Decision-Tree-Based Classification of Lifetime Maximum Intensity of Tropical Cyclones in the Tropical Western North Pacific

Sung-Hun Kim [1], Il-Ju Moon [2],[*], Seong-Hee Won [3], Hyoun-Woo Kang [1] and Sok Kuh Kang [1]

[1]   Korea Institute of Ocean Science and Technology, Busan 49111, Korea; sh.kim@kiost.ac.kr (S.-H.K.); hwkang@kiost.ac.kr (H.-W.K.); skkang@kiost.ac.kr (S.K.K.)
[2]   Typhoon Research Center, Jeju National University, Jeju City 63241, Korea
[3]   National Typhoon Center, Korea Meteorological Administration, Jeju City 63614, Korea; shwon11@korea.kr
[*]   Correspondence: ijmoon@jejunu.ac.kr

Abstract: The National Typhoon Center of the Korea Meteorological Administration developed a statistical–dynamical typhoon intensity prediction model for the western North Pacific, the CSTIPS-DAT, using a track-pattern clustering technique. The model led to significant improvements in the prediction of the intensity of tropical cyclones (TCs). However, relatively large errors have been found in a cluster located in the tropical western North Pacific (TWNP), mainly because of the large predictand variance. In this study, a decision-tree algorithm was employed to reduce the predictand variance for TCs in the TWNP. The tree predicts the likelihood of a TC reaching a maximum lifetime intensity greater than 70 knots at its genesis. The developed four rules suggest that the pre-existing ocean thermal structures along the track and the latitude of a TC's position play significant roles in the determination of its intensity. The developed decision-tree classification exhibited 90.0% and 80.5% accuracy in the training and test periods, respectively. These results suggest that intensity prediction with the CSTIPS-DAT can be further improved by developing independent statistical models for TC groups classified by the present algorithm.

Keywords: tropical cyclone; depth-averaged temperature; decision tree; lifetime maximum intensity

1. Introduction

The accurate prediction of tropical cyclone (TC) intensity is a major task in operational forecasting. Regarding intensity prediction, the capabilities of the widely used traditional statistical approaches have improved considerably more than those of the dynamical models [1]. A new statistical–dynamical model, the CSTIPS-DAT [2], which uses a clustering technique and depth-averaged ocean temperature (DAT)-based predictors, has facilitated significant improvements in intensity prediction in the western North Pacific (WNP). However, the CSTIPS-DAT shows relatively large errors for specific clusters, particularly those with a large predictand variance [2].

The tropical western North Pacific (TWNP) TCs, which belong to Cluster 2 in the CSTIPS-DAT model, spend most of their lifetimes over the tropics, where the environmental factors are favorable for their development (Figure 1a). Therefore, Cluster 2 is characterized by the strongest mean TC intensity in the WNP, and many TCs in the said cluster are distinguished by noticeable intensification. However, a considerable number of TCs in the said cluster still do not intensify even under favorable conditions, which produces a large breadth of intensity distribution (Figure 2a) and a large predictand variance. The distribution of the lifetime maximum intensity (LMI) in the TWNP is bimodal, characterized by a local minimum (at about 70 knots LMI) that separates the two groups between weakly (1st mode) and strongly developing TCs (2nd mode). Because the CSTIPS-DAT is a multiple-linear-regression-based model, the TWNP cluster was trained to fit well with strong TCs

with a high-density distribution; thus, major errors can occur in the prediction of weak TCs. For example, the intense Typhoon Phanfone in 2014, with an LMI of 95 kt, was well predicted by the CSTIPS-DAT. However, the relatively weak Typhoon Faxai (2014) was not accurately predicted, mostly because of overestimation (Figure 1b,c). These results suggest that with prior knowledge of the LMI type at the genesis of a TC, intensity prediction in the TWNP could be improved through the development of independent statistical models for each classified group.

The LMI, which is an integrated metric of TC intensification, can be used to present basic TC climatology characteristics [3–5]. Several studies have noted that the global distribution of the LMI is bimodal [6–8]. However, there is no consensus on why this bimodal LMI distribution occurs. Torn and Snyder [9] argued that the bimodality is the result of an artificially low number of Category 3 hurricanes in the Atlantic, and that this may be linked to the low resolution of the Dvorak technique which has been used to estimate their intensity. Soloviev et al. [10] attempted to explain the bimodal distribution of the LMI by using the ratio of surface exchange coefficients as a function of wind speed. They suggested that a local maximum of the ratio is favorable for rapid intensification (RI), and thereby increases the number of TCs in the second high-intensity peak. Lee et al. [8] reported that RI is be a key factor in the bimodality in the LMI distribution of two types of TCs: those that undergo RI during their lifetimes (RI TCs) and those that do not (non-RI TCs). They found that the LMI had a normal distribution with a unimodal peak for each TC type, at approximately 120 kt and 45 kt for RI TCs and non-RI TCs, respectively. The establishment of classification criteria to determine the types of TCs (weakly or strongly developing TCs) in the early developing stages of the TWNP TCs will contribute to a better understanding of the global bimodal LMI distribution.
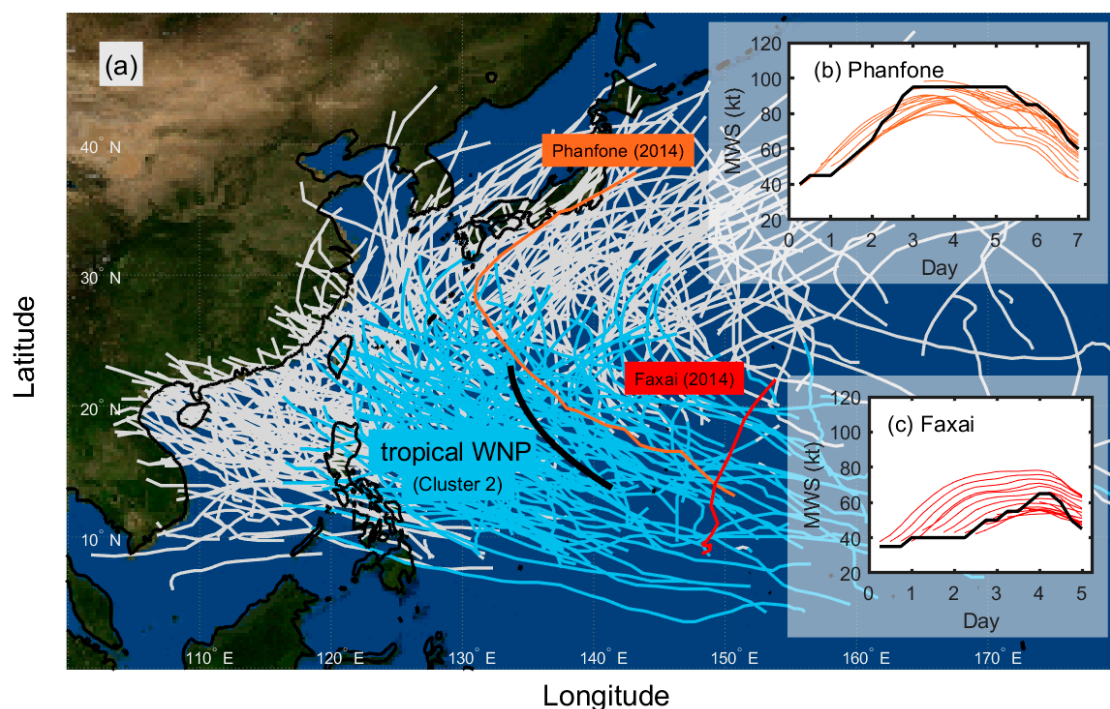

Figure 1. (**a**) All tracks in the tropical western North Pacific (Cluster 2, blue lines), and all tracks in the western North Pacific (gray lines) in 2004–2014. The orange and red lines indicate the tracks of Typhoon Phanfone and Faxai in 2014, respectively. The thick black line is the mean track for Cluster 2 in CSTIPS-DAT. Results of individual intensity predictions from CSTIPS-DAT for Typhoons (**b**) Phanfone and (**c**) Faxai in 2014. A thick black line is an observation (Regional Specialized Meteorological Center best track data), and the colored lines are individual CSTIPS-DAT predictions.
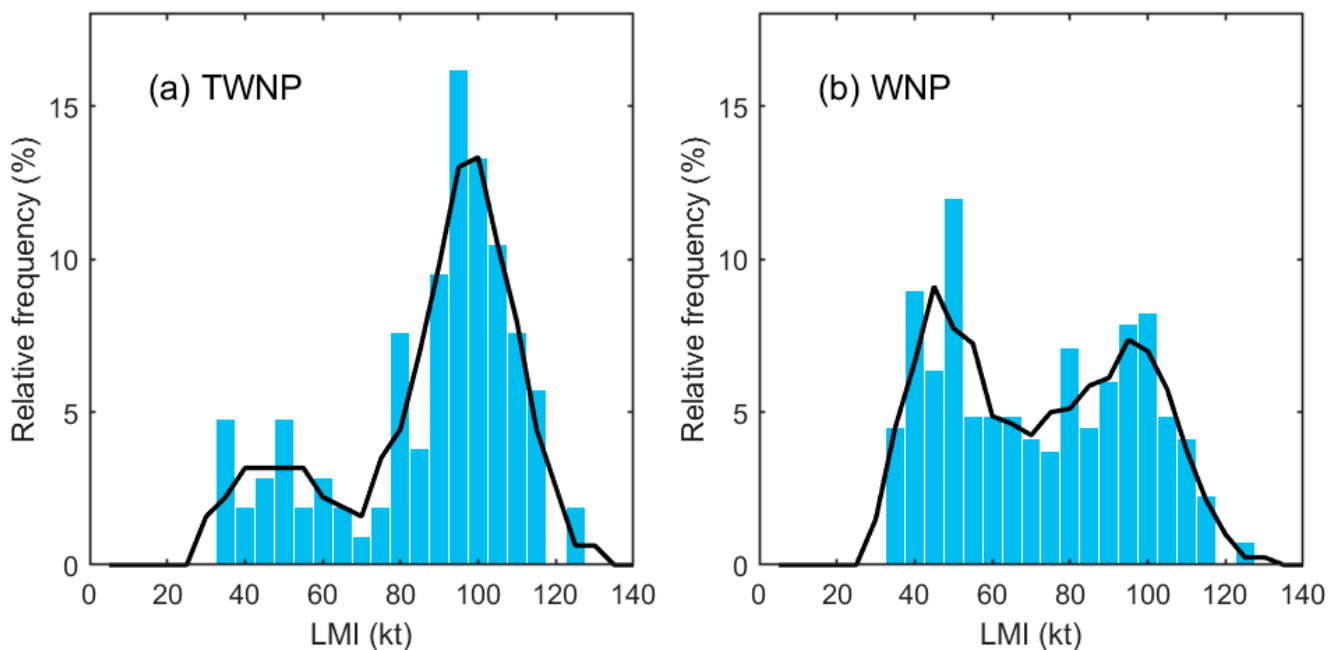
Figure 2. Distribution of lifetime maximum intensity. The relative frequencies presented were calculated on the basis of the 2004–2016 tropical cyclones in (**a**) the tropical western North Pacific (i.e., Cluster 2 in CSTIPS-DAT) and (**b**) the western North Pacific. The blue bars show the raw data binned into 5 kt bins. The black lines present the smoothed relative frequencies with a window width of 15 kt.

The analysis of climate information on TCs has socioeconomic implications and scientific significance because it leads to a better understanding of TC activity and the related mechanisms [11–13]. However, the large volume of varied data on TCs has continued to increase significantly, at a pace that has seemed to outstrip the capabilities of traditional analytical methods [14,15]. The decision tree, as a data-mining technique, is a process of finding useful rules, patterns, and knowledge in large, diverse archived databases to facilitate decision making [16].

Recently, the decision tree, as a useful tool for schematic classification, has been widely employed to investigate the mechanisms of TC development and impact in the WNP [14,17–23] and the North Atlantic [24–27]. Li et al. [17] employed a decision-tree algorithm to investigate the collective contributions to Atlantic hurricanes from sea surface temperature (SST), water vapor, vertical wind shear, and zonal stretching deformation. Zhang et al. [14] applied a decision tree to the binary classification of TCs as intensifying or weakening within 24 h. The decision tree, which used only three variables, exhibited remarkable prediction accuracy: 90.2%. Zhang et al. [18] used a decision tree to investigate the classification of tropical disturbances that did or did not develop into tropical storms in the WNP. The classification accuracies of the developed model were 81.7% for training and 84.6% for validation. Gao et al. [19] used a decision-tree algorithm to develop an RI prediction model that classified intensity changes as RI and non-RI events. They showed that the prestorm ocean coupling potential intensity index, which uses DATs instead of SST to calculate the maximum potential intensity (MPI), improved the RI classification accuracy by approximately 6% during the test period. Park et al. [20] developed a decision-tree-based WNP TC genesis detection algorithm using satellite observation-based predictors. They found that circulation symmetry and intensity were the most critical parameters for characterizing the development of tropical disturbances. Lee et al. [21] developed a scheme for TC formation using machine learning in the WNP and applied it for operational prediction of TC formation. Kim et al. [22] compared the prediction performance of three machine-learning algorithms (decision tree, random forest, support vector machines) and a linear-discriminant-analysis-based model in WNP TC genesis detection. They showed that

machine-learning-based models were more capable than conventional linear approaches at detecting TC formation. Yang et al. [24] showed that using the association rule algorithm, the RI prediction performance of the model using only three predictors was better than that of the model consisting of five predictors proposed by Kaplan and DeMaria [28]. Yang [25] performed RI prediction using various classifiers based on the Statistical Hurricane Intensity Prediction Scheme (SHIPS) database. Su et al. [26], using satellite-observation-based storm internal structure and the predictors of the National Hurricane Center probabilistic forecast guidance, developed an RI prediction model for Atlantic hurricanes based on a machine-learning method. Wei and Yang [27] built an artificial intelligence system based on the SHIPS database, significantly improving the RI prediction performance for Atlantic hurricanes. These studies have shown that decision trees are useful for binary classification related to TC genesis and intensification, which further suggests that a decision tree could be a useful tool to split the components in the bimodal distribution of the LMI.

The bimodal distribution results in a large variance of TC intensity, which makes accurate intensity predictions difficult. Therefore, if we can successfully classify the type of LMI at the point when a TC occurs, the statistical TC intensity prediction can be improved by reducing the variance of the predictand. To check such a possibility, this study aimed to build a decision-tree classifier that can predict the intensification type when a target TC occurs. Section 2 describes the dataset and the classification method. In Section 3, the potential predictors are examined, and the classification and model verification results are discussed. A summary and conclusion are provided in Section 4.

2. Data and Methodology

2.1. Data

A decision tree was trained using the 2004–2013 TWNP TCs, which belong to Cluster 2 as classified by the TC track pattern clustering method [2]. Meanwhile, the tree was validated using the 2014–2016 TCs. The TC information was obtained from the Regional Specialized Meteorological Center's best track data. The environmental data were derived from two dynamical models' analysis data. The atmospheric variables were obtained from the National Centers for Environmental Prediction Global Forecast System analysis data, with a 1×1 degree of horizontal resolution at 6 h intervals. The oceanic variables were calculated with three-dimensional ocean data derived from the Hybrid Coordinate Ocean Model (HYCOM) + Navy Coupled Ocean Data Assimilation Global Analysis (GLBa0.08) provided by the U.S. Naval Research Laboratory.

2.2. Methodology

2.2.1. Static and Synoptic Potential Predictors

A total of 38 variables were used to build the decision tree, and are listed in Table 1 with their correlations with LMI. The potential variables considered in this study are factors known to be related to TC intensity [2], and are similar to those considered for the development of the CSTIPS-DAT. Four static variables were included: the absolute Julian day number, TC latitude (LAT) and longitude, and TC translation speed. There were 34 synoptic variables: divergence at 200 hPa (D200), the relative vorticity at 500 hPa (RV500) and 850 hPa (RV850), 200 hPa zonal wind (U200) and air temperature (T200), 500–300 hPa layer mean relative humidity (RHHI), 850–700 hPa layer mean relative humidity (RHLO), 200–850 hPa vertical wind shear (SH200), 500–850 hPa vertical wind shear (SH500), ocean heat content (OHC), depth-averaged temperature at various depths (DAT; [29,30]), and DAT-based MPI (DMPI; [31,32]). Lin et al. [31] suggested DMPI using DAT instead of prestorm sea surface temperature to consider negative feedback by TC-induced sea surface cooling on existing SST-based MPI. DMPI has significantly reduced the overestimation of maximum intensity of the existing SST-based MPI and has frequently been used to predict TC intensity and RI [2,19,32,33]. The variables based on intensification potential (POT; MPI − initial intensity) were the essential factors in the CSTIPS-DAT model. However, in this study, TC genesis was defined as the first moment of at least 35 kt intensity; thus, the

POT and DMPI had the same correlation coefficient. Because the current study focused on classifying the LMI of TCs at their genesis, the POT and DAT-based POT were excluded from the pool of potential variables. DATs and DMPIs had the highest correlation among all variables, reaching 0.54 and 0.56, respectively. OHC, a widely used index for upper-ocean thermal conditions, also had a high correlation coefficient (r = 0.52). Price [29] showed that OHC and DAT are well correlated in the high OHC range and deep water, but they are poorly correlated in low OHC and shallow continental shelves. Since the TWNP is mostly deep and has high OHC, the correlation coefficients of OHC and DAT are not very different there. All the variables were averaged from the genesis to 3.25 days along the TC track—the sum of the average time (1.7 days) and standard deviation (1.55 days) to reach LMI after TWNP TCs' occurrence.

Table 1. Potential variables in the present model and their correlation coefficients (r) with the lifetime maximum intensity for the 2004–2013 TWNP TCs (Cluster 2 in CTIPS-DAT). All the variables were averaged along the TC track from the genesis to 3.25 days.

Variable	Description	r
JDAY	The absolute value of Julian day—248	−0.27
LAT	Latitude of typhoon location	−0.33
LON	Longitude of typhoon location	0.07
SPD	Storm moving speed	−0.23
D200	Area-averaged (0 km to 1000 km) divergence at 200 hPa	0.05
RV500	Area-averaged (0 km to 1000 km) relative vorticity at 500 hPa	0.16
RV850	Area-averaged (0 km to 1000 km) relative vorticity at 850 hPa	0.04
U200	Area-averaged (200 km to 800 km) zonal wind at 200 hPa	−0.28
T200	Area-averaged (200 km to 800 km) air temperature at 200 hPa	−0.39
RHHI	Area-averaged (200 km to 800 km) relative humidity 500–300 hPa	0.32
RHLO	Area-averaged (200 km to 800 km) relative humidity 850–700 hPa	0.29
SH200	Area-averaged (200 km to 800 km) 200 hPa to 850 hPa vertical wind shear	−0.17
SH500	Area-averaged (200 km to 800 km) 500 hPa to 850 hPa vertical wind shear	−0.32
OHC	Area-averaged (0 km to 200 km) ocean heat contents	0.52
DAT10—DAT120	Ocean temperatures averaged from the near-surface down to the various depth (10 to 120 m, 10-m interval)	0.48–0.54
DMPI10—DMPI120	Maximum potential intensity using DAT10—DAT120	0.47–0.56

2.2.2. Classification and Regression Tree

The classification and regression tree (CART) is one of the decision-tree algorithms that are used for categorical and continuous variables [34]. The rules generated by the CART are easy to interpret, and overfitting can be avoided by postpruning a fully grown tree. The CART is a binary partitioning algorithm with only two child nodes from the parent node. The Gini index, the sum of the misclassification probabilities, can be used as an impurity or diversity measure in each node. It is expressed as follows:

$$G = 1 - \sum_{j=1}^{c} \left(\frac{n_j}{n}\right)^2 \tag{1}$$

where n is the number of observations in the node, c is the number of categories of target variables, and n_j is the number of observations belonging to the jth category of the target variable. The CART algorithm selects the best predictor to minimize the Gini index for each split and finds the optimal separation of each node, and this division process is repeated for each node to construct a decision tree. For example, in order to classify TC intensity using environmental variables, it is necessary to perform the classification by repeatedly changing the classification reference value (e.g., the sea surface temperature, 26 °C), to calculate the Gini index of the classified group, and to determine the optimal reference value which has a minimum Gini index. The above process is repeatedly performed as many times as the

specified number of nodes. In this study, a classifier was developed based on the "fitctree" function included in Matlab's "statistics and machine learning toolbox".

2.2.3. The k-Fold Cross-Validation

The *k*-fold cross-validation is one of the most popular resampling techniques for increasing the statistical reliability of model performance measurements [35]. The procedure is as follows. First, the entire sample is divided into *k* equally sized subsamples in which one subsample is reserved as validation data. Second, the model is trained with $k - 1$ subsample, tested (or validated) with the retained subsample, and cross-validated *k* times until each subsample has been used for validation only once. Finally, the results of each step of the process are averaged to form an evaluation index, which can be used to perform forecast verification. The advantage of cross-validation is that all the cases are used for both training and validation, and each case is used for validation once. In this study, 10-fold cross-validation was used.

2.2.4. Synthetic Minority Oversampling Technique

When the binary classification model is trained with inequality data, a classifier will be biased toward the more frequently occurring class. The accuracy of the majority class is likely to be inflated in training, thus resulting in inappropriate predictive accuracy in testing. In the present study, the synthetic minority oversampling technique (SMOTE; [36]), one of the most commonly used oversampling techniques, was used to avoid the inequality sample problem. It randomly extracts samples from the minority class and increases the number of samples by generating synthetic samples with the ambient values of the extracted samples. In this study, the number of nearest neighbors to consider was set to five.

3. Results

The 2004–2016 distribution of LMI in the TWNP had two local maxima at approximately 50 kt and 100 kt, and a local minimum at 70 kt (Figure 2a). A bimodal distribution of the relative frequency of the LMI was also found in the WNP (Figure 2b). However, unlike the TWNP, the first peak in the WNP was higher than the second. The TWNP is a sub-basin in which the strongest TCs in the WNP occur, so the relative frequency of the strong TCs (2nd peak) was higher than that of the weak TCs (1st peak). In this study, the TWNP TCs were classified into two types: those with LMI above 70 kt (strongly developing TCs; A70) and those with LMI below 70 kt (weak TCs; B70).

Intensity prediction using the CSTIPS-DAT [2] revealed large mean absolute error (MAE) values and bias for the two classified groups, A70 and B70 (Figure 3a). As predicted, this was because the model was trained with the entire TWNP TCs that contain both weakly and strongly developing TCs, resulting in a negative bias (underestimation; see the red solid line in Figure 3a) for A70 and a positive bias (overestimation; see the blue solid line in Figure 3a) for B70. Indeed, most of the MAE values in the TWNP were related to the large biases, suggesting that the bias correction using individual models for A70 and B70 reduced the intensity prediction error. Overall, the MAE and bias were greater in B70 than in A70. This was related to the fact that during training, the model fit A70 better than B70 because A70 had about four times more samples than B70. In fact, the numbers of samples of the A70 and B70 groups were 60 and 17 TCs, respectively, during the training period, and 26 and 10 TCs, respectively, during the test periods (Table 2). To resolve the inequality in the training data set, SMOTE was used to increase the number of samples for B70 to 60, as in A70.
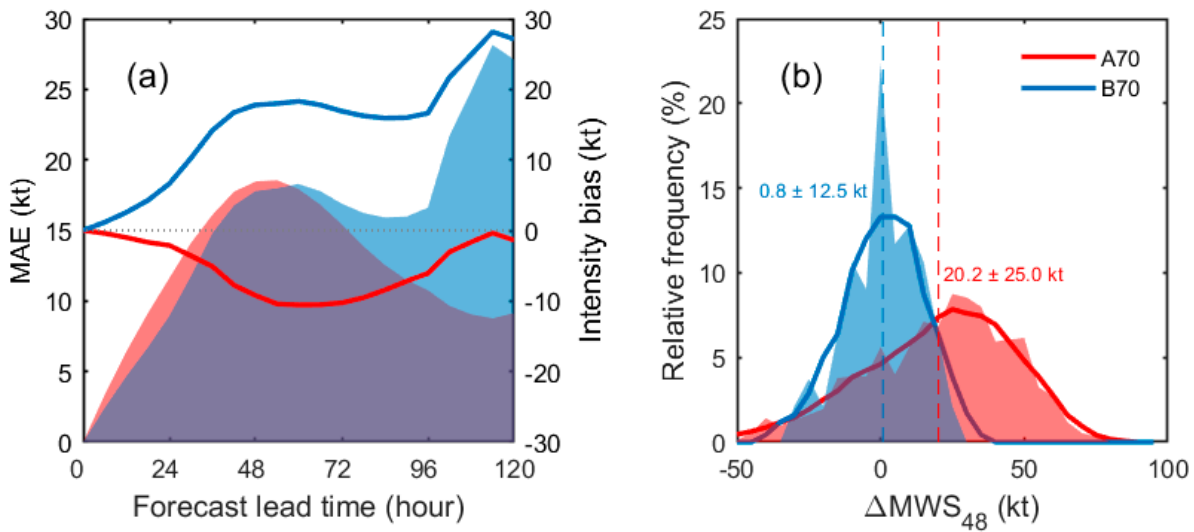
Figure 3. Comparison of (**a**) mean absolute errors (shading) and biases (solid lines) for intensity predictions of A70 (TCs with LMI greater than 70 kt) and B70 (TCs with LMI less than 70 kt) at each lead time for the 2013–2014 TWNP TCs. (**b**) Comparison of the relative frequencies of intensity change in 48 h in the classified groups (red: A70; blue: B70). The shaded areas show the raw data binned into 5 kt bins. The thick lines are the smoothed relative frequencies with a window width of 15 kt. The dashed lines indicate the means of each group. The mean values with $\pm\sigma$ are represented in colored text.

Table 2. Number of A70 and B70 tropical cyclones in 2004–2013, 2014–2016, and 2004–2016.

Period	A70	B70	Total
2004–2013	60	17	77
2014–2016	26	10	36
2004–2016	86	27	113

Figure 3b compares the relative frequencies of the intensity change for A70 and B70. The mean intensity change within 48 h was 20.2 ± 25.0 kt for A70 and 0.8 ± 12.5 kt for B70. The two-tailed Student's *t*-test revealed that the difference between the means of the two groups was statistically significant at the 5% test level. Therefore, it was expected that an LMI-based classification could reduce the variance of the intensity change in Cluster 2 and that the intensity prediction would be improved by the development of specific prediction models for each intensity type.

A confusion matrix [37] was used to calculate verification measures, namely the probability of detection (POD), false alarm rate (FAR), and accuracy. The POD is the ratio of the number of times a correct warning is issued for a target event to the total number of target events. The FAR is the number of times a warning is issued but an event does not occur divided by the number of times the warning is issued. The POD, FAR, and accuracy were calculated as follows:

$$\mathrm{POD} = \frac{TP}{TP + FN} \tag{2}$$

$$\mathrm{FAR} = \frac{FP}{FP + TP} \tag{3}$$

$$\mathrm{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

where *TP* is the true positive, *TN* is the true negative, *FP* is the false positive, and *FN* is the false negative. In this study, A70 was defined as the target class.

A decision tree generates the rule until the number of samples in a leaf drops below a specified size, i.e., the minimum leaf (min-leaf) size. The min-leaf size determines when splitting should be stopped; therefore, it is an important parameter that needs to

be carefully tuned. Figure 4a presents the classification performance of the decision tree during the training period with various min-leaf sizes. The skill scores can be used to set the parameters. Naturally, the highest accuracy and POD were achieved at the min-leaf size of 1, and the performance score decreased with increased min-leaf sizes. The FAR varied by 0–12% with the min-leaf size; however, no significant trend was associated with the min-leaf size.
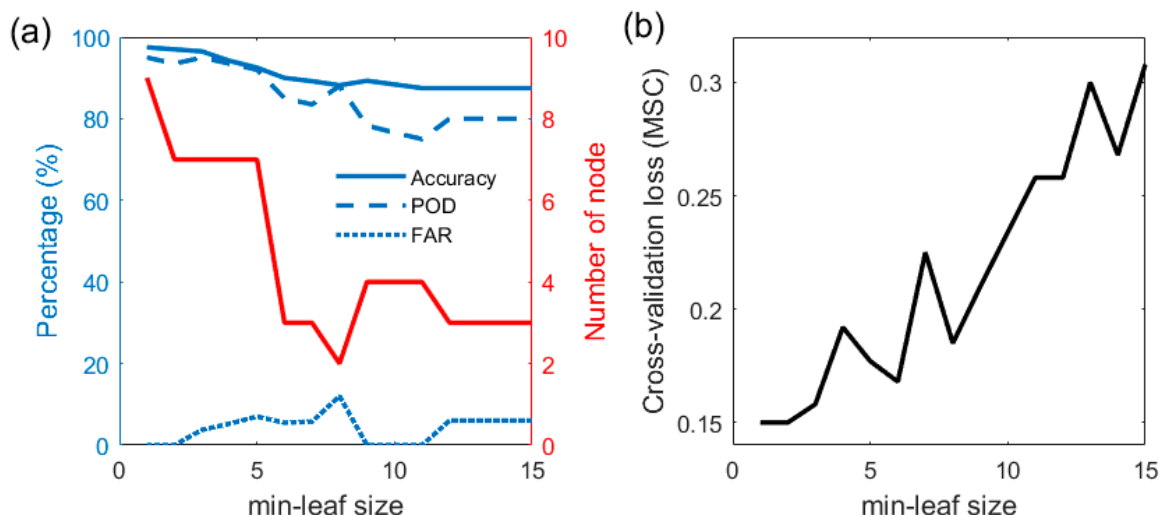

Figure 4. (**a**) Skill scores (blue lines) and the number of nodes (red line) at each minimum leaf size used in the decision-tree algorithm. (**b**) Distribution of cross-validation loss (mean misclassification rate) on the basis of the minimum leaf size using the *k*-fold cross-validation method.

A decision tree with a smaller min-leaf size usually has better performance. However, a small min-leaf size generates a complicated tree with many nodes, making a physical interpretation difficult. In addition, complicated trees can cause overfitting problems in classifications with insufficient sample sizes. A model should be trained to make reliable predictions for the test data. Overfitting is the result of modeling with noise instead of the underlying relationship. An excessively overfitted model performs poorly in real-time predictions because it is tuned to overreact to minor fluctuations in the training data. To avoid the prediction instability of overfitting, we determined the optimal min-leaf size using comparisons of the cross-validation (CV) loss. In this study, *k*-fold cross-validation was used to obtain the CV loss by averaging the misclassification rate (MSC), as shown below:

$$CV = \frac{1}{k} \sum_{i=1}^{k} MSC_i \tag{5}$$

$$MSC_i = \frac{n_{miss, \, i}}{n_i} \tag{6}$$

where *k* is the number of fold (here is set to be 10), $n_{miss,i}$ is the number of misclassification samples in *i*th test set, and n_i is the total number of samples in *i*th test set.

Figure 4b shows the change of the CV loss with min-leaf sizes. The CV loss tended to increase as the min-leaf size increased. The CV was the smallest at min-leaf sizes of 1 and 2, followed by local minima at 6 and 8. Min-leaf sizes of 1 and 2 required nine and seven nodes (fairly complex structure), respectively, and min-leaf sizes of 6 and 8 required three and two nodes, respectively (red line in Figure 4a). In this study, the min-leaf size was set to 6 to make the decision tree structure relatively simple with a small CV loss.

The trained decision tree included three nodes with four decision rules. Table 3 lists the decision rules governing the decision tree. Rule 1 shows that it is difficult for a TC in a low DMPI20 environment to intensify as it develops. MPI has been the most critical

predictor in previous statistical intensity prediction models [38–42]. In Rule 1, shallow (i.e., 20 m deep) DMPI was selected as a classification factor, and this informed the classification of many weak TCs. Weak TCs cannot interact with the deep ocean; thus, the shallow-depth ocean-temperature-based MPI can be a good criterion for categorizing weak TCs.

Table 3. Description and the confidence of the rule of the developed decision tree. Note that all variables here were averaged along the TC track from genesis to 3.25 days.

Rule NO.	Decision Rules	The Confidence of the Rule
1	If DMPI20 < 114 kt, then TC will not develop above 70 kt.	45/51 = 88.2%
2	If DMPI20 $\geq$ 114 kt and LAT $\geq$ 22.1° N, then TC will not develop above 70 kt.	8/9 = 88.9%
3	If DMPI20 $\geq$ 114 kt, LAT < 22.1° N and DAT100 < 26.3 °C, TC will not develop above 70 kt.	4/6 = 66.7%
4	If DMPI20 $\geq$ 114 kt, LAT < 22.1° N and DAT100 $\geq$ 26.3 °C, TC will develop above 70 kt.	51/54 = 94.4%

Rule 2 states the following: If DMPI20 $\geq$ 114 kt and LAT $\geq$ 22.1° N, TCs are less likely to intensify to more than 70 kt. This suggests that it is difficult for a TC that stays at high latitudes on average during development to be classified as A70. This is because TCs with higher LAT tend to move northward and thus their tracks become closer to the polar westerlies, resulting in increased vertical wind shear that suppress TC intensification. The selection of LAT explains why vertical wind shear (SH200 and SH500), a well-known dynamic index related to TC intensity, was not singled out in the rule.

Rule 3 states the following: If DMPI20 $\geq$ 114 kt, LAT < 22.1° N, and DAT100 < 26.3 °C, a TC cannot intensify to more than 70 kt. This suggests that a high DMPI20 and a low LAT are favorable for intensity; however, TCs are less likely to develop with strong intensity if DAT100 is less than 26.3 °C. Price [21] suggested that 100 m is the typical vertical mixing depth that major TCs induce; thus, DAT100 is the realistic temperature that represents the sea surface thermal conditions under intense TCs. If DAT100 is less than 26.3 °C, which is close to the 2 m dew point temperature of the tropics [43], the ocean can no longer supply heat to the TC, thus reducing the likelihood of strong intensification.

Rule 4 states the following: If DMPI20 $\geq$ 114 kt, LAT < 22.1° N, and DAT100 $\geq$ 26.3 °C, a TC can intensify to more than 70 kt. This rule suggests that the development of intense TCs generally occurs when all three conditions are satisfied. The confidence of this rule was 94.4%.

To evaluate the capability of the decision tree to classify intensity, we analyzed the accuracy during the training and test period. The results showed a classification accuracy of 90.0% for training (Table 4) and 80.5% for testing (Table 5). According to the confusion matrix for the test period (Table 5), 24 of 26 TCs were correctly classified as A70, and 5 of 29 that were classified as A70 were B70. Thus, the POD had 92.3%, and the FAR had only 17.2%. These results exhibited high enough accuracy to build an independent statistical model for the TC groups classified on the basis of this algorithm.

Table 4. Confusion matrix for the training period.

		Model	
		A70	B70
Observed	A70	51	9
	B70	3	57

Table 5. Confusion matrix for the test period.

		Model	
		A70	B70
Observed	A70	24	2
	B70	5	5

4. Discussion

Kim et al. [2] classified TCs on the basis of their track patterns, by which the intensity characteristics could be classified. They showed that the prediction performance could be improved by reducing the variance of the predictand through the development of an individual model for each cluster. This study attempted to further reduce the predictand variance on the basis of the LMI classification, especially for Cluster 2 (TWNP TCs) of CSTIPS-DAT. The TWNP TCs show a bimodal LMI distribution, which can be classified as weakly (B70) and strongly developing TCs (A70). Because of this bimodality, the intensity prediction estimated using the CSTIPS-DAT showed large MAEs for the two groups. The large MAEs are mostly attributed to significant positive and negative biases for B70 and A70, respectively. This implies that correcting the biases through binary classification and developing independent prediction models for the classified groups can reduce the predictand variance and ultimately improve TC intensity prediction.

To improve the performance of the CSTIPS-DAT and to increase the understanding of LMI bimodality, this study developed a CART-algorithm-based decision tree which classifies the TC type at the time of genesis, based on whether or not it will reach an intensity of 70 kt or more during its lifetime. Among the 38 potential predictors, CART selected three variables that reached an accuracy of 90.0% in the training period (2004–2013) and 80.5% in the testing period (2014–2016). The selected variables were DMPI20, LAT, and DAT100. The splitting values were 114 kt for DMPI20, 22.1° N for LAT, and 26.3 °C for DAT100. The four developed rules indicate that the prestorm ocean thermal conditions (DMPI20 and DAT100) and latitude play a key role in determining the LMI in the TWNP.

It should be noted that DAT100 played an essential role in the decision tree developed for binary classification. For the unclassified TWNP TCs (black line in Figure 5), the correlation coefficients between various DATs and LMI were highest at DAT50. However, for strongly developing TCs (red line in Figure 5), the correlation was highest in DAT100. Price [21] proposed DAT100 as an oceanic index reflecting the sea surface cooling induced by Saffir–Simpson Category 3 TCs (96–113 kts). Interestingly, the Category 3 intensity belonged to the second peak of the LMI distribution (Figure 2a) and accounted for about 40% of the TWNP TCs. In contrast, for weak TCs (blue line in Figure 5) the correlation was very low at all DATs. This suggests that the pre-existing ocean thermal structures along the track are not essential in determining the LMI for weak TCs. Again, this highlights the need to develop individual models that consider key environmental factors differently depending on the classified groups.
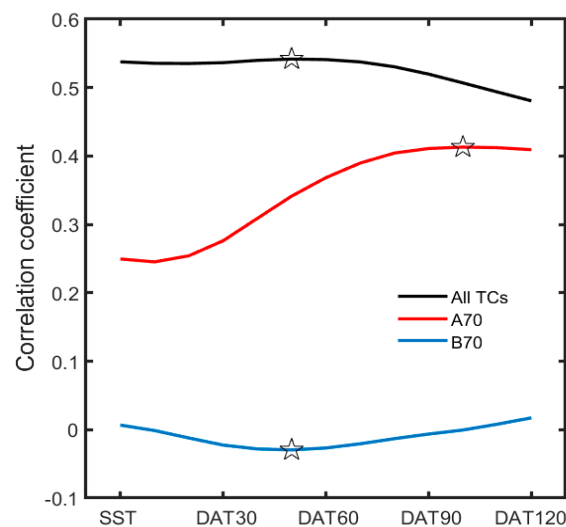

Figure 5. Comparison of correlation coefficients between various DATs and LMI for A70, B70, and all TCs. Black, red, and blue lines indicate unclassified TWNP TCs (i.e., all TCs), A70, and B70, respectively. Open stars represent the locations with the maximum values for each group.

5. Conclusions

Understanding the bimodal LMI distribution is important for improving TC intensity prediction. Previously known causes of this bimodality are the reduction of air–sea roughness at a particular wind speed range [10] and the presence or absence of rapid intensification events [8]. However, due to the lack of observational data in extreme winds, it is still difficult to fully understand the cause of the bimodal distribution. This study cannot directly explain the mechanism of the bimodality with the rules discovered, but it does present environmental parameters and their thresholds that can distinguish the two modes. This will make some contribution to a better understanding of the causes of the bimodal LMI distribution.

In this study, the CART algorithm, a machine-learning algorithm, was used for classification. Although the CART algorithm is widely used for binary classification, it cannot be affirmed that it is the optimal classification algorithm for classification of intensification types. Therefore, as in previous studies [22,25] that compared and evaluated several machine-learning algorithms for binary classification, research to find the optimal classification method by applying new classification tools must be conducted.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. DeMaria, M.; Sampson, C.R.; Knaff, J.A.; Musgrave, K.D. Is tropical cyclone intensity guidance improving? *Bull. Am. Meteorol. Soc.* **2014**, *95*, 387–398. [CrossRef]
2. Kim, S.-H.; Moon, I.-J.; Chu, P.-S. Statistical-dynamic typhoon intensity predictions in the western North Pacific using Track Pattern Clustering and Ocean Coupling Predictors. *Weather Forecast.* **2018**, *33*, 347–365. [CrossRef]
3. Emanuel, K.A. A statistical analysis of tropical cyclone intensity. *Mon. Weather Rev.* **2000**, *128*, 1139–1152. [CrossRef]
4. Park, D.-S.R.; Ho, C.-H.; Kim, J.-H. Growing threat of intense tropical cyclones to East Asia over the period 1977–2010. *Environ. Res. Lett.* **2014**, *9*, 014008. [CrossRef]
5. Moon, I.-J.; Kim, S.-H.; Klotzbach, P.; Chan, J.C.L. Roles of interbasin frequency changes in the poleward shifts of the maximum intensity location of tropical cyclones. *Environ. Res. Lett.* **2015**, *10*, 104004. [CrossRef]
6. Manganello, J.V.; Hodges, K.I.; Kinter, J.L., III; Cash, B.A.; Marx, L.; Jung, T.; Achuthavarier, D.; Adams, J.D.; Altshuler, E.L.; Huang, B.; et al. Tropical cyclone climatology in a 10-km global atmospheric GCM: Toward weather-resolving climate modeling. *J. Clim.* **2012**, *25*, 3867–3893. [CrossRef]
7. Zhao, M.; Held, I.M.; Lin, S.-J.; Vecchi, G.A. Simulations of global hurricane climatology, interannual variability, and response to global warming using a 50-km resolution GCM. *J. Clim.* **2009**, *22*, 6653–6678. [CrossRef]
8. Lee, C.-Y.; Tippett, M.K.; Sobel, A.H.; Camargo, S.J. Rapid intensification and the bimodal distribution of tropical cyclone intensity. *Nat. Commun.* **2016**, *7*, 10625. [CrossRef] [PubMed]
9. Torn, R.D.; Snyder, C. Uncertainty of tropical cyclone best-track information. *Weather Forecast.* **2012**, *27*, 715–729. [CrossRef]
10. Soloviev, A.V.; Lukas, R.; Donelan, M.A.; Haus, B.K.; Ginis, I. The air-sea interface and surface stress under tropical cyclones. *Sci. Rep.* **2014**, *4*, 5306. [CrossRef]
11. Bengtsson, L.; Botzet, M.; Esch, M. Will greenhouse-induced warming over the next 50 years lead to higher frequency and greater intensity of hurricanes? *Tellus* **1996**, *48A*, 57–73. [CrossRef]

12. Emanuel, K.A. Increasing destructiveness of tropical cyclones over the past 30 years. *Nature* **2005**, *436*, 686–688. [CrossRef] [PubMed]

13. Webster, P.J.; Holland, G.; Curry, J.A.; Chang, H.-R. Changes in tropical cyclone number, duration, and intensity in a warming environment. *Science* **2005**, *309*, 1844–1846. [CrossRef] [PubMed]

14. Zhang, W.; Gao, S.; Chen, B.; Cao, K. The application of decision tree to intensity change classification of tropical cyclones in western North Pacific. *Geophys. Res. Lett.* **2013**, *40*, 1883–1887. [CrossRef]

15. Kim, H.-S.; Kim, H.S. Development of scheme for tropical cyclone genesis using machine learning. In Proceedings of the Autumn Meeting, Busan, Korea, 31 October–2 November 2016; KMS: Seoul, Korea, 2016; pp. 847–848.

16. Quinlan, J. *C4.5: Programs for Machine Learning*; Morgan Kaufmann: Burlington, MA, USA, 1993; 302p.

17. Li, W.; Yang, C.; Sun, D. Mining geophysical parameters through decision-tree analysis to determine correlation with tropical cy-clone development. *Comput. Geosci.* **2009**, *35*, 309–316. [CrossRef]

18. Zhang, W.; Fu, B.; Peng, M.S.; Li, T. Discriminating developing versus non-developing tropical disturbances in the western North Pacific through decision tree analysis. *Weather Forecast.* **2015**, *30*, 446–454. [CrossRef]

19. Gao, S.; Zhang, W.; Liu, J.; Lin, I.-I.; Chiu, L.S.; Cao, K. Improvements in typhoon intensity change classification by incorporating an ocean coupling potential intensity index into decision trees. *Weather Forecast.* **2016**, *31*, 95–106. [CrossRef]

20. Park, M.S.; Kim, M.; Lee, M.I.; Im, J.; Park, S. Detection of tropical cyclone genesis via quantitative satellite ocean surface wind pattern and intensity analyses using decision trees. *Remote Sens. Environ.* **2016**, *183*, 205–214. [CrossRef]

21. Lee, H.-M.; Won, S.-H.; Cha, E.-J.; Jung, J.-U. Development of technique for tropical cyclone formation using machine learning. In Proceedings of the Autumn Meeting, Gyeongju, Korea, 30 October–1 November 2019; KMS: Seoul, Korea, 2019; pp. 544–545.

22. Kim, M.; Park, M.-S.; Im, J.; Park, S.; Lee, M.-I. Machine learning approaches for detecting tropical cyclone formation using satellite data. *Remote Sens.* **2019**, *11*, 1195. [CrossRef]

23. Nam, C.C.; Park, D.-S.R.; Ho, C.-H.; Chen, D. Dependency of tropical cyclone risk on track in South Korea. *Nat. Hazards Earth Syst. Sci.* **2018**, *18*, 3225–3234. [CrossRef]

24. Yang, R.; Tang, J.; Kafatos, M. Improved associated conditions in rapid intensifications of tropical cyclones. *Geophys. Res. Lett.* **2007**, *34*, L20807. [CrossRef]

25. Yang, R. A Systematic Classification Investigation of Rapid Intensification of Atlantic Tropical Cyclones with the SHIPS Database. *Weather Forecast.* **2016**, *31*, 495–513. [CrossRef]

26. Su, H.; Wu, L.; Jiang, J.H.; Pai, R.; Liu, A.; Zhai, A.J.; Tavallali, P.; DeMaria, M. Applying satellite observations of tropical cyclone internal structures to rapid intensification forecast with machine learning. *Geophys. Res. Lett.* **2020**, *47*, e2020GL089102. [CrossRef]

27. Wei, Y.; Yang, R. An Advanced Artificial Intelligence System for Investigating Tropical Cyclone Rapid Intensification with the SHIPS Database. *Atmosphere* **2021**, *12*, 484. [CrossRef]

28. Kaplan, J.; De Maria, M. Large-Scale Characteristics of Rapidly Intensifying Tropical Cyclones in the North Atlantic Basin. *Weather Forecast.* **2003**, *18*, 1093–1108. [CrossRef]

29. Price, J.F. Metrics of hurricane-ocean interaction: Vertically-integrated or vertically-averaged ocean temperature? *Ocean Sci.* **2009**, *5*, 351–368. [CrossRef]

30. Park, J.H.; Yae, D.E.; Lee, K.J.; Lee, H.J.; Lee, S.W.; Noh, S.; Kim, S.J.; Shin, J.Y.; Nam, S.H. Rapid decay of slowly moving typhoon Soulik (2018) due to interactions with the strongly stratified Northern East China Sea. *Geophys. Res. Lett.* **2019**, *46*, 14595–14603. [CrossRef]

31. Lin, I.-I.; Black, P.; Price, J.F.; Yang, C.Y.; Chen, S.S.; Lien, C.C.; Harr, P.; Chi, N.H.; Wu, C.C.; D'Asaro, E.A. An ocean coupling potential intensity index for tropical cyclones. *Geophys. Res. Lett.* **2013**, *40*, 1878–1882. [CrossRef]

32. Balaguru, K.; Foltz, G.R.; Leung, L.R.; D'Asaro, E.; Emanuel, K.A.; Liu, H.; Zedler, S.E. Dynamic Potential Intensity: An improved representation of the ocean's impact on tropical cyclones. *Geophys. Res. Lett.* **2015**, *42*, 6739–6746. [CrossRef]

33. Lee, W.; Kim, S.H.; Chu, P.S.; Moon, I.J.; Soloviev, A.V. An index to better estimate tropical cyclone intensity change in the western North Pacific. *Geophys. Res. Lett.* **2019**, *46*, 8960–8968. [CrossRef]

34. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*; Wadsworth & Brooks/Cole Advanced Books & Software: Monterey, CA, USA, 1984; ISBN 978-0-412-04841-8.

35. McLachlan, G.J.; Do, K.-A.; Ambroise, C. *Analyzing Microarray Gene Expression Data*; Wiley: Hoboken, NJ, USA, 2004.

36. Chawla, N.V. C4.5 and imbalanced data sets: Investigating the effect of sampling method, probabilistic estimate, and decision tree structure. In Proceedings of the International Conference on Machine Learning, Washington, DC, USA, 21–24 August 2003; International Machine Learning Society: Princeton, NJ, USA, 2003; Volume 3, pp. 66–73.

37. Powers, D.M. Evaluation: From precision, recall and f-measure to roc, informedness, markedness and correlation. *J. Mach. Learn. Technol.* **2011**, *2*, 37–63.

38. 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]

39. Kaplan, J.; DeMaria, M.; Knaff, J.A. A revised tropical cyclone rapid intensification index for the Atlantic and eastern North Pacific basins. *Weather Forecast.* **2010**, *25*, 220–241. [CrossRef]

40. Gao, S.; Chiu, L.S. Development of statistical typhoon intensity prediction: Application to satellite observed surface evaporation and rain rate (STIPER). *Weather Forecast.* **2012**, *27*, 240–250. [CrossRef]

41. Kaplan, J.; Rozoff, C.M.; DeMaria, M.; Sampson, C.R.; Kossin, J.P.; Velden, C.S.; Cione, J.J.; Dunion, J.P.; Knaff, J.A.; Zhang, J.A.; et al. Evaluating environmental impacts on tropical cyclone rapid intensification predictability utilizing statistical models. *Weather Forecast.* **2015**, *30*, 1374–1396. [CrossRef]
42. Knaff, J.A.; Sampson, C.R.; Musgrave, K.D. An Operational Rapid Intensification Prediction Aid for the Western North Pacific. *Weather Forecast.* **2018**, *33*, 799–811. [CrossRef]
43. Wada, A. Reexamination of tropical cyclone heat potential in the western north pacific. *J. Geophys. Res. Atmos.* **2016**, *121*, 6723–6744. [CrossRef]