



Article Evaluation of Tropical Cyclone Disaster Loss Using Machine Learning Algorithms with an eXplainable Artificial Intelligence Approach

Shuxian Liu¹, Yang Liu¹, Zhigang Chu^{2,*}, Kun Yang¹, Guanlan Wang¹, Lisheng Zhang¹ and Yuanda Zhang^{3,4}

- ¹ National Meteorological Center, China Meteorological Administration, Beijing 100081, China; liushuxian@cma.gov.cn (S.L.)
- ² Key Laboratory for Aerosol-Cloud-Precipitation of China Meteorological Administration, Nanjing University of Information Science and Technology, Nanjing 210044, China
- ³ Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science and Technology, Nanjing 210044, China
- ⁴ State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, Beijing 100081, China
- * Correspondence: chuzhigang@nuist.edu.cn

Abstract: In the context of global warming, tropical cyclones (TCs) have garnered significant attention as one of the most severe natural disasters in China, particularly in terms of assessing the disaster losses. This study aims to evaluate the TC disaster loss (TCDL) using machine learning (ML) algorithms and identify the impact of specific feature factors on the prediction of model with an eXplainable Artificial Intelligence (XAI) approach, SHapley Additive exPlanations (SHAP). The results show that LightGBM outperforms Random Forest (RF), Support Vector Machine (SVM), and Naive Bayes (NB) for estimating the TCDL grades, achieving the highest accuracy value of 0.86. According to the SHAP values, the three most important factors in the LightGBM classifier model are proportion of stations with rainfall exceeding 50 mm (ProRain), maximum wind speed (MaxWind), and maximum daily rainfall (MaxRain). Specifically, in the estimation of high TCDL grade, events characterized with MaxWind exceeding 30 m/s, MaxRain exceeding 200 mm, and ProRain exceeding 30% tend to exhibit a higher susceptibility to TC disaster due to positive SHAP values. This study offers a valuable tool for decision-makers to develop scientific strategies in the risk management of TC disaster.

Keywords: tropical cyclones; disaster loss; machine learning; XAI; SHAP

1. Introduction

Tropical cyclones (TCs) are among the most severe natural disasters in the world [1]. TCs trigger extreme winds, torrential rains, high waves, and storm surges, posing significant threats to human life, property, and coastal ecosystems [2–4]. China is frequently affected by TCs every year owing to its proximity to the northwest Pacific Ocean, which is one of the largest TC genesis regions in the world. Statistical data from 2001 to 2020 indicate that the direct economic loss and fatalities induced by TCs in China accounted for 17% and 10% of the total losses from meteorological disasters, respectively [5]. Moreover, the occurrence of extreme natural disasters has become more and more common, which is attributed to global warming and shifting climates [6,7]. Consequently, effective TC disaster management has emerged as a critical component in achieving sustainable development and resilience in the face of evolving risks in China.

Based on the well-established concept that TC disaster loss (TCDL) is primarily determined by hazard, vulnerability, and resilience, extensive studies have been conducted to examine the role of these three factors in TCDL assessment [8–12]. However, significant uncertainty still remains concerning the relevant conclusions. Some studies suggest that



Citation: Liu, S.; Liu, Y.; Chu, Z.; Yang, K.; Wang, G.; Zhang, L.; Zhang, Y. Evaluation of Tropical Cyclone Disaster Loss Using Machine Learning Algorithms with an eXplainable Artificial Intelligence Approach. *Sustainability* **2023**, *15*, 12261. https://doi.org/10.3390/ su151612261

Academic Editors: Paraskevas Tsangaratos, Ioanna Ilia, Wei Chen and Ioannis Matiatos

Received: 9 June 2023 Revised: 6 August 2023 Accepted: 8 August 2023 Published: 11 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). the impact of socio-economic development on TCDL is more significant than TC intensity. For instance, Schmidt et al. [13] employed the widely used nonlinear least squares algorithm, Levenberg–Marquardt, to investigate the influence of socio-economic factors and climate change on TCDL in the United States. Their findings revealed that losses attributed to socio-economic factors were approximately three times greater than those caused by climatic factors. Yonson et al. [14] utilized statistical methods to assess the impact of socio-economic vulnerability and hazard on TC-related fatalities. It was found that the number of deaths appeared to be more influenced by the poverty incidence rate rather than the rainfall amount during TC events. On the other hand, some studies argue that the impact of TC intensity change caused by climatic factors on disaster losses is more substantial. Ye et al. [15] used a negative binomial regression model to quantify the relationship between direct economic losses caused by TC and maximum wind speed, asset value, and per capita Gross Domestic Product (GDP), and the results showed that the effect of maximum wind speed on economic losses was greater than that of asset value and per capita GDP.

While physically based models have proven effective in solving weakly non-linear problems of low dimensionality, they are inadequate for accurate prediction of TCDL, which is complex, high-dimensional, and strongly non-linear in nature. Therefore, an effective assessment model for natural disasters should encompass multiple factors and reflect the complicated non-linear relationship between these factors and TCDL [16].

In this context, Artificial Intelligence (AI) models have been successfully applied in earth system science and hazard assessment, yielding more encouraging results compared to physical models [17–22]. Zhang et al. [23] employed five different models, including Back Propagation Neural Network (BPNN), 1D convolutional neural network, Decision Tree (DT), Random Forest (RF), and XGBoost, to examine the correlation between debrisflow-triggering factors and disaster losses. They found that the XGBoost model based on Gradient Boosting Decision Trees (GBDT) exhibited a significantly higher accuracy than the RF and other models. In 2017, LightGBM was introduced as an improved model of XGBoost by Microsoft and recognized as one of the most successful and advanced implementations of GBDT due to its exceptional speed and accuracy [24]. However, the use of AI models in natural hazard assessment is limited by the hindrance of lack of transparency and explainability, which stems from the inherent "black box" nature for most AI models [25,26].

Thus, it is of utmost significance that the model outputs can be explained and interpreted. The emergence of eXplainable AI (XAI) algorithms, such as SHapley Additive exPlanations (SHAP) [27], the Local Interpretable Model-agnostic Explanations (LIME) [28], etc., provides analyses to identify the contribution of each conditioning factor to the probability of natural hazard occurrences at a sample-wise scale, thereby enhancing the transparency of complex AI models. By representing feature attributions as a linear model, SHAP offers a unified framework for interpreting machine learning (ML) models that combines the strengths of both Shapley values and LIME. Felsche and Ludwi [29] used SHAP to understand the factors contributing to droughts and found that variables like the North Atlantic oscillation index and air pressure 1 month before the event prove essential for prediction. Aydin and Iban [30] employed SHAP to explain the generated ML-based flood susceptibility maps, and the results showed that lower elevations, lower slopes, and areas closer to river banks are more prone to flooding. Iban and Bilgilioglu [31] utilized SHAP to provide insights into how each factor affects the occurrence of snow avalanches and drew the conclusion that ski resorts with elevations of more than 2000 m and slopes of less than 30 degrees have a higher sensitivity to avalanches, as indicated by higher positive SHAP values.

As demonstrated above, XAI has gained widespread use recently and serves as a valuable instrument for devising innovative strategies to mitigate the harmful consequences of natural hazards. Despite the potential benefit of XAI, the current state of its application, its achievements, and the challenges it faces remain underexplored. Recent studies have extensively investigated the application of XAI in various natural disasters, including

droughts, floods, snow avalanches, and others. However, XAI methods for TC disaster management have yet to be fully evaluated and implemented. Therefore, in response to this gap, this study aims to further explore the potential of XAI methods for TCDL assessment.

The novelty of this study lies in the application of ML and XAI algorithms to predict TCDL and to further ascertain the factors that contribute to the predictive model and their relative significance. The study is structured as follows. Section 2 introduces the data and methods used in this study. Section 3 evaluates the performance of ML models and utilizes SHAP to provide interpretation and explanation for the predictions. Section 4 discusses the results and Section 5 draws the conclusion.

2. Data and Methods

2.1. Data Sources

This paper focuses on 492 disaster events caused by TC that occurred from 2000 to 2020 in different provinces in China, as depicted in Figure 1. Within the domain of ML research, the predictive performance of ML models heavily depends on the input features [32,33]. Constructing a comprehensive and scientific indicator system for the estimation of TCDL is of great significance, yet there is currently no unified system for TCDL indicators in China. Therefore, this study extensively collects open-source data and develops a relatively comprehensive indicator system covering three aspects of TCDL: the hazard of disaster-causing factors (maximum daily rainfall, maximum wind speed, etc.) [34], the vulnerability of the disaster-bearing body (provincial GDP, population, etc.) [35], and the resilience (beds of medical institutions, telephones, etc.) [36] (Table 1). Furthermore, the system incorporates multiple factors of society, economy, population, medical treatment, transportation, etc.



Figure 1. A total of 492 disaster events caused by TC that occurred from 2000 to 2020 in different provinces of China.

Category	Indicator	Short Name	Data Source	
Hazard	Maximum daily rainfall (mm)	MaxRain	Dataset of basic	
	Proportion of stations with rainfall exceeding 50 mm (%)	ProRain	meteorological elements from surface meteorological	
	Maximum wind speed (m/s)	MaxWind	stations in China (v3.0)	
	Proportion of stations with wind speed exceeding 14 m/s (%)	ProWind	(http://idata.cma/, accessed on 1 May 2023)	
Vulnerability	Provincial GDP (billion)	GDP		
	Population	POP		
	Population density per km ²	POPDens	National bureau of statistics (http://www.stats.gov.cn/,	
	Area of agricultural crop sown (hm ²)	CropArea		
	Area of buildings constructed (m ²)	ConsArea	accessed on 1 May 2023)	
	Area of buildings completed (m ²)	ComArea		
	Total line length of bus and trolley bus operation lines (km)	BUS		
Resilience	Beds of medical institutions per 10,000 people	MedBeds	National bureau of statistics (http://www.stats.gov.cn/, accessed on 1 May 2023)	
	Telephones per 100 people	TEL		
	Internet per 10,000 people	NET		
	Per capita GDP	PCGDP		
TCDL	Direct economic loss (billion)		Yearbook of meteorological disasters in China during 2000–2020 [37]	
	Casualties	—		
	Affected area (hm ²)	—		
	Collapsed houses	_		

Table 1. The categories and sources of samples.

2.2. Data Preparations

2.2.1. Adjustment of Economic Indicators

Considering the impact of inflation, it is not advisable to directly compare the same economic indicators between different years. Thus, the inflation should be eliminated to get the real indicator which can reflect the actual economy level by the GDP deflator [15]. The actual economic loss can be obtained according to Equation (1) as follows:

Actual economic loss = Nominal economic loss/GDP Deflator (1)

The GDP deflator data is from the website of World Bank (http://data.worldbank. org/datacatalog/world-development-indicators, accessed on 1 May 2023). The trend of China's GDP deflator from 2000 to 2020 is shown in Figure 2.

2.2.2. Normalization

As indicators usually have different units and orders of magnitude in a multi-indicator system it is necessary to normalize the indicators to ensure the reliability of the results [36]. Each indicator was normalized using Equation (2).

$$X_{ij}^* = (X_{ij} - min) / (max - min)$$
⁽²⁾

where X_{ij} and X_{ij}^* represent the values of indicator *j* in the *i*-th TC event before and after normalization, respectively, and *min* and *max* represent the minimum and maximum value of the given indicators among all TC events, respectively.



Figure 2. The trend of China's GDP deflator from 2000 to 2020.

2.2.3. Comprehensive Disaster Grade

In order to comprehensively and quantitatively evaluate the four disaster indicators casualties, actual economic losses, affected area, and collapsed houses, this study employs a combination of subjective and objective weighting methods to determine their respective weights. Specifically, the subjective weighting method utilized in this study is the expert scoring method [38], while the objective weighting method is the entropy method. The combined weight is calculated as follows:

$$w_j = \frac{\sqrt{\alpha_j \beta_j}}{\sum_{j=1}^4 \sqrt{\alpha_j \beta_j}} \tag{3}$$

where w_j represents the combined weight of indicator j, α_j is the weight obtained using the expert scoring method, and β_j is the weight calculated using the entropy method.

As shown in Table 2, the weight values for the four loss indicators, casualties, actual economic loss, collapsed houses, and affected area, are determined as $w_j = (0.33, 0.27, 0.21, 0.19)$ (j = 1, 2, 3, 4), respectively. The formula for calculating the comprehensive disaster index D_i is expressed as follows:

$$D_i = \sum_{j=1}^4 w_j \times X_{ij}^* \tag{4}$$

Table 2. The combined weight of each disaster index.

	Casualty	Actual Economic Loss	Collapsed Houses	Affected Area
Weight	0.33	0.27	0.21	0.19

The K-means algorithm was utilized to classify the 492 samples into low (73), moderate (216) and high-class (203) based on the comprehensive disaster index, denoted by green, blue, and red markers, respectively, in Figure 3.



Figure 3. The comprehensive disaster grades of samples, with the green, blue, and red markers representing the low, moderate and high-class of TCDL, respectively.

2.3. TCDL Evaluation System

In this study, the assessment of TCDL was conducted using four ML algorithms, LightGBM, Random Forest (RF), Support Vector Machine (SVM), and Naive Bayes (NB). SVM and NB are widely used single ML models, while RF and LightGBM are typical representatives of ensemble ML models based on bagging and boosting, respectively. Indicators of hazard, vulnerability, and resilience are employed as feature variables, and the comprehensive disaster grade is considered as the predictive variable for training and testing in ML models (Figure 4).



Figure 4. The workflow of the methodology in this paper.

80% of the total samples are randomly selected as the training set or cross-validation set (CV set), while the remaining 20% are designed as the test set (not involved in training). In order to enhance the robustness and ensure the stability of model, a 5-fold cross-validation method was utilized to train and fine-tune the model for optimal performance. Specifically, the training set was equally divided into 5 parts, with one part selected as the validation set in a non-repetitive manner, while the other four parts were used as the training set for parameter adjustment.

To assess the sensitivity of the feature variables to the label index, the probability density function (PDF) distributions of MaxRain and MaxWind are presented in Figure 5. It shows that the distributions of PDF across different categories are noticeably distinct both for MaxRain and MaxWind, which indicates a promising potential for the prediction. Similarly, this characteristic is observed for other feature variables as well. Furthermore, in comparison with MaxRain, the PDF of MaxWind shows more obvious peaks, displaying its greater significance in distinguishing the categories.



Figure 5. PDF distribution of (**a**) MaxRain and (**b**) MaxWind under low, moderate, and high classes of TCDL.

2.3.2. Model Tuning

To achieve the best performance of the LightGBM model, 7 parameters were selected for tuning, with the ranges exhibited in Table 3. A grid search method was subsequently employed to determine the optimal combination of parameters, involving a total of 37,500 iterations ($5 \times 5 \times 5 \times 5 \times 3 \times 4 \times 5$). The optimal model was selected based on the minimum value of Log loss, and the corresponding best parameter combination is presented in Table 3. Additionally, the parameter "is_unbalance" in LightGBM is set to "true" to effectively address the issue of data imbalance and enhance the model's generalization performance. The optimal parameter combinations for the other three ML models are omitted here.

Parameter	Dynamic Range	Optimal Value
num_leaves	Max number of leaves in one tree [10, 15, 20, 25, 30]	15
max_depth	Maximum depth of the tree [5, 6, 7, 8, 9]	7
max_bin	Max number of bins [5, 10, 15, 20, 25]	10
min_leaf	Minimal number of data in one leaf [10, 15, 20, 25, 30]	20
fea_frac	Fraction of features randomly selected on each tree [0.6, 0.8, 1.0]	1.0
learn_rate	Shrinkage rate [0.01, 0.03, 0.05, 0.1]	0.01
n_estimators	Number of boosting iteration [50, 100, 150, 200, 250]	100

Table 3. The range and optimal combination of parameters in LightGBM.

2.3.3. Evaluation Metrics of Models

The model evaluation was conducted with several widely used metrics in classification problems to quantitatively assess and compare the performance of models. These metrics include precision, accuracy, recall, and F1 score, which are calculated based on the true positive (*TP*), true negative (*TN*), false positive (*FP*), and false negative (*FN*) values.

Precision refers to the ratio of *TP* to the total number of positive predictions. It measures the ability of model to accurately identify positive instances. The formula is expressed as follows:

$$Precision = \frac{IP}{TP + FP}$$
(5)

Accuracy represents the ratio of correctly predicted instances (both *TP* and *TN*) to the total number of instances. It provides an overall measure of how well the model performs. The formula is defined as follows:

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN}$$
(6)

Recall, also known as sensitivity or true positive rate, calculates the ratio of *TP* to the total number of actual positive instances. It measures the ability of model to identify all positive instances correctly. The formula is shown as follows:

$$Recall = \frac{TP}{TP + FN}$$
(7)

The F1 score is a harmonic mean of precision and recall. It provides a balanced evaluation of the model's performance, considering both precision and recall simultaneously. The formula is as follows:

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall}$$
(8)

By utilizing these metrics, the results of a model can be quantitatively evaluated and compared, allowing for a comprehensive assessment of its performance.

2.4. SHapley Additive exPlanations (SHAP)

SHAP was initially introduced in game theory by Shapley [27] as a method to assess the individual contributions of players in a collaborative game. Its primary objective is to distribute the overall gain among players in proportion to their respective contributions to the final outcome. By introducing SHAP values, a solution is provided to address the challenge of fairly rewarding each player while assigning a distinct value that considers local accuracy, consistency, and null effect [27]. In contrast to other models for computing global feature importance, such as information gain ratio or permutation feature importance, SHAP allows for a sample-wise evaluation of the impact of each conditioning factor. It has been successfully employed in various studies related to natural hazard susceptibility mapping, including water erosion [39], wildfires [40], and landslides [41]. Due to its outstanding performance, SHAP was utilized in this study to reveal the reasoning behind TCDL prediction.

The Python-based SHAP library developed by Lundberg and Lee [42] was utilized for calculating SHAP values. A larger mean absolute Shapley value (|SHAP|) indicates a conditioning factor's greater importance for the output feature. The direction of a conditioning factor's contribution can be determined by its positive or negative SHAP values [30]. Scholars have employed a range of SHAP plots and visualizations, including force plots, summary graphs, and dependence plots, to effectively showcase the global and local significance of specific factors and samples for the model's output.

The recent advancements in machine learning algorithms, as demonstrated by Lundberg and Lee [42], have paved the way for gaining deeper insights into model outputs, thereby enhancing transparency in traditionally opaque black box models.

3. Results

3.1. Model Evaluation

The evaluation metrics used in this study to assess the performance of ML models include accuracy, recall, precision, and F1 score. As presented in Table 4, the single-model algorithms, SVM and NB, show considerably lower performance compared to the ensemble-model algorithms, RF and LightGBM. Notably, the LightGBM model based on boosting outperforms the RF model based on bagging and exhibits the best performance among the four models. The accuracy and precision of LightGBM reach 0.86 and 0.83, respectively, indicating its ability to accurately predict comprehensive disaster losses in TC events. Moreover, the recall value of 0.83 demonstrates that the LightGBM model effectively identifies positive cases of high disaster losses in TC events. The F1 score, which considers both precision and recall, also reaches 0.83, suggesting a well-balanced performance for LightGBM between the two metrics. Overall, these results strongly support the suitability of LightGBM for the prediction of TCDL.

AccuracyRecallPrecisionLightGBM0.860.830.83

Table 4. Assessment of TCDL based on four ML algorithms.

0.71

0.64

0.64

3.2. Interpretation of the LightGBM Model

As illustrated in Section 3.1, LightGBM has superior performance compared to the single-model algorithms SVM and NB, as well as the ensemble model RF, for the prediction of TCDL. Consequently, the LightGBM model was selected to be explained and interpreted using the SHAP approach in Section 3.2.

0.7

0.54

0.63

3.2.1. SHAP Summary Plots

RF

SVM

NB

Figure 6 presents the sample-wise SHAP summary plot of input feature factors derived from the LightGBM classifier. The feature factors are ranked based on their contributions. The X-axis represents the SHAP value, while the Y-axis represents the feature factors. Each dot on the plot corresponds to a sample of a TC disaster event from the test dataset, with the color indicating the value of a specific factor. Sky blue signifies a lower value, while magenta denotes a higher value. The horizontal position of the dot indicates whether the feature factor has a positive or negative influence on the prediction.

F1

0.83

0.7

0.53

0.62

0.72

0.52

0.67



Figure 6. Local explanation of the LightGBM classifier using sample-wise SHAP values. (a) Low class, (b) moderate class, (c) high class.

As depicted in Figure 6a, in the low class of TCDL, samples with higher values of ProRain (proportion of stations with rainfall exceeding 50 mm), MaxWind (maximum wind speed), NET (internet per 10,000 people), and CropArea (area of agricultural crop sown) display negative SHAP values. Conversely, samples with higher values of ProWind (proportion of stations with wind speed exceeding 14 m/s), TEL (telephones per 100 people), and MedBeds (beds of medical institutions per 10,000 people) exhibit positive SHAP values. This indicates that ProRain, MaxWind, NET, and CropArea have an adverse impact on the likelihood of the low TCDL class, while ProWind, TEL, and MedBeds have a favorable influence on it. In the moderate class of TCDL, PCGDP (per capita GDP) and CropArea have positive SHAP values (Figure 6b), which illustrates that the likelihood of TCDL increases as PCGDP and CropArea increase. It can be seen from Figure 6c that the magenta dotted MaxWind, ProRain, and MaxRain (maximum daily rainfall) values have positive impacts on the prediction ability for the high TCDL class. However, the situation is reversed for

the low class, in which MaxWind, ProRain, and MaxRain have negative impacts on TCDL. Moreover, in comparison with the minor and positive impacts on the low and moderate TCDL classes, PCGDP shows obvious negative impacts on the high TCDL class. This also reveals that a higher PCGDP will reduce the risk of severe TC disasters.

Figure 7 displays the mean of the absolute SHAP (|SHAP|) values for all input feature factors in the test dataset. The |SHAP| values provide insight into the magnitude of the impact for each feature factor in the LightGBM classifier model. The higher the mean |SHAP| value, the more significant the contribution of the respective feature factor to the overall prediction process.



Figure 7. Local explanation of LightGBM classifiers using mean |SHAP| values for each conditioning factor.

It can be seen from Figure 7 that ProRain (proportion of stations with rainfall exceeding 50 mm) and MaxWind (maximum wind speed) play a significant role in all three classes of TCDL. Their contributions to the prediction of TCDL grades are almost twice those of the other feature factors. Conversely, the contribution of the vulnerability factors is relatively lower when compared to hazard and resilience in general. In the moderate class of TCDL, the overall contribution of all feature factors is smaller in comparison with their contribution in the low and high classes. This indicates that the impact of feature factors on the model's prediction varies across different classes of TCDL. For instance, PCGDP (per capita GDP) presents a mean |SHAP| value close to 0 in low-class predictions. However, it exhibits a relatively substantial contribution to moderate and high-class predictions, with mean |SHAP| values reaching approximately 0.4.

3.2.2. SHAP Dependence Plots

Figure 8 depicts the SHAP dependence plot for the four most significant contributing factors (MaxWind, ProRain, MaxRain, and MedBeds) in the high TCDL class. The SHAP dependence plot can identify the relationship between a single factor (X-axis) and the corresponding SHAP values generated (Y-axis) to evaluate the effect of each feature factor on prediction accuracy.



Figure 8. SHAP dependence plots of sample-wise factor values vs. the corresponding SHAP values for (**a**) MaxWind, (**b**) ProRain, (**c**) MaxRain, and (**d**) MedBeds.

As shown in Figure 8a, samples with MaxWind (maximum wind speed) values of less than approximately 30 m/s have negative SHAP values, which implies a negative contribution to the likelihood of TCDL. Conversely, samples with MaxWind values greater than 30 m/s have positive SHAP values, highlighting a positive contribution to the probability of TCDL. In addition, a quasi-linear relationship exists between MaxWind and its corresponding SHAP values. Regarding ProRain (proportion of stations with rainfall exceeding 50 mm), samples with values below approximately 30% have negative SHAP values, while samples with values above 30% have positive SHAP values (Figure 8b). In general, the SHAP value rises as the value of ProRain increases. It can be observed from Figure 8c that samples with MaxRain (maximum daily rainfall) values of more than approximately 200 mm exhibit positive SHAP values, revealing that the model is more likely to predict a higher probability of TCDL when it encounters an extreme rainfall event. From Figure 8d, it is evident that there appears to be a quasi-linear relationship between NET (internet per 10,000 people) and its corresponding SHAP values. The SHAP value decreases as the value of NET increases when the NET value is less than approximately 48. Figure 9a–c display the probability waterfall plots for three samples of low, moderate, and high TCDL classes, respectively. The total probability value (f(x)) for each sample is marked in black at the top right and calculated using the SHAP value. Additionally, the factors that have positive influences on the total probability are depicted in magenta, while the factors that have negative influences are represented in light blue, along with their corresponding probability values.



Figure 9. Probability waterfall plots for three different cases for (**a**) low class, (**b**) moderate class, and (**c**) high class of TCDL. The magenta and blue colors denote the positive and negative impacts, respectively.

In Figure 9a, the highest positive probability value of 0.46 for a test sample in the low class of TCDL is produced by ProRain (proportion of stations with rainfall exceeding 50 mm) with the value of 0. Additionally, values of 75.48 for MedBeds (beds of medical

institutions per 10,000 people) and 33.4 m/s for MaxWind (maximum wind speed) also result in high positive probability values. On the other hand, for a test sample in the moderate class (Figure 9b), the ProRain value of 42.86% produces a negative probability value when compared to that in the low class. For the sample in the high TCDL class (Figure 9c), factors such as MaxWind with a value of 49.3 m/s and MaxRain (maximum daily rainfall) with a value of 303.5 mm generate positive probability values of 0.37 and 0.18, respectively, illustrating a prediction of high TCDL susceptibility. Overall, it can be concluded from Figure 9c that samples with a value of MaxWind exceeding 30 m/s, a value of MaxRain exceeding 200 mm, and a value of ProRain exceeding 30% generally have a high risk of TC disaster, which is also confirmed in the SHAP dependence plot (Figure 8).

4. Discussion

In natural disaster research, ML algorithms have gained prominence as one of the most successful strategies. The prediction capabilities of single-model and ensemble-model classifiers (NB, SVM, RF, LightGBM) for generating the TCDL grade are compared in this study. LightGBM, based on the GBDT algorithm, exhibits superior performance compared to the other classifiers in all performance criteria. Other scholars have also indicated that GBDT-based ensemble classifiers surpass the other tree-based ensemble classifiers [23,30]. However, the results of Zhang et al. [36] showed that RF, based on the bagging algorithm, demonstrates the best performance when compared to other tree-based models. Hence, more comparisons are necessary to determine the suitable classifiers for natural disaster assessment.

The indicators of hazard, vulnerability, and resilience are incorporated into the system for estimating TCDL grades. According to the SHAP value, ProRain (proportion of stations with rainfall exceeding 50 mm) and MaxWind (maximum wind speed) are the two most important contributing factors, followed by MaxRain (maximum daily rainfall), MedBeds (beds of medical institutions per 10,000 people), and ProWind (proportion of stations with wind speed exceeding 14 m/s). Moreover, the factors of hazard and resilience have larger SHAP values than vulnerability in general, indicating a greater contribution to TCDL grade prediction. Similarly, Ye et al. [15] have found that the effect of maximum wind speed during TC on economic losses is greater than that of asset value and per capita GDP. Nevertheless, this claim will vary in different regions with respect to different natural hazards [13,14,36].

For the low class of TCDL, events characterized by higher values of ProRain (proportion of stations with rainfall exceeding 50 mm), MaxWind (maximum wind speed), NET (internet per 10,000 people), and CropArea (area of agricultural crop sown) display negative SHAP values. Conversely, events with higher values of ProWind (proportion of stations with wind speed exceeding 14 m/s), TEL (telephones per 100 people), and MedBeds (beds of medical institutions per 10,000 people) exhibit positive SHAP values. As a result, ProRain, MaxWind, NET, and CropArea have an adverse impact on the likelihood of the low TCDL class, while ProWind, TEL, and MedBeds have a favorable influence on it. For the moderate class, PCGDP (per capita GDP) and CropArea have positive SHAP values, suggesting a positive impact on the likelihood of a TC disaster event. It is noted for the high class of TCDL that MaxWind, ProRain, and MaxRain (maximum daily rainfall) values have a positive impact on the prediction ability, which is contrary to the low class. Moreover, events with values of MaxWind exceeding 30 m/s, ProRain greater than about 30%, and MaxRain of more than about 200 mm tend to produce positive SHAP values, implying a positive contribution to the probability of TCDL.

5. Conclusions

Tropical cyclones are among the most challenging natural hazards to be predicted due to multiple factors and the complex nonlinear relationships between them. Therefore, the assessment of TCDL is essential for TC disaster prevention, risk mitigation, and decisionmaking. The primary objective of this study is to develop a model for estimating TCDL grades based on ML algorithms and enhance the transparency and explainability of prediction process by using XAI approaches. This will allow decision-makers to transform their perception of ML as a black box into a transparent and explainable technology, enabling them to make informed judgments based on XAI interpretation. The main findings of the study are as follows:

- Among the four ML models (LightGBM, RF, SVM, NB), LightGBM demonstrates superior performance, achieving the highest values for accuracy (0.86), recall (0.83), precision (0.83), and F1 score (0.83).
- For the estimation of all three classes (low, moderate, high) of TCDL, ProRain (proportion of stations with rainfall exceeding 50 mm) and MaxWind (maximum wind speed) exhibit notable significance. And their contributions to TCDL grade prediction are approximately twice as substantial as those of other feature factors. In contrast, the impact of vulnerability factors is relatively lower when compared to hazard and resilience factors in general.
- Specifically, the impact of each feature factor on the model's prediction varies across in the low, moderate, and high classes of TCDL. In terms of the high class, events characterized by MaxWind (maximum wind speed) with values exceeding 30 m/s, MaxRain (maximum daily rainfall) with values exceeding 200 mm, and ProRain (proportion of stations with rainfall exceeding 50 mm) with values exceeding 30% tend to present a higher risk of TCDL.
- Future work will focus on incorporating remote sensing data for enhanced coverage and spatial resolution, along with exploring other additive SHAP properties for TCDL assessment.

Author Contributions: Conceptualization, S.L. and Z.C.; Methodology, S.L. and L.Z.; Software, K.Y.; Validation, K.Y. and Y.Z.; Investigation, S.L. and L.Z.; Data curation, S.L., Z.C. and G.W.; Writing—original draft, S.L. and Y.L.; Writing—review and editing, S.L. and Z.C.; Visualization, S.L. and Y.Z.; Supervision, Z.C. All authors have read and agreed to the published version of the manuscript.

Funding: The work was supported by the National Natural Science Foundation of China (42075190) and the Youth Fund Project of the National Meteorological Center (Q202212).

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

- Bakkensen, L.A.; Mendelsohn, R.O. Global tropical cyclone damages and fatalities under climate change: An updated assessment, hurricane risk. In *Hurricane Risk*; Collins, J.M., Walsh, K., Eds.; Springer: Berlin/Heidelberg, Germany, 2019; pp. 179–197.
- Dube, S.; Jain, I.; Rao, A.; Murty, T. Storm surge modelling for the bay of Bengal and Arabian Sea. *Nat. Hazards* 2009, 51, 3–27. [CrossRef]
- 3. Krapivin, V.F.; Soldatov, V.Y.; Varotsos, C.A.; Cracknell, A.P. An adaptive information technology for the operative diagnostics of the tropical cyclones; solar-terrestrial coupling mechanisms. *J. Atmos. Sol. Terr. Phys.* **2012**, *89*, 83–89. [CrossRef]
- Sahoo, B.; Bhaskaran, P.K. Multi-hazard risk assessment of coastal vulnerability from tropical cyclones—A GIS based approach for the Odisha coast. J. Environ. Manag. 2018, 206, 1166–1178. [CrossRef] [PubMed]
- 5. Li, Y.; Zhao, S.; Wang, G. Spatiotemporal variations in meteorological disasters and vulnerability in China during 2001–2020. *Front. Earth Sci.* **2021**, *9*, 789523. [CrossRef]
- 6. Moon, I.J.; Kim, S.H.; Chan, J. Climate change and tropical cyclone trend. *Nature* 2019, 570, 3–5. [CrossRef]
- Knutson, T.; Camargo, S.J.; Chan, J.C.L.; Emanuel, K.; Ho, C.-H.; Kossin, J.; Mohapatra, M.; Satoh, M.; Sugi, M.; Walsh, K.; et al. Tropical cyclones and climate change assessment: Part ii: Projected response to anthropogenic warming. *Bull. Amer. Meteor. Soc.* 2020, 101, 303–322. [CrossRef]
- 8. Schiermeier, Q. Hurricane link to climate change is hazy. *Nature* 2005, 437, 461. [CrossRef]

- 9. Mendelsohn, R.; Emanuel, K.; Chonabayashi, S.; Bakkensen, L. The impact of climate change on global tropical cyclone damage. *Nat. Clim. Chang.* **2012**, *2*, 205–209. [CrossRef]
- 10. Peduzzi, P.; Chatenoux, B.; Dao, H.; De, B.A.; Herold, C.; Kossin, J.; Mouton, F.; Nordbeck, O. Global trends in tropical cyclone risk. *Nat. Clim. Chang.* 2012, *2*, 289–294. [CrossRef]
- 11. Gettelman, A.; Bresch, D.N.; Chen, C.C.; Truesdale, J.E.; Bacmeister, J.T. Projections of future tropical cyclone damage with a high-resolution global climate model. *Clim. Chang.* **2018**, *146*, 575–585. [CrossRef]
- 12. Nam, C.C.; Park, D.R.; Ho, C.; Chen, D. Dependency of tropical cyclone risk on track in South Korea. *Nat. Hazard Earth Sys.* 2018, 18, 3225–3234. [CrossRef]
- Schmidt, S.; Kemfert, C.; Hoeppe, P. The impact of socio-economics and climate change on tropical cyclone losses in the USA. *Reg. Environ. Chang.* 2010, 10, 13–26. [CrossRef]
- Yonson, R.; Noy, I.; Gaillard, J.C. The measurement of disaster risk: An example from tropical cyclones in the Philippines. *Rev. Dev. Econ.* 2018, 22, 736–765. [CrossRef]
- 15. Ye, M.; Wu, J.; Liu, W.; He, X.; Wang, C. Dependence of tropical cyclone damage on maximum wind speed and socioeconomic factors. *Environ. Res. Lett.* **2020**, *15*, 094061. [CrossRef]
- 16. Sun, H.; Wang, J.; Ye, W. A Data Augmentation-Based Evaluation System for Regional Direct Economic Losses of Storm Surge Disasters. *Int. J. Environ. Res. Public Health* **2021**, *18*, 2918. [CrossRef]
- 17. Abbot, J.; Marohasy, J. Input selection and optimisation for monthly rainfall forecasting in Queensland, Australia, using artificial neural networks. *Atmos. Res.* **2014**, *138*, 166–178. [CrossRef]
- 18. Sahin, E.K. Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest. *SN Appl. Sci.* **2020**, *2*, 1308. [CrossRef]
- Saber, M.; Boulmaiz, T.; Guermoui, M.; Abdrabo, K.I.; Kantoush, S.A.; Sumi, T.; Boutaghane, H.; Nohara, D.; Mabrouk, E. Examining LightGBM and CatBoost models for wadi flash flood susceptibility prediction. *Geocarto Int.* 2021, 37, 7462–7487. [CrossRef]
- Zhang, H.; Song, Y.; Xu, S.; He, Y.; Li, Z.; Yu, X.; Liang, Y.; Wu, W.; Wang, Y. Combining a class-weighted algorithm and machine learning models in landslide susceptibility mapping: A case study of Wanzhou section of the Three Gorges Reservoir, China. *Computers. Geosci.* 2022, 158, 104966. [CrossRef]
- 21. Darvishi, B.A.; Neysani, S.N.; Papi, R.; Soleimani, M. Dust source susceptibility mapping in Tigris and Euphrates basin using remotely sensed imagery. *CATENA* 2022, 209, 105795. [CrossRef]
- Panahi, M.; Rahmati, O.; Rezaie, F.; Lee, S.; Mohammadi, F.; Conoscenti, C. Application of the group method of data handling (GMDH) approach for landslide susceptibility zonation using readily available spatial covariates. *CATENA* 2022, 208, 105779. [CrossRef]
- 23. Zhang, Y.; Ge, T.; Tian, W.; Liou, Y.-A. Debris flow susceptibility mapping using machine-learning techniques in Shigatse Area, China. *Remote Sens.* **2019**, *11*, 2801. [CrossRef]
- Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS), Long Beach, CA, USA, 4–9 December 2017; pp. 3149–3157.
- 25. Chakraborty, D.; Basagaoglu, H.; Winterle, J. Interpretable vs. noninterpretable machine learning models for data-driven hydro-climatological process modelling. *Expert Syst. Appl.* **2021**, *170*, 114498. [CrossRef]
- Dikshit, A.; Pradhan, B. Why interpretable machine learning algorithms should be used in drought forecasting? In Proceedings of the Natural Hazards Alerts, NSF Convergence Workshop, Online, 24–28 May 2021.
- 27. Shapley, L.S. Stochastic games. Proc. Natl. Acad. Sci. USA 1953, 39, 1095–1100. [CrossRef] [PubMed]
- Ribeiro, M.T.; Singh, S.; Guestrin, C. "Why should I trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 1135–1144.
- 29. Felsche, E.; Ludwi, R. Applying machine learning for drought prediction in a perfect model framework using data from a large ensemble of climate simulations. *Nat. Hazards Earth Syst. Sci.* **2021**, *21*, 3679–3691. [CrossRef]
- Aydin, H.E.; Iban, M.C. Predicting and analyzing flood susceptibility using boosting-based ensemble machine learning algorithms with SHapley Additive ExPlanations. *Nat. Hazards* 2023, *116*, 2957–2991. [CrossRef]
- Iban, M.C.; Bilgilioglu, S.S. Snow avalanche susceptibility mapping using novel tree-based machine learning algorithms (XGBoost, NGBoost, and LightGBM) with eXplainable Artificial Intelligence (XAI) approach. *Stoch. Environ. Res. Risk Assess.* 2023, 37, 2243–2270. [CrossRef]
- An, S.; Wang, J.; Wei, J. Local-Nearest-Neighbors-Based Feature Weighting for Gene Selection. IEEE/ACM Trans. Comput. Biol. Bioinform. 2017, 15, 1538–1548. [CrossRef]
- An, S.; Wang, J.; Wei, J.; Yang, Z. Unsupervised Feature Selection with Joint Clustering Analysis. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, Singapore, Singapore, 6–10 November 2017; pp. 1639–1648.
- Lou, W.; Chen, H.; Shen, X.; Sun, K.; Deng, S. Fine assessment of tropical cyclone disasters based on GIS and SVM in Zhejiang Province, China. *Nat. Hazards* 2012, 64, 511–529. [CrossRef]
- Fricker, T.; Elsner, J.B.; Jagger, T.H. Population and energy elasticity of tornado casualties. *Geophys. Res. Lett.* 2017, 44, 3941–3949. [CrossRef]

- 36. Zhang, S.; Zhang, J.; Li, X.; Du, X.; Zhao, T.; Hou, Q.; Jin, X. Estimating the grade of storm surge disaster loss in coastal areas of china via machine learning algorithms. *Ecol. Indic.* 2022, *136*, 108533. [CrossRef]
- 37. China Meteorological Administration (CMA). *Yearbook of Meteorological Disasters in China* 2000–2020; China Meteorological Press: Beijing, China, 2021. (In Chinese)
- Wang, X.R.; Zhang, L.S.; Li, W.B. Improvement and application analysis of the comprehensive grade evaluation model of typhoon disaster. *Meteor. Mon.* 2018, 44, 304–312. (In Chinese)
- 39. Mohammadifar, A.; Gholami, H.; Comino, J.R.; Collins, A.L. Assessment of the interpretability of data mining for the spatial modelling of water erosion using game theory. *CATENA* **2021**, 200, 105178. [CrossRef]
- 40. Iban, M.C.; Sekertekin, A. Machine learning based wildfire susceptibility mapping using remotely sensed fire data and GIS: A case study of Adana and Mersin provinces. *Turk. Ecol. Inf.* **2022**, *69*, 101647. [CrossRef]
- 41. Zhou, X.; Wen, H.; Li, Z.; Zhang, H.; Zhang, W. An interpretable model for the susceptibility of rainfall-induced shallow landslides based on SHAP and XGBoost. *Geocarto Int.* **2022**, *37*, 13419–13450. [CrossRef]
- 42. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 4765–4774.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.