset, and (b) the core DL model based on the ResNet-18 architecture used for TCG reconstruction in this study.

exclude all data layers with substantial incompleteness and limit our analysis to layers below the 100-hPa level during the pre-
185 processing stage. The potential impacts of these missing values on model performance will be further evaluated in the machine learning experiments presented in the following section.

3 TCG-Net: Deep Learning Framework

3.1 Model

Recent advances in DL research have demonstrated substantial potential across a range of fields, including image recognition,
190 natural language processing, autonomous driving, and weather prediction (Gao et al., 2018b; Kim et al., 2019a; Giffard-Roisin et al., 2020; Miller et al., 2017b; Su et al., 2020; Bi et al., 2023; Weyn et al., 2021; Nguyen and Kieu, 2024). In the context of TCG applications, several DL approaches using satellite imagery data and TCG predictors have been proposed (Zhang et al., 2015; Park et al., 2016; Matsuoka et al., 2018; Zhang et al., 2019; Kim et al., 2019a). To make further use of climate data in a broader context, NK2024 employed a convolutional neural network (CNN)-based DL framework for TCG application, which

https://doi.org/10.5194/egusphere-2025-4333
Preprint. Discussion started: 9 December 2025

showed promising results despite limited training data. Their study underscores DL's image recognition capabilities, which enable the detection and analysis of relevant environmental patterns from climate meteorological fields for TCG prediction. Given such promising performance of the CNN-based models as well as the features tailored specifically for TCG reported in NK2024, we use a similar DL architecture based on the 18-layer Residual Net (ResNet-18) model in this study. ResNet-18 is a widely used CNN architecture for image recognition tasks. Its key advantage is the use of residual blocks, each consisting of two convolutional layers, a batch normalization layer, and a ReLU activation function, along with a '*shortcut*' connection that directly adds the block's input to its output. This technique helps mitigate the '*vanishing gradient*' problem, which commonly occurs in deep CNN architectures when the number of layers becomes excessively large.

In this study, we adapt the original ResNet-18 architecture for our problem of reconstructing the climatology of TCG. The modified network consists of eight residual blocks, preceded by an initial convolutional layer for input embedding, and followed by a fully connected layer with a softmax activation to predict the probability of storm occurrence, forming a total of 18 layers (Fig. 2b). This architecture differs slightly from that used in NK2024, as our study employs the MERRA-2 dataset, which lacks several environmental variables such as CAPE and tropopause-level features. Consequently, the ResNet-18 model required refinement and testing to achieve optimal performance, which differs somewhat from the architecture tailored for the NCEP Final reanalysis dataset in NK2024. The entire workflow for reading MERRA-2 data to final visualization is shown in Fig. 2a. It is noted that our experiments with alternative DL architectures such as more convolutional layers, vision transformers, or pre-trained models all showed minimal improvement over the performance of the adapted ResNet-18 used in this study. While this observation is obtained from the few models that we have tried, it is possible that the TCG problem has limited predictability or MERRA-2 dataset may contain limited information for TCG at 0.5 degree resolution that more sophisticated models or architectures could not help learn further. This is a known problem in DL training, which explains why a simple CNN model could reach as good a performance as other models (Brigato and Iocchi, 2020). Note also that complex models with more parameters generally have a very high "capacity" to learn. When the training data does not contain a variety of patterns, a high-capacity model can easily memorize the training data, including its noise and idiosyncrasies, instead of learning generalizable features. This leads to good performance on the training set but poor performance on unseen data (validation/test set), which is known as overfitting. Thus, the ResNet-18 model was adapted herein for our TCG reconstruction problem.

3.2 Evaluation metrics

To monitor and verify our DL models, we employ several metrics from traditional classification problems as well as other meteorological metrics specific for TCG climatology such as seasonal or spatial distributions. For the DL model training, the performance of our DL model is based on Precision (P), Recall (R), and F1 score. These metrics are useful due to the inherent class imbalance, where TC occurrences (positive samples) are significantly fewer than non-TC cases (negative samples). By definition, Precision measures the proportion of correctly identified TC instances out of all instances predicted as TC:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{1}$$

https://doi.org/10.5194/egusphere-2025-4333
Preprint. Discussion started: 9 December 2025

Data subset	Period	Number of TCG events
Training	1980 - 2016 (Random 90%)	1117
Validation	1980 - 2016 (Random 10%)	124
Test	2017 - 2023	188

Table 3. The default sampling ratio for training, validation, and test data based on the dataset MERRA-2 on experiments

where TP (True Positives) represents correctly detected TCs, and FP (False Positives) denotes non-TC cases incorrectly classified as TC. A high precision ensures that the model minimizes false alarms, which is critical for reliable early warning systems.

In contrast, Recall quantifies the proportion of actual TC occurrences that the model successfully identifies:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{2}$$

where FN (False Negatives) represents actual TCs that the model fails to detect. A high recall ensures that most TC events are captured, reducing the risk of missing critical storm formations.

To balance take into account both P and R, we use the F1 score defined as:

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{3}$$

A high F1 score indicates a good trade-off between false alarms and missed detections, making it a crucial metric for assessing the reliability of TC detection models. Since missed TCs can lead to severe consequences, while excessive false alarms may reduce trust in predictions, optimizing both Precision and Recall is essential in operational forecasting.

With the above verification metrics, we could train a DL model to maximize the model F1 score performance for any period, which is set by default to be 5 years of data from 2017-2023 for testing and 10% of data from 1980-2016 for validation in this study (see Table 3). Note that these category verification metrics are needed to evaluate the model's performance during training and testing. Whether the model can perform the required task depends further on its performance over other climate evaluations that are specific to the TCG problem such as seasonal distribution or spatial distribution, for which we will present in the next section.

3.3 TCG data imbalance

Given the severe class imbalance of TCG data caused by the limited number of TCG events, i.e., ∼30 events per year in the WNP basin, it is essential to apply some additional techniques to mitigate this imbalance before training a DL model. One approach, as described in Section 2.3, involves temporal data enrichment using past windows, which is mostly suitable for the PD labeling strategy. This same past data window enrichment can also be extended to the DD labeling strategy by using the data from previous cycles to help increase the number of positive labels. Note however that the imbalance ratio for the DD labeling strategy is always 1:8, and so adding more past data would only help increase the overall number of TCG samples but not the imbalance ratio.

https://doi.org/10.5194/egusphere-2025-4333
Preprint. Discussion started: 9 December 2025

In this study, we also introduce a complementary method to address class imbalance, which is the Random UnderSampling (RUS) approach. Specifically, the RUS method controls the ratio between negative and positive TCG samples in the training dataset. For example, a RUS value of 1:4 maintains four negative samples for every positive one. Since this undersampling strategy may still leave a slight imbalance, we further apply class weighting in the loss function to emphasize the minority class during training. This technique increases the loss penalty for misclassifying rare (positive) cases, encouraging the model to better capture them. Two types of class weights are used in this study: the balanced class weight (default) and the proportional class weight, defined as 1/5 RUS. In addition to the temporal data enrichment, the RUS-based sampling and class weighting strategies can improve the performance of our DL models for TCG reconstruction, surpassing the simple hyperparameter optimization based on the F1-score criterion.

3.4 TCG detection

With the output from our ResNet-18 model, we can finally produce a TC probability map for climate reconstruction. Given two different strategies for data labeling based on the DD and PD methods, detecting a TCG event and assigning it to a point in space and time require some specific details that we need to discuss further here. The negative labels are designed differently for different practical purposes.

For the PD strategy, it focuses on the temporal aspect of TCG prediction, which generates negative samples at the same location P as positive samples. Thus, TCG detection for this approach is straightforward, as one can simply choose a location P at time t, obtain data at that same location from previous times $t-i$ to time t, where i represents previous time steps, and then generate TCG probability for that location P. With this strategy, one can obtain TCG probability for any area of interest, so long as the model is trained with the same data frames from earlier times $t-i$ up to time t. In this regard, PD can provide TCG information for any location at a given time t.

With the DD strategy, note that one wants to extract the map information of TCG for the entire WNP basin at any given time t, which explicitly displays the spatial variation of TCG. So, our approach for reconstructing TCG for this DD strategy is to divide the entire WNP basin into a box of $5^\circ \times 5^\circ$, and apply the ResNet-18 model on each box separately. This way, we can detect TCG for all boxes simultaneously at time t. Assuming that a positive TCG event is found when TCG probability is larger than a certain threshold (e.g., 0.5), one can generate a map of all TCG locations at time t as expected.

Because of these different purposes for the PD and DD strategies, our evaluation and interpretation of the model performance for TCG climatology have to be based on separate metrics and criteria as presented in the section 4.

4 Results

4.1 DL model benchmarking

Fig. 3 shows the precision, recall, and F1 score obtained from the PD and DD strategies. It is noted that for the PD strategy, we focus on a fixed domain over a part of the WNP basin whose TC activity has the most influence on Vietnam's coastal region

https://doi.org/10.5194/egusphere-2025-4333
Preprint. Discussion started: 9 December 2025

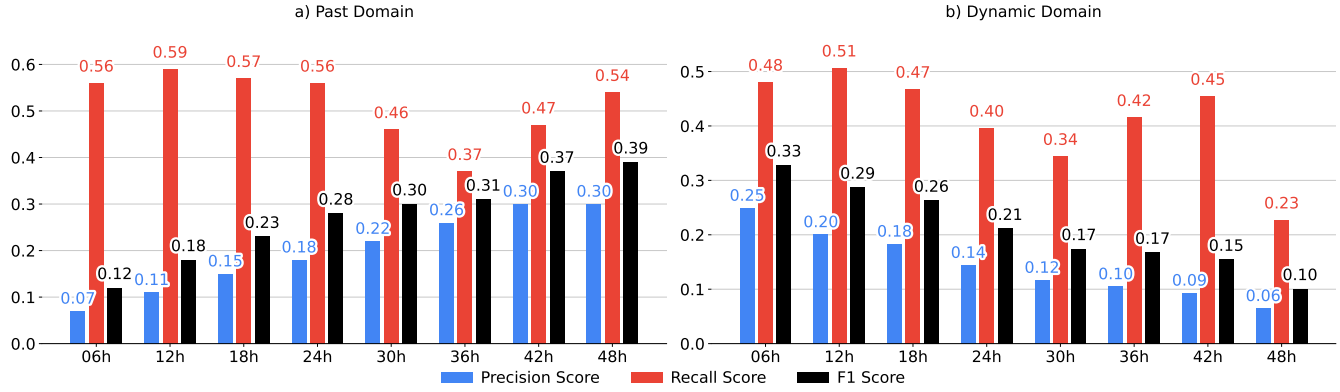

Figure 3. Precision (P, blue columns), Recall (R, red columns), and F1 score (black columns) for the TCG prediction as obtained from the ResNet-18 model with the test set as a function of the temporal data enrichment window at an interval of 6 hours, using a) the PD sampling strategy, and b) the DD sampling strategy.

([0-30°N]-[100-150°W]). This domain choice is flexible and can be adjusted for each specific region in the operational forecast or climatology. Given such an arbitrary choice for the PD strategy, this section will report all results over the fixed domain.

Clearly, the temporal feature enrichment plays a significant role in the performance of the ResNet-18 model. For the PD strategy, a longer period of feature enrichment generally gives a better precision score without much reduction in the recall performance (0.57-0.62), thus allowing for a higher F1 score for longer enrichment windows. Physically, this means that including more past information at one specific location generally helps improve the accuracy of the model performance in capturing TCG. However, the overall P and R scores are relatively low even when including all past information up to 48 hours prior to the formation of a TC. Thus, the F1 score is much less as compared to detecting TCG directly from satellite image data (see, e.g., Chen et al., 2020; Wimmers et al., 2019)

As discussed in Nguyen and Kieu (2024), such a low performance of DL models in detecting TCG from climate reanalysis datasets could reflect the fact that the TCG problem has limited predictability that including more information would not help improve its performance, consistent with the high false alarm rate in physical-based models. Another potential reason for such a low F1 score is the MERRA-2 data itself, which may not cover all possible environmental patterns for rare extreme events like TCG. This highlights the difficulty in reconstructing TCG climatology from climate reanalysis dataset.

For the DD strategy, the dependence of the model performance on the temporal enrichment is somewhat opposite. Specifically, the F1-score is optimal for very short enrichment during a 6 to 12-hour window, and it quickly decreases when the time window is longer. Practically, this means using past information to explain why a TC forms at one place but not at other places will not help if one includes more past information. This result is somewhat expected if one recalls that the main aim of the DD strategy is to search for where a TCG occurs at any given of time. So, the model is trained to detect TCG signals from the spatial information rather than from past information. Enriching the DD strategy by including more past information over

https://doi.org/10.5194/egusphere-2025-4333
Preprint. Discussion started: 9 December 2025

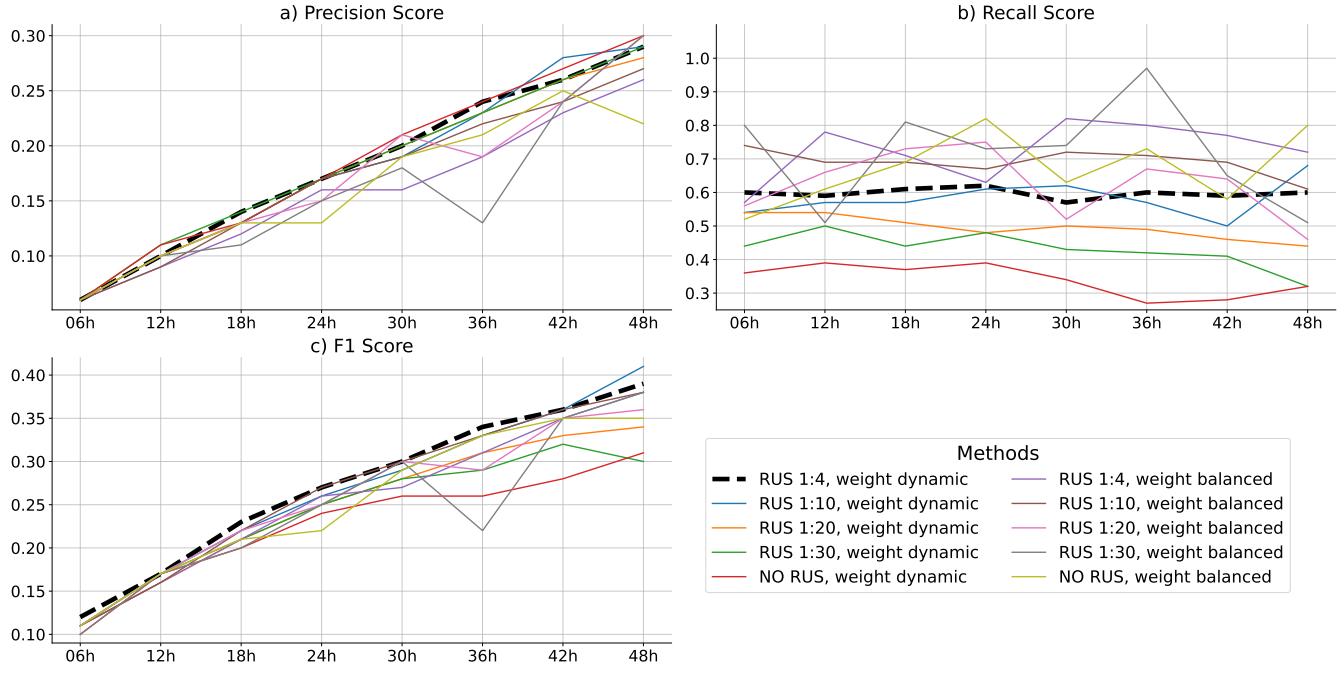

Figure 4. (a) The P score for the TCG prediction of the ResNet-18 model with the test set as a function of the temporal data enrichment window using a range of the RUS ratio and class weight (solid colors) for the PD sampling strategy, (b)-(c) similar to (a) but for the R and F1 scores, respectively. Dashed black line denotes the values used for our best-tuned model. *Weight balanced* assigns fixed importance to each class based on frequency whilst *weight dynamics* adaptively adjusts sample or class importance.

the entire WNP basin therefore introduces more irrelevant environmental information from the past (i.e., negative labels) that causes some bias in the model towards negative samples. This explains why the R score (blue columns in Fig. 3) decreases quickly with a longer enrichment window. As a result, the DD strategy exhibits a declining F1-score, indicating that the model's overall reliability in detecting TCG decreases when too much past information is used.

Given the fact that the TCG dataset is highly imbalanced due to the rarity of positive TCG labels, it is also important to examine how the ResNet-18 model could be optimized under various training data scenarios. For this purpose, Figure 4 shows the ResNet-18 model performance with different RUS ratios as described in Section 3.3. For the sake of clarity, we show here the absolute RUS ratios with a fixed number of positive TCG labels (minority) while changing the number of negative labels (majority) according to each ratio displayed in Fig. 4. For each RUS ratio, a class weight value that adjusts the loss function is provided, which controls the importance towards the positive TCG labels during training.

Figures 4-5 display a range of the R, P, and F1 scores for different RUS ratios and class weights, using both the PD and DD sampling strategies. One notices that a larger RUS ratio (i.e., more balance between positive and negative labels) tends to give higher P and R scores, thus resulting in a higher F1 score for the PD strategy. The optimal RUS ratio of 1:4 (one positive

https://doi.org/10.5194/egusphere-2025-4333
Preprint. Discussion started: 9 December 2025

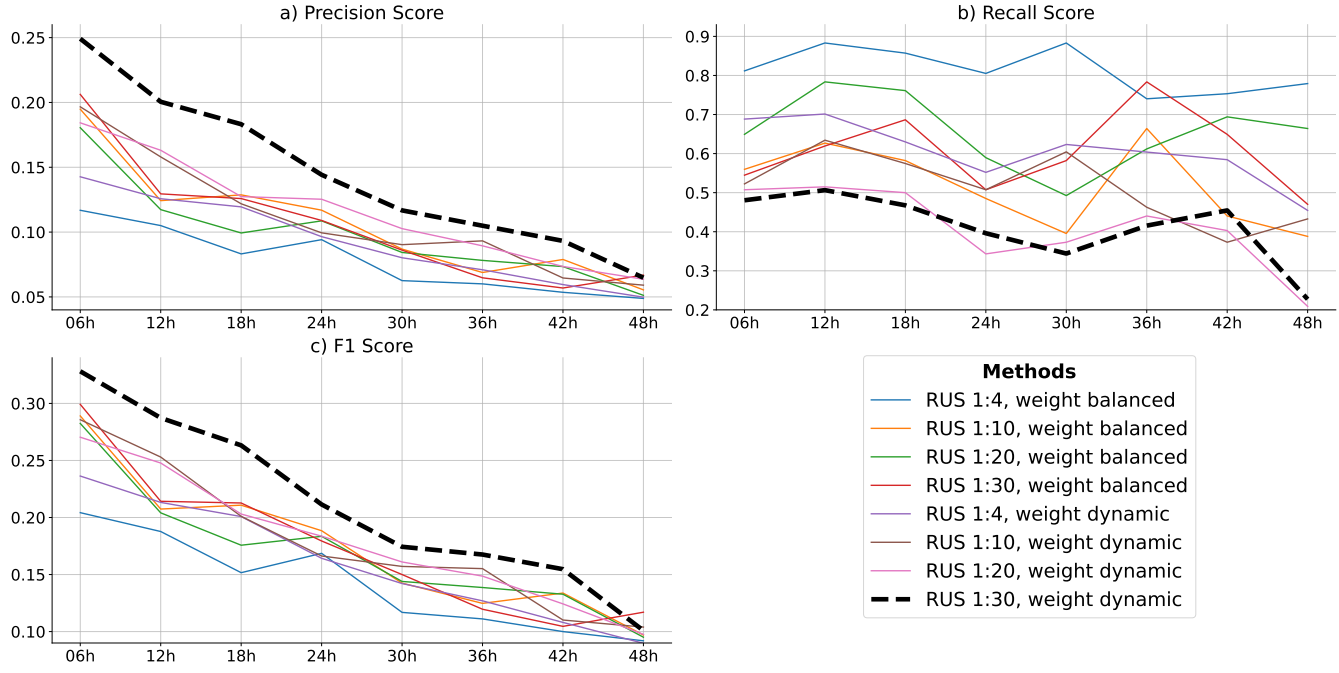

Figure 5. Similar to Figure 4 but for the DD sampling strategy.

label corresponds to 4 negative labels) combined with a class weight of 0.5 provides the best performance for the PD strategy in terms of detecting TCG, which is chosen as a default value in Fig. 3.

For the DD strategy, the model behavior is again somewhat different because of the constrain that one positive TCG location is surrounded by 8 negative labels by design. Therefore, the data imbalance at any given time is always fixed for the training. When applying the temporal enrichment with more data taken from the past for the test period, the number of negative TCG labels, however, increases rapidly because most of the days in the test period have no TCG. As a result, the imbalance ratio of the positive TCG over negative TCG labels becomes very small. For a reference, we show here only three RUS ratios of 1:10, 1:20, and 1:30 so one can compare their performance for the DD labeling strategy.

Unlike the PD strategy, we notice that the DD strategy performs best when the RUS ratio is 1:30, with a weight class of 0.2. Too small or too large RUS ratios both degrade the model performance. Consistent with the control performance shown in Fig. 3, all RUS ratio and class weight experiments also show best performance when the data enrichment window is around 0-12 hours, beyond which the performance of the DD strategy starts decaying rapidly. Such behavior is likely because the inclusion of larger negative samples from surrounding environment at farther time back into the past cannot help the model distinguish the positive labels, leading to a much lower R score. In constrast, too large RUS ratio means less over data for training. As a result, the overall F1 score is better for a RUS ratio of 1:30 and a shorter time window as seen in Fig. 3.

https://doi.org/10.5194/egusphere-2025-4333
Preprint. Discussion started: 9 December 2025

In addition to the model optimizations based on the RUS ratio, feature enrichment time windows, and class weights as presented above, there are of course many other factors related to model architecture and hyperparameter settings that must also be considered to achieve optimal performance. Many of these aspects are excessively granular and so we cannot discuss every single one of them here. However, it is of interest to note that no single DL model is universally optimal across all weather features and spatial-temporal scales, particularly given the current limitations of available climate datasets and DL architectures.

In the context of the real-time TCG prediction task, we should note that optimizing a DL model for this task involves more than just the typical hyperparameter tuning. In fact, the optimization process is tied to the specific aim of predicting either the spatial distribution of TCG or the timing of TCG at one location, which corresponds to different data labeling strategies as presented above. Regardless of the strategies, our experiments with a range of model architectures from simple CNN to more complex DL models yielded similar performance in terms of the F1, precision, and recall scores (not shown). This observation supports our decision to fix the ResNet-18 architecture and focus on optimizing it with different model settings as presented herein, rather than exploring a broad array of alternative DL models for TCG climatology reconstruction purposes.

In the following section, we apply the best-performing model to reconstruct two important aspects of TCG climatology in the WNP basin, with the aim of ultimately extending this approach to climate downscaling in our future climate change studies.

4.2 TCG reconstructed climatology

With the optimized ResNet-18 model for TCG climatology, our first attempt to reconstruct TCG climatology is the TCG seasonal distribution. For this purpose, the seasonal distribution is defined as the averaged ratio of the count of detected TCG events over the entire WNP basin in a given month to the total number of TCG events detected each year, which can be interpreted as the monthly TCG frequency. To avoid the arbitrary choice of the domain location in the PD strategy, we will generate this monthly TCG frequency for the DD strategy only.

As seen in Figure 6, the ResNet-18 model could reproduce well the overall seasonal distribution of TCG during the test period from 2017-2022, with a peak of TC activities in July-October, followed by an inactive period in January-April, similar to the observed TCG frequency. In particular, the ResNet-18 model could also reproduce the double peaks of TCG numbers in August and October consistent with the observed distribution, albeit the dip of TCG frequency in September is not as clear as that from the best track. During the peak period, note that the trainings with longer enrichment windows (i.e., 36-48 hours) tend to generate more positive TCG events than those with shorter windows (6-24 hour), suggesting that early signals of TC formation become more detectable as more past information is included.

Towards the end of the peak season (during November-December), note that our DL model tends to produce more TCG than the observation, while it tends to underestimate the TCG frequency during May-June. These differences indicate a common fact in TC climate research that optimizing a DL model based on one specific metric such as F1 score, precision, or recall would lead to biases in other metrics (See, e.g., da Silva et al., 2025). Regardless of such differences due to different tunings, the consistency of the seasonal TCG reconstruction by the ResNet-18 model is seen across the time enrichment windows, thus providing confidence in the robustness of our model. This result also highlights the ResNet's broader ability to understand

https://doi.org/10.5194/egusphere-2025-4333
Preprint. Discussion started: 9 December 2025

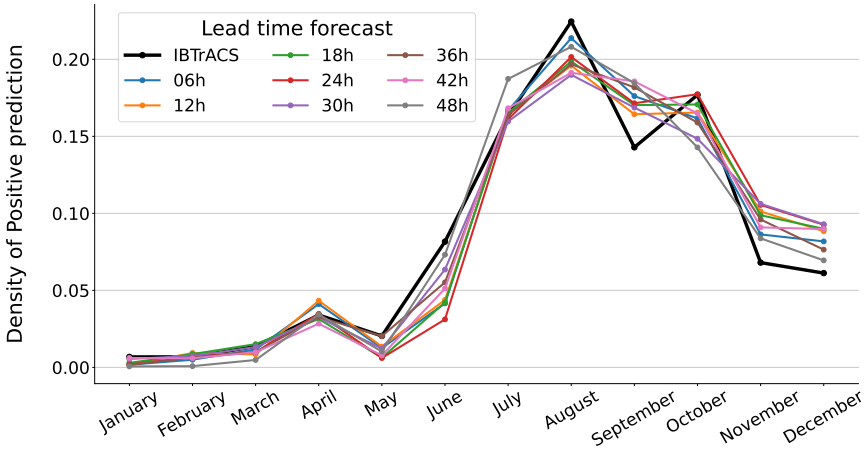

Figure 6. Monthly distribution of the TCG frequency detected in the WNP basin from the test data (2017-2022), using the best-tuned ResNet-18 model for the DD strategy with 8 different time windows from 6 to 48 hours. The black solid curve denotes the TCG density obtained from the best track during the same time period.

the seasonal changes in large-scale environments for TCG, which are needed for practical TCG prediction from global model outputs or climate change analyses.

To further look into the spatial distribution of TCG climatology by the ResNet-18 model, Figure 7 shows the horizontal map of the predicted TCG probability over the entire WNP basin, using the DD strategy with data enrichment windows from 6 to 48 hours. Here, the shading in each $5° \times 5°$ box represents the averaged probability of positive TCG predictions from the ResNet-18 model over the test period from 2017-2022. For the observed TCG, the shading denotes the actual number of TCG events occurs in each box, divided by the total number of TCGs during the 2017-2022 period (also commonly known as TCG density in literature). Thus, the predicted and observed TCG distribution are proportional to each other, and can serve as a metrics to evaluate our DL model's performance from a different angle.

Overall, the ResNet-18 model could again reproduce well the TCG distribution in the WNP basin during the 2017-2022 period, with clear TCG hotspots in both the East Philippine Sea and the South China Sea (SCS). For long enrichment windows between 18-42 hours, note that the ResNet-18 model tends to extend the TCG region too far east of the East Philippine Sea as compared to the observation. However, the overall spatial distribution of TCG probability is still distinct and concentrated in the central SCS, the eastern Philippine Sea, and the Vietnam coastal region. For the shorter data enrichment window of 6-12 hours, the model could also provide a reasonable fit as compared to the observed distribution (Fig. 7b), despite not being as good as the longer windows or the seasonal distribution in Fig. 6.

It is worth noting that while the predicted TCG probabilities consistently peak in the eastern Philippine Sea, some localized maxima in the central SCS and along the Vietnamese coastline lack consistency across data enrichment windows. This variability reflects the inherent challenge of detecting early TCG signals in the SCS, which are typically weak and highly variable.

https://doi.org/10.5194/egusphere-2025-4333
Preprint. Discussion started: 9 December 2025

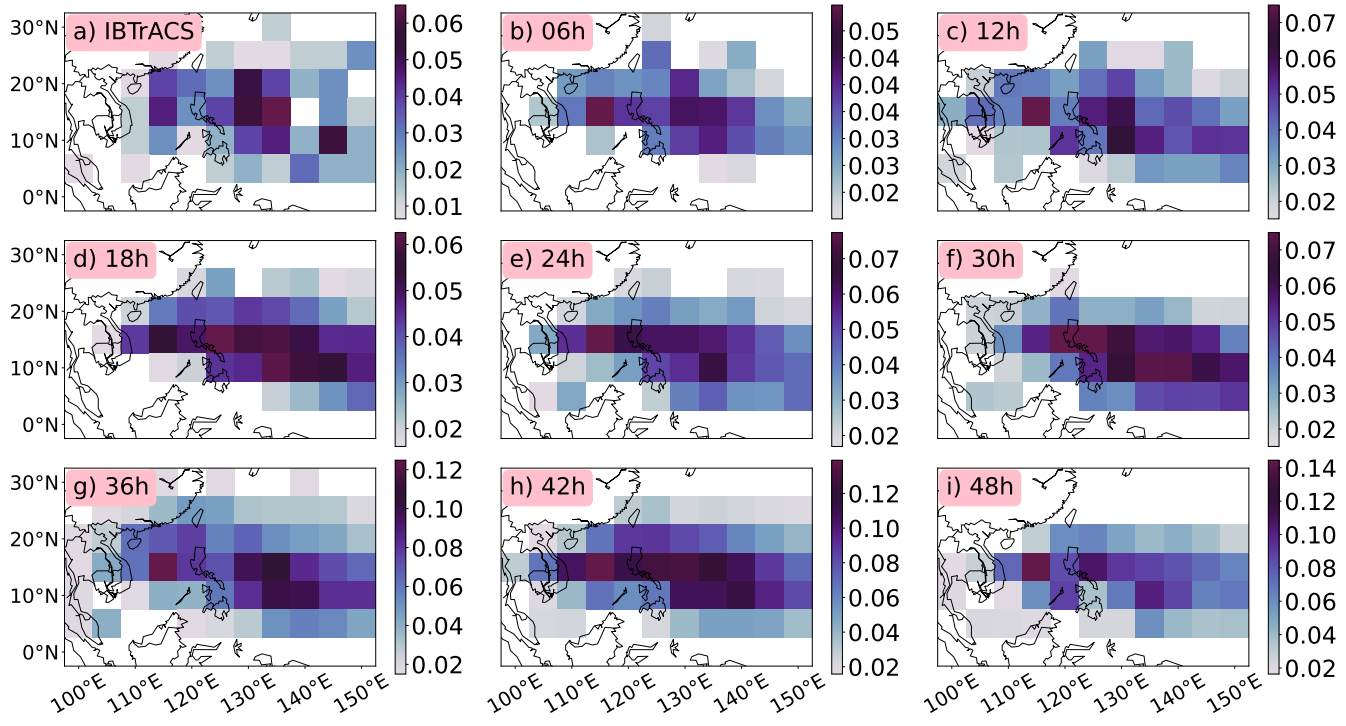

Figure 7. a) The spatial distribution of the observed TCG density (shaded) during the 2017-2022 period as obtained from the best track; (b)-(i) 5-year average of TCG probability prediction that is obtained from the ResNet-18 model with different data enrichment windows from 6-48 hours during the same test period as in (a). Note that different shading scales are used for different data enrichment windows so that one can better see the contrast between the areas of maximum probability for TCG predicted by the ResNet-18 model.

As a result, the ResNet-18 model, when optimized based on the F1 score over the broader WNP basin, struggles to capture these localized signals effectively. Unfortunately, the amount of available TCG data in the SCS region alone is insufficient for meaningful DL model training, making it difficult to resolve such inconsistencies given the current limitations of climate datasets. Because of this, we could only reconstruct the TCG climatology using the DD strategy as presented in this subsection, even though the PD strategy is more directly applicable for real-time forecasting purposes.

Aside from these local issues, the ability of our DL model in capturing the broad spatiotemporal patterns of TCG suggests that atmospheric signals associated with TCG become increasingly detectable when more relevant information is incorporated. From a practical standpoint, these results are very significant, because not only do they validate the performance of our DL model, but they also demonstrate that TCG can be predicted from large-scale environmental information, even at a spatial resolution of 0.5 degrees. This result has two important consequences: 1) as long as climate models can reliably simulate the large-scale environment, it is possible to learn TCG patterns with DL models and derive TCG climatology without resorting to the more computationally expensive high-resolution dynamical downscaling, and 2) changes in TCG climatology can be

https://doi.org/10.5194/egusphere-2025-4333
Preprint. Discussion started: 9 December 2025

captured through changes in large-scale environments, which are generally much more robust and reliable in climate projections than individual storm-scale features.

4.3 Sensitivity analyses

One critical question in developing a DL model for applications when the training data does not contain sufficient information is how to make full use of the data to optimize the model's performance. This process, known as feature engineering, becomes more significant when one tries to understand why we get what we see from DL model's output. Instead of running a DL model as a black box with all possible input data channels, exploring the importance of different input data can help better understand the role of different physical information in the model prediction that we wish to examine from a physical standpoint.

This subsection provides several additional analyses that employ a different set of input channels to see how effectively the ResNet-18 model could perform with limited information. Specifically, we examine two additional analyses that 1) use a set of features known to be of importance for TCG from previous studies, and 2) apply an automatic feature filter based on the rank of input channels. While this is often considered to be a part of model tuning, we treat them separately in this study. The choosing the right input channels will have significant implications in our further understanding of TCG processes.

For the first approach (hereafter referred to as feature engineering), input channels are selected based on their importance as obtained from previous observational and modeling studies (See, e.g., Gray, 1968; Riehl and Malkus, 1958; Yanai, 1964; Zhang and Bao, 1996; Bister and Emanuel, 1997; Ritchie and Holland, 1997; Simpson et al., 1997; Molinari et al., 2000; Nolan et al., 2007; Nguyen and Kieu, 2024; Kieu et al., 2023). Since not all meteorological data have strong control on TCG, only a subset of atmospheric variables that display the strongest TCG signals will be taken into consideration. Note that data on adjacent pressure levels are often correlated with each other. Therefore, our feature engineering approach will be applied to a few layers that can be categorized into low, middle and high pressure levels as shown in Table 4 instead of every single levels.

For the feature filtering approach (hereafter feature ranking), data channels are automatically selected by their contribution to the prediction score instead of depending on specialized knowledge, as for the feature engineering approach. This can be done by ranking all input features in terms of the mean activation value of the first filter from our best-tuned ResNet-18 model. Selection is then proceeded iteratively as follows. First, the channel with the highest score in the scoreboard is chosen and removed from the pool. Next, we eliminate any remaining layers that are highly correlated with the chosen channel (i.e., exceeding a Pearson R threshold). Finally, the progress is continued until all remaining features are either selected from their score or discarded, depending on the threshold that is used to stop the selection.

Table 4 compares the features after applying the two feature selection methods on the MERRA-2 data. If is of interest to note that the selected features between these two methods share a great number of overlaps, indicating that previous findings on TCG factors such as vertical wind shear (zonal wind components at 800 and 200-hPa level), low-level moisture (900-700 hPa relative humidity, RH), mid-level vertical motion (OMEGA at 500 hPa), or low-to-mid level temperature all play an important role in TCG prediction.

In addition, we observe that feature ranking tends to select a broader range of pressure levels than feature engineering. For example, NK2024's feature engineering uses RH at 750 hPa, while feature ranking captures a larger number of levels for the

https://doi.org/10.5194/egusphere-2025-4333
Preprint. Discussion started: 9 December 2025

ID	Name of Features	Past Domain		Dynamic Domain	
		Engineering	Ranking	Engineering	Ranking
1	QL		400, 700, 825, 900, 950		100, 1000. 150, 200, 300, 400, 500, 600, 700, 800, 875, 900, 950, 975
2	H	500	200, 925	500	100, 550, 950
3	QI		250, 450, 600, 800, 900, 925, 950, 1000		100, 1000, 150, 500, 600, 700, 900
4	OMEGA	500	450, 875	500	100, 150, 250, 600, 925, 1000
5	T	500, 900	725	500, 900	150, 200, 900
6	U	200, 800	825, 1000	200, 800	1000, 200, 550
7	V	200, 800	150, 550	200, 800	100, 1000, 150, 400, 600
8	RH	750	950	750	100, 200, 400, 700, 825, 875, 925, 1000
9	QV				100, 150, 900
10	VOR	200, 700, 900		200, 700, 900	
11	DIV	200		200	

Table 4. List of features selected using the feature engineering and feature ranking filter approach as obtained for each labeling strategy during the training period.

DD strategy (see last column in Table 4). Likewise, zonal wind components in feature engineering require only 200 and 800 hPa, but the feature ranking shows several levels at 1000, 600, 400, 150, and 100 hPa. However, the fact that both feature engineering and feature ranking could share a large group of features apparently indicate that the DL model is capable of learning large-scale environments correctly for the TCG problem. From this regard, our DL model not only justifies the use of the feature engineering method in previous studies, but also presents a way to help enhance our understanding of the key environments governing TCG when the ranking is expanded to include more environmental factors.

Among all selected features, we should note that there are several features from the feature ranking method that are not included in the feature engineering such as the specific liquid (QL) and ice content (QI), which represent the cloud information. In their early study, NK2024 did not include these features as they are linked to vertical motion and relative humidity via CAPE channel. Similarly, the specific humidity (QV) is just an equivalent representation of RH and so it is not included in feature

https://doi.org/10.5194/egusphere-2025-4333
Preprint. Discussion started: 9 December 2025

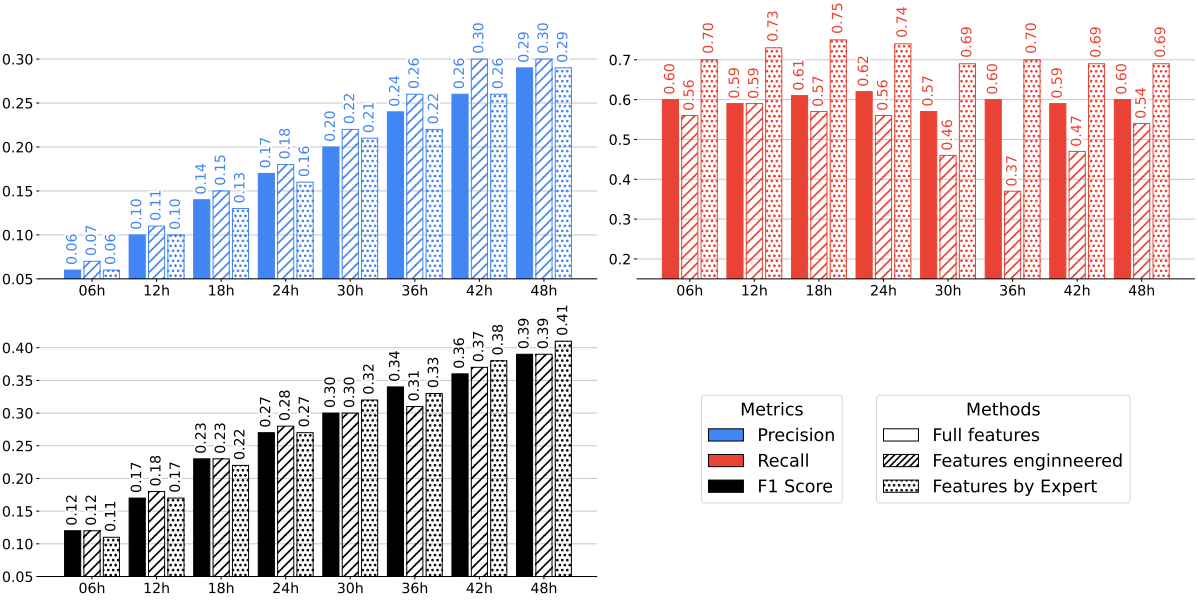

Figure 8. Evaluation of the model performance for several different feature selection methods including all features (solid color columns), 13 selected features based on feature engineering in Nguyen and Kieu (2024) (striped columns), and feature ranking of top 10% (dotted columns) using the Past Domain strategy for a) Precision, b) Recall, and c) F1 score.

engineering. These variables, however, emerge in the feature ranking, probably because of their stronger roles in capturing TCG during the training, even though they do not provide any new physical implications. Note also that the original MERRA-2 data does not contain some fields such as vorticity or divergence used in NK2024. Thus, several features in NK2024's study cannot be obtained from the ranking feature method shown in Table 4.

In terms of the model performance, Figures 8-9 compare the F1, P, and R scores among the full-feature, feature engineering, and feature ranking methods for both PD and DD labeling strategies. Among these feature selection methods, the feature ranking appears to deliver the most stable and effective results for the PD labeling strategy, particularly for long enrichment windows. The feature engineering and full-feature methods also perform well, but with slightly lower overall performance as compared to the feature engineering.

In contrast, the DD labeling strategy shows a declining trend in performance, with P, R and F1 score all decreased over longer time windows when using the all-feature method. However, the feature engineering method still maintains relatively higher and more stable R values, indicating its potential utility in capturing more relevant signals in predicting TCG spatial distribution without the need of all data channels.

Regardless of the labeling strategy, one can see that both feature engineering and feature ranking provide similar performance for a range of data enrichment windows, despite much fewer features than the full-feature method. For a long window of 36-48 hours, feature engineering and feature ranking could deliver even better performance for the PD approach, thus confirming that

https://doi.org/10.5194/egusphere-2025-4333
Preprint. Discussion started: 9 December 2025

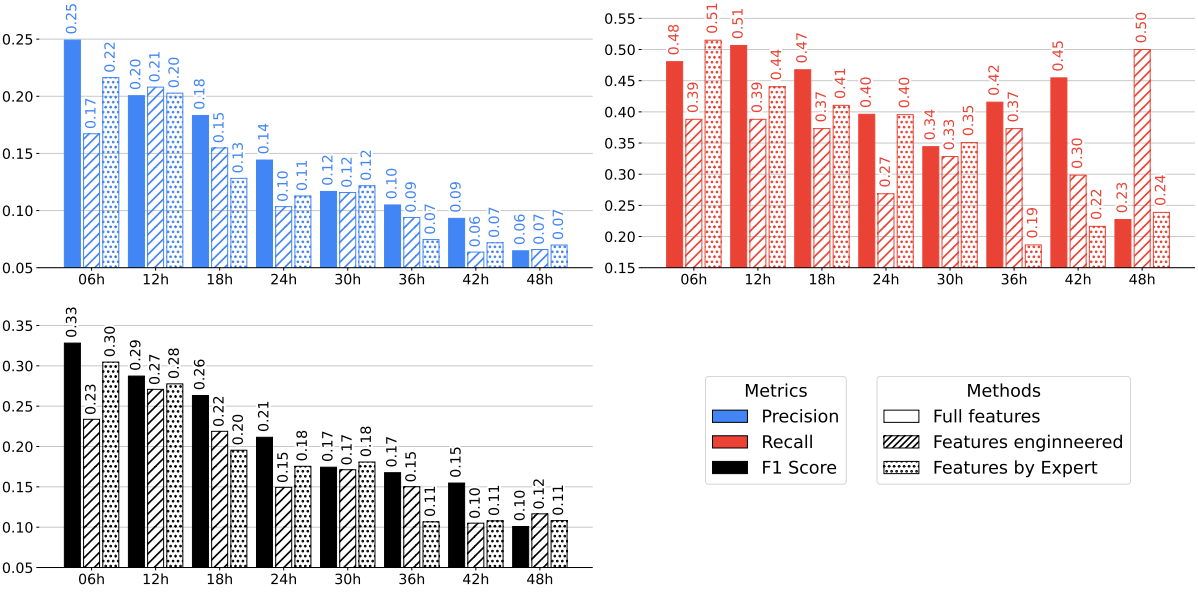

Figure 9. Similar to Fig. 8 but for the Dynamic Domain strategy.

predicting TCG would mostly rely on a subset of variables/channels instead of full data at all levels. In fact, examining the spatial distributions of TCG climatology from both feature engineering and feature ranking (Figs 10-11) shows little change in the overall patterns as compared to the full-feature output (cf. Fig. 7). The consistency among all feature selection methods is also captured for the seasonal TCG density (not shown), thus giving us some foundation for further understanding and predicting TCG based on a set of selected features for future DL improvement or implementations.

5 Conclusion

In this study, we presented a deep learning (DL) framework for reconstructing the climatology of tropical cyclone genesis (TCG) from climate reanalysis datasets. Recognizing that the definition of TCG climatology may vary depending on specific purposes and practical needs, our DL approach was designed and evaluated from multiple perspectives, based on different TCG labeling strategies for model training. Due to the limited number of TCG events available for training, we also implemented different data enrichment and feature selection methods to optimize our DL model for both TCG climatology reconstruction and potential prediction tasks.

Using the MERRA-2 reanalysis as the training dataset and the ResNet-18 architecture as our DL model, we demonstrated that ResNet-18 exhibits promising capability in detecting TCG from climate data at 0.5° resolution. Although the F1 score for TCG prediction remains relatively low, which is partly due to the inherent low predictability of TCG, the limited TCG

https://doi.org/10.5194/egusphere-2025-4333
Preprint. Discussion started: 9 December 2025

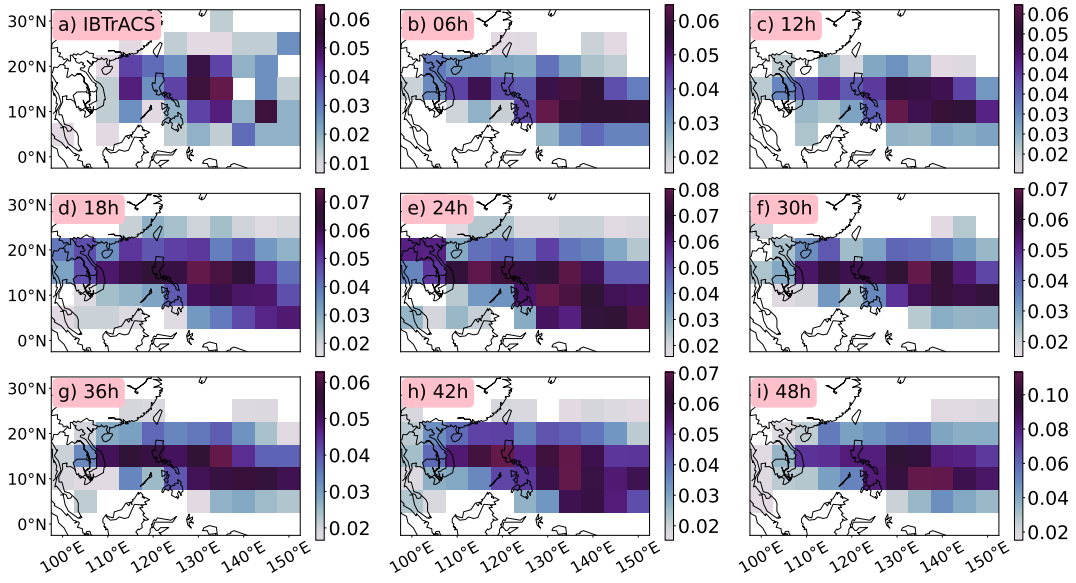

Figure 10. The spatial distributions of TCG climatology on map using the feature engineering approach.

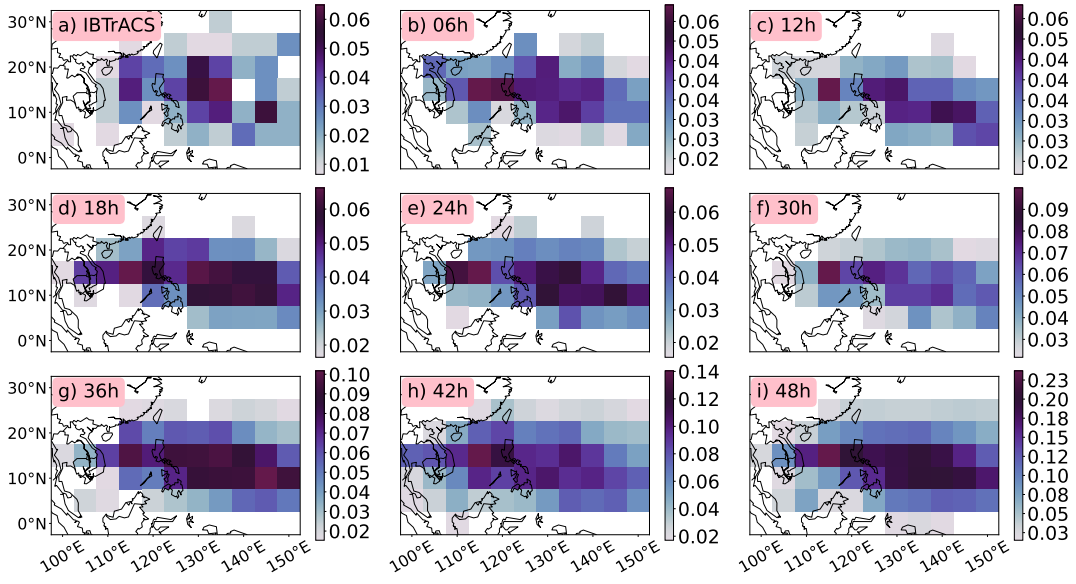

Figure 11. The spatial distributions of TCG climatology on map using the automatic feature ranking approach.

samples, and related information available in MERRA-2, we showed that the F1 score can be improved through appropriate hyperparameter tuning, labeling strategies, class weighting, and feature selection.

https://doi.org/10.5194/egusphere-2025-4333
Preprint. Discussion started: 9 December 2025

Comparing the reconstructed TCG seasonality and spatial distribution to the best track during the 2017-2022 period showed several noteworthy results. First, our ResNet-18 design could reproduce the seasonality of monthly TCG frequency, with double peaks of TCG frequency in August and October as well as the inactive period from January to May, consistent with those obtained from observations. Second, ResNet-18 could recover also the spatial distribution of TCG climatology, with main areas in the Eastern Philippine Sea and SCS. While there are some fluctuations in the TCG distribution for the areas along the coastal regions of the WNP basin, the overall well-recovered map of TCG in the WNP basin indicates that large-scale environments from the MERRA-2 dataset contain some important hidden signals of TCG processes that DL models can be trained and learn.

Further sensitivity experiments with different feature selection methods revealed that reconstructing TCG from any climate datasets is possible due to the existence of some key channels that contain the required TCG signals for DL models to learn. Specifically, our use of feature engineering based on a set of features reported in previous studies and feature ranking that filters input features based on the model impacts both capture some common data channels (variables) needed for TCG processes. Several of these key features obtained from the MERRA-2 dataset include vertical wind shear, low-to-mid level moisture, mid-level vertical motion, or mid-level geopotential height. The feature ranking method could detect some additional features such as high-level liquid or ice content that may include some cloud signal information beyond what used in previous models.

The results from these feature-sensitivity experiments support that combining expert knowledge with automated feature engineering can help enhance feature representation when training data is not sufficient or when running full features is too costly. Moreover, while feature engineering helps avoid overfitting and promotes interpretability, the feature ranking approach can help uncover hidden signals beyond what known from previous studies. For our TCG problem here, at least both the feature engineering and feature ranking approaches could share similar factors, which help build more confidence in using those input features to understand the variability of TCG climatology. Future improvements could involve integrating approaches using ranking mechanisms, attention-based models, or dimensionality reduction to retain a compact yet informative feature set for TCG prediction tasks.

Along with the primary focus on reconstructing TCG climatology, this study suggests that our approach holds potential for real-time TCG prediction in operational settings. Depending on the specific forecasting objective, i.e., whether predicting TCG at a fixed location (as in the PD labeling strategy) or forecasting the spatial distribution of TCG at a given time (as in the DD labeling strategy), a DL model can be designed and trained for each task separately. While the relatively low F1 scores observed for both strategies indicate that real-time forecasts of TCG from climate or global model output would be subject to considerable uncertainty, our DL approach offers a valuable, independent alternative to traditional physical-based models, with the capability to provide early warnings of TCG 1–3 days in advance. In this context, our work contributes to a deeper understanding of TCG processes and offers another practical guidance for improving DL model in real-time applications, particularly for extreme events where minority classes are the main focus.

Despite promising capabilities, this study reveals also several key challenges in applying DL models to TCG research. First, DL performance is highly sensitive to data preprocessing methods, particularly in labeling positive and negative TCG events, an issue that is exacerbated by the limited number of observed TCG occurrences. Second, the pronounced class imbalance

https://doi.org/10.5194/egusphere-2025-4333
Preprint. Discussion started: 9 December 2025

between positive and negative TCG labels during training remains a significant barrier. Addressing this challenge requires a careful integration of undersampling techniques, data augmentation strategies, or dynamic class weighting to ensure more robust and consistent performance across evaluation metrics. These challenges highlight the complexity of DL applications to extreme weather events. Moving forward, we aim to refine these approaches to extend TCG climatology reconstructions further back into the pre-satellite era for which we will present in our upcoming study.

Code availability. The complete source code and pipeline of the TCG-Net framework used for reconstructing TCG climatology in this study are publicly available in Le et al. (2025), https://doi.org/10.5281/zenodo.15640334. The repository includes all scripts for data preprocessing, model training, evaluation metrics (including sensitivity analysis), and visualization. Our code structure is designed in a way to support not only the reproducibility of results herein but also further experimentation with different climate datasets or TCG-related tasks. A detailed user guide for running the full workflow can be found in our repository.

Data availability. This research explores two publicly available datasets including: the NASA Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2, URL: https://disc.gsfc.nasa.gov/datasets?project=MERRA-2) and the International Best Track Archive for Climate Stewardship (IBTrACS, URL: https://www.ncei.noaa.gov/products/international-best-track-archive). Owing to the large size of the MERRA-2 dataset (approximately 20 TB), only a reference link to the raw data is provided. The complete data processing workflow, from downloading the raw data to generating the final training and testing datasets, is described in detail in the accompanying repository (https://doi.org/10.5281/zenodo.15640334), which also contains the IBTrACS dataset as a compressed file.

Author contributions. **Duc-Trong Le**: Formal analysis, Conceptualization, Methodology, Validation, Supervision, Writing-review&editting. **Tran-Binh Dang**: Investigation, Software, Validation, Writing-original draft. **Anh-Duc Hoang Gia**: Methodology, Software, Writing-original draft. **Duc-Hai Nguyen**: Software, Validation. **Minh-Hoa Tien**: Software, Validation. **Xuan-Truong Ngo**: Software, Validation. **Quang-Trung Luu**: Software, Validation, Writing-review&editting. **Quang-Lap Luu**: Software, Validation. **Tai-Hung Nguyen**: Validation, Supervision. **Thanh T.N. Nguyen**: Formal analysis, Conceptualization, Funding acquisition, Methodology, Supervision. **Chanh Kieu**: Conceptualization, Methodology, Validation, Supervision, Writing-review&editting.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Acknowledgements. This research is supported by Vingroup Innovation Foundation (VINIF, code VINIF.2023.DA019) and the U.S. National Science Foundation (NSF, Grant/Award Number 135999)

https://doi.org/10.5194/egusphere-2025-4333
Preprint. Discussion started: 9 December 2025

References

Bengtsson, L., Hodges, K. I., Esch, M., Keenlyside, N., Kornblueh, L., Luo, J.-J., and Yamagata, T.: How may tropical cyclones change in a warmer climate?, Tellus A: Dynamic Meteorology and Oceanography, 59A, 539–561, https://doi.org/10.1111/j.1600-0870.2007.00251.x, 2007.

Bi, K., L, X., H, Z., X, C., X, G., and Q., T.: Accurate medium-range global weather forecasting with 3D neural networks, Nature, 619, 533–538, https://doi.org/doi: 10.1038/s41586-023-06185-3, 2023.

Bister, M. and Emanuel, K. A.: The Genesis of Hurricane Guillermo: TEXMEX Analyses and a Modeling Study, Monthly Weather Review, 125, 2662–2682, https://doi.org/10.1175/1520-0493(1997)125<2662:TGOHGT>2.0.CO;2, 1997.

Brigato, L. and Iocchi, L.: A Close Look at Deep Learning with Small Data, https://arxiv.org/abs/2003.12843, 2020.

Camargo, S. J. and Zebiak, S. E.: Improving the Detection and Tracking of Tropical Cyclones in Atmospheric General Circulation Models, Weather and Forecasting, 17, 1152–1162, https://doi.org/10.1175/1520-0434(2002)017<1152:ITDATO>2.0.CO;2, 2002.

Camargo, S. J., Murakami, H., Bloemendaal, N., Chand, S. S., Deshpande, M. S., Dominguez-Sarmiento, C., González-Alemán, J. J., Knutson, T. R., Lin, I.-I., Moon, I.-J., Patricola, C. M., Reed, K. A., Roberts, M. J., Scoccimarro, E., Tam, C. Y. F., Wallace, E. J., Wu, L., Yamada, Y., Zhang, W., and Zhao, H.: An update on the influence of natural climate variability and anthropogenic climate change on tropical cyclones, Tropical Cyclone Research and Review, 12, 216–239, https://doi.org/https://doi.org/10.1016/j.tcrr.2023.10.001, 2023.

Cha, E. J., Knutson, T. R., Lee, T.-C., Ying, M., and Nakaegawa, T.: Third assessment on impacts of climate change on tropical cyclones in the Typhoon Committee Region – Part II: Future projections, Tropical Cyclone Research and Review, 9, 75–86, https://doi.org/https://doi.org/10.1016/j.tcrr.2020.04.005, 2020.

Chen, A. and Yuan, C.: Deep learning-based spatial downscaling and its application for tropical cyclone detection in the western North Pacific, Frontiers in Earth Science, Volume 12 - 2024, https://doi.org/10.3389/feart.2024.1345714, 2024.

Chen, R., Zhang, W., and Wang, X.: Machine Learning in Tropical Cyclone Forecast Modeling: A Review, Atmosphere, 11, https://doi.org/10.3390/atmos11070676, 2020.

da Silva, S. T., Gabrick, E. C., de Moraes, A. L. R., Viana, R. L., Batista, A. M., Caldas, I. L., and Kurths, J.: Predicting temperatures in Brazilian states capitals via Machine Learning, The European Physical Journal Special Topics, pp. 1–20, 2025.

Defforge, C. L. and Merlis, T. M.: Observed warming trend in sea surface temperature at tropical cyclone genesis, Geophysical Research Letters, 44, 1034–1040, https://doi.org/doi.org/10.1002/2016GL071045, 2017.

Gao, S., Zhao, P., Pan, B., Li, Y., Zhou, M., Xu, J., Zhong, S., and Shi, Z.: A nowcasting model for the prediction of typhoon tracks based on a long short term memory neural network, Acta Oceanologica Sinica, 37, 8–12, 2018a.

Gao, S., Zhao, P., Pan, B., Li, Y., Zhou, M., Xu, J., Zhong, S., and Shi, Z.: A nowcasting model for the prediction of typhoon tracks based on a long short term memory neural network, Acta Oceanologica Sinica, 37, 8–12, 2018b.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B.: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), Journal of Climate, 30, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1, 2017.

Giffard-Roisin, S., Yang, M., Charpiat, G., Kumler Bonfanti, C., Kégl, B., and Monteleoni, C.: Tropical cyclone track forecasting using fused deep learning from aligned reanalysis data, Frontiers in big Data, 3, 1, 2020.

https://doi.org/10.5194/egusphere-2025-4333
Preprint. Discussion started: 9 December 2025

Gray, W. M.: Global View of The Origin of Tropical Disturbances, Mon. Wea. Rev., 96, 669–700, https://doi.org/10.1175/1520-0493(1968)096<0669:GVOTOO>2.0.CO;2, 1968.

Horn, M., Walsh, K., Zhao, M., Camargo, S. J., Scoccimarro, E., Murakami, H., Wang, H., Ballinger, A., Kumar, A., Shaevitz, D. A., Jonas, J. A., and Oouchi, K.: Tracking Scheme Dependence of Simulated Tropical Cyclone Response to Idealized Climate Simulations, Journal of Climate, 27, 9197–9213, https://doi.org/10.1175/JCLI-D-14-00200.1, 2014.

Kieu, C. and Nguyen, Q.: Binary dataset for machine learning applications to tropical cyclone formation prediction, Scientific Data, 11, 446, 2024.

Kieu, C., Zhao, M., Tan, Z., Zhang, B., and Knutson, T.: On the Role of Sea Surface Temperature in the Clustering of Global Tropical Cyclone Formation, Journal of Climate, pp. 1–39, https://doi.org/10.1175/JCLI-D-22-0623.1, 2023.

Kim, M., Park, M.-S., Im, J., Park, S., and Lee, M.-I.: Machine Learning Approaches for Detecting Tropical Cyclone Formation Using Satellite Data, Remote Sensing, 11, https://doi.org/10.3390/rs11101195, 2019a.

Kim, M., Park, M.-S., Im, J., Park, S., and Lee, M.-I.: Machine Learning Approaches for Detecting Tropical Cyclone Formation Using Satellite Data, Remote Sensing, 11, https://doi.org/10.3390/rs11101195, 2019b.

Knapp, K. R., Kruk, M. C., Levinson, D. H., Diamond, H. J., and Neumann, C. J.: The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying Tropical Cyclone Data, Bulletin of the American Meteorological Society, 91, 363–376, https://doi.org/10.1175/2009BAMS2755.1, 2010.

Knutson, T. R., Tuleya, R. E., and Kurihara, Y.: Simulated increase of hurricane intensities in a CO2-warmed climate., Science, 279, 1018–1021, https://doi.org/10.1126/science.279.5353.1018, 1998.

Kossin, J. P., Emanuel, K. A., and Camargo, S. J.: Past and Projected Changes in Western North Pacific Tropical Cyclone Exposure, Journal of Climate, 29, 5725 – 5739, https://doi.org/10.1175/JCLI-D-16-0076.1, 2016.

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W., Merose, A., Hoyer, S., Holland, G., Vinyals, O., Stott, J., Pritzel, A., Mohamed, S., and Battaglia, P.: Learning skillful medium-range global weather forecasting, Science, 382, 1416–1421, https://doi.org/10.1126/science.adi2336, 2023.

Le, D. T., Dang, T.-B., Hoang Gia, A.-D., Nguyen, D.-H., Tien, M.-H., Luu, Q.-T., Luu, Q.-L., Nguyen, T.-H., Nguyen, T. N. T., and Kieu, C.: Data Processing for TCG-Net: Reconstructing Tropical Cyclogenesis Climatology, https://doi.org/10.5281/zenodo.15640334, 2025.

Lee, C.-Y., Camargo, S. J., Vitart, F., Sobel, A. H., Camp, J., Wang, S., Tippett, M. K., and Yang, Q.: Subseasonal Predictions of Tropical Cyclone Occurrence and ACE in the S2S Dataset, Weather and Forecasting, 35, 921 – 938, https://doi.org/doi.org/10.1175/WAF-D-19-0217.1, 2020.

Matsuoka, D., Nakano, M., Sugiyama, D., and Uchida, S.: Deep learning approach for detecting tropical cyclones and their precursors in the simulation by a cloud-resolving global nonhydrostatic atmospheric model, Progress in Earth and Planetary Science, https://doi.org/10.1186/s40645-018-0245-y, 2018.

Miller, J., Maskey, M., and Berendes, T.: Using deep learning for tropical cyclone intensity estimation, in: AGU Fall Meeting Abstracts, vol. 2017, pp. IN11E–05, 2017a.

Miller, J., Maskey, M., and Berendes, T.: Using deep learning for tropical cyclone intensity estimation, in: AGU Fall Meeting Abstracts, vol. 2017, pp. IN11E–05, 2017b.

Molinari, J., Vollaro, D., Skubis, S., and Dickinson, M.: Origins and Mechanisms of Eastern Pacific Tropical Cyclogenesis: A Case Study, Monthly Weather Review, 128, 125–139, https://doi.org/10.1175/1520-0493(2000)128<0125:OAMOEP>2.0.CO;2, 2000.

Nguyen, Q.: Deep Learning for Tropical Cyclone Formation Detection, ProQuest Dissertations Publishing, Indiana University, 120p, 2023.

https://doi.org/10.5194/egusphere-2025-4333
Preprint. Discussion started: 9 December 2025

Nguyen, Q. and Kieu, C.: Predicting Tropical Cyclone Formation with Deep Learning, Weather and Forecasting, 39, 241–258, https://doi.org/10.1175/WAF-D-23-0103.1, 2024.

Nolan, D., Rappin, E. D., and Emanuel, K. A.: Tropical cyclogenesis sensitivity to environmental parameters in radiative–convective equilibrium, Quart. J. Roy. Meteor. Soc., 133, 2085–2107, 2007.

Park, M.-S., Kim, M., Lee, M.-I., Im, J., and Park, S.: Detection of tropical cyclone genesis via quantitative satellite ocean surface wind pattern and intensity analyses using decision trees, Remote sensing of environment, 183, 205–214, 2016.

Pathak, J., Subramanian, S., Harrington, P., Raja, S., Chattopadhyay, A., Mardani, M., Kurth, T., Hall, D., Li, Z., Azizzadenesheli, K., Hassanzadeh, P., Kashinath, K., and Anandkumar, A.: FourCastNet: A Global Data-driven High-resolution Weather Model using Adaptive Fourier Neural Operators, 2022.

Peduzzi, P., Chatenoux, B., Dao, H., Bono, A. D., Herold, C., Kossin, J., Mouton, F., and Nordbeck, O.: Tropical cyclones: Global trends in human exposure, vulnerability and risk, Nat. Climate Change, 2, 289–294, https://doi.org/10.1038/NCLIMATE1410, 2012.

Riehl, H. and Malkus, J. S.: On the heat balance in the equatorial trough zone, Geophysica, 6, 503–538, 1958.

Ritchie, E. A. and Holland, G. J.: Scale Interactions during the Formation of Typhoon Irving, Monthly Weather Review, 125, 1377–1396, https://doi.org/10.1175/1520-0493(1997)125<1377:SIDTFO>2.0.CO;2, 1997.

Scher, S. and Messori, G.: Weather and climate forecasting with neural networks: using general circulation models (GCMs) with different complexity as a study ground, Geoscientific Model Development, 12, 2797–2809, https://doi.org/10.5194/gmd-12-2797-2019, 2019.

Schultz, M. G., Betancourt, C., Gong, B., Kleinert, F., Langguth, M., Leufen, L. H., Mozaffari, A., and Stadtler, S.: Can deep learning beat numerical weather prediction?, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 379, 20200 097, https://doi.org/10.1098/rsta.2020.0097, 2021.

Simpson, J., Ritchie, E., Holland, G. J., Halverson, J., and Stewart, S.: Mesoscale Interactions in Tropical Cyclone Genesis, Monthly Weather Review, 125, 2643–2661, https://doi.org/10.1175/1520-0493(1997)125<2643:MIITCG>2.0.CO;2, 1997.

Strachan, J., Vidale, P. L., Hodges, K., Roberts, M., and Demory, M.-E.: Investigating Global Tropical Cyclone Activity with a Hierarchy of AGCMs: The Role of Model Resolution, Journal of Climate, 26, 133–152, https://doi.org/10.1175/JCLI-D-12-00012.1, 2013.

Su, H., Wu, L., Jiang, J. H., Pai, R., Liu, A., Zhai, A. J., Tavallali, P., and DeMaria, M.: Applying satellite observations of tropical cyclone internal structures to rapid intensification forecast with machine learning, Geophysical Research Letters, 47, e2020GL089 102, 2020.

Tan, P., T., L. T., H., H. B., and Kieu, C.: Seasonal forecasting of tropical cyclone activity in the coastal region of Vietnam using RegCM4.2, Climate Research, 62, 115–129, https://doi.org/doi.org/10.3354/cr01267, 2015.

Thanh, N. T., Cuong, H. D., Hien, N. X., and Kieu, C.: Relationship between sea surface temperature and the maximum intensity of tropical cyclones affecting Vietnam's coastline, International Journal of Climatology, 40, 2527–2538, https://doi.org/https://doi.org/10.1002/joc.6348, 2020.

Tien, T. T., Hoa, D. N.-Q., Thanh, C., and Kieu, C.: Assessing the Impacts of Augmented Observations on the Forecast of Typhoon Wutip's (2013) Formation Using the Ensemble Kalman Filter, Weather and Forecasting, 35, 1483–1503, https://doi.org/10.1175/WAF-D-20-0001.1, 2020.

Tran-Quang, D., Pham-Thanh, H., Vu, T.-A., Kieu, C., and Phan-Van, T.: Climatic Shift of the Tropical Cyclone Activity Affecting Vietnam's Coastal Region, Journal of Applied Meteorology and Climatology, 59, 1755 – 1768, https://doi.org/10.1175/JAMC-D-20-0021.1, 2020.

Trinh, D. H., Cuong, H. D., Kham, D. V., and Kieu, C.: Remote Control of Sea Surface Temperature on the Variability of Tropical Cyclone Activity Affecting Vietnam's Coastline, Journal of Applied Meteorology and Climatology, 60, 323 – 339, https://doi.org/10.1175/JAMC-D-20-0170.1, 2021.

https://doi.org/10.5194/egusphere-2025-4333
Preprint. Discussion started: 9 December 2025

Ullrich, P. and Zarzycki, C.: TempestExtremes v1.0: A framework for scale-insensitive pointwise feature tracking on unstructured grids, Geosci. Model. Dev., 10, 1069–1090, 2017.

Vu, T.-A., Kieu, C., Chavas, D., and Wang, Q.: A Numerical Study of the Global Formation of Tropical Cyclones, Journal of Advances in Modeling Earth Systems, 13, https://doi.org/10.1029/2020MS002207, 2021.

Walsh, K. J. E., Fiorino, M., Landsea, C. W., and McInnes, K. L.: Objectively Determined Resolution-Dependent Threshold Criteria for the Detection of Tropical Cyclones in Climate Models and Reanalyses, Journal of Climate, 20, 2307–2314, https://doi.org/10.1175/JCLI4074.1, 2007.

Walsh, K. J. E., Camargo, S. J., Vecchi, G. A., Daloz, A. S., Elsner, J., Emanuel, K., Horn, M., Lim, Y.-K., Roberts, M., Patricola, C., Scoccimarro, E., Sobel, A. H., Strazzo, S., Villarini, G., Wehner, M., Zhao, M., Kossin, J. P., LaRow, T., Oouchi, K., Schubert, S., Wang, H., Bacmeister, J., Chang, P., Chauvin, F., Jablonowski, C., Kumar, A., Murakami, H., Ose, T., Reed, K. A., Saravanan, R., Yamada, Y., Zarzycki, C. M., Vidale, P. L., Jonas, J. A., and Henderson, N.: Hurricanes and Climate: The U.S. CLIVAR Working Group on Hurricanes, Bulletin of the American Meteorological Society, 96, 997–1017, https://doi.org/10.1175/BAMS-D-13-00242.1, 2015.

Wang, Z., Zhao, J., Huang, H., and Wang, X.: A Review on the Application of Machine Learning Methods in Tropical Cyclone Forecasting, Frontiers in Earth Science, Volume 10 - 2022, https://doi.org/10.3389/feart.2022.902596, 2022.

Weyn, J. A., Durran, D. R., Caruana, R., and Cresswell-Clay, N.: Sub-Seasonal Forecasting With a Large Ensemble of Deep-Learning Weather Prediction Models, Journal of Advances in Modeling Earth Systems, 13, e2021MS002 502, https://doi.org/https://doi.org/10.1029/2021MS002502, 2021.

Wimmers, A., Velden, C., and Cossuth, J. H.: Using Deep Learning to Estimate Tropical Cyclone Intensity from Satellite Passive Microwave Imagery, Monthly Weather Review, 147, 2261 – 2282, https://doi.org/10.1175/MWR-D-18-0391.1, 2019.

Yanai, M.: Formation of tropical cyclones, Rev. Geophys., 2, 367–414, 1964.

Zarzycki, C. and Ullrich, P.: Assessing sensitivities in algorithmic detection of tropical cyclones in climate data, Geophys. Res. Lett., 44, 1141–1149, 2017.

Zhang, D.-L. and Bao, N.: Oceanic Cyclogenesis as Induced by a Mesoscale Convective System Moving Offshore. Part I: A 90-h Real-Data Simulation, Monthly Weather Review, 124, 1449–1469, https://doi.org/10.1175/1520-0493(1996)124<1449:OCAIBA>2.0.CO;2, 1996.

Zhang, T., Lin, W., Lin, Y., Zhang, M., Yu, H., Cao, K., and Xue, W.: Prediction of Tropical Cyclone Genesis from Mesoscale Convective Systems Using Machine Learning, Weather and Forecasting, 34, 1035–1049, https://doi.org/10.1175/WAF-D-18-0201.1, 2019.

Zhang, W., Fu, B., Peng, M. S., and Li, T.: Discriminating developing versus nondeveloping tropical disturbances in the western North Pacific through decision tree analysis, Weather and Forecasting, 30, 446–454, 2015.

Zhao, M., Held, I. M., Lin, S.-J., and Vecchi, G. A.: Simulations of Global Hurricane Climatology, Interannual Variability, and Response to Global Warming Using a 50-km Resolution GCM, Journal of Climate, 22, 6653–6678, https://doi.org/10.1175/2009JCLI3049.1, 2009.