Article

Relationship between Early-Stage Features and Lifetime Maximum Intensity of Tropical Cyclones over the Western North Pacific

Ren Lu and Xiaodong Tang *

Key Laboratory of Mesoscale Severe Weather, Ministry of Education, and School of Atmospheric Sciences, Nanjing University, Nanjing 210023, China; mg1828011@smail.nju.edu.cn
* Correspondence: xdtang@nju.edu.cn

Abstract: The relationship between early-stage features and lifetime maximum intensity (LMI) of tropical cyclones (TCs) over the Western North Pacific (WNP) was investigated by ensemble machine learning methods and composite analysis in this study. By selecting key features of TCs' vortex attributes and environmental conditions, a two-step AdaBoost model demonstrated accuracy of about 75% in distinguishing weak and strong TCs at genesis and a coefficient of determination (R^2) of 0.30 for LMI estimation from the early stage of strong TCs, suggesting an underlying relationship between LMI and early-stage features. The composite analysis reveals that TCs with higher LMI are characterized by lower latitude embedded in a continuous band of high low-troposphere vorticity, more compact circulation at both the upper and lower levels of the troposphere, stronger circulation at the mid-troposphere, a higher outflow layer with stronger convection, a more symmetrical structure of high-level moisture distribution, a slower translation speed, and a greater intensification rate around genesis. Specifically, TCs with greater "tightness" at genesis may have a better chance of strengthening to major TCs (LMI ≥ 96 kt), since it represents a combination of the inner and outer-core wind structure related to TCs' rapid intensification and eyewall replacement cycle.

Keywords: tropical cyclone; lifetime maximum intensity; machine learning; AdaBoost; decision tree; composite analysis

Academic Editors: Sundararaman Gopalakrishnan and Ghassan Alaka

1. Introduction

Tropical cyclones (TCs), one of the most catastrophic weather events over the Western North Pacific (WNP), have caused huge damage with strong winds and heavy precipitation for decades [1,2]. Great effort has been put into improving TC intensity prediction for a certain lead time through the development of statistical and dynamical models [3–8]. However, there is a lack of research on influential factors of a TC's lifetime maximum intensity (LMI), a measurement related to its upper boundary of destructiveness.

LMI might be affected by multiple factors during a TC's lifetime, including its genesis conditions. Previous studies on the physical mechanisms and favorable conditions of TC genesis have been conducted [9–13]. The genesis process of a TC can be divided into two consecutive stages [14,15]: first, from a tropical disturbance to a tropical depression (TD) with the formation of initial circulation; and second, from a tropical depression to a tropical storm (TS) when its warm-core structure is established. Gray [16,17] noted several favorable factors for TC genesis, including thermodynamic factors of sufficient ocean thermal energy, conditional instability throughout the low troposphere and high relative humidity in the mid-troposphere, and dynamic factors of a large enough Coriolis parameter, above-normal low-level vorticity, and weak vertical wind shear near the center of a TC's circulation. He further emphasized the key roles of climate conditions (e.g., region, season, etc.), certain synoptic flow patterns (e.g., monsoon trough), and active mesoscale convective systems (MCSs) in TC genesis. Based on that, the genesis potential

index (GPI) [18,19] was developed to quantitively assess the probability of TC genesis at a certain location, which suggests some key factors for LMI as well.

In addition to the genesis conditions, the development stage also plays an important role in determining LMI when a formed TC interacts with the environment and changes its own structure. From the dynamic aspect, vertical wind shear is commonly detrimental to TC intensification [20–22]; the interaction of a TC with an upper-level trough can lead to intensification [23–25] and other factors such as the distribution of environmental vorticity and a TC's inertial stability can also influence TC intensification [26,27]. From the thermodynamic aspect, variations of ocean surface temperature, heat content, and exchange coefficients of air–sea fluxes can significantly affect a storm's intensification rate [28–32], and ambient dry air may inhibit TC intensification [33–35]. A TC's internal features and processes are also found to be associated with its intensity change. For instance, a TC's inertial stability contributes a lot to its growth by effective local warming with cumulus convection [26,36]; distribution of rainfall and convection is related to a TC's rapid intensification (RI) [37,38] and the eyewall replacement cycles (ERC) can result in re-intensification [39,40].

For an individual TC, as the time and location of LMI are both uncertain, it is difficult to "forecast" LMI using traditional numerical models. Considering that the two key stages mentioned above (genesis and development) have a great influence on a TC's intensity change, LMI could be regarded as the result of various factors during these two stages. Ditchek et al. [41] investigated the relationship between the maximum attained intensity and the genesis environment of a TC over the North Atlantic (NA). They used a stepwise regression method to select the most important genesis variables for LMI and then established a linear function to assess their relationship. The regression had an overall R^2 of 0.41, indicating that even if maximum attained intensity was not fully determined by a TC's genesis conditions, the relationship did exist. TCs reaching higher intensity are associated with stronger, more compact low-level vortices, better-defined outflow jets, a more compact region of high midlevel relative humidity, and higher water vapor content at genesis over the NA.

However, the corresponding relationship was never proved over the WNP, and the issue is full of challenges, as TCs over the WNP are subjected to more complex environmental factors (e.g., monsoon trough, monsoon gyre, etc.) [34,42,43]. For this reason, a statistical model with better nonlinear fitting capability is required to explain the contributions of factors to LMI over the WNP. Recently, machine learning methods have been found to be capable of handling complicated issues in earth sciences [44–46]. For example, K-means clustering is used to segment maps of radar echoes [47], decision trees work well in classifying convection areas [48], and artificial neural networks (ANNs) are applied to make short-term predictions of TC intensity [49]. Among these algorithms, decision tree has an outstanding interpretability and can be easily utilized for classification or regression, which is applicable to the LMI attribution issue here.

The purpose of this study is to discover how much the LMI of TCs over the WNP is related to their vortex attributes and environmental conditions near genesis, and search for the key factors that will affect LMI. For this purpose, features of the vortex and environment around TC genesis are firstly extracted using reanalysis and best track datasets, then the relationship between these features and LMI is investigated by ensemble machine learning methods for two separate steps (one for rough classification and the other for specific regression). After the model's parameters are well tuned, a composite analysis of the leading features that have largest impact on LMI is conducted to find the distinctions between TCs with different LMI. In Sections 2.1–2.5, we briefly describe the data source as well as the ways to extract the features, and show the workflow of the whole model. The results of the model fitting and composite analysis of features are presented in Sections 3.1 and 3.2. Finally, an overall summary and a further discussion are provided in Section 4.

2. Materials and Methods

2.1. Data Source

Combining information from numerous TC best-track datasets, version 4.0 of the International Best Track Archive for Climate Stewardship (IBTrACS) [50–52] provides multiple attributes of TCs (e.g., location, wind speed, translation speed, etc.) in every basin. To avoid bias from datasets produced by different agencies as much as possible, IBTrACS data of early-stage features of storms over the WNP basin from July to November over 41 years (1979–2019) in 3 h intervals were obtained from the Joint Typhoon Warning Center (JTWC). These TCs are sorted into 3 groups according to their LMI: (1) never intensified beyond tropical storm ($\leq$63 kt, TD/TS); (2) reached minor hurricane intensity but never achieved major hurricane intensity (64–95 kt, minor TC); and (3) reached major hurricane intensity ($\geq$96 kt, major TC). TD/TS is also called weak TC and major/minor TC are collectively named strong TC. For convenience, TCs are labeled by their LMI level hereafter. Environmental features are derived from ERA5 hourly reanalysis provided by the European Centre for Medium-Range Weather Forecasts (ECMWF), with a horizontal resolution of $0.25° \times 0.25°$.

2.2. Preprocessing of the Original Dataset

Preprocessing of the TC data was conducted to make the model work properly, including spatial restriction to focus on a certain scope of genesis and temporal filtering to remove short-lived TCs. First, the studied genesis area was restricted to a rectangular region over the WNP in a range of latitude of 0–30° N and longitude 130–180° E to exclude effects from land during the TC genesis stage (Figure 1). Then, TCs with a lifetime less than 48 h were removed, since TCs with longer lifetimes are more noteworthy in general. The dataset was still large enough for traditional machine learning tasks after preprocessing (Table 1) [53]. Further, information on features in each case was complete, so the model would not suffer from drawbacks caused by missing values.
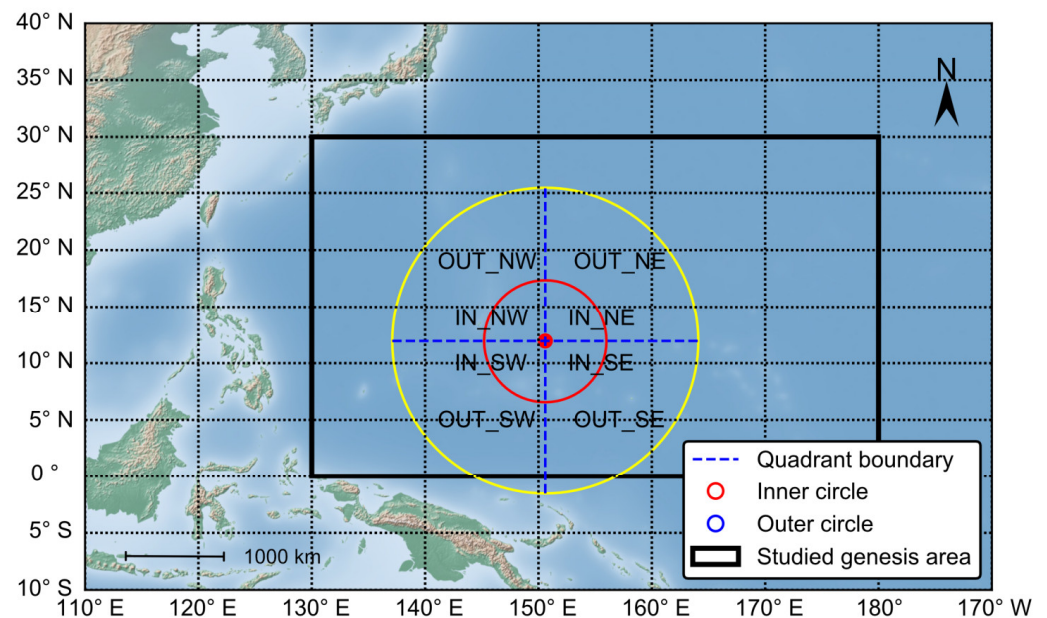

Figure 1. Illustration of studied genesis area (black rectangle) and eight storm-centered sectors where variables are averaged. Red point is location of TC center, dashed blue lines are boundaries of four quadrants, and red and yellow circles represent radii of 600 km (inner circulation) and 1500 km (outer environment), respectively.

Table 1. Number of TC cases in original dataset and after preprocessing. Preprocessing includes spatial restriction and temporal filter.

	TD/TS	**Minor TC**	**Major TC**	**Total**
Original	506	311	557	1474
After preprocessing	186	151	256	593

2.3. Calculation of Features

Given that the definitions of TC genesis are not always the same according to different scientific research and operational agencies [42,54], in this study, we define that a TC forms when its 1 min maximum sustained wind speed reaches 21 knots (1 knot equals about 0.5144 m s^{-1}) for the first time, which approaches the lower bound of TD (10.8 m s^{-1}) defined by the China Meteorological Administration (CMA).
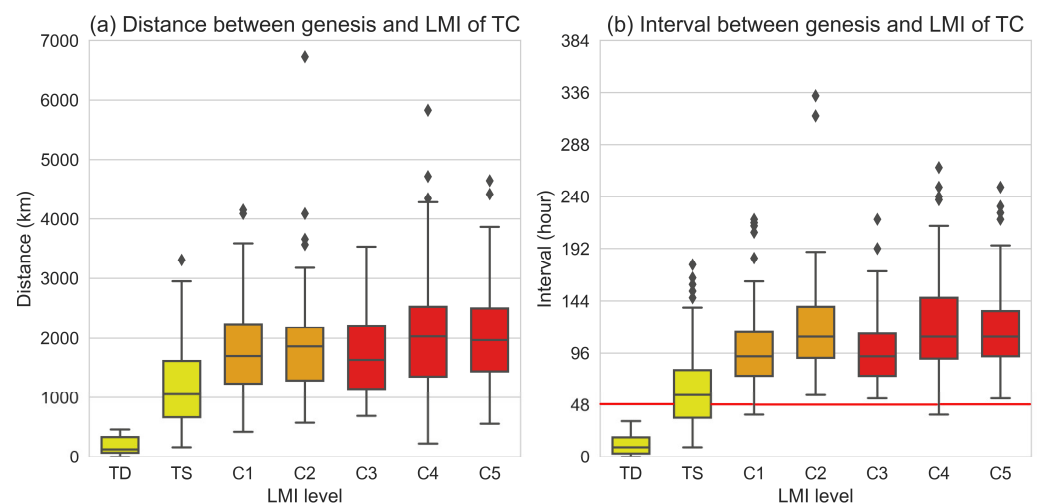

Figure 2. Box-and-whisker plots of (**a**) distance (km) and (**b**) interval (hour) between locations of TC genesis and LMI over the WNP grouped by LMI level (yellow for TD/TS, orange for minor TC, and red for major TC). Boxplot displays median (horizontal black line near box center), interquartile range (box perimeter; $[q_1, q_3]$), whiskers (black lines; $[q_1 - 1.5 (q_3 - q_1), q_3 + 1.5 (q_3 - q_1)]$), and outliers (rhombic points). The red horizontal line in (**b**) is a reference for the interval of 48 h.

In order to determine the temporal and spatial range of features, the occurrence time of LMI and location of TCs are investigated. As Figure 2a shows, storms with an LMI level of TD or TS (weak TC) will not travel too far from their genesis location at maximum intensity, especially TDs (<500 km). However, the medians of minor and major TCs (strong TC) are all in the range of 1500–2000 km, with quite small differences among them. The mean LMI location of weak TC is about 5° north and 5° west of the genesis location, while that for strong TC is about 7.5° latitude and 14° longitude (Table 2 and Figure 3). The mean genesis location of strong TC (13.470° N) is to the south of weak TC (17.101° N), but there is not much difference between the mean latitude of their LMI location (21.078° N and 21.925° N, respectively). These results are understandable, since strong TCs are created under more favorable environmental conditions (e.g., warmer sea surface temperature (SST)) and are potentially fueled by more energy to travel after genesis. On the other hand, higher latitude usually comes with worse environmental conditions for intensification, thus TCs reach LMI at a similar latitude no matter how strong they are. In order to obtain as much useful information as possible during their early lifetime and considering the asymmetrical structure of TCs, the corresponding variables are averaged within 8 arc-shaped sectors of different radii (600 km for the inner circle, referring to a TC's main circulation, and 600–1500 km for the outer circle, referring to the surrounding environment) and orientations in a storm-centered area (Figure 1). Compared with the

calculation method introduced by Ditchek et al. [41], this method better considers the round shape of TC circulation, and features are independent of each other. Moreover, we also found including an axisymmetric average (i.e., a circle over the TC center) as one of the features in the machine learning cannot change the results materially in terms of what variables are the most important for LMI, but otherwise performs badly on testing.

Table 2. Mean genesis and LMI locations of TCs and differences in two categories. All differences of averages are significant at a confidence level of 99%.

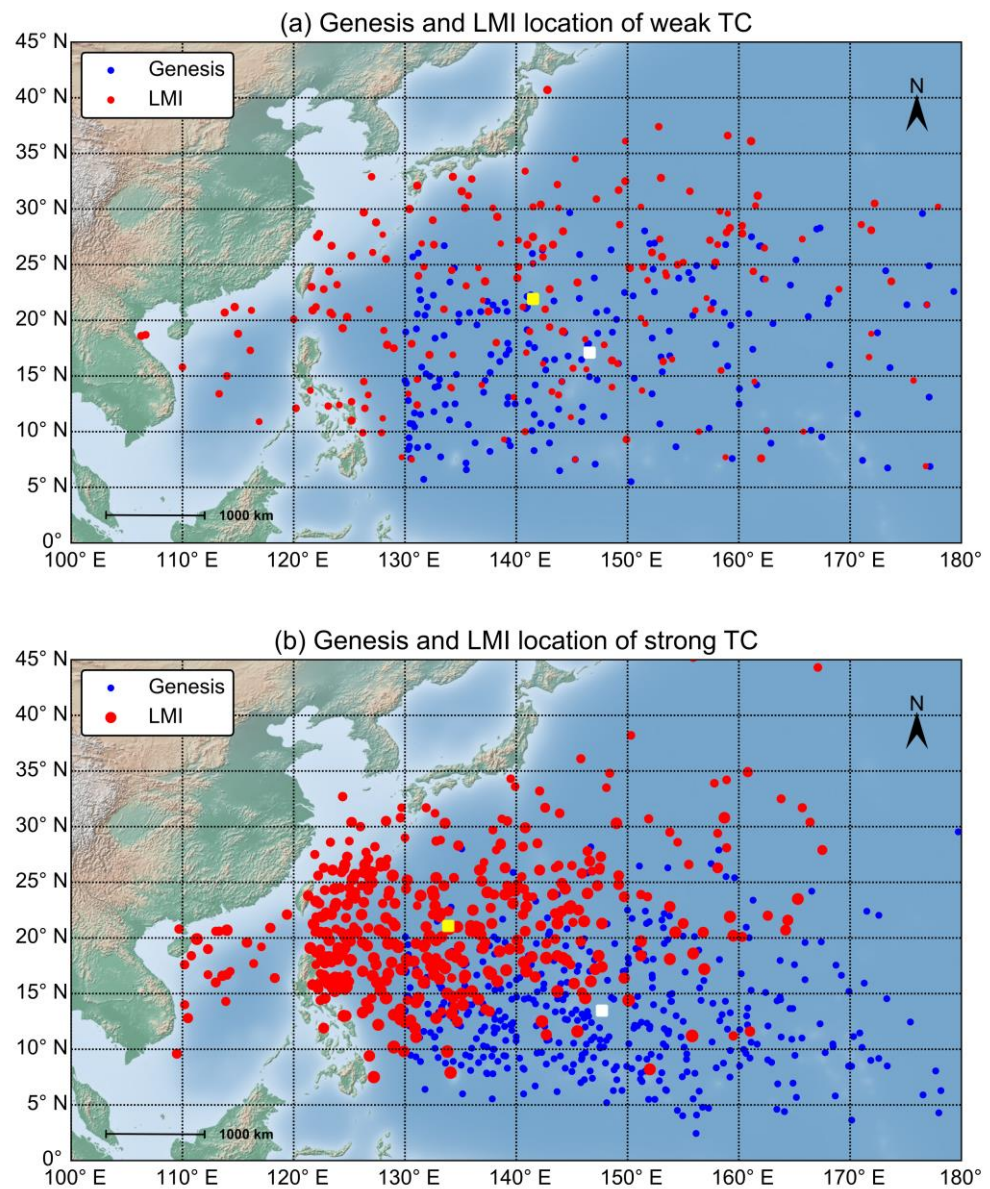|	Weak TC		Strong TC	
	Latitude	Longitude	Latitude	Longitude
Genesis	17.101° N	146.594° E	13.470° N	147.699° E
LMI	21.925° N	141.508° E	21.078° N	133.890° E
Difference	4.824°	−5.086°	7.608°	−13.809°

Figure 3. Locations of TC genesis (blue points) and LMI (red points) for (**a**) weak TCs (LMI of TD/TS) and (**b**) strong TCs (LMI of major and minor TCs). White and yellow squares indicate mean locations of genesis and LMI, respectively. Size of each point indicates intensity (knots) at that time.

Figure 2b shows the interval between TC genesis and LMI. The distribution is similar to that in Figure 2a, indicating that generally the stronger the LMI, the longer the TC interval. It is interesting that almost every strong TC (only 3 exceptional cases) experienced a "developing stage" for at least 2 days before reaching LMI after genesis. For weak TCs, the interval was quite short (e.g., less than 48 h for all TDs). Therefore, information from the first 48 h is available to represent early-stage conditions of strong TCs.

Similar to the process in the Statistical Hurricane Intensity Prediction Scheme (SHIPS) [6,7,55], features are divided into 2 groups in this study: (1) TC state features, which are scalars that describe the current status or variation trend of a TC such as size, moving direction, and translation speed (Table 3); and (2) environmental features, which are multidimensional variables that depict the dynamic or thermodynamic conditions of a TC, such as air temperature, relative humidity, and vertical wind shear (Table 4). Some of these parameters are crucial predictors in SHIPS for intensity prediction (e.g., SHRS and SHRD) [55], and some have a huge impact on TC genesis (e.g., translation speed) [11,12,56]. All of them are derived from ERA5 hourly reanalysis and the IBTrACS dataset. The method using reanalysis and actual best track data to establish a statistical model is known as the "perfect prognostic" methodology [57].

Table 3. Variables used as TC state features in the model.

Variable	Abbreviation	Unit
Day of year when TC generates *	JDAY	—
Intensity variation in past 6 h **	DV	Knot
Translation speed	SPD	Knot
Translation direction	DIR	Degree
Coriolis parameter	F	10^{-6} s^{-1}
Difference in Coriolis parameter **	DF	10^{-6} s^{-1}
Radius of maximum wind	RMW	Km
Radius of 3 m s^{-1} wind	R3	Km
Tightness	TI	—

* JDAY is made as a single feature since it does not vary with time. ** Due to the lack of TC information before genesis in IBTrACS, DV, and DF are not included in the classifier (step 1).

Table 4. Variables used as environmental features in the model.

Variable	Abbreviation	Unit	Vertical Level
Sea surface temperature	SST	°C	Surface
Maximum potential intensity	MPI	Knot	Surface
Relative humidity	RH	%	200/500/850 hPa
Air temperature	T	°C	200 hPa
Relative vorticity	VOR	10^{-5} s^{-1}	850 hPa
Divergence	DIV	10^{-5} s^{-1}	200 hPa
U-component of wind speed	U	m s^{-1}	200 hPa
Vertical wind shear of deep layer	SHRD	m s^{-1}	200–850 hPa
Vertical wind shear of shallow layer	SHRS	m s^{-1}	500–850 hPa
U-component vertical wind shear of deep layer	USHRD	m s^{-1}	200–850 hPa
U-component vertical wind shear of shallow layer	USHRS	m s^{-1}	500–850 hPa

As for the variables mentioned in Table 3, each one is averaged every 12 h during the first 2 days of a TC's lifetime to be a feature, except for JDAY (absolute value of genesis year-day minus 248). Specifically, the variables related to TC size are computed from 10 m wind data from ERA5 [58], since the corresponding information in IBTrACS is incomplete. In this way, the piecewise cubic Hermite interpolating polynomial (PCHIP) method [59] is employed to extract the radius of 3 m s^{-1} wind speed (R_3) and the radius of maximum wind (*RMW*) from the storm-relative azimuthal-mean radial profiles. They represent the storm

sizes of the inner and outer core, respectively. Similar to the concept of TC fullness [60] as the ratio of the TC's outer-core wind skirt to outer-core size, tightness is calculated by:

$$TI = \frac{R_3 - RMW}{R_3} = 1 - \frac{RMW}{R_3}. \tag{1}$$

By quantitively measuring the TC's outer-core wind structure, this variable describes the destructiveness of the storm to some extent.

Variables listed in Table 4 are also averaged within 8 sectors to be a feature in the model after temporal averaging (Figure 1). The maximum potential intensity (*MPI*) used in this study is calculated by an empirical function derived from the observed maximum intensity of TC with respect to SST [55,61], rather than the theoretical form raised by Emanuel [62]:

$$MPI = A + Be^{C\ (T - T_0)}. \tag{2}$$

The coefficients in this exponential function are given by A = 38.21 kt, B = 170.72 kt, C = 0.1909 C^{-1}, and T_0 = 30.0 C^{-1}, and 185 kt is set as the upper boundary of MPI.

2.4. Ensemble Learning Method

The decision tree model mimics how people think about a problem and finally make decisions, based on the rules organized in a tree shape [63]. It has a variety of forms, and one of them is the classification and regression tree (CART), which typically uses the Gini index as the rule to choose the best splitting feature at each node in classification [64]:

$$Gini(D) = \sum_{k=1}^{K} p_k(1 - p_k) = 1 - \sum_{k=1}^{K} p_k^2, \tag{3}$$

$$Gini\ index(D,\ a) = \sum_{v=1}^{V} \frac{|D_v|}{|D|}\ Gini(D_v). \tag{4}$$

where D and a refer to the original dataset and the selected feature, respectively, K is the total number of features, V represents the number of possible values of a, p_k is the probability of the sample belonging to class k, and D_v is the subset split by a. The *Gini index* shows the "impurity" of the subsets by calculating the possibility that two randomly chosen samples in a subset have different actual labels. A low Gini index suggests that the subset split by a is quite homogeneous, hence it is useful for classification [65]. After the training is finished, the model will be able to classify new samples into certain categories by judging their features step-by-step. CART can handle both classification and regression issues well, with good capacity for interpretation, and acquires less training data than artificial neural networks [66]. The detailed algorithms for CART are provided in Appendix A.

Since a single decision tree is prone to overfit the training data by generating too many branches [67], we use "pre-pruning" procedures (e.g., restricting the maximum depth of a single tree) to prevent an unnecessarily complicated structure, and use ensemble to resist overfitting. Ensemble learners contain sets of weak learners, and three ensemble learning methods based on CART were applied in this study: Adaptive Boosting (AdaBoost), Extreme Gradient Boosting (XGBoost), and random forest [68–70]. AdaBoost and XGBoost are boosting models that train base models in series to reduce the bias by changing the weight distribution of samples at each step. Random forest is a typical "bagging" algorithm that has a parallel framework to reduce variance by constructing many decision trees. The detailed algorithms for tree-based ensemble models are provided in Appendix B. Generally speaking, ensemble learning methods are much more accurate and robust than individual decision tree models [71,72].

Similar to the individual decision tree model, the tree-based ensemble model not only has good performance on classification and regression tasks, but is also available to trace the contribution of each feature. Along with node division by the values of splitting

features, the decreased impurity in subsets is maximized at each step. Mean decrease impurity (MDI) is employed to judge the importance of feature x_m when the splitting point is set as s at node t, whose value equals the mean decrease of selected metric i over all nodes and all trees [70]:

$$MDI(x_m) = \frac{1}{N_T} \sum_{T_i} w(T_i) \sum_{t \, \in \, T_i \; : \; v(s_t) = x_m} p(t) \, \Delta i(s, \, t) \, , \qquad (5)$$

where N_T is the number of decision trees in ensemble model T, $p(t)$ is the fraction of subset at node t in a decision tree, $\Delta i(s, t)$ refers to the decreased impurity measured by the selected splitting criterion, $w(T_i)$ is the weight of decision tree T_i ($w(T_i) \equiv 1$ in random forest), and $v(s_t)$ is the value of the feature used in partition. Since we chose the Gini index as the splitting criterion for all models, we call the normalized MDI the Gini importance index (GII; not the same as the Gini index).

However, critical features assessed by only one criterion may be misleading, as the GII will be abnormally high when applied to high cardinality features [64]. To ensure the robustness of selected features, two other criteria, mean minimum tree depth (MMTD) and total split time (TST), are also considered quantitative indicators of feature importance. In tree-based models, the earlier and more frequently a feature is selected, the more important it is. Therefore, if the feature has a high GII, a small MMTD, and a large TST, then it is significant for LMI estimation.

2.5. Workflow of the Model

In order to better capture the detailed factors of LMI for TCs with different intensity, we developed a two-step model to estimate the LMI of a formed TC based on a classifier and a regressor (Figure 4). The first step of the model is to judge whether or not a storm will become a strong TC by learning its genesis features (step 1). Since we are less interested in the specific intensity that a weak TC will finally reach, the next step of the model further explores the exact intensity of strong TCs only (step 2), where features during the first 48 h after genesis are considered.
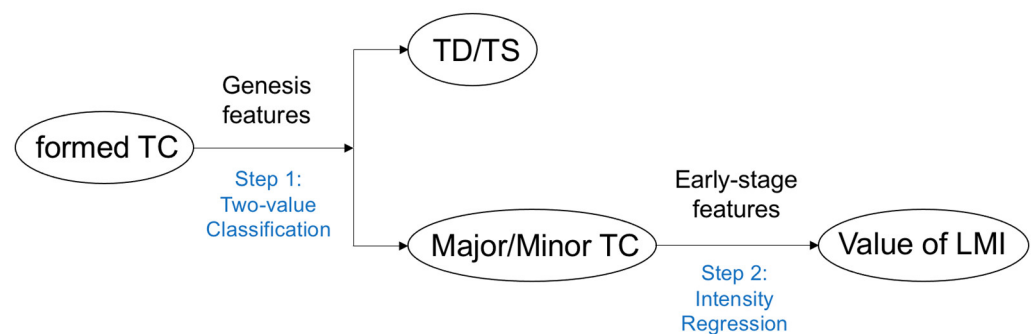

Figure 4. Flow diagram of two-step LMI analysis in this study. TD/TS is also called weak TC and major/minor TCs are collectively named strong TCs.

TC cases are randomly divided into two parts to establish the model: the training set, used to tune the parameters of the model, and a testing set, used to evaluate its performance. The ratio of the two subsets is 5/1 in this study. During the training process, the three ensemble methods mentioned above are applied to the training set to tune its critical parameters in the two steps (Appendix B). Meanwhile, k-fold cross-validation [73] is applied to the training set to verify the capability of the model (k = 10 in this study). The training set is divided equally into k subsets; then, training and testing are performed for k iterations. During each iteration, one subset is selected for validation while the remaining k–1 subsets are used to tune the parameters without overlap, so that each sample of the dataset can be used for training and validation. Finally, the well-tuned model is assessed in

the testing set. In step 1, we use accuracy and F1-score as the metrics to evaluate the fitting capability of classifier:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{6}$$

$$F1 = \frac{2}{\left(\frac{1}{P} + \frac{1}{R}\right)}, \tag{7}$$

where P is precision and R is recall, calculated as:

$$P = \frac{TP}{TP + FP}, \tag{8}$$

$$R = \frac{TP}{TP + FN}. \tag{9}$$

The meanings of the double-letter variables in Equations (6)–(9) are explained in the confusion matrix (Figure 5). Accuracy indicates the correctness of all decisions, and the F1-score is a comprehensive term that judges the robustness of a classifier. The model will get a high F1-score only when precision and recall are both high, with precision measuring the quality of predicting true positive cases and recall measuring the completeness of the classifier's judgment. In step 2, the coefficient of determination (R^2) and root mean square error (RMSE) are the two main metrics to evaluate the fitting capability of the regressor.
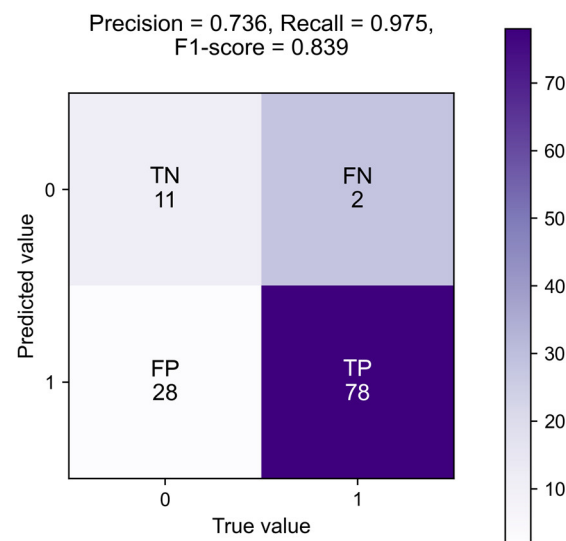

Figure 5. Confusion matrix with best score in testing set by AdaBoost classifier in step 1. Numbers in squares indicate amount of corresponding TC cases. For both true and predicted value, 0 refers to TD/TS (negative) and 1 refers to major TC/minor TC (positive).

After the three ensemble methods are well tuned for their optimum parameters, the one showing the best performance on the testing set is selected as the benchmark in steps 1 and 2. To better understand the contributions of different features to the LMI of TCs, the GII, MMTD, and TST of the benchmark are assessed to determine the most important features. After that, the leading features are analyzed through storm-centered composites of different LMI groups. Comparing their horizontal distribution and temporal variation can show how the differences happen at the early stage of a TC's lifetime.

3. Results

3.1. Features Related to LMI at TC Genesis

In step 1, 111 features of 593 samples at genesis were applied to establish the classification model distinguishing whether a storm will develop into a weak or strong TC, and

the fitting results for the three ensemble methods are shown in Table 5. It is clear in the table that whether accuracy or F1-score is chosen as the criterion, the classifier based on AdaBoost ensemble is ranked first (accuracy of 0.7479 and F1-score of 0.8387). It is also optimum in terms of robustness (Appendix C). This suggests that a TC's LMI is related to its vortex attributes and environmental conditions at genesis over the WNP. As the aim of this study is to discuss the impact factors of LMI rather than to provide operational forecasting of LMI, the result of the AdaBoost classifier is good enough to ensure that the following factor diagnosis is reliable. Therefore, it serves as the benchmark of step 1.

Table 5. Best results of fitting by three ensemble methods in steps 1 and 2.

	Step 1		Step 2	
	Accuracy	**F1-Score**	**RMSE (Knots)**	**R^2**
AdaBoost	0.7479	0.8387	23.7697	0.3004
XGBoost	0.6975	0.8105	24.0103	0.2861
Random Forest	0.7227	0.8156	25.3060	0.2070

Figure 5 shows the confusion matrix of the result produced by the AdaBoost classifier. It gets an F1-score of 0.839 with a high recall of 0.975 and a low precision of 0.736, mainly caused by the large amount of FP cases (28 of 119). It is not surprising that the model tends to overestimate the LMI of weak TCs but rarely underestimates that of strong ones. Most TCs that form under favorable conditions suffer from disadvantageous factors after genesis along their tracks (e.g., close to land) and will not attain a high LMI. However, this situation cannot be captured by the model, since it learns information at genesis only. Furthermore, the imbalance of the dataset induced by the relatively small proportion of TD/TS cases (186 of 593) makes it harder for the classifier to learn the genesis features of weak TCs. Nevertheless, the model does show some skill in classification.

The relative importance of features at genesis in step 1 assessed by GII, MMTD, and TST is depicted in Figure 6; the most important features are highlighted by red points in the upper left of the figure (MMTD $\leq$ 6.0, GII $\geq$ 0.015, TST $\geq$ 250) and ordinary features are in blue. As the figure shows, TC vortex vorticity at genesis has the biggest impact on LMI, with the most significant region northwest of the TC's inner circulation. Vertical wind shear of deep and shallow layers, relative humidity at the upper troposphere, and translation speed at genesis are also key features. It is notable that two points in the lower left of Figure 6 are far from the cluster (MPI_OUT_NW and USHRD_IN_NW), suggesting that features judged by only one criterion may be misleading, so it is necessary to assess their relative importance by multiple metrics.

3.1.1. Relative Vorticity at 850 hPa

The most important feature in step 1, storm-centered composites of relative vorticity at 850 hPa in two groups and their differences, are shown in Figure 7a–c. For both weak and strong TCs, the storm is situated in a continuous large vorticity band connecting to the west (greater than 2×10^{-5} s^{-1}), and the gradient near the storm center is also large. However, as was found in NA [41], the eastern side of a weak TC's outer environment is covered by negative vorticity, with two features showing their evident difference (VOR850_OUT_NE and VOR850_OUT_SE). Because most TCs over the WNP form to the south of the subtropical high, this suggests that TCs that reach high LMI tend to generate at a distance from the subtropical high, or when it is weak.
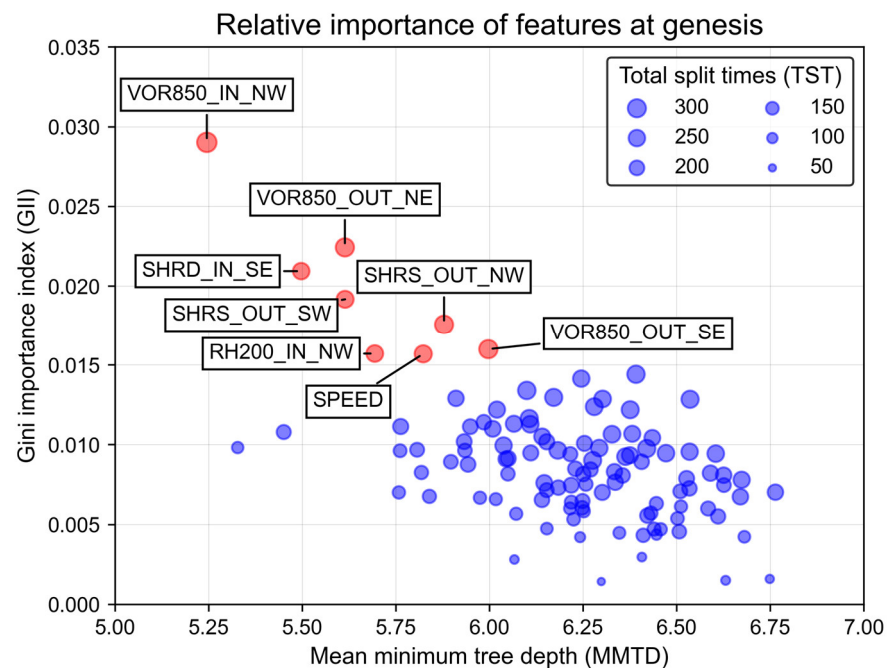
Figure 6. Scatter plot of features at genesis showing their relative importance to LMI in step 1. x- and y-axes refer to weighted mean minimum tree depth (MMTD) of all base estimators and Gini importance index (GII), respectively, with size of points controlled by total split time (TST) in all trees of the ensemble model. Red points are features with the greatest importance judged by the three criteria, and blue scatters are ordinary features. Text after feature names in rectangles denote sectors indicated in Figure 1.

From the difference field (Figure 7c), we can detect a region with homogeneous positive values of 0.5–$1.0 \times 10^{-5}\,\mathrm{s}^{-1}$ northwest of the TC's inner circulation in accordance with the most important feature in step 1 (VOR850_IN_NW), and a region with negative values at the southwest of the inner circulation. As the wind vectors show, the main circulation of strong TCs (within a radius of 600 km) seems more symmetrical about the zonal axis than that of weak TCs. This might be a signal that storms that organize with a symmetrical circulation at genesis have a greater chance to reach higher LMI.

3.1.2. Local Vertical Wind Shear

Previous studies have recognized the remarkable impact of vertical wind shear on the generation and intensity variation of TCs [20,74]. Figure 8 depicts the local vertical wind shear in two groups and their differences. In terms of the local shear of deep layer (Figure 8a–c), the patterns in weak and strong TCs are quite similar, both characterized by a narrow zonal band with low values about 8–$10\,\mathrm{m\,s}^{-1}$ across the storm center, and higher values at the northern and southern sides. The most significant difference is the wider region of strong shear at the north and east of the storm in weak TCs, where the maximum difference exceeds $4\,\mathrm{m\,s}^{-1}$. Since there is little difference in wind fields at 850 hPa (Figure 7a,b), this is mainly induced by the smaller range of anticyclonic flow to the east of the storm center at 200 hPa in strong TCs. A more compact anticyclone nearer to the storm center is observed in strong TCs, while in weak TCs the outflow extends farther northward before wrapping back southward, leading to the ventilation of energy away from the circulation [16]. Therefore, it can be inferred that a compact circulation in the outflow layer at genesis is indicative of better conditions for TCs to attain higher LMI. However, only one feature related to deep-layer shear is vital in step 1 (SHRD_IN_SE). This may result from some extremes, which can dramatically influence the composite fields, but the corresponding feature may be not indicative for classification.
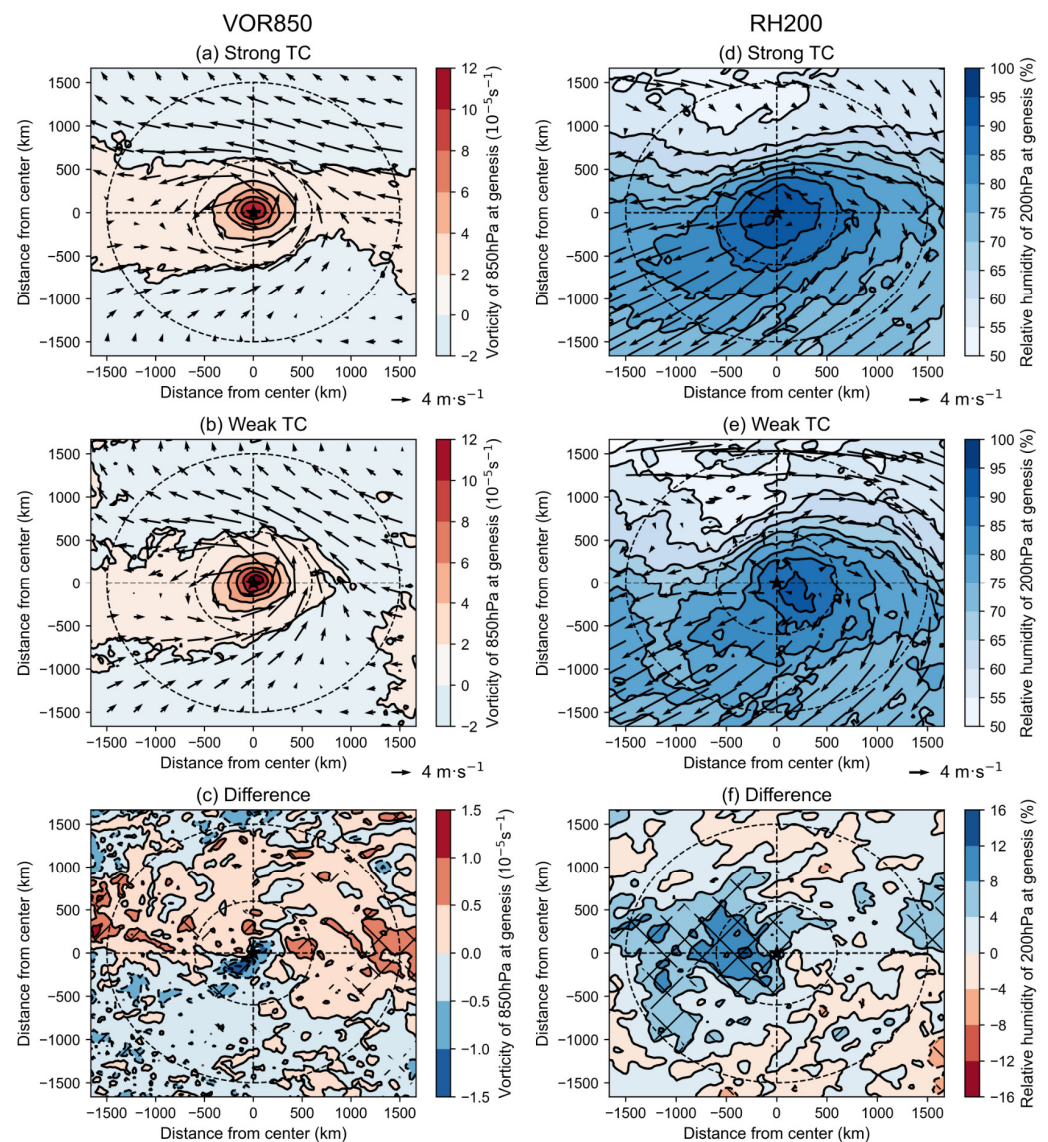
Figure 7. Storm-centered composites of (**a**) relative vorticity (VOR850) by contours with wind vectors at 850 hPa (10^{-5} s^{-1}) and (**d**) relative humidity at 200 hPa (RH200) by contours with wind vectors at 200 hPa (%) for strong TCs at genesis. (**b**,**e**) are the same as (**a**,**d**), but for weak TCs. (**c**,**f**) Difference fields of strong minus weak TCs. Black star represents storm center, and dotted black lines are boundaries of eight sectors discussed in Figure 1. Areas with crossing lines in (**c**,**f**) depict where differences between two categories are statistically significant at the 95% confidence level.

Both strong and weak TCs feature a cyclonic vortex in the middle layer of the troposphere, and there is a region with weak shallow-layer wind shear at the north of the storm center (Figure 8d,e). In terms of the wind shear of the shallow layer, SHRS_OUT_SW and SHRS_OUT_NW are selected as key features in step 1, which roughly conform to the two statistically significant regions in Figure 8f. Due to the similarity in wind field at 850 hPa between the two groups (Figure 7a,b), we attribute this difference to the storm's circulation at 500 hPa. Comparing Figure 8d,e, it is shown that weak TCs feature stronger southwest winds to the southwest of the outer environment and weaker easterlies to the north of the storm center, which means the circulation in the middle layer is also weaker compared with strong TCs. As a result, weak TCs have a greater chance to draw in more dry air with low potential vorticity at the middle level from the surrounding environment; hence, their intensification is hindered [75].
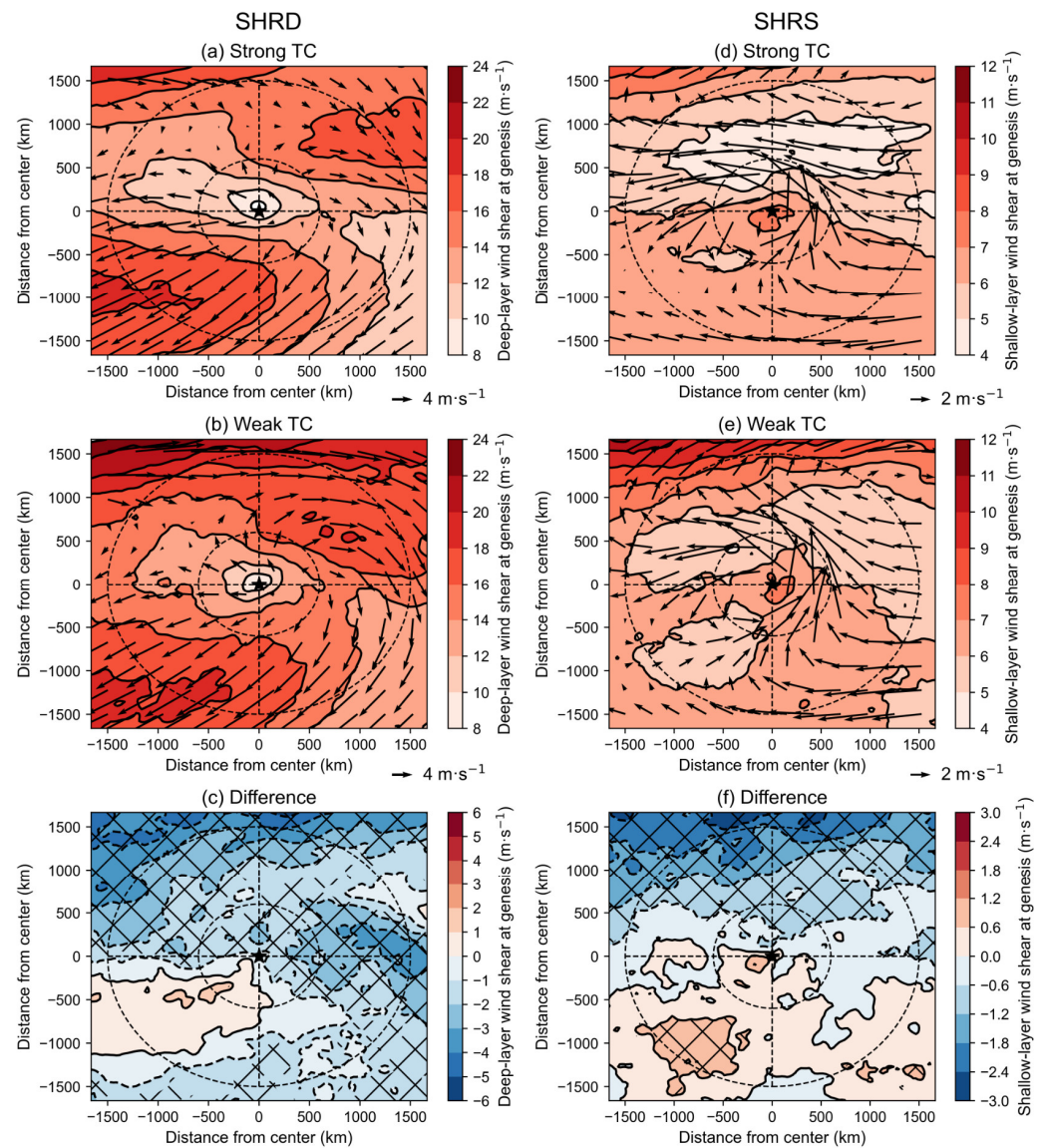
Figure 8. As in Figure 7, but for (**a**–**c**) deep-layer vertical wind shear (SHRD) of 850–200 hPa with wind vectors at 200 hPa (m s^{-1}) and (**d**–**f**) shallow-layer vertical wind shear (SHRS) of 850–500 hPa with wind vectors at 500 hPa (m s^{-1}).

3.1.3. Relative Humidity at 200 hPa

Only one key feature related to relative humidity (RH200_IN_NW) is selected in step 1, which indicates the valid difference in moisture conditions at the upper level of the troposphere between the two groups (Figure 7d–f). In general, there is little moisture at the upper troposphere because ordinary convections can barely reach there [76]. However, high relative humidity (nearly 100%) covers the storm center in both strong and weak TCs, due to the low saturated water pressure. There are some similarities between TCs in the two groups. There is greater moisture to the south and its gradient is quite large at the north of the storm center. However, moisture at the west of the storm center in weak TCs is not as abundant as in strong TCs (the largest difference exceeds 12%), while strong TCs have round-shaped and symmetrical wet areas around the storm center. In addition, the gradient of relative humidity at the key region (northwest of a TC's inner circulation) is greater in weak TCs, which means the storm is embedded in a drier environment. This difference implies that TCs with high LMI may have stronger and deeper convection at genesis, which humidifies the outflow on the northwestern side.

3.1.4. Translation Speed

The translation speed of storms is also found to be indicative in step 1. Overall, strong TCs move a little slower than weak TCs at genesis (average speed 9.34 kt versus 10.35 kt), and the difference is statistically significant at the 95% confidence level. This is contrary to a previous study indicating that the enhancement of TC intensity is restrained by cold water upward from the deep ocean due to the pumping effect when the storm remains in a certain location for a long time [77]. On the other hand, TCs are usually formed in the tropics with a warm underlying surface, so interaction with warmer seawater for a longer time around genesis provides a better chance for the storm to gain heat flux from the ocean and develop quickly. Moreover, since the study focuses on the early lifetime of TCs when the wind speed of circulation is very low, the latter factor may have an advantage over the former in affecting LMI. That is to say, TCs with a slower translation speed at genesis have a greater chance to attain higher LMI.

3.1.5. Other Features

Some features related to the critical factors in the generation and intensity variation of TCs are not selected in step 1 (e.g., SST, relative humidity at middle troposphere, divergence at upper troposphere) because they do not differ much between the two groups. Taking SST for instance, all of the TC cases investigated in this study form under similar thermodynamic conditions of the ocean (Figure 3), so it is hard to distinguish their LMI by features computed from a region-averaged SST. Similarly, MPI is also filtered by two metrics, although it has a particularly small MMTD (Figure 6). This does not mean that it does not contribute to TC genesis and intensity variation, but it is not a key feature affecting LMI. A similar explanation may also be applied in step 2.

3.2. Features Related to LMI at Early Stage

In step 2, only minor and major TCs are investigated, and 449 early-stage features of 407 samples are applied to establish the regression model (step 2) estimating the LMI of strong TCs. The results of fitting by the three ensemble methods are shown in Table 5; it is clear that the AdaBoost ensemble method again ranks first (RMSE of 23.7700 kt and R^2 of 0.3004). Figure 9 depicts the comparison between estimated and actual LMI in the testing set, which resembles Figure 5 in Ditchek et al. [41]. The regression line of estimated values has a smaller slope than line $y = x$, suggesting that step 2 is effective but has poor performance on the extremes, similar to most machine learning models [78]. It implies that the LMI of strong TCs could be affected by early-stage factors. Since we are seeking a reasonable relationship between these factors and LMI rather than a perfect prediction, the results produced by the AdaBoost-based model are considered credible and were used to further discuss the relative importance of features.

As in Figure 6, the relative importance of features during the first 48 h after TC genesis is depicted with GII, MMTD, and TST in Figure 10. Unlike the close positions of scatters in Figure 6, the features in step 2 are dispersed in Figure 10 and have an approximately linear distribution from the upper left to the lower right, suggesting that the key features selected by the three metrics in this step are quite robust. Many TC state features are considered to be crucial in step 2, which is a signal that vortex attributes of TCs begin to differentiate during this period. On the other hand, the most critical environmental features are nearly the same as those in step 1: deep-layer vertical wind shear, high-level relative humidity, and low-level vorticity, with the key interval of 24–48 h after TC genesis. This implies that these features have a great influence on LMI at the TC development stage as well as at genesis.
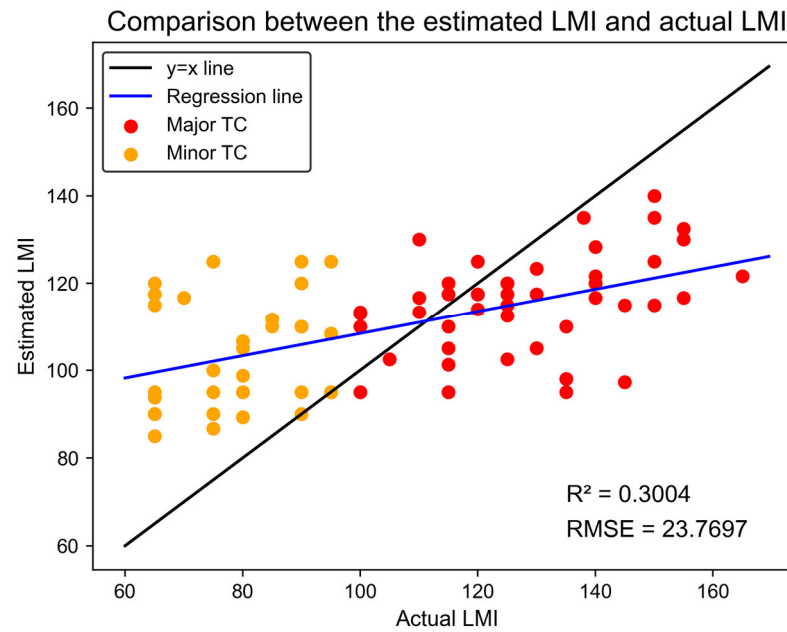
Figure 9. Comparison between estimated and actual LMI (knots) in the testing set, colored according to actual LMI (red for major and yellow for minor TC). Black line is $y = x$ and blue line is regression line of estimated values. Values of metrics for evaluating fitting are in the lower right. Correlation is statistically significant at the confidence level of 99%.
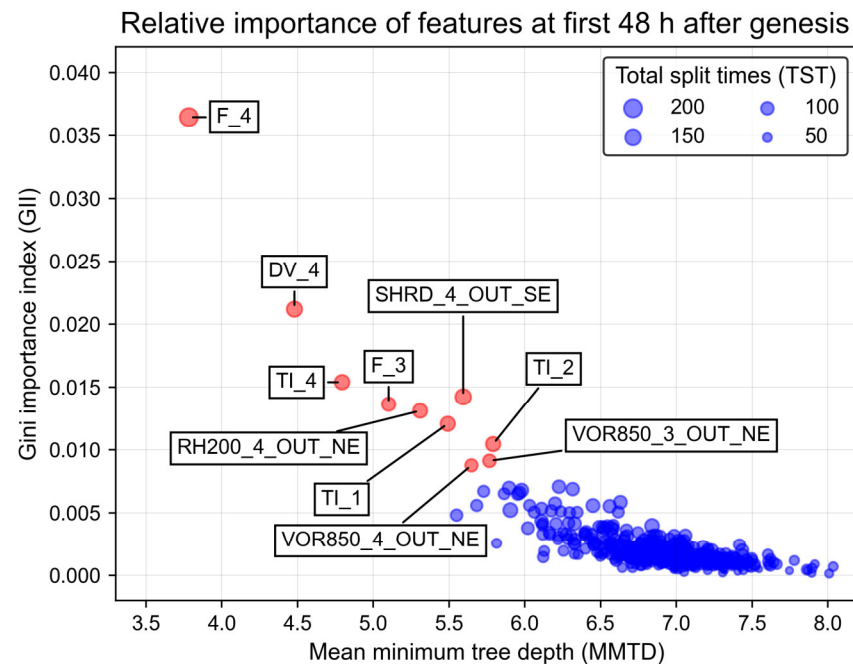
Figure 10. As in Figure 6, but for features during the first 48 h after genesis in step 2.

3.2.1. TC State Features

Variations of critical TC state features and differences between two groups during the first 48 h after genesis are illustrated in Figure 11. The averaged Coriolis parameters of two intervals (24–36 h and 36–48 h; Figure 11a) are found to be effective in step 2 (F_3 and F_4). It can be inferred that major TCs tend to stay in a lower latitude with a slower poleward motion, and the difference accumulates as time elapses (beyond 7.5 over 36–48 h). This agrees with step 1, in that TCs with larger LMI spend more time in the tropics obtaining energy from warmer seawater around genesis. As the difference in averaged translation

speed between major and minor TCs gets bigger (0.5 m s^{-1} over 0–12 h but 0.78 m s^{-1} over 36–48 h after genesis, not shown), the difference in the Coriolis parameter also becomes larger, making the feature indicative for LMI estimation.
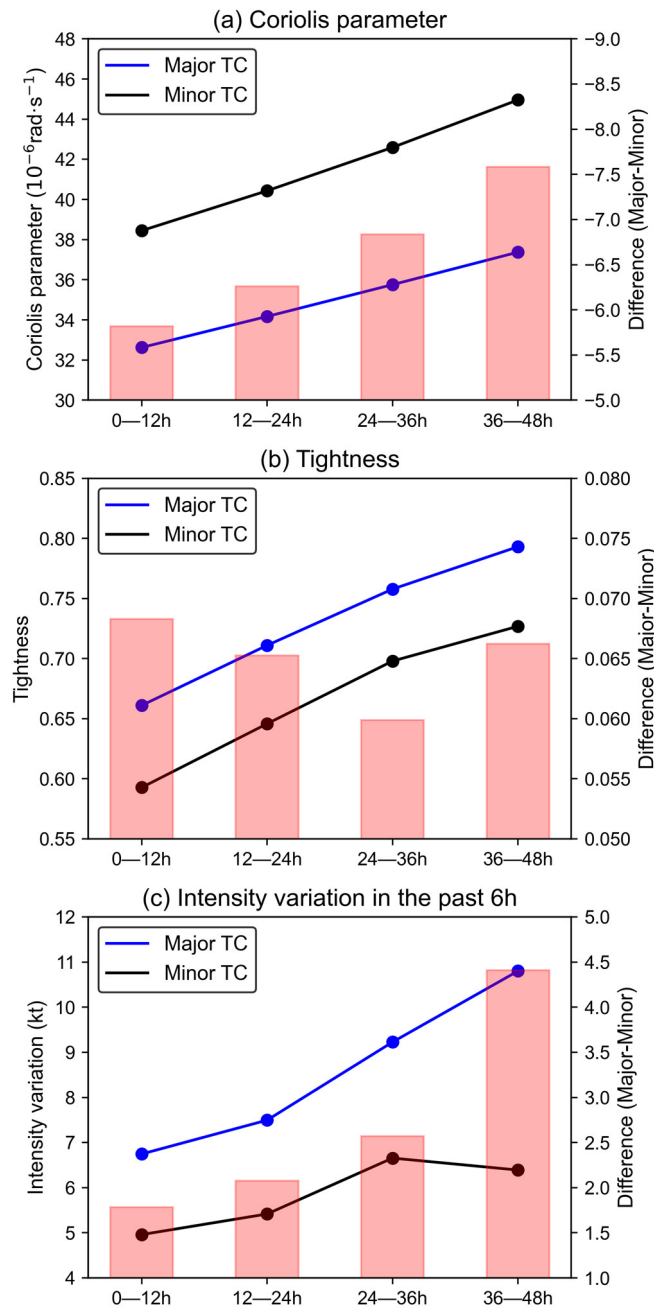

Figure 11. Composite evolution of critical TC state features (blue and black lines) of (**a**) Coriolis parameter (F), (**b**) tightness (TI), and (**c**) intensity variation in the past 6 h (DV6) and their corresponding differences between major and minor TCs (red bars) during first 48 h after genesis. Except for first three intervals in (**c**), differences of all intervals are significant at a confidence level of 99%.

Tightness has a similar increasing trend with the Coriolis parameter during the early lifetime of TCs (Figure 11b). As mentioned above, tightness is a term that describes the extent of a "valid" wind structure showing its destructiveness; a greater tightness value (major TC) indicates a better-defined storm circulation. During this period, major TCs have greater tightness than minor TCs, but the difference between the two decreases sharply at the interval of 24–36 h after genesis, possibly as result of the eyewall replacement cycle

(ERC) process, which often takes place after a TC's rapid intensification (RI; i.e., intensity increasing more than 30 knots in 24 h). During the ERC process, the RMW of the storm suddenly enlarges, leading to a decrease in tightness [40]. Among all the cases in step 2, 46 major TCs (17.97%) experienced RI during this period, but only 11 minor TCs (7.28%) did, which supports our hypothesis. As a result, tightness at three intervals (TI_1, TI_2, and TI_4) shows its importance to LMI at the early development stage of TCs over the WNP.

As for 6 h intensity variation, the difference between major and minor TCs is only notable at the interval of 36–48 h (nearly 4.5 knots every 6 h), which is matched by a key feature selected in step 2 (DV_4). During this period, major TCs keep developing fast, but minor TCs have a drop in the intensification rate (from about 2.4 m s^{-1} to 2.2 m s^{-1}), which makes the difference suddenly increase (Figure 11c). Similar to the evolution of tightness, this is probably related to the RI process. During 30–42 h after genesis, 47 major TCs (18.36%) began to rapidly intensify, but only 9 minor TCs (5.96%) did. Because major TCs are more likely to go through the RI process, the result indicates a key interval when most major TCs will begin to intensify rapidly.

3.2.2. Local Vertical Wind Shear of Deep Layer

Despite the lack of features describing shallow-layer shear, local wind shear of the deep layer is found to be critical in step 2 (SHRD_4_OUT_SE). There is no obvious difference between composites of major and minor TCs (Figure 12a,b), both of which resemble the genesis field in Figure 8a. The biggest difference (about 3 m s^{-1}) at the southeast of the outer environment is due to weaker deep-layer shear in major TCs. This difference is mainly caused by the weaker anticyclonic flow at 200 hPa of major TCs, since the difference in the wind field at 850 hPa is very small between the two groups (Figure 13d,e). Since the difference takes place around genesis, this may result from the faster organization of deeper convection and higher outflow layer by major TCs.

3.2.3. Relative Humidity at 200 hPa

Figure 12d,e respectively depict the composite fields of relative humidity at 200 hPa of major and minor TCs, and Figure 12f shows their difference. Except for the wetter environment around the storm center, major TCs have similar moisture distribution to minor TCs. As implied by the key feature of RH200_4_OUT_NE, there is a key region northeast of a TC's outer environment for LMI (the biggest difference exceeds 8 m s^{-1}). Here, the environmental air of major TCs is extremely dry, where the gradient of relative humidity reaches its maximum. This could be the consequence of stronger compensating subsidence in the environment. Meanwhile, the anticyclonic circulation of major TCs is also stronger. These characteristics imply that the upper-layer structure of major TCs is more compact, with higher inertial stability, which is favorable for TCs to intensify continuously [36]. This difference is not obvious in the genesis field when the circulation is not well established.

3.2.4. Relative Vorticity at 850 hPa

There are two key features describing the low-level vorticity of TCs in step 2 (VOR850_3_OUT_NE and VOR850_4_OUT_NE). Since they are calculated from two successive intervals, their composites and difference fields are quite similar (Figure 13a,b,d,e). Similar to the situation in genesis fields, stronger TCs are situated in more continuous vorticity bands with greater convergence of southwest wind and easterlies to the east of the storm center. There are two significant regions with large values in the difference fields (Figure 13c,f): a negative one lying in the inner circulation around the storm center, and positive one at the east of the outer environment. The former can be explained by the fact that major TCs usually have a smaller inner core than minor TCs. Therefore, the difference fields are covered by positive values within a radius of about 200 km, but the values outside are negative. As a result, the mean vorticity of the inner circulation is similar in the two groups; thus, the corresponding features are not selected in step 2. The latter region could be

attributed to the stronger low-level easterlies at the northeast of a storm, which can interact with some tropical systems such as monsoon troughs to make a TC intensify continuously.
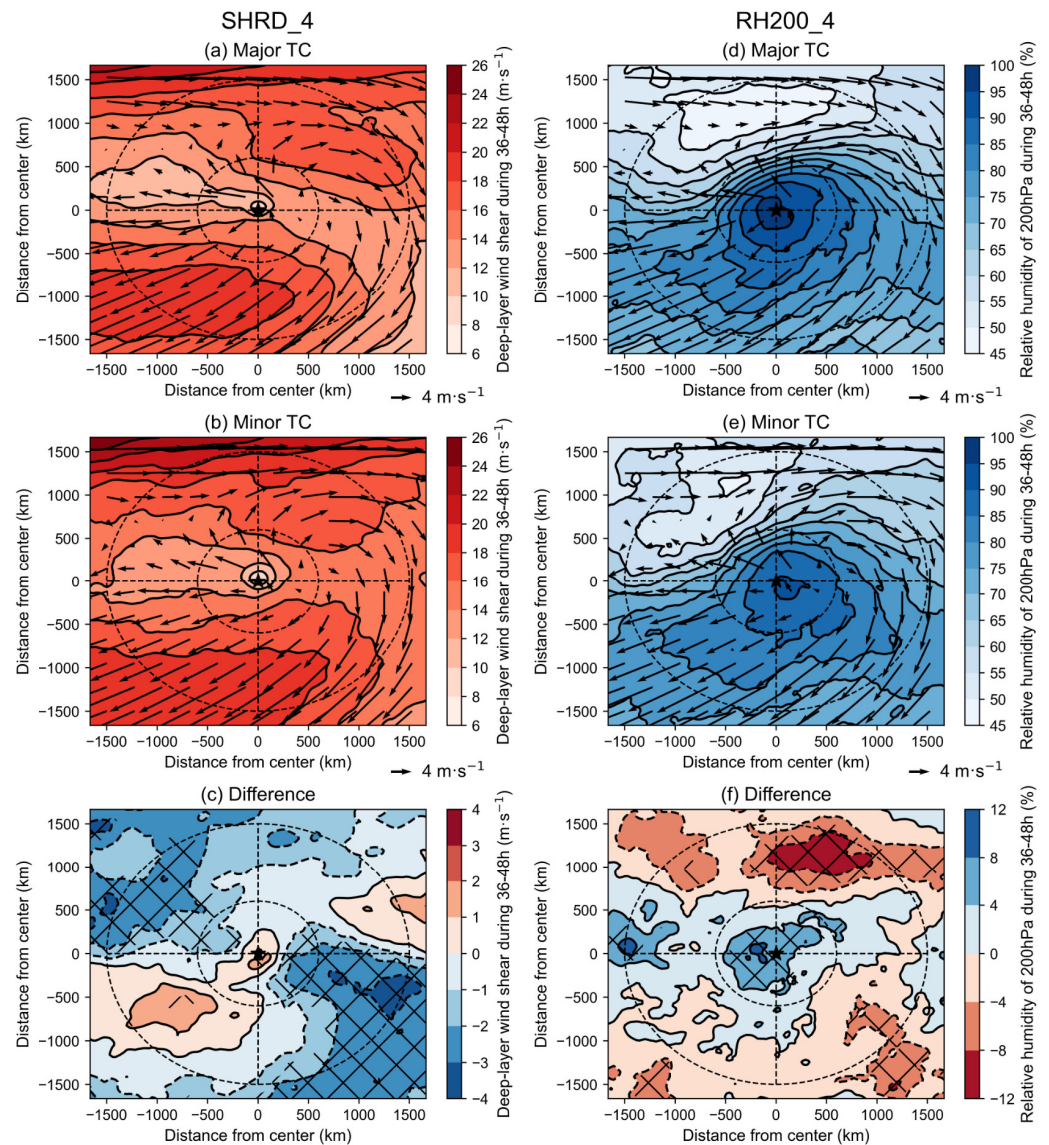

Figure 12. Storm-centered composites of (**a**) deep-layer vertical wind shear of 850-200 hPa during first 36–48 h after genesis (SHRD_4) by contours with wind vectors at 200 hPa (10^{-5} m s^{-1}) and (**d**) relative humidity at 200 hPa during first 36–48 h after genesis (RH200_4) by contours with wind vectors at 200 hPa (%) for major TCs. (**b**,**e**) Same as (**a**,**d**), but for minor TC. (**c**,**f**) Differences in major minus minor TCs. Black star represents storm center and dotted black lines are boundaries of eight sectors discussed in Figure 1. Areas with crossing lines in (**c**,**f**) depict where differences between the two categories are statistically significant at the 95% confidence level.
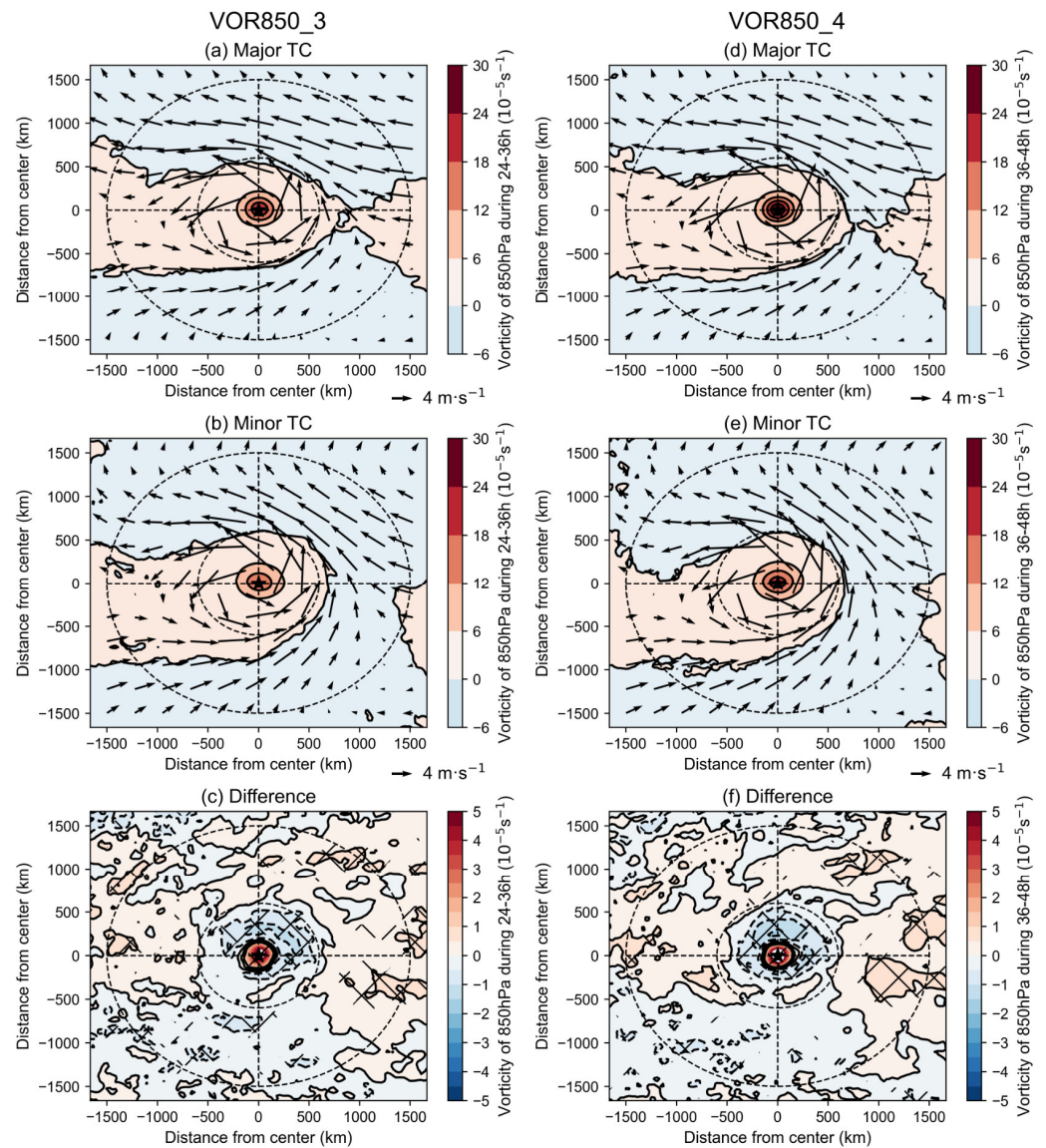
Figure 13. As in Figure 12, but for relative vorticity with wind vectors at 850 hPa (10^{-5} s^{-1}) during (**a**–**c**) 24–36 h and (**d**–**f**) 36–48 h after genesis.

4. Summary and Discussion

A two-step statistical model to investigate the relationship between the early-stage features and LMI of TCs over the WNP was established by the AdaBoost ensemble learning method in this research. The first step was to discriminate between TS/TD and stronger TCs at genesis, and the second step was to estimate the intensity of major and minor TCs. Composite analysis was then conducted to compare the differences in critical features between TCs with different intensities. Features used in the statistical models were obtained from ERA-5 daily reanalysis and IBTrACS datasets; the studied TCs were generated from June to November over 41 years (1979–2019) over a region ranging from latitudes of 0–30° N and longitudes of 130–180° E. Through the procedures described above, critical features of the LMI of TCs and their relative importance were identified. The key intervals and quadrants of critical features are highlighted in Table 6 and Figure 14.

Table 6. Critical TC state features and their key intervals and quadrants.

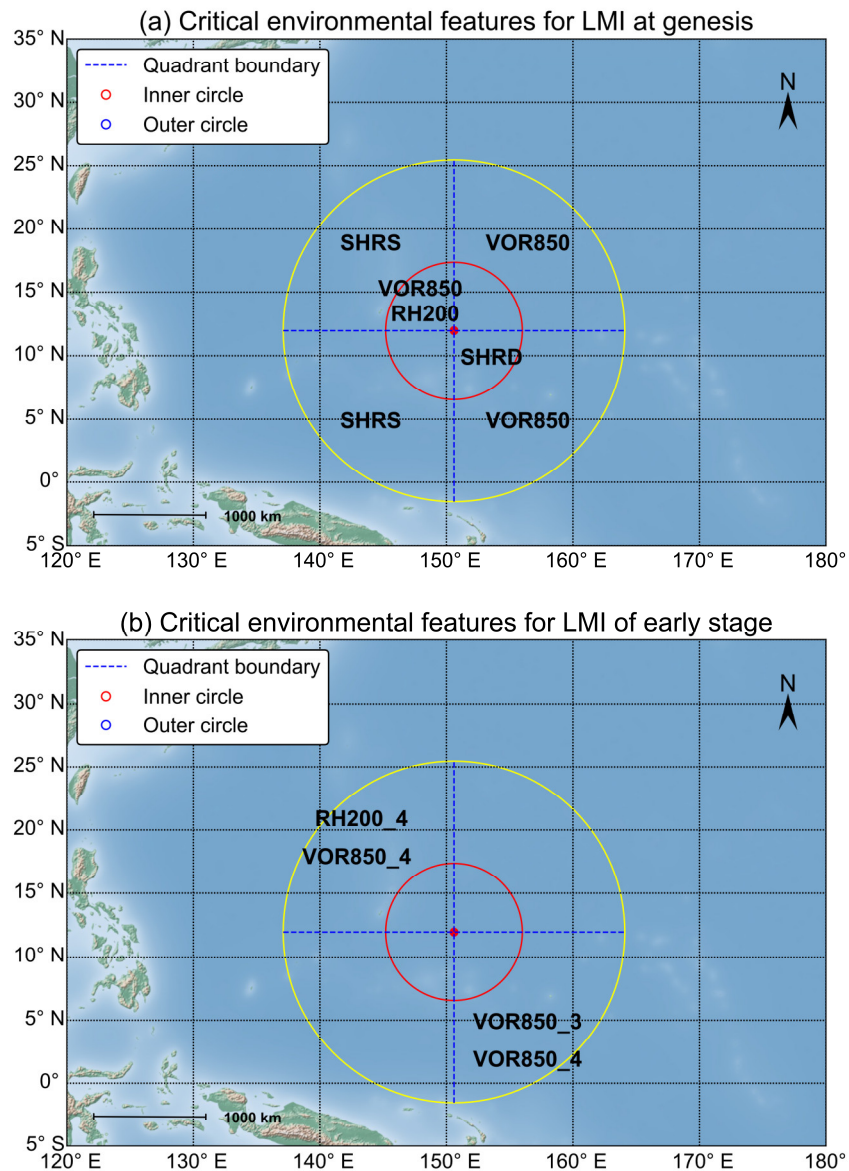|	**Genesis**	**0–12 h**	**12–24 h**	**24–36 h**	**36–48 h**
Translation speed	✓				
Coriolis parameter				✓	✓
Tightness		✓	✓		✓
6 h intensity variation					✓

Figure 14. Critical environmental features and their key intervals and quadrants.

The classification model based on the AdaBoost algorithm in step 1 had an accuracy of 0.7479 and an F1-score of 0.8387 on the testing set, implying that LMI is related to the vortex attributes and environmental conditions of a TC at genesis over the WNP. Among these features, several were found to be critical to estimate the range of LMI: (1) vorticity at the northwest of the inner circulation, and northeast and southeast of a TC's outer environment at 850 hPa; (2) deep-layer shear at the southeast of a TC's inner circulation; (3) shallow-layer shear at the southwest and northwest of a TC's inner circulation; (4) relative humidity at the northwest of a TC's inner circulation at 200 hPa; and (5) translation speed. From the composite analysis, we infer that strong TCs (LMI $\leq$ 63 kt) feature genesis location farther

away from the subtropical high embedded in a continuous band of high low-troposphere vorticity, zonal symmetrical circulation at the low troposphere, stronger circulation at the mid-troposphere, more compact circulation in the outflow layer, more symmetrical distribution of high-level moisture, and slower translation speed at genesis. However, other features of TC states are not obviously related to LMI. Some of these findings are similar to findings shown over the NA, such as the distribution of low-level vorticity.

At the second step, the AdaBoost based regressor again showed the best performance, with an RMSE of 23.7700 kt and R^2 of 0.3004 on the testing set, suggesting an underlying relationship between the early-stage features and LMI of TCs. Critical features include: (1) the Coriolis parameter during 24–48 h after genesis; (2) the 6-h intensity variation during 36–48 h after genesis; (3) TC tightness during 0–24 h and 36–48 h after genesis; (4) the deep-layer wind shear at the southeast of a TC's outer environment during 36–48 h after genesis; (5) the relative humidity at the northeast of a TC's outer environment at 200 hPa during first 36–48 h after genesis; and (6) the vorticity at the northwest of a TC's outer environment at 850 hPa during first 24–48 h after genesis. The important role of tightness at the TC's early development stage in varying LMI is revealed, which may be applied to intensity forecasting. In conclusion, a storm will have a greater opportunity to strengthen into major a TC when it moves slowly at low latitude and maintains tightness, intensifies continuously or even more quickly, has a high outflow layer with strong convection, has a compact structure at the top of the troposphere, has stronger easterlies at the outer environment in the lower troposphere, and has a smaller inner core during its early lifetime.

Even though the two-step model found a close and reasonable relationship between the LMI and early-stage features of TCs, there are still some issues to explore in the future. First, the study discusses TCs generating in a restricted area, so other TCs, especially those that form in the South China Sea, need to be further studied. Second, due to the shortage of TC data and the low relevance between genesis features and LMI, the model tends to overestimate weak TCs whose environmental conditions at genesis seem favorable. This problem may be overcome by using a longer series of TC information to make the model learn better. In addition, features used in this study do not cover all potential factors of LMI, such as the distribution of convection and rainfall in a TC [37]. This is because statistical models can barely contain all the features while keeping the model simple and efficient. Finally, this study updates our understanding of LMI and can be regarded as a qualitative reference for intensity prediction.

Appendix A

The decision tree model has many types of algorithms: ID3 [67,79], C4.5 [80], and classification and regression trees (CARTs) [64]. Figure A1 shows the typical structure of a CART for classification in this study (step 1), which consists of a root node, leaf nodes, and branches. The only root node suggests the very first feature to consider, each branch represents a possible decision according to the value of the "splitting feature", and it will reach a leaf node providing the final choice. During the node splitting process, the dataset is gradually divided, and finally a class label is assigned to each case corresponding to the label of the leaf node that it belongs to.
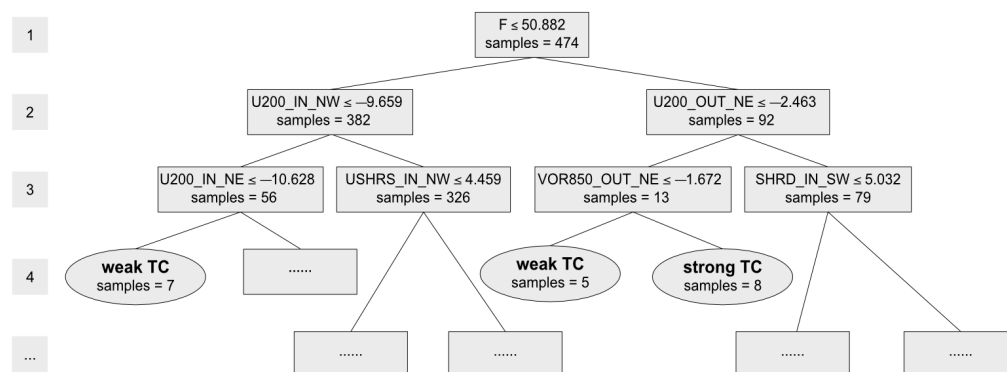

Figure A1. Classification and regression tree (CART) constructed by features at genesis in this study. Rectangles are parent nodes, ellipses are leaf nodes. Inequality in rectangles suggests judging condition of each leaf node; the left branch leads to positive conditions and right branch leads to negative, and the number of cases split on each branch is shown by "samples". Sequence numbers in the left column refer to depth of each dividing feature.

The biggest difference among decision tree algorithms is the rule to choose the "best splitting features" at nodes (e.g., "information gain" for ID3 and "information gain ratio" for C4.5). After selecting the splitting feature at the node, we try every possible value of the feature to pick up the best one with the smallest Gini index. However, in regression tasks, since the features are continuous variables, the rule to find splitting features in classification trees is no longer applicable. Thus, a heuristic algorithm is employed to find the best splitting point in the range of feature values. First, two sectors of the dataset are defined when it comes to the *j*th feature:

$$R_1(j,\,s) = \left\{ x \middle| x^{(j)} \leq s \right\}, \qquad R_2(j,\,s) = \left\{ x \middle| x^{(j)} > s \right\}, \tag{A1}$$

where feature $x^{(j)}$ and its value s are set as the splitting feature and splitting point, respectively. After that, we seek the optimum $(j,\,s)$ to make:

$$\min_{j,\,s} \left[\min_{c_1} \sum_{x_i \in R_1(j,\,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{x_i \in R_2(j,\,s)} (y_i - c_2)^2 \right], \tag{A2}$$

where y_i is the true value of input data, while c_1 and c_2 are the mean values of y_i in the divided sectors. The subset is thus separated into two sectors at every step, and each sector has an output value of:

$$\hat{c}_m = \frac{1}{N_m} \sum_{x_i \in R_m(j,\,s)} y_i, \qquad x_i \in R_m, \qquad m = 1,\,2, \tag{A3}$$

where N_m is the number of samples in R_m. In this way, the input space is finally divided into M sections, and a regression tree is presented to fit the function between x_i and y_i by:

$$f(x) = \sum_{m=1}^{M} \hat{c}_m I (x \in R_m), \tag{A4}$$

where I is the length of space R_m. Although there is a difference between the forms of the loss function in classification and regression, the intrinsic purpose is the same: to decrease the "impurity" of the subsets.

Appendix B

Ensemble learning methods construct sets of individual learners and then combine them with a specific strategy that is applicable to various machine learning models (e.g., decision tree, support vector machine (SVM), artificial neural network (ANN), etc.). Usually, they have much stronger generalization capability than base estimators, since they have their own ways to reduce overfitting [81].

AdaBoost and XGBoost are boosting models that train base learners in series, so each one will affect the next one. Taking AdaBoost, for instance (Figure A2), after a base estimator is trained, the weight distribution of the original samples is changed according to a loss function with the intention to pay more attention to misclassified cases (deviated cases in the regression task). After the training is finished, the ensemble model will make a decision by the linear combination of weights on every base estimator, where the estimator with lower error will be assigned a higher weight. XGBoost is a recently developed model based on the gradient boosting algorithm [82] that adds regularizations like those in ANN to prevent overfitting. It also has a more flexible framework for the parallel calculation of blocks, but at the cost of more complexity and memory.

In contrast, random forest is a typical bagging algorithm that has a framework parallel to decision tree models (Figure A3). The subset of each learner is extracted by bootstrap sampling [83], and the base estimator does not apply all features of the original data, but only considers a random section so as to elevate the ensemble's generalization capability by adding disturbance to both samples and features. Finally, the decision is made by major voting for classification and simple average for regression of all the base estimators, with no weights considered.

In order to adapt to our datasets, the crucial parameters of these ensemble models need to be tuned during the training process. The parameters tuned in this study are listed in Table A1.
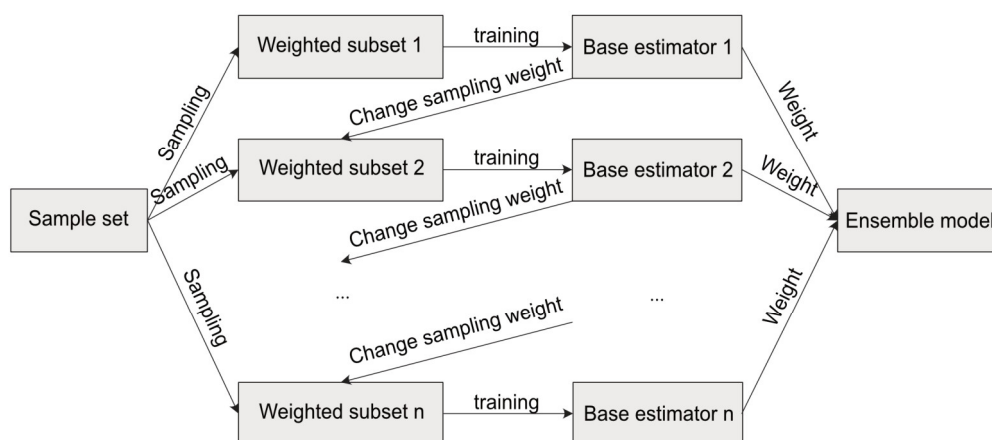

Figure A2. Workflow of AdaBoost model. Sampling weight is adjusted after every iteration and final decision is made by weighted average.
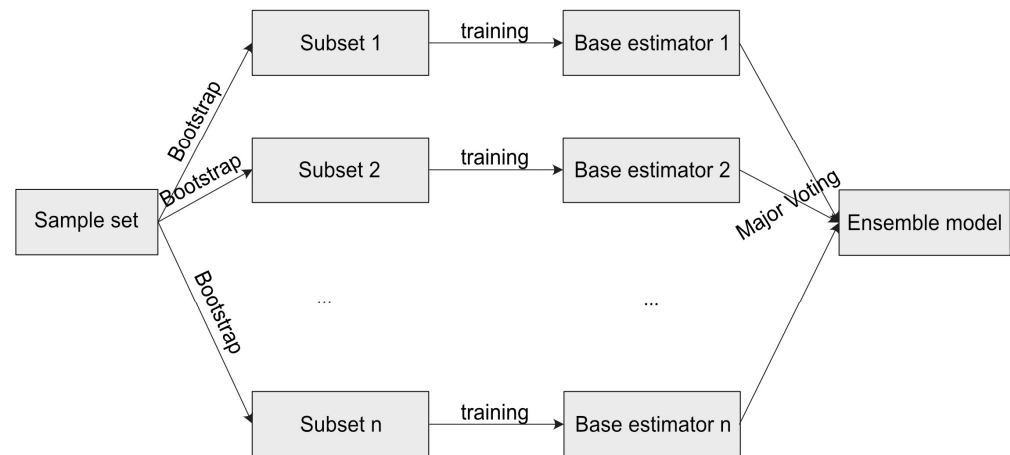
Figure A3. Workflow of random forest model. Subset used to train base estimator is sampled from original set by bootstrap method and final decision is made by major voting.

Table A1. Parameters tuned in three ensemble learning methods for classification and regression models in this study. All were tested by a range of values with the midpoint of reference value.

AdaBoost	XGBoost	Random Forest
Number of estimators	Number of estimators	Number of estimators
Learning rate	Learning rate	Maximum depth
Maximum depth	Maximum depth	Maximum features
Maximum features	Maximum features	Minimum samples to split
Minimum samples to split	Minimum sum of weights at child nodes	Minimum samples at leaf nodes
Minimum samples at leaf nodes	Minimum decrease of loss function to split	
	Subsample ratio	
	Coefficient of Lasso regularization	
	Coefficient of Ridge regularization	

Appendix C

The receiver operation characteristic (ROC) [84] curves (Figure A4a) and the precision-recall (P-R) curves (Figure A4b) were used to test the robustness of classifiers in this study. The x- and y-axes of the ROC curve refer to false positive rate (FPR) and true positive rate (TPR), respectively:

$$TPR = \frac{TP}{TP + FN} \, , \tag{A5}$$

$$FPR = \frac{FP}{TN + FP} \, . \tag{A6}$$

TPR shows the possibility to correctly distinguish positive cases among all positive cases, and FPR describes the possibility to mistake negative cases for positive ones among all negative cases. The integral of the ROC curve is the area under the ROC curve (AUC), whose value is an indicator of the classifier's performance (Figure A4a). Average precision (AP) is calculated by the area under the smoothed P-R curve, and the break-even point (BEP) indicates where precision equals recall (Figure A4b), both of which measure the quality of classification. Generally, the higher the AUC, AP, and BEP, the more robust the classifier.
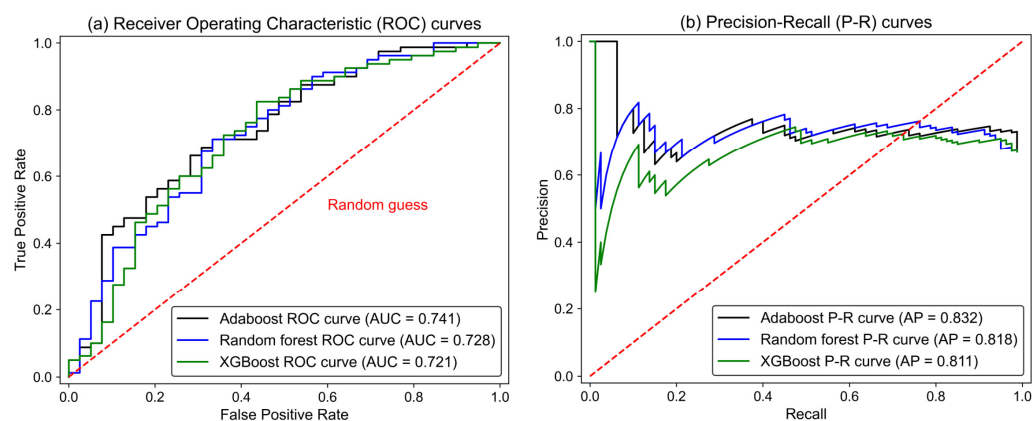
Figure A4. Receiver operating characteristic (ROC) curves and precision-recall (P-R) curves of results produced by three ensemble methods. Dotted red line in (**a**) refers to results of random guesses and in (**b**) is a reference line for break-even point (BEP) where precision equals recall. Values of area under ROC curve (AUC) and average precision (AP) are shown in lower right.

Comparing three ROC curves in Figure A4a, the AdaBoost curve spreads the farthest from the diagonal line of random guesses and has the largest AUC (0.741). Comparing the P-R curves (Figure A4b), the BEP values for the three models are quite close (around 0.7), but AdaBoost obtains the highest AP (0.832), indicating its outstanding robustness for LMI level estimation.

References

1. Peduzzi, P.; Chatenoux, B.; Dao, H.; De Bono, A.; Herold, C.; Kossin, J.; Mouton, F.; Nordbeck, O. Global trends in tropical cyclone risk. *Nat. Clim. Chang.* **2012**, *2*, 289–294. [CrossRef]
2. Arthur, W.C. A statistical-parametric model of tropical cyclones for hazard assessment. *Nat. Hazards Earth Syst. Sci.* **2021**, *21*, 893–916. [CrossRef]
3. Tsai, H.C.; Elsberry, R.L. Seven-Day Intensity and Intensity Spread Predictions in Bifurcation Situations with Guidance-On-Guidance for Western North Pacific Tropical Cyclones. *Asia-Pac. J. Atmos. Sci.* **2018**, *54*, 421–430. [CrossRef]
4. Mehra, A.; Tallapragada, V.; Zhang, Z.; Liu, B.; Wang, W.; Zhu, L.; Kim, H.S.; Iredell, D.; Liu, Q.; Zhang, B.; et al. Recent and future advances in tropical cyclone modeling at NOAA's national weather service national center for environmental prediction (NWS/NCEP). In Proceedings of the 33rd Conference on Hurricanes and Tropical Meteorology, Ponte Vedra, FL, USA, 16–20 April 2018; American Meteorological Society: Boston, MA, USA, 2018.
5. Heming, J.T.; Vellinga, M. The impact of recently implemented and planned changes to the Met Office Global Model on tropical cyclone performance. In Proceedings of the 33rd Conference on Hurricanes and Tropical Meteorology, Ponte Vedra, FL, USA, 16–20 April 2018; American Meteorological Society: Boston, MA, USA, 2018.
6. DeMaria, M.; Kaplan, J. A Statistical Hurricane Intensity Prediction Scheme (SHIPS) for the Atlantic Basin. *Weather Forecast.* **1994**, *9*, 209–220. [CrossRef]
7. DeMaria, M.; Kaplan, J. An updated Statistical Hurricane Intensity Prediction Scheme (SHIPS) for the Atlantic and eastern North Pacific basins. *Weather Forecast.* **1999**, *14*, 326–337. [CrossRef]
8. Pan, B.; Xu, X.; Shi, Z. Tropical cyclone intensity prediction based on recurrent neural networks. *Electron. Lett.* **2019**, *55*, 413–415. [CrossRef]
9. Wang, Z. Role of cumulus congestus in tropical cyclone formation in a high-resolution numerical model simulation. *J. Atmos. Sci.* **2014**, *71*, 1681–1700. [CrossRef]
10. Kilroy, G.; Smith, R.K.; Montgomery, M.T. An idealized numerical study of tropical cyclogenesis and evolution at the Equator. *Q. J. R. Meteorol. Soc.* **2020**, *146*, 685–699. [CrossRef]
11. Peng, M.S.; Fu, B.; Li, T.; Stevens, D.E. Developing versus nondeveloping disturbances for tropical cyclone formation. Part I: North Atlantic. *Mon. Weather Rev.* **2012**, *140*, 1047–1066. [CrossRef]
12. Fu, B.; Peng, M.S.; Li, T.; Stevens, D.E. Developing versus nondeveloping disturbances for tropical cyclone formation. Part II: Western north pacific. *Mon. Weather Rev.* **2012**, *140*, 1067–1080. [CrossRef]
13. Zhang, W.; Fu, B.; Peng, M.S.; Li, T. Discriminating developing versus nondeveloping tropical disturbances in the Western North Pacific through decision tree analysis. *Weather Forecast.* **2015**, *30*, 446–454. [CrossRef]
14. Briegel, L.M.; Frank, W.M. Large-scale influences on tropical cyclogenesis in the western North Pacific. *Mon. Weather Rev.* **1997**, *125*, 1397–1413. [CrossRef]

15. Ritchie, E.A.; Holland, G.J. Large-scale patterns associated with tropical cyclogenesis in the western Pacific. *Mon. Weather Rev.* **1999**, *127*, 2027–2043. [CrossRef]

16. Gray, W.M. Global view of the origin of tropical disturbances and storms. *Mon. Weather Rev.* **1968**, *96*, 669–700. [CrossRef]

17. Gray, W.M. The formation of tropical cyclones. *Meteor. Atmos. Phys.* **1998**, *67*, 37–69. [CrossRef]

18. Emanuel, K.; Nolan, D.S. Tropical cyclone activity and the global climate system. In Proceedings of the 26th Conference on Hurricanes and Tropical Meteorology, Miami, FL, USA, 3–7 May 2004; American Meteorological Society: Boston, MA, USA, 2004.

19. Camargo, S.J.; Emanuel, K.A.; Sobel, A.H. Use of a genesis potential index to diagnose ENSO effects on tropical cyclone genesis. *J. Clim.* **2007**, *20*, 4819–4834. [CrossRef]

20. Demaria, M. The effect of vertical shear on tropical cyclone intensity change. *J. Atmos. Sci.* **1996**, *53*, 2076–2087. [CrossRef]

21. Finocchio, P.M.; Majumdar, S.J.; Nolan, D.S.; Iskandarani, M. Idealized tropical cyclone responses to the height and depth of environmental vertical wind shear. *Mon. Weather Rev.* **2016**, *144*, 2155–2175. [CrossRef]

22. Wei, N.; Zhang, X.H.; Chen, L.; Hu, H. Comparison of the effect of easterly and westerly vertical wind shear on tropical cyclone intensity change over the western North Pacific. *Environ. Res. Lett.* **2018**, *13*. [CrossRef]

23. Leroux, M.D.; Plu, M.; Roux, F. On the sensitivity of tropical cyclone intensification under upper-level trough forcing. *Mon. Weather Rev.* **2016**, *144*, 1179–1202. [CrossRef]

24. Wei, N.; Li, Y.; Zhang, D.L.; Mai, Z.; Yang, S.Q. A statistical analysis of the relationship between upper-tropospheric cold low and tropical cyclone track and intensity change over the western North Pacific. *Mon. Weather Rev.* **2016**, *144*, 1805–1822. [CrossRef]

25. Fischer, M.S.; Tang, B.H.; Corbosiero, K.L. Assessing the influence of upper-tropospheric troughs on tropical cyclone intensification rates after genesis. *Mon. Weather Rev.* **2017**, *145*, 1295–1313. [CrossRef]

26. Rappin, E.D.; Morgan, M.C.; Tripoli, G.J. The impact of outflow environment on tropical cyclone intensification and structure. *J. Atmos. Sci.* **2011**, *68*, 177–194. [CrossRef]

27. Wu, Y.; Chen, S.; Li, W.; Fang, R.; Liu, H. Relative vorticity is the major environmental factor controlling tropical cyclone intensification over the Western North Pacific. *Atmos. Res.* **2020**, *237*, 104874. [CrossRef]

28. Shen, W.; Ginis, I. Effects of surface heat flux-induced sea surface temperature changes on tropical cyclone intensity. *Geophys. Res. Lett.* **2003**, *30*, 1933. [CrossRef]

29. Lin, I.I.; Chen, C.H.; Pun, I.F.; Liu, W.T.; Wu, C.C. Warm ocean anomaly, air sea fluxes, and the rapid intensification of tropical cyclone Nargis (2008). *Geophys. Res. Lett.* **2009**, *36*, L03817. [CrossRef]

30. Gao, S.; Chiu, L.S. Surface latent heat flux and rainfall associated with rapidly intensifying tropical cyclones over the western North Pacific. *Int. J. Remote Sens.* **2010**, *31*, 4699–4710. [CrossRef]

31. Jaimes, B.; Shay, L.K.; Brewster, J.K. Observed air-sea interactions in tropical cyclone Isaac over Loop Current mesoscale eddy features. *Dyn. Atmos. Ocean.* **2016**, *76*, 306–324. [CrossRef]

32. Ma, Z.; Fei, J.; Liu, L.; Huang, X.; Li, Y. An investigation of the influences of mesoscale ocean eddies on tropical cyclone intensities. *Mon. Weather Rev.* **2017**, *145*, 1181–1201. [CrossRef]

33. Ge, X.; Li, T.; Peng, M. Effects of vertical shears and midlevel dry air on tropical cyclone developments. *J. Atmos. Sci.* **2013**, *70*, 3859–3875. [CrossRef]

34. Lin, N.; Jing, R.; Wang, Y.; Yonekura, E.; Fan, J.; Xue, L. A Statistical Investigation of the Dependence of Tropical Cyclone Intensity Change on the Surrounding Environment. *Mon. Weather Rev.* **2017**, *145*, 2813–2831. [CrossRef]

35. Wang, Y. How do outer spiral rainbands affect tropical cyclone structure and intensity? *J. Atmos. Sci.* **2009**, *66*, 1250–1273. [CrossRef]

36. Schubert, W.H.; Hack, J.J. Inertial stability and tropical cyclone development. *J. Atmos. Sci.* **1982**, *39*, 1687–1697. [CrossRef]

37. Tao, C.; Jiang, H. Distributions of shallow to very deep precipitation-convection in rapidly intensifying tropical cyclones. *J. Clim.* **2015**, *28*, 8791–8824. [CrossRef]

38. Xu, W.; Rutledge, S.A.; Zhang, W. Relationships between total lightning, deep convection, and tropical cyclone intensity change. *J. Geophys. Res.* **2017**, *122*, 7047–7063. [CrossRef]

39. Sitkowski, M.; Kossin, J.P.; Rozoff, C.M. Intensity and structure changes during hurricane eyewall replacement cycles. *Mon. Weather Rev.* **2011**, *139*, 3829–3847. [CrossRef]

40. Fischer, M.S.; Rogers, R.F.; Reasor, P.D. The rapid intensification and eyewall replacement cycles of Hurricane Irma (2017). *Mon. Weather Rev.* **2020**, *148*, 981–1004. [CrossRef]

41. Ditchek, S.D.; Nelsona, T.C.; Rosenmayer, M.; Corbosiero, K.L. The relationship between tropical cyclones at genesis and their maximum attained intensity. *J. Clim.* **2017**, *30*, 4897–4913. [CrossRef]

42. Fudeyasu, H.; Yoshida, R. Western North Pacific tropical cyclone characteristics stratified by genesis environment. *Mon. Weather Rev.* **2018**, *146*, 435–446. [CrossRef]

43. Ma, C.; Peng, M.; Li, T.; Sun, Y.; Liu, J.; Bi, M. Effects of background state on tropical cyclone size over the Western North Pacific and Northern Atlantic. *Clim. Dyn.* **2019**, *52*, 4143–4156. [CrossRef]

44. Chen, R.; Zhang, W.; Wang, X. Machine Learning in Tropical Cyclone Forecast Modeling: A Review. *Atmosphere* **2020**, *11*, 676. [CrossRef]

45. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N. Prabhat Deep learning and process understanding for data-driven Earth system science. *Nature* **2019**, *566*, 195–204. [CrossRef]

46. McGovern, A.; Elmore, K.L.; Gagne, D.J.; Haupt, S.E.; Karstens, C.D.; Lagerquist, R.; Smith, T.; Williams, J.K. Using artificial intelligence to improve real-time decision-making for high-impact weather. *Bull. Am. Meteorol. Soc.* **2017**, *98*, 2073–2090. [CrossRef]

47. Lakshmanan, V.; Zhang, J.; Howard, K. A technique to censor biological echoes in radar reflectivity data. *J. Appl. Meteorol. Climatol.* **2010**, *49*, 453–462. [CrossRef]

48. Gagne, D.J.; McGovern, A.; Brotzge, J. Classification of convective areas using decision trees. *J. Atmos. Ocean. Technol.* **2009**, *26*, 1341–1353. [CrossRef]

49. Cloud, K.A.; Reich, B.J.; Rozoff, C.M.; Alessandrini, S.; Lewis, W.E.; Delle Monache, L. A feed forward neural network based on model output statistics for short-term hurricane intensity prediction. *Weather Forecast.* **2019**, *34*, 985–997. [CrossRef]

50. Knapp, K.R.; Kruk, M.C.; Levinson, D.H.; Diamond, H.J.; Neumann, C.J. The international best track archive for climate stewardship (IBTrACS). *Bull. Am. Meteorol. Soc.* **2010**, *91*, 363–376. [CrossRef]

51. Knapp, K.R.; Diamond, H.J.; Kossin, J.P.; Kruk, M.C.; Schreck, C.J. *International Best Track Archive for Climate Stewardship (IBTrACS) Project, Version 4*; NOAA National Centers for Environmental Information: Washington, DC, USA, 2018.

52. Zhong, Q.; Li, J.; Zhang, L.; Ding, R.; Li, B. Predictability of tropical cyclone intensity over the Western North Pacific using the IBTrACS dataset. *Mon. Weather Rev.* **2018**, *146*, 2741–2755. [CrossRef]

53. Zhou, Z.H. *Machine Learning*; Tsinghua University Press: Beijing, China, 2016; 425p.

54. Horn, M.; Walsh, K.; Zhao, M.; Camargo, S.J.; Scoccimarro, E.; Murakami, H.; Wang, H.; Ballinger, A.; Kumar, A.; Shaevitz, D.A.; et al. Tracking scheme dependence of simulated tropical cyclone response to idealized climate simulations. *J. Clim.* **2014**, *27*, 9197–9213. [CrossRef]

55. Knaff, J.A.; Sampson, C.R.; DeMaria, M. An operational statistical typhoon intensity prediction scheme for the western North Pacific. *Weather Forecast.* **2005**, *20*, 688–699. [CrossRef]

56. Tippett, M.K.; Camargo, S.J.; Sobel, A.H. A poisson regression index for tropical cyclone genesis and the role of large-scale vorticity in genesis. *J. Clim.* **2011**, *24*, 2335–2357. [CrossRef]

57. Kalnay, E. *Atmospheric Modeling, Data Assimilation and Predictability*; Cambridge University Press: Cambridge, UK, 2003; 341p.

58. Chavas, D.R.; Vigh, J. *QSCAT-R: The QuikSCAT Tropical Cyclone Radial Structure Dataset (No. NCAR/TN-513+STR)*; National Center for Atmospheric Research: Boulder, CO, USA, 2014; 25p.

59. Gander, W.; Gander, M.J.; Kwok, F. *Scientific Computing—An Introduction Using Maple and MATLAB*; Springer: New York, NY, USA, 2010; 912p.

60. Guo, X.; Tan, Z.M. Tropical cyclone fullness: A new concept for interpreting storm intensity. *Geophys. Res. Lett.* **2017**, *44*, 4324–4331. [CrossRef]

61. Levitus, S. *Climatological Atlas of the World Oceans*; NOAA Prof. Paper 13; U.S. Government Printing Office: Washington, DC, USA, 1982; 173p.

62. Emanuel, K.A. The maximum intensity of hurricanes. *J. Atmos. Sci.* **1988**, *45*, 1143–1155. [CrossRef]

63. Quinlan, J.R. Simplifying decision trees. *Int. J. Man. Mach. Stud.* **1987**, *27*, 221–234. [CrossRef]

64. Breiman, L.; Friedman, J.; Stone, C.J.; Olshen, R.A. *Classification and Regression Trees*; Chapman & Hall/CRC: Boca Raton, FL, USA, 1984; 368p.

65. Li, H. *Statistical Learning Method*, 2nd ed.; Tsinghua University Press: Beijing, China, 2019; 464p.

66. Rasp, S.; Lerch, S. Neural networks for postprocessing ensemble weather forecasts. *Mon. Weather Rev.* **2018**, *146*, 3885–3900. [CrossRef]

67. Quinlan, J.R. Induction of decision trees. *Mach. Learn.* **1986**, *1*, 81–106. [CrossRef]

68. Freund, Y.; Schapire, R.E. A desicion-theoretic generalization of on-line learning and an application to boosting BT—Computational learning theory. *Comput. Learn. Theory.* **1995**, *904*, 23–37. [CrossRef]

69. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the KDD'16: 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. [CrossRef]

70. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

71. Sharma, A.K.; Chaurasia, S.; Srivastava, D.K. Supervised Rainfall Learning Model Using Machine Learning Algorithms. In *The International Conference on Advanced Machine Learning Technologies and Applications (AMLTA2018)*; Advances in Intelligent Systems and, Computing; Hassanien, A., Tolba, M.F., Elhoseny, M., Mostafa, M., Eds.; Springer: Cham, Switzerland, 2018; Volume 723. [CrossRef]

72. Zhang, T.; Lin, W.; Lin, Y.; Zhang, M.; Yu, H.; Cao, K.; Xue, W. Prediction of tropical cyclone genesis from mesoscale convective systems using machine learning. *Weather Forecast.* **2019**, *34*, 1035–1049. [CrossRef]

73. Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In Proceedings of the IJCAI'95 14th International Joint Conference on Artificial Intelligence, Montreal, QC, Canada, 20–25 August 1995; AAAI: Palo Alto, CA, USA, 1995; pp. 1137–1145.

74. Zhang, F.; Tao, D. Effects of vertical wind shear on the predictability of tropical cyclones. *J. Atmos. Sci.* **2013**, *70*, 975–983. [CrossRef]

75. Frank, W.M.; Ritchie, E.A. Effects of vertical wind shear on the intensity and structure of numerically simulated hurricanes. *Mon. Weather Rev.* **2001**, *129*, 2249–2269. [CrossRef]

76. Robinson, F.J.; Sherwood, S.C. Modeling the impact of convective entrainment on the tropical tropopause. *J. Atmos. Sci.* **2006**, *63*, 1013–1027. [CrossRef]

77. Pasquero, C.; Desbiolles, F.; Meroni, A.N. Air-Sea Interactions in the Cold Wakes of Tropical Cyclones. *Geophys. Res. Lett.* **2021**, *48*, 1–6. [CrossRef]

78. Hill, A.J.; Herman, G.R.; Schumacher, R.S. Forecasting severe weather with random forests. *Mon. Weather Rev.* **2020**, *148*, 2135–2161. [CrossRef]

79. Quinlan, J.R. Discovering rules by induction from large collections of examples. In *Expert Systems in the Micro-Electronic Age*; Michie, D., Ed.; Edinburgh University Press: Edinburgh, UK, 1979; pp. 168–201.

80. Quinlan, J.R. *C4.5: Programs for Machine Learning*; Morgan Kaufmann: San Francisco, CA, USA, 1993; 49p.

81. Dietterich, T.G. Ensemble methods in machine learning. In Proceedings of the 1st International Workshop on Multiple Classifier Systems (MCS), Cagliari, Italy, 21–23 June 2000; pp. 1–15.

82. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]

83. Efron, B.; Gong, G. A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation. *Am. Stat.* **1983**, *37*, 36. [CrossRef]

84. Spackman, K.A. Signal Detection Theory: Valuable Tools for Evaluating Inducted Learning. In Proceedings of the Sixth International Workshop on Machine Learning, Ithaca, NY, USA, 26–27 June 1989; Elsevier: Amsterdam, The Netherlands, 1989; pp. 160–163.