Earth's Future

Key Points:

- Global hotspots for marine heatwave-induced rapid intensifications identified, accounting for over $360 billion in losses only in 2021–2023
- An ensemble machine learning model is developed to predict this destructive subset of RI events
- We quantify enhanced predictive skill achieved by using marine heatwave intensity metrics as predictors compared to sea surface temperature

Global Predictability of Marine Heatwave Induced Rapid Intensification of Tropical Cyclones

Soheil Radfar[1,2] , Ehsan Foroumandi[1,2] , Hamed Moftakhari[1,2] , Hamid Moradkhani[1,2] , Gregory R. Foltz[3] , and Alex Sen Gupta[4]

[1]Center for Complex Hydrosystems Research, The University of Alabama, Tuscaloosa, AL, USA, [2]Department of Civil, Construction and Environmental Engineering, The University of Alabama, Tuscaloosa, AL, USA, [3]Atlantic Oceanographic and Meteorological Laboratory, The National Oceanic and Atmospheric Administration (NOAA), Miami, FL, USA, [4]Climate Change Research Centre and ARC Centre of Excellence for Climate Extremes, University of New South Wales, Sydney, NSW, Australia

Abstract Prediction of the rapid intensification (RI) of tropical cyclones (TCs) is crucial for improving disaster preparedness against storm hazards. These events can cause extensive damage to coastal areas if occurring close to landfall. Available models struggle to provide accurate RI estimates due to the complexity of underlying physical mechanisms. This study provides new insights into the prediction of a subset of rapidly intensifying TCs influenced by prolonged ocean warming events known as marine heatwaves (MHWs). MHWs could provide sufficient energy to supercharge TCs. Preconditioning by MHW led to RI of recent destructive TCs, Otis (2023), Doksuri (2023), and Ian (2022), with economic losses exceeding $150 billion. Here, we analyze the TC best track and sea surface temperature data from 1981 to 2023 to identify hotspot regions for compound events, where MHWs and RI of tropical cyclones occur concurrently or in succession. Building upon this, we propose an ensemble machine learning model for RI forecasting based on storm and MHW characteristics. This approach is particularly valuable as RI forecast errors are typically largest in favorable environments, such as those created by MHWs. Our study offers insight into predicting MHW TCs, which have been shown to be stronger TCs with potentially higher destructive power. Here, we show that using MHW predictors instead of the conventional method of using sea surface temperature reduces the false alarm rate by 30%. Overall, our findings contribute to coastal hazard risk awareness amidst unprecedented climate warming causing more frequent MHWs.

Plain Language Summary Tropical cyclones (TCs) that rapidly intensify pose a significant threat to coastal areas. Current prediction models struggle to accurately forecast these events due to the complex nature of underlying physical processes. This study focuses on improving predictions for a specific type of rapidly intensifying TCs influenced by prolonged ocean warming events known as marine heatwaves. By developing an ensemble machine learning model that considers storm and marine heatwaves characteristics, we provide a new tool for detecting rapid intensification events that are particularly challenging to forecast in favorable environments created by MHWs. These events tend to produce stronger TCs. The developed model demonstrated higher accuracy in detecting rapid intensification while reducing false alarms. This research enhances coastal hazard preparedness in the face of rising ocean temperatures.

1. Introduction

Tropical cyclones (TCs) that strike coastal areas have the potential to cause severe flooding, torrential rains, and strong winds often resulting in severe economic disruption, property damage and casualties. According to the list of documented catastrophic events from the Emergency Events Database (EM-DAT), the annual average damage of TCs between 2000 and 2023 exceeds $63 billion (Delforge et al., 2023). The extensive impacts can be more severe when TCs undergo rapid intensification (RI) (Lok et al., 2021). TC RI is defined as a process wherein the maximum sustained wind speed of a TC increases by at least 30 kt over a 24 hr period (Kaplan & DeMaria, 2003).

Predicting RI remains challenging because as it involves the understanding of internal dynamics, interplay of environmental parameters, and nonlinear processes (Kieu & Zhang, 2009; Neetu et al., 2020; L. Zhang & Oey, 2019). RI forecast errors are estimated to be 2–3 times larger due to limited observational data and lack of understanding and prediction of inner-core processes (Bhatia et al., 2022; Trabing & Bell, 2020). This positions RI at the forefront of the TC forecasting field (Gall et al., 2013) and signals the urgent need for enhanced

Validation: Soheil Radfar,
Ehsan Foroumandi, Gregory R. Foltz,
Alex Sen Gupta
Visualization: Soheil Radfar
Writing – original draft: Soheil Radfar,
Ehsan Foroumandi
Writing – review & editing:
Hamed Moftakhari, Hamid Moradkhani,
Gregory R. Foltz, Alex Sen Gupta

forecasting capabilities and adaptive strategies to mitigate the heightened risks of TCs. To date, researchers have predicted RI using a variety of statistical methods and numerical models (W. Wang et al., 2023; Z. Zhang et al., 2023). Among the widely used models in these categories are the Statistical Hurricane Intensity Prediction Scheme Rapid Intensification Index (SHIPS-RII) (Kaplan & DeMaria, 2003; Kaplan et al., 2010) and the Hurricane Weather Research and Forecasting (HWRF) (Tallapragada et al., 2014). In addition, substantial effort has gone into using data assimilation techniques such as ensemble Kalman filter for numerical weather prediction or forecasting models to enhance their RI predictability skill (Hartman et al., 2023; Munsell et al., 2017; Tao et al., 2022).

The primary challenge of data-driven RI forecasting models is choosing predictors that inform RI formation and evolution (Narayanan et al., 2023; Rozoff et al., 2015). Therefore, there is a growing interest in developing a more robust set of environmental predictors (Balaguru et al., 2020; Emanuel et al., 2004; Kowch & Emanuel, 2015). While high sea surface temperature (SST) and ocean heat content, low vertical wind shear, and abundant atmospheric moisture are generally recognized as favorable environmental conditions for triggering RI (Mawren et al., 2022), their coexistence is not necessarily required for RI. For example, Hurricane Irene (1999), Hurricane Edouard (2014) (J. A. Zhang et al., 2017), Hurricane Irma (2017) (Fischer et al., 2018), and Hurricane Michael (2017) (Le Hénaff et al., 2021) developed and underwent RI despite high vertical wind shear. Similarly (Rathore et al., 2022), showed that cyclone Amphan (2020) experienced RI in the Bay of Bangal despite high vertical wind shear. This was largely attributed to the presence of a strong collocated marine heatwave (MHW) before and during the development phase of Amphan. The role of MHW as a precursor to TC RI has been highlighted in recent studies (Dzwonkowski et al., 2020; Mawren et al., 2022; Pun et al., 2023). One of the most recent studies over the western North Pacific and North Atlantic basins showed that the TCs that passed over MHWs intensified roughly three times more rapidly than the TCs that did not (Choi et al., 2024). Although it is commonly known that SST is a reliable indicator of the RI process (Foltz et al., 2018), the influential characteristics of MHW for initiating RI events have not yet been fully explored. Additionally, to the best of our knowledge, there is no research done on utilizing MHW characteristics as predictors of RI events.

Over the last decade, machine learning (ML) and deep learning (DL) methods have emerged as viable means of predicting TC RI. For example, the long short-term memory (LSTM) and the convolutional neural network (CNN) are among the most widely used DL methods in experimental RI prediction models (Griffin et al., 2024; Li et al., 2017; C.-H. Zhang & Zhang, 2023; Zhou et al., 2022). Aside from the computational costs of DL-based models, additional challenges could arise in selecting a candidate architecture and designing layers. By contrast, shallow ML-based algorithms are more popular for RI predictions because of their simplicity and low computational demands. The supervised ML models typically attempt to use individual or ensemble algorithms to solve a classification or regression problem and identify RI occurrences. The first ML effort in this field focused on the application of the Naive Bayesian and logistic regression techniques. Rozoff and Kossin (2011) used SHIPS to train these models for the probabilistic prediction of rapid intensity change over the North Atlantic and eastern North Pacific Ocean basins. Apart from the logistic regression algorithm, which is the most frequently used algorithm in the literature (Kaplan et al., 2015; Knaff et al., 2023; Ko et al., 2023; Narayanan et al., 2023; Su et al., 2020; Tam et al., 2021), several studies have used tree-based algorithms; for instance, Decision Tree (Narayanan et al., 2023; Su et al., 2020), Random Forest (Ko et al., 2023; A. Mercer & Grimes, 2017; Su et al., 2020), Extremely Randomized Trees (Su et al., 2020), and eXtreme Gradient Boosting (XGBoost) (Wei & Yang, 2021). Common alternatives included using support vector machines (A. Mercer & Grimes, 2015, 2017; A. E. Mercer et al., 2021) or artificial neural networks (Ko et al., 2023; A. Mercer & Grimes, 2017).

ML-based methods, as data-driven models, primarily work by utilizing the existing information within the data, making them reliant on the quality and representativeness of the data. Another limitation for predicting RI with more accuracy is the imbalanced data issue, which makes it difficult to train accurate models (DeMaria et al., 2021; Yang et al., 2020). In ML, imbalanced data often leads to models being overexposed to the majority samples during training, causing a bias toward those instances and neglect of the minority samples. Considering that RI events are generally limited compared to non-RI situations, resampling techniques should be used to prevent models from being skewed toward the majority class. These techniques adjust the dataset composition, either by augmenting the minority class (referred to as oversampling) or by reducing the majority class (known as undersampling), or by combining both methods. Ko et al. (2023) adopted a type of the Synthetic Minority Oversampling Technique (SMOTE), called Borderline SMOTE, as an oversampling scheme and edited nearest neighbor as an undersampling method to rebalance the training datasets for RI forecasting. In RI studies, SMOTE-

based techniques for synthesizing new samples are more popular than other methods, and it has been demonstrated that SMOTE can successfully address and rectify the imbalance issue (Kim et al., 2024; Li et al., 2017; Shaiba & Hahsler, 2016; Wei et al., 2023; Wei & Yang, 2021; Yang et al., 2020).

The overarching goal of this study is to develop a global predictive ML model for a subset of RI events impacted by nearby MHWs. Notably, the RI forecast errors are largest in favorable environments like high SSTs (Trabing & Bell, 2020). Given that MHWs create conditions of anomalously high SSTs, portions of TC tracks influenced by MHWs have a greater potential for significant forecast errors. These conditions not only challenge accurate forecasting but also tend to lead to stronger TC intensities (Choi et al., 2024; Radfar et al., 2024). Therefore, it is crucial to further improve models to better predict TC behavior in MHW-affected areas. For this purpose, global RI and MHW events are identified over the 1981–2023 period. According to the framework by (Bevacqua et al., 2021), MHWs and RIs may form temporally and spatially compounding events. This means when they occur close together in time and space, their combined presence can facilitate the rapid intensification process of TCs. To identify compound MHW and RI events, the spatiotemporal association of MHW events to the onset of RI events is analyzed. The global pattern of these events is discussed and then used to identify hotspot regions for the formation of MHW-RI events in Section 3.2. Finally, four tree-based algorithms are trained, validated, and tested to build a robust ensemble machine learning algorithm that performs well in predicting MHW-induced RI events globally. The performance of the model in capturing the costly TC events of 2021–2023 is examined in Section 3.3.

2. Materials and Methods

2.1. Data Sources

We use TC best-track data to identify RI events. The dataset is sourced from the publicly available International Best Track Archive for Climate Stewardship (IBTrACS) dataset (K. R. Knapp et al., 2018). This dataset contains 3-hourly interpolated TC track data from 1841 to the present. The latitude and longitude values in the dataset are rounded to the nearest 0.1°, which corresponds to approximately 10 km in spatial resolution. The data includes the maximum sustained wind speed, radius, pressure, storm translational speed, and direction. The NOAA Optimum Interpolation Sea Surface Temperature version 2.1 (NOAA OISST v2.1) dataset is used in this study for MHW detection. This product provides the daily SST data (Huang et al., 2021) from 1 September 1981 through 19 October 2023 at a spatial resolution of 25 km.

2.2. Data Preprocessing and Event Detection

2.2.1. TC Data

The IBTrACS dataset is a collection of data from multiple agencies. Therefore, the data needs to be preprocessed in order to remove any inconsistencies, including unifying longitudes and merging wind speed formats from different agencies. For a single latitude and longitude pair, the dataset may contain multiple wind speeds from different agencies, while we need a single data for each grid point. To solve this issue, we select a unique wind speed for each location. We use wind speed provided by the USA agencies where data is available and non-zero. This choice arises from the fact that this type of information in IBTrACS has the best coverage (around 85% coverage for all reported TC track locations). If this wind speed is not available, we find the distance from the row's coordinates to other agencies' reported locations and select the wind speed of the closest agency. Subsequently, the processed data are used to detect RI events according to the United States' National Hurricane Center definition whereby RI occurs if the maximum sustained wind speed increases at least 30 knots in 24 hr (Kaplan & DeMaria, 2003). In our analysis, we included only those points in the best track data that were classified as tropical or subtropical cyclones, excluding all other storm types such as extratropical, post-tropical, waves, and disturbances.

2.2.2. MHW Events

We conducted a pointwise analysis to characterize global MHW events at the grid points of the NOAA OISST v2.1 dataset, which features a 0.25° grid resolution. The detection period is from 1 September 1981, to 19 October 2023. For this, a MHW event is defined as a period in which SST is above the local and seasonally evolving 90th percentile threshold of SSTs for at least five consecutive days (Hobday et al., 2016). For each grid point, we

identified the days with SSTs above the 90th percentile. Additionally, two events less than 3 days apart are considered a single MHW event.

A 30 year baseline period, from 1991 to 2020, is used in our analysis. The climatological mean and percentile threshold are calculated for each calendar day from the pool of daily SSTs within an 11 day window centered on the certain day to ensure sufficient sample sizes for estimating these values (Oliver et al., 2018, 2019). The climatological means and thresholds for each day are further smoothed by applying a 31 day moving average centered on each day to the 30 year historical data, to eliminate high-frequency noise (Sen Gupta et al., 2020). This MHW characterization is implemented using an R package developed by (Schlegel & Smit, 2018).

2.3. Methodology

The general framework implemented in this study is presented in Figure 1. Our goal is to create a global ML model that can predict RI based on TC characteristics and prior MHWs that occurred close to the associated TCs. To achieve this, a set of structurally different forecast models is developed to predict if RI occurs for a particular TC over each 24 hr period over the TC lifetime. The models' predictors include characteristics of the TC in question as well as information of ocean state, including any collocated MHWs (the list of all predictors is included in Table S1 in Supporting Information S1). The models' hyperparameters (model settings) are first tuned using a grid search approach. We then combine the separate forecast models into a single 'ensemble' model to enhance the forecast skill. Finally, we use the model to predict RI globally and regionally (i.e., basin-wise).

MHWs are identified based on the Hobday definition (Section 2.2.2). We next identify TCs that pass near to MHW events to create a compound MHW-TC dataset. A compound event occurs if a TC track lies within 200 km of a MHW. This distance spans the inner and outer core radii of a TC (Weatherford & Gray, 1988). In addition, a temporal condition must also be met. In this regard, a MHW must be present 10 days or less before the TC record time. This is the typical period used for analyzing MHW conditions before the RI phase (Dzwonkowski et al., 2020; Rathore et al., 2022).

Using our database of TCs that pass close to MHWs, we next examine all locations along the track every 3 hr (the temporal resolution of the IBTrACS dataset). Each location/time is considered an independent sample. For each location we determine certain MHW metrics (Table 1) based on the proximal MHW events. We also determine if RI occurs over the subsequent 24 hr (this is labeled a MHW-RI event). After this step, we have an updated MHW-TC dataset that has two classes (positive class or RI class vs. negative class or non-RI class). This dataset is used as input for our ML models. Based on TC and MHW metrics for each location and time along the TC track we want to predict whether RI occurs or not. Figure 2 illustrates how we build the compound MHW-TC dataset using the track of Hurricane Ian (2022). The red points represent the MHW locations that meet our spatiotemporal requirements (within 200 km and within 10 days of the TC track), while the yellow stars indicate intensification points, and the blue line shows the path of Hurricane Ian. If MHW information is paired with a TC track location where intensification happened, it is flagged as a positive class; otherwise, it is flagged as a negative class.

In developing ML algorithms, it is common practice to partition the dataset into training, validation, and test subsets to build the model, refine its performance during training, and evaluate its prediction accuracy on unseen data, respectively. In our analysis, the testing period (2021–2023) covers approximately 20% of the total compounding MHW-TC events in the dataset (Figure 3). It should be noted that this timeframe covers the most recent compounding events and provides a clearer representation of the warming patterns observed in recent years. The rest of the dataset is used to train and validate the model. Our analysis shows that the average number of TC Best Track points per year associated with MHWs during 2021–2023 increased to approximately 3036, compared to an average of 2177 per year during the training/validation period (1981–2020). This represents about a 40% increase and highlights the potential impact of rising SSTs and greater prevalence of MHWs in a warmer world. For training and validation, we divide the input data from 1981 to 2020 into five equal time periods (a.k.a. folds). We use four folds for training and one for validation.

Four tree-based ML classification algorithms (namely XGBoost, Random Forest (RF), ExtraTrees (ET) and LightGBM) are selected to develop a predictive RI binary model (see SI for more details on these algorithms). As discussed earlier, RI forecast is an imbalanced problem, where the number of non-RI events along a TC path is significantly larger than RI cases. In our compound MHW-TC dataset, to rebalance the training datasets and improve predictions of RF and ET algorithms for the minority class (RI class), we rely primarily on the original
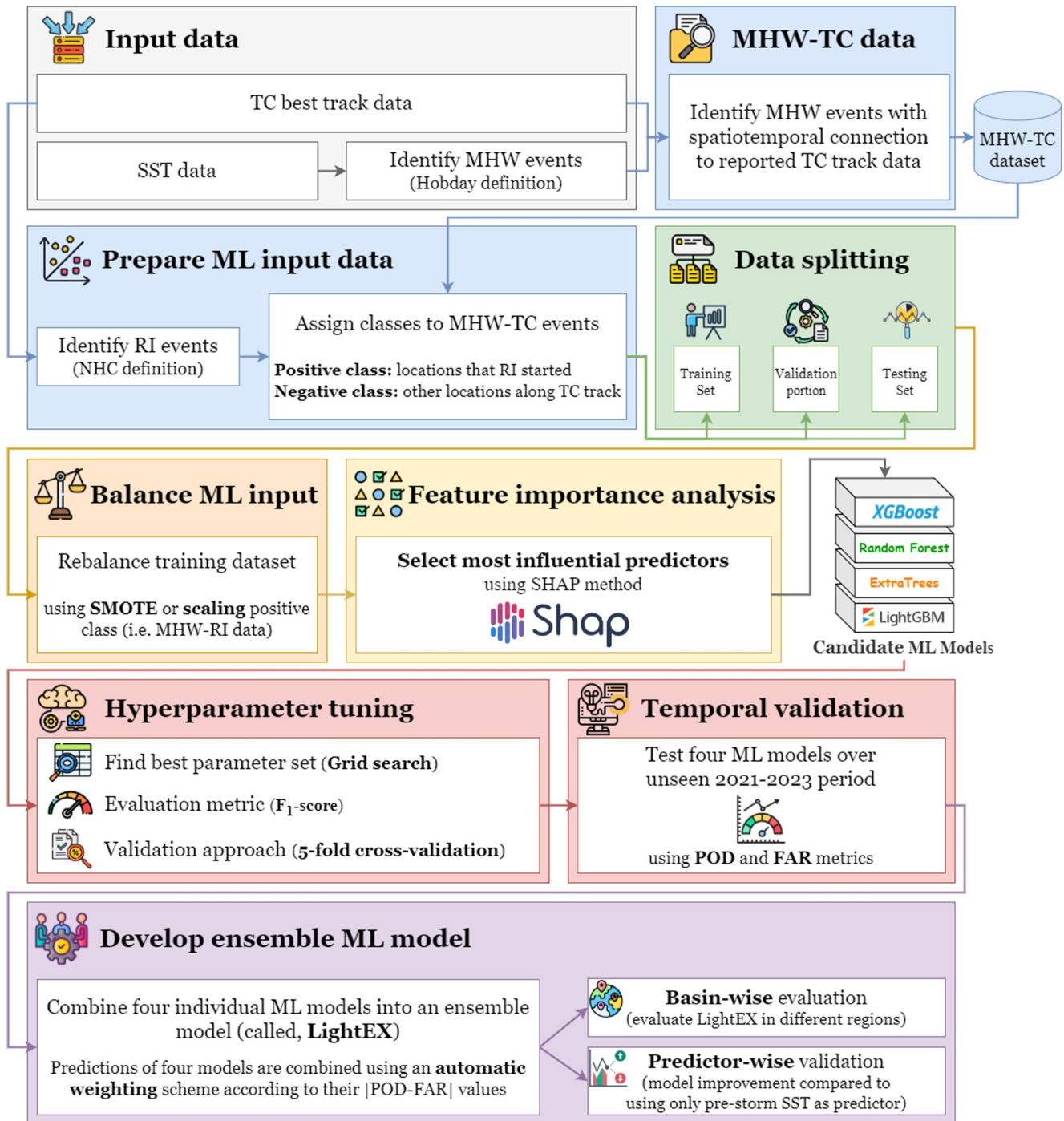
Figure 1. Schematic representation of the proposed methodology.

Synthetic Minority Over-sampling Technique (a.k.a. SMOTE) proposed by (Chawla et al., 2002). In this method instead of simply replicating minority samples, synthetic samples are generated by interpolating along the line connecting several minority class instances within a defined neighborhood (Fernández et al., 2018; L. Wang et al., 2021). In our study, following a trial-and-error approach, we set the neighborhood as the 5 closest instances around the sample of interest. The mathematical representation of SMOTE is as follows (Pradipta et al., 2021):

$$X_{new} = X_i + (X_j - X_i) \times rand(0,1) \tag{1}$$

Table 1
Contributions of Important Individual Features on the Performance of the Predictive ML Model, Ranked From Top to Bottom

| Predictor	Description	Mean |SHAP| value	Data source
start_wind_speed	Reported wind speed at the specific latitude and longitude on the TC track from the IBTrACS dataset	1.05	TC
HI_LAT	Latitude of the TC point	0.62	TC
intensity_mean_abs	Absolute mean MHW intensity during the event	0.47	MHW
dist2land	Distance of TC point to land	0.37	TC
intensity_max_abs	Maximum absolute MHW temperature	0.35	MHW
storm_speed	Translational speed of TC	0.19	TC
intensity_mean	Mean MHW intensity above the climatological mean	0.18	MHW
storm_dir	Direction of TC	0.17	TC
HI_LON	Longitude of the TC point	0.15	TC
lat	Latitude of the MHW event	0.14	MHW
HI_MM	Month in which TC event occurred	0.13	TC
lon	Longitude of the MHW event	0.08	MHW

where X_i is a minority sample, and X_j is one of its 5 nearest neighbors. Their difference is calculated and then multiplied by a random number between 0 and 1, to be added to the vector to the selected sample X_i. This implementation results in a synthetic data along the line segment connecting two samples.

In addition to the SMOTE rebalancing, XGBoost and LightGBM algorithms also have an internal parameter, called "scale_pos_weight", for dealing with imbalanced datasets. The "scale_pos_weight" parameter, determined via hyperparameter tuning, acts as a multiplier for the loss function of the positive class, which is the minority class. The general representation of this parameter in the loss function of a binary classifier is as follows:

$$L(y_i, \hat{y}_i) = -\left[\alpha \cdot y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)\right] \tag{2}$$

where, $L(y_i, \hat{y}_i)$ is the loss for a single instance, y_i is the true label for the instance, $\hat{y}_i$ is the predicted probability for the positive class ($y_i = 1$), and α is scale_pos_weight parameter that acts as a multiplier for the loss associated with the minority class. We assign a value greater than one in order to make the model prioritize the positive class, imposing heavier penalties for inaccurate classification of this class. This parameter aids these algorithms to be less biased toward negative (or majority) class. We also evaluated the computational cost of using this method instead of SMOTE for XGBoost and LightGBM. In our model configuration, using the scale_pos_weight parameter results in about 20% faster runtime compared to the SMOTE option and also a simpler training process since scale_pos_weight is an internal parameter of these models and can be adjusted iteratively.

Next, the set of predictors describing the compound MHW-TC relationship for the ML models are determined by performing feature importance analysis. This approach is used to determine which of the potential predictors (i.e., MHW and TC metrics determined previously) are most useful for making RI predictions. We employ the Shapley additive explanations (SHAP) model explainer (Lundberg & Lee, 2017) to rank the importance of various features and select the most influential ones. This metric measures the contribution of each predictor by comparing the predictive performance of the model in the absence and presence of that predictor. This method is built upon calculating Shapely values as formulated in Equation 3 (Molnar, 2022):
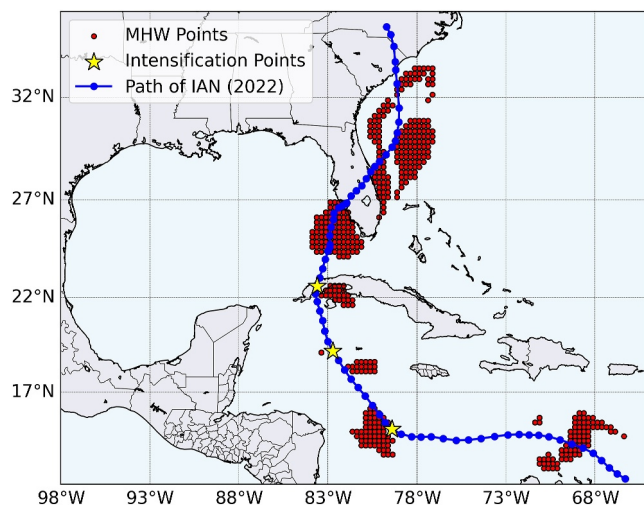
Figure 2. Spatial distribution of MHW points relative to the track of Hurricane Ian (2022). The plot shows the locations of all MHWs that occurred within 200 km and within 10 days of the TC track. Hurricane Ian (2022) experienced three distinct RI phases along its path, marked by yellow stars in the plot.
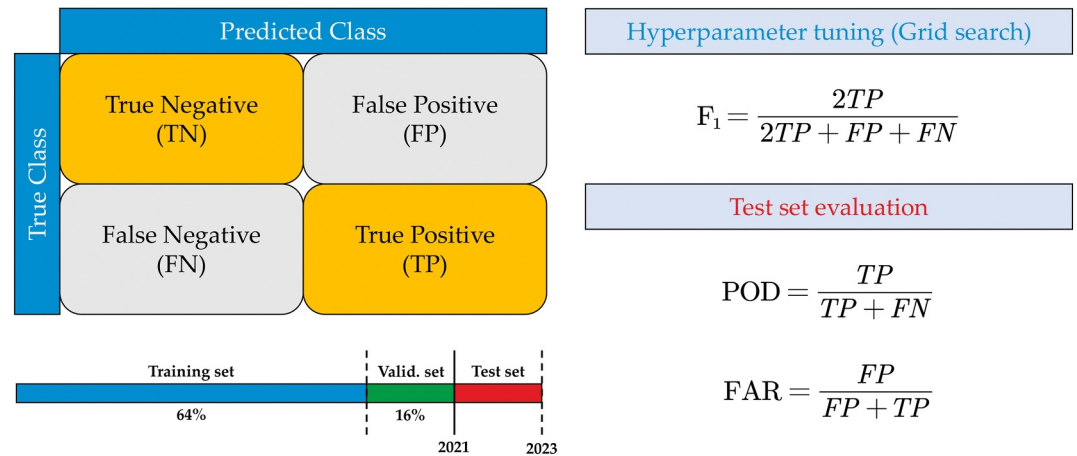
Figure 3. Confusion matrix and performance metrics formula. The F_1-score is used as the evaluation metric of fivefold cross validation during the grid search process for tuning the hyperparameters of implemented ML algorithms. Performance of test-set is measured using POD and FAR. This test-set covers the compounding MHW-RI events from 2021 to 2023.

$$\phi_j = \sum_{S \subseteq N\backslash\{j\}} \frac{|S|!(N - |S| - 1)!}{N!}(v(S \cup \{j\}) - v(S)) \tag{3}$$

where, j is the given feature, S is any subset of features that does not include the jth feature, and $|S|$ is its size, N is the number of all features, $v(S \cup \{j\})$ represents all combinations with adding jth feature, $v(S)$ is all combinations without j's presence, $v()$ function defines the prediction output of the model, and $\sum_{S \subseteq N\backslash}$ denotes summation over all possible combinations without j.

Every ML algorithm has hyperparameters or a set of algorithm parameters that control its learning process. During the tuning process, different values for the hyperparameters are tested to find a set of optimal hyperparameters for a learning algorithm. The list of tuned hyperparameters of the four candidate ML algorithms are presented in Tables S2-S5 in Supporting Information S1. The trained ML models are then evaluated over the 2021–2023 period, which is after the training period and represents an unseen and independent MHW-TC dataset. Various performance metrics are used during the validation and evaluation steps. When dealing with imbalanced datasets, a common approach is to assess how well a model performs in terms of precision (P) and recall (R) (Cavaiola et al., 2024). In particular, evaluation of RI forecasting models often relies on Probability of Detection (POD) and False Alaram Ratio (FAR) (Figure 3) as standard metrics for testing the performance of both operational and proposed models (Narayanan et al., 2023). FAR is related to P or the Success Ratio (SR), with P being defined as 1- FAR. Additionally, POD or Sensitivity is equal to R (DeMaria et al., 2021). suggest that a model is valuable if it has a non-zero POD and a FAR of 50% or less, since it detects some RI cases and is at least half of the time correct. Another important metric in RI prediction models is bias score. Bias is related to POD and SR values and is calculated as follows:

$$\text{bias} = \frac{\text{POD}}{\text{SR}} \tag{4}$$

A bias score of 1.0 indicates that the fraction of RI events predicted by a model is the same as the observed number of RI cases (DeMaria et al., 2021). Therefore, bias scores greater than 1 suggest overprediction, while scores less than 1.0 indicate underprediction.

To achieve a trade-off between both POD and FAR metrics, the models are trained using solely the training sets, and their performance on the validation set is evaluated based on F_1-score during hyperparameter tuning process. This is accomplished by selecting the hyperparameter set with the highest 5-fold mean F_1-score for the four candidate ML models. F_1-score is the harmonic mean of P and R and is calculated as $F_1 = 2PR/(P + R)$. Based on three entries listed in the confusion matrix of Figure 3, these metrics can be calculated: "TP" is True Positive and
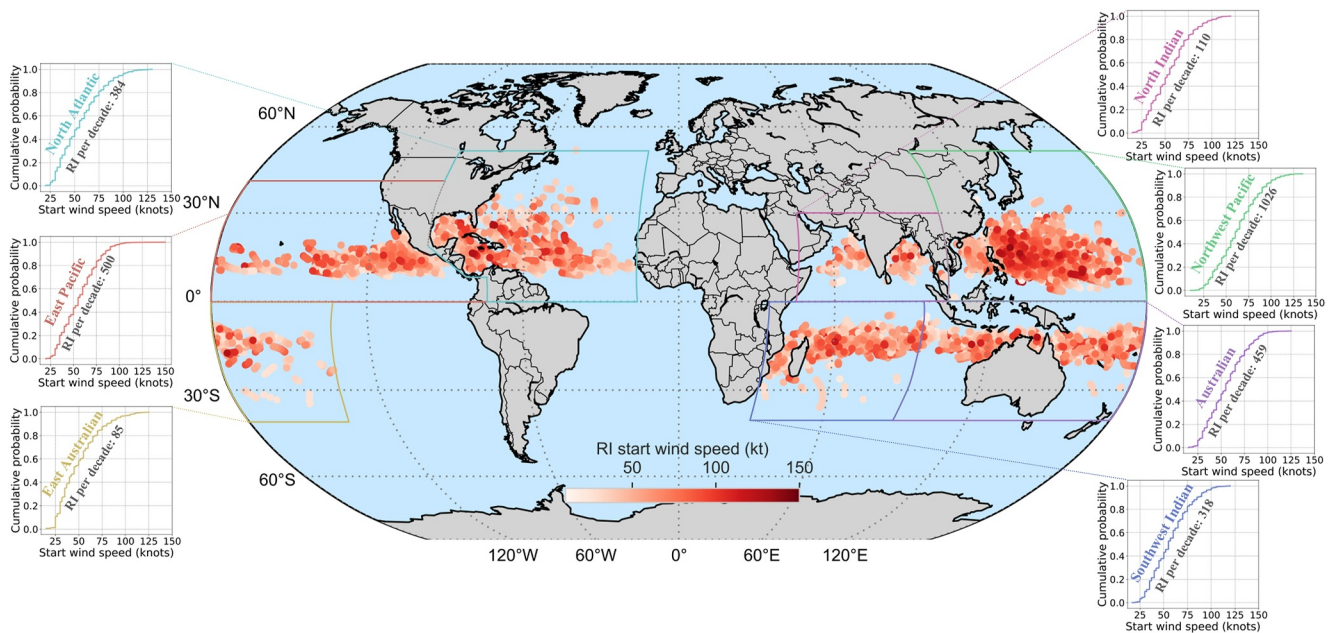
Figure 4. Global distribution of RI starting points across active TC basins from 1981 to 2023. On the map, the initiation zones of RI events are color-coded according to the start wind speed intensity. The number of RI starting points meeting the definition (Kaplan & DeMaria, 2003) in each basin is also presented as part of the starting wind speed cumulative distribution plots. Note that RI counts are based on a 3 hr time gap between two consecutive RI onset points.

refers to an outcome where the model correctly predicts the positive class or the event occurrence class, "FP" denotes False Positive and refers to an outcome where the model incorrectly predicts the positive class, and "FN" represents False Negative, which is an outcome where the model incorrectly predicts the negative class or the event nonoccurrence class.

Ensemble learning is a ML technique that leverages the complementary strengths of multiple algorithms to improve overall performance (more detail is included in the SI). We explore the possibility of integrating a combination of the four ML models into an ensemble ML model to benefit from the predictive capability of the combination of the individual models. In an ensemble model, each individual ML algorithm contributes to the final prediction of the model according to its weight. The weights of this ensemble model were automatically assigned based on their "POD—FAR" values. Our algorithm tries different combinations of weights for each model (sum of the weights is 1), calculates the metric, and finally picks the combination that achieves the best score. We select the best ensemble ML model so that it performs on the testing set with a FAR <50% and there is a larger margin between its POD and FAR (i.e., better POD—FAR). The developed ensemble model will be referred to as "LightEX". Next, we evaluate the LightEX's performance across various basins. Here, we select the portion of test period events belonging to each region and run the developed model for them to evaluate the performance of the global model for different TC basins. Finally, we evaluate the actual improvement achieved by using MHW predictors compared to using pre-storm SST as a predictor. The motivation for this comparison is that SST is the most common predictor in the available RI forecast models, and we want to determine whether using MHW predictors instead of SST is sufficiently informative for the model or not.

3. Results and Discussion

3.1. Global Distribution of RI Points

There are well-known regions around the world that provide favorable conditions for the formation and strengthening of TCs, also known as tropical cyclogenesis (Hsu, 2003). These active TC basins are the North Atlantic, East Pacific, Northwest Pacific, North Indian, Southwest Indian, Australian (including two subregions in the southeastern Indian Ocean and the western Oceania), and East Australian (located at the southwestern Pacific). Figure 4 illustrates the locations for the starting points of historical RI events along with the cumulative distributions of RI starting wind speeds across active TC basins. Results show that RI events generally do not
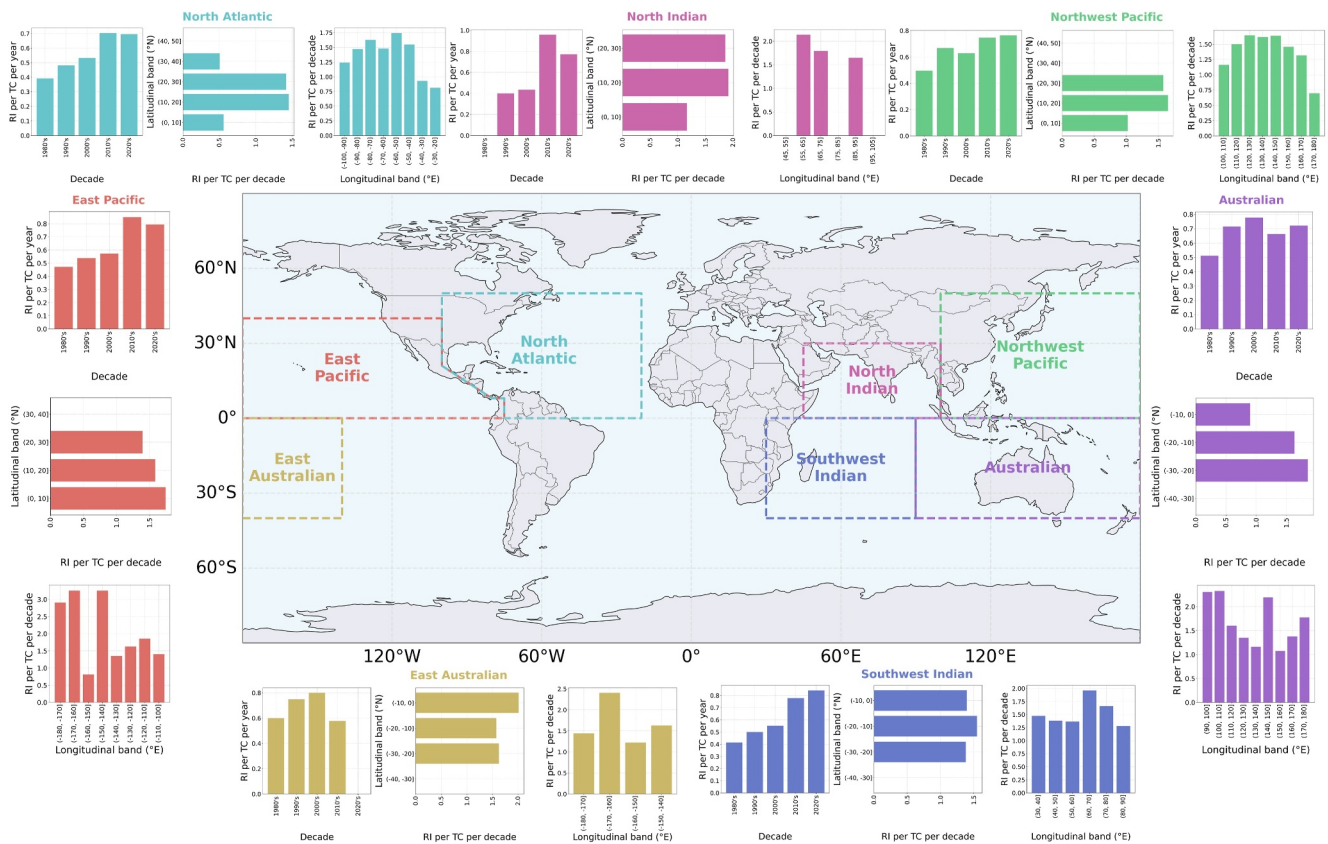
Figure 5. **Global characteristics of MHW-induced RI events from 1981 to 2023 and the seven active TC basins across the globe.** The number of RI events in each basin is represented for each decade as well as for latitudinal and longitudinal bands. The number of RI events in each basin is normalized by the number of TCs. Each latitude and longitude bands cover 10° slice of that specific band.

occur when wind speeds exceed approximately 125 knots in all basins except for the East Pacific where the limit is around 155 knots. The wind speed of 155 knots was observed during Hurricane Patricia, which experienced an RI on 23 October 2015, where the wind speed increased to 185 knots in just 6 hr. Patricia stands out as the most powerful TC ever recorded globally in terms of maximum sustained wind speed (Kimberlain et al., 2016).

Results show that RI events generally start at higher wind speeds over the Northwest Pacific, where the 90th percentile of wind speed at RI initiation is 100 knots, 10–20 knots more than in any other basin. This is also confirmed by a denser population of dark red dots over that basin (Figure 3). The tropical Northwest Pacific is the most active TC region, with more than twice as many events as the East Pacific basin. According to the percentile analysis, TCs start undergoing RI between 30 and 90 knots (10th to 90th percentiles).

3.2. Global Compound MHW and RI Events

A comparative analysis of the global compound MHW-RI characteristics in active TC basins is presented in Figure 5 by depicting the decadal, latitudinal and longitudinal variations in the number of RI events, normalized by the number of TCs, associated with MHWs. These values represent the normalized count of RI events, adjusted by the number of TCs. It is important to note that the frequency of RI events is based on a 3 hr time gap between two consecutive RI onset points because our ML model analyzes all recorded points on TC tracks for RI prediction. Therefore, these numbers represent overlapping RI events. The decadal plots illustrate an increase in the ratio of MHW-induced RI events over most basins. In particular, the number of RI events in the 2020s alone reached or surpassed the 2010s in all basins except the East Australian. For the Australian and the Northwest Pacific basins, we observe a nearly constant frequency since the 1990s at around 0.6–0.7. In the East Pacific, data for the 1990s, 2000s and 2020s exhibit consistently high activity of MHW-induced RI events, with RI frequencies of 06–0.8 per year, which is close to the peak for other basins. In the East Australian basin, peak activity was

recorded in the 2000s, then the MHW-induced RIs fell by half and recently disappeared for the last 4 years. This trend is due to the basin's relatively low activity, with only 51 rapidly intensified TCs in the last four decades and only one TC, WASI (2020), in the past 4 years, which was not influenced by MHWs.

From the spatial distribution of MHW-induced RI events, it is evident that latitude-wise, the highest RI frequencies are distributed across a wide range of latitudes for the East Pacific, the Southwest Indian, and the East Australian. For other basins, we observe a more localized distribution of RI events, with peak frequencies occurring within specific latitude ranges that vary by basin. Closer inspection of the Northern Hemisphere basins reveals that the (10, 30) °N band has the strongest activity. The highest RI frequencies of about 2 are observed in the North Indian and the East Australian basins. On the other hand, the North Atlantic basin has the lowest latitude ratios of less than 1.5, indicating the lower chance of relative RI occurrence among all TCs paths over this region. Considering the longitudes of the events, the highest RI frequencies greater than 2 are observed in the East Pacific, the North Indian, and both Australian basins. Overall, the lowest RI frequencies around 1.5 are in the Northwest Pacific Basin. However, it can be seen that for this basin and the North Atlantic basin, the calculated RI frequencies have a broader distribution across the longitudinal bands, suggesting diverse regional MHW influences. Generally, the spatial distribution results (Figure 4) suggest that the Caribbean Sea (near the Lesser Antilles), westernmost Pacific waters, coastlines of India in the Bay of Bengal, eastern coasts of Philippines, northwest coasts of Australia, and east of the Southwest Indian basin are the regional hotspots for the formation of MHW-induced RI events.

3.3. Performance of the Predictive ML Model

Using the SHAP method, we select the most important features to develop ML algorithms more efficiently. Among the total 31 features considered in this analysis (Table S1 in Supporting Information S1), the SHAP results indicate that the 12 listed in Table 1 are the most influential parameters (i.e., highest mean absolute SHAP) affecting the predictive ability of the ML model. The results suggest that aside from the geographical coordinates of MHW and TC events, the most important storm characteristics are wind speed and distance to land. Additionally, mean absolute intensity is the most important MHW metric.

Using the most influential predictors, the hyperparameter tuning is performed for four individual algorithms using the grid search method on a High-Performance Computing (HPC) system, equipped with 120 cores (3.0 GHz). The runtimes for executing the models are presented in Figure 5. The results show that the LightGBM model runs significantly faster compared to other models (8 times faster than XGBoost, 12 times faster than ET, and 24 times faster than RF) while producing predictions with comparable skill to the other algorithms. This showcases the superiority of LightGBM in terms of computational efficiency for large datasets (in our case, more than 4 million MHW-TC events; see Supplementary Information for further details about sample size). The predictions from these models were then combined to build an ensemble model. For this purpose, the implemented algorithm randomly assigns a weight between 0 and 1 to each candidate model to find the best weights in terms of *POD-FAR* metric. The best ensemble model adopts a weight set of 0.3, 0.0, 0.1, and 0.6 for XGBoost, RF, ET, and LightGBM, respectively. As can be seen, RF does not add value to the ensemble model therefore it is excluded from the ensemble model using the implemented algorithm. The weighted average of the probabilities assigned to each class by all models is then calculated, and the output class is predicted based on the Argmax of these weighted averages. The resultant performance metrics for the individual models and the ensemble model (LightEX) are presented in Figure 6. These metrics support that LightEX exhibits a good balance between detection accuracy and precision (POD = 54% and FAR = 49%), and the results are provided without over- or underprediction (bias = 1.06).

To measure the applicability of the LightEX model over active TC basins, we performed basin-wise (regional) evaluation across seven active TC basins and evaluated their corresponding POD and FAR values. Figure 7 demonstrates how the model performed across various basins. It should be noted that we evaluated the performance of the LightEX model using the unseen data (2021–2023 events) from each individual basin without retraining the model for each basin. In the North Atlantic, the model shows a relatively high POD, indicating a good ability to detect RI events. However, FAR is as high as POD and this represents a common challenge for all available RI prediction models, highlighting the urgent need to improve the predictor set. The model exhibits a close but slightly weaker performance over the Northwest Pacific basins. POD of the model in the East Pacific exceeds that of other basins, while FAR remains comparable. This is in agreement with (Narayanan et al., 2023), which obtained the best performance for the East Pacific and the lowest for the Atlantic and Southern Hemisphere basins. Similarly, our model resulted in POD = 44% and FAR = 55% for the Australian basin, which shows the
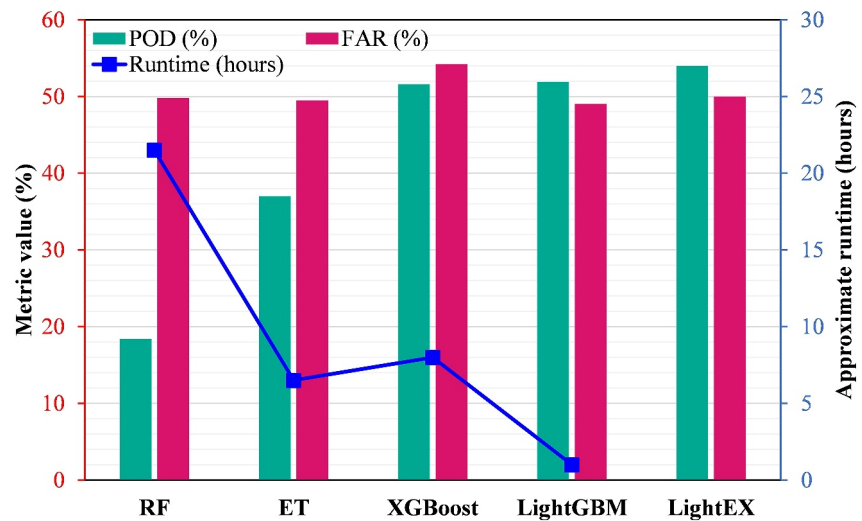
Figure 6. POD and FAR performance metrics of the four candidate ML models.

worst performance among all basins. It is worth noting that some previous studies did not distinguish between Southwest Indian, main and eastern Australian basins and considered them part of the Southern Hemisphere. We believe this is important because the model in this study performs very well in the East Australian and the Southwest Indian basins, which shows the differences between the characteristics of these basins. The high POD value in the East Pacific is partly related to the fact that the probability of RI occurrence (minority class of the ML model) is higher in this basin than in other basins (Bhatia et al., 2022). This fact reduces the model's difficulty in dealing with imbalanced classes and provides a broader training set. In addition, since our model relies on SST characteristics, the differences in the SST impacts between basins would greatly affect the performance of the model across basins. Foltz et al. (2018) compared SST variabilities with vertical wind shear and outflow temperature (temperature at 200 hPa), and their contributions to variance in TC intensification rates. This approach led them to conclude that SST is a better predictor of intensification in the East Pacific than in the North Atlantic and Northwest Pacific basins. Lastly, the model demonstrates similar POD in the North Indian basin compared to the North Atlantic and Northwest Pacific basins, but with roughly 20% lower FAR. This suggests that the RI class in this basin has a lower complexity and therefore has a lower number of false positives. It is crucial for resource
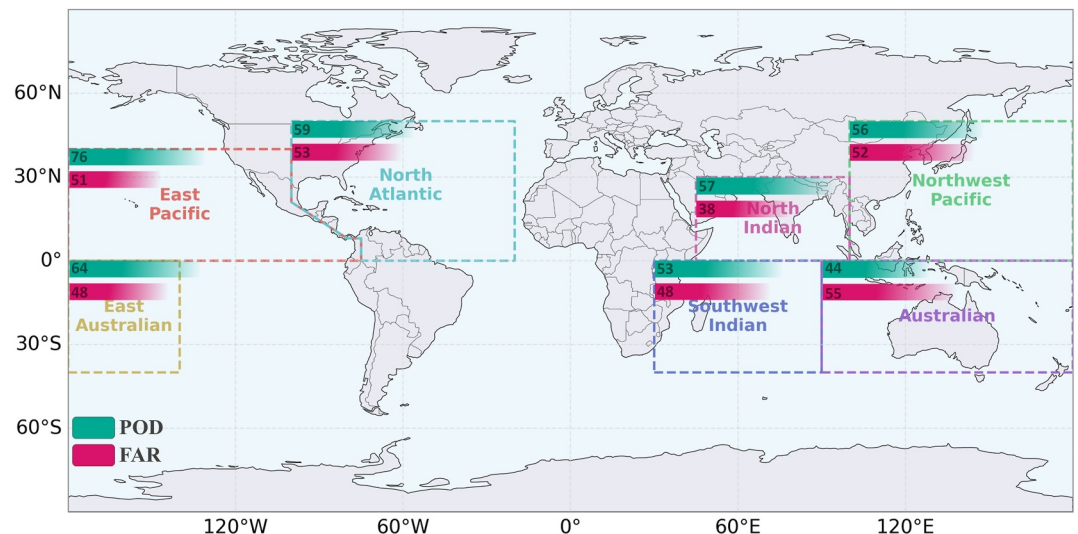
Figure 7. Performance of the predictive ML model for RI forecast based on close MHW events. POD and FAR values are displayed for the seven active basins for TC formation.
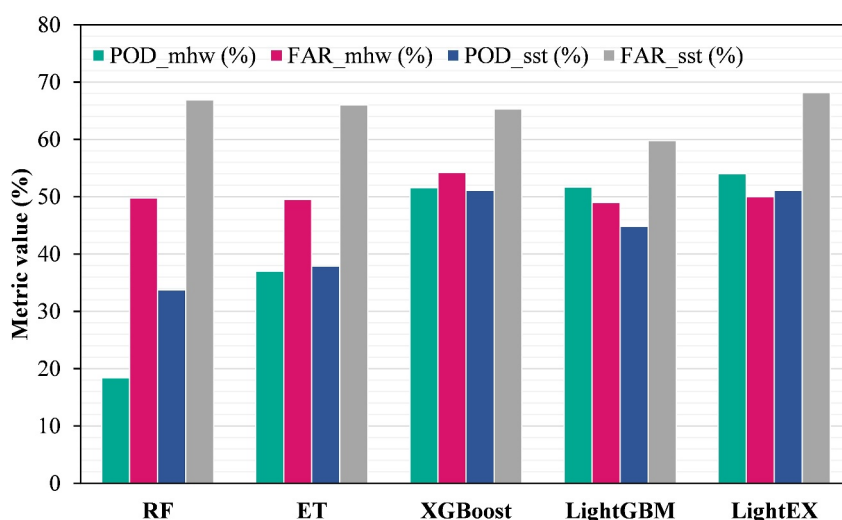
Figure 8. Impact of using MHW and pre-storm SST as predictors on POD and FAR metrics of the RI forecast models.

allocation and risk management in TC-prone regions to make better risk-informed decisions. In this regard, lower FAR values lead to fewer unnecessary preparations or alerts for RI that do not actually occur.

To get a clearer idea of value added by including MHW features in RI prediction models, it is necessary to compare our models with ML models developed with SST as a predictor. Therefore, we repeated the training and testing processes for all previously mentioned models, this time using SST at the location/time of the TC plus the seven TC predictors from Table 1 (eight predictors in total). The obtained best parameter sets and weights for the new ensemble model are listed in Table S7 in Supporting Information S1. Figure 8 compares the POD and FAR metrics for different machine learning models based on MHW characteristics versus only using pre-storm SST (see Supplementary Information for further details about the alternative model). Although we observe a slight improvement in LightEX's POD (~3%), its FAR value deteriorates by 36% (50% compared to 68%). The ML models exhibit a significant increase in false alarms, with FAR values exceeding 60%. For the ET and XGBoost models, the POD values remain at the same level, while we observe a better POD for RF and a lower POD for LightGBM when using SST as a predictor. The relatively high FAR values are unfavorable for RI forecast models and bolster the advantage of MHW predictors compared to widely used SST for controlling false alarms. A practical approach for incorporating MHWs into RI models is to aggregate MHW characteristics within a 200 km radius of the TC center into a single set of predictors. This aggregation ensures operational feasibility while maintaining the essential physical relationships between MHWs and TC intensification. Specifically, area-accumulated intensity, defined as the sum of intensities weighted by the respective areas of all MHWs within the radius, can serve as a key predictor. Additionally, the total area of MHWs within the 200 km radius provides a metric for the spatial extent of anomalously warm waters influencing the TC. These aggregated metrics allow for a single forecast at each synoptic time, aligning with existing operational RI model frameworks. Follow-up studies can examine the operational value of this new feature set by integrating it into existing models.

4. Conclusions

This study aims to develop a global predictive ML model for the occurrence of RI events that are influenced by proximal MHWs. Notably, a considerable number of recent catastrophic TCs, such as Otis (2023) in the East Pacific basin, Doksuri (2023) in the Northwestern Pacific basin, and Ian (2022) in the North Atlantic basin, belong to the category of MHW-induced RI events. This motivated us to develop an alternative RI forecast model that considers the underlying characteristics of this category. In addition to storm characteristics such as wind speed and distance to land, mean and maximum absolute SSTs during MHWs emerged as reliable predictors of compounding MHW-TC events. Building upon four decision tree-based models, we developed an ensemble ML model, called LightEX. A practical implication of LightEX is the possibility of developing a global RI prediction model applicable to various active TC basins with an acceptable performance during the recent period characterized by a warming climate. The ensemble model showed promising performance over the most active TC

basins worldwide, that is, the Pacific and the Atlantic basins. Additionally, LightEX was trained/validated over a long period of time and tested on the 2021–2023 period. Supplementary Figure 1 in Supporting Information S1 illustrates that for the majority of RI events (around 70%), it takes approximately 24 hr to achieve a 30-knot increase in maximum sustained wind speeds. Therefore, based on the availability of TC and MHW data, the developed model can aid 24 hr forecasts of RI. This is the minimum threshold for classifying the event as RI. It is important to note that if there is a lag in obtaining TC data or if rare RI events with relatively short time spans occur, the performance of the model may be affected.

We have demonstrated that including MHW characteristics as predictors not only helps achieve favorable POD levels, but also maintains a FAR below the 50% threshold. On the other hand, using SST as a predictor instead of MHW metrics, which is common in the available models, leads to undesirably higher FAR values (more than 30%). This opens up opportunities for further efforts to incorporate specific MHW metrics in the existing models.

To the best of our knowledge, this is the first study that used LightGBM to classify TC intensity changes. We have shown that LightGBM outperforms other decision tree algorithms with significantly lower computational costs (about one-eighth of XGBoost and one-twentieth of RF). The potential of this ML algorithm can be further explored by implementing it on other dynamic or statistical TC datasets. As the dataset grows larger and/or as more predictors are included, the computational demand required for training can escalate rapidly for other algorithms, whereas LightGBM's optimized gradient boosting framework ensures relatively minimal increases in training time.

TCs experience MHW and RI along their path tend to become stronger and potentially more destructive a few days (Choi et al., 2024; Radfar et al., 2024). During our testing period, there were 51 TCs with a total damage of $360 B (see Supplementary Figure 2 in Supporting Information S1), of which the RI phase of 28 TCs (55%) was captured by our model. The inability to predict more TC RI events is rooted in the inherent limitations and complexities of ML classification models. The reasons can be categorized into two main areas: (a) input quality, which includes data availability, data quality, and feature representations (see (Narayanan et al., 2023; Rozoff et al., 2015; Trabing & Bell, 2020) among others), and (b) model inadequacies, which encompass model generalization, algorithmic limitations, and the bias-variance tradeoff (see (Chen et al., 2020; Guan & Burton, 2022; Kim et al., 2024; Trabing & Bell, 2020) among others). Improving data quality, incorporating more environmental features, and refining algorithms could enhance the model's predictive capabilities.

Given the objectives of this study, we included MHW characteristics as the predictors of RI. However, other important environmental predictors, such as vertical wind shear, ocean heat content, convective potential energy, divergence at 200 hPa from the storm center, and vorticity, may also be important (Choi et al., 2024; Radfar et al., 2024). have discussed the impact of MHWs on surface latent heat flux and precipitation rates, suggesting that these factors also influence the intensity of TCs. Therefore, expanding the feature set to include these and other relevant environmental parameters will allow for a more robust assessment of the proposed approach through a comprehensive representation of the physical underpinnings of RI events. It is also important to address uncertainties that impact forecast capabilities and adapt approaches to deal with error propagation due to intensity forecast errors and insufficient observations (Emanuel & Zhang, 2016; Lok et al., 2021). Incorporating forecast data from reliable numerical weather prediction models enables real-time applicability of the model. However, this introduces complexities such as choosing the appropriate model, dealing with track forecast uncertainties, and assessing the sensitivity of results to these factors. These challenges are beyond the scope of our current study. The IBTrACS data also has uncertainties in its reported position and intensity (K. Knapp, 2019). The position uncertainty is estimated to be approximately 10–15 km for strong TCs ($W_s > 100$ kt). The range of intensity uncertainties across ocean basins is ±7–20% (lower values since 2000). Additionally, different agencies use varying averaging times for maximum wind estimates (usually 1 or 10 min), and wind speeds reported by different agencies are not directly convertible due to procedural and observational differences (Harper et al., 2008; K. R. Knapp et al., 2010).

In terms of MHW detection criteria, we adhered to the commonly used definition of MHWs to be consistent with the common practice when conveying key takeaways. However, there are suggestions to use a shifting temperature baseline rather than a fixed one to avoid detecting more frequent and more intense MHWs (Amaya et al., 2023). Also (X. Zhang et al., 2022), found that MHWs become shorter and less intense in the absence of smoothing techniques. In light of the circumstances imposed by the unprecedented global warming conditions, future studies should be aware of these alternative definitions and examine how they may influence compounding MHW-TC events.

Data Availability Statement

All data that support the findings of this study are publicly available to download and use. The time span covered by our analysis extends from 1 September 1981, to 19 October 2023. TC best track data were obtained from the International Best Track Archive for Climate Stewardship (IBTrACS) dataset (K. R. Knapp et al., 2018), which is freely accessible in CSV format via the National Centers for Environmental Information (NCEI) website at https://www.ncei.noaa.gov/data/international-best-track-archive-for-climate-stewardship-ibtracs-v04r00/access/csv/. SST data were obtained from the NOAA Optimum Interpolation Sea Surface Temperature version 2.1 (NOAA OISST v2.1) dataset (Huang et al., 2021), available in NetCDF format and can be accessed through the ERDDAP data server at https://www.ncei.noaa.gov/erddap/info/index.html?page=1&itemsPerPage=1000.

All codes developed for conducting analyses, generating results, and creating plots are published on Zenodo (Radfar et al., 2024). Additionally, they are also available on our GitHub pages: https://github.com/sradfar/MHW-RI_GlobalML, https://github.com/CHL-UA, and https://github.com/hmoradkhani.

References

Amaya, D. J., Jacox, M. G., Fewings, M. R., Saba, V. S., Stuecker, M. F., Rykaczewski, R. R., et al. (2023). Marine heatwaves need clear definitions so coastal communities can adapt. *Nature*, *616*(7955), 29–32. https://doi.org/10.1038/d41586-023-00924-2

Balaguru, K., Foltz, G. R., Leung, L. R., Kaplan, J., Xu, W., Reul, N., & Chapron, B. (2020). Pronounced impact of salinity on rapidly intensifying tropical cyclones. *Bulletin of the American Meteorological Society*, *101*(9), E1497–E1511. https://doi.org/10.1175/bams-d-19-0303.1

Bevacqua, E., De Michele, C., Manning, C., Couasnon, A., Ribeiro, A. F. S., Ramos, A. M., et al. (2021). Guidelines for studying diverse types of compound weather and climate events. *Earth's Future*, *9*(11), e2021EF002340. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021EF002340

Bhatia, K., Baker, A., Yang, W., Vecchi, G., Knutson, T., Murakami, H., et al. (2022). A potential explanation for the global increase in tropical cyclone rapid intensification. *Nature Communications*, *13*(1), 6626. https://doi.org/10.1038/s41467-022-34321-6

Cavaiola, M., Cassola, F., Sacchetti, D., Ferrari, F., & Mazzino, A. (2024). Hybrid AI-enhanced lightning flash prediction in the medium-range forecast horizon. *Nature Communications*, *15*(1), 1188. https://doi.org/10.1038/s41467-024-44697-2

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Chen, R., Zhang, W., & Wang, X. (2020). Machine learning in tropical cyclone forecast modeling: A review. *Atmosphere*, *11*(7), 676. https://doi.org/10.3390/atmos11070676

Choi, H.-Y., Park, M.-S., Kim, H.-S., & Lee, S. (2024). Marine heatwave events strengthen the intensity of tropical cyclones. *Communications Earth and Environment*, *5*(1), 69. https://doi.org/10.1038/s43247-024-01239-4

Delforge, D., Wathelet, V., Below, R., Sofia, C. L., Tonnelier, M., van Loenhout, J., & Speybroeck, N. (2023). EM-DAT: The emergency events database

DeMaria, M., Franklin, J. L., Onderlinde, M. J., & Kaplan, J. (2021). Operational forecasting of tropical cyclone rapid intensification at the National Hurricane Center. *Atmosphere*, *12*(6), 683. https://doi.org/10.3390/atmos12060683

Dzwonkowski, B., Coogan, J., Fournier, S., Lockridge, G., Park, K., & Lee, T. (2020). Compounding impact of severe weather events fuels marine heatwave in the coastal ocean. *Nature Communications*, *11*(1), 4623. https://doi.org/10.1038/s41467-020-18339-2

Emanuel, K., DesAutels, C., Holloway, C., & Korty, R. (2004). Environmental control of tropical cyclone intensity. *Journal of the Atmospheric Sciences*, *61*(7), 843–858. https://doi.org/10.1175/1520-0469(2004)061<0843:ecotci>2.0.co;2

Emanuel, K., & Zhang, F. (2016). On the predictability and error sources of tropical cyclone intensity forecasts. *Journal of the Atmospheric Sciences*, *73*(9), 3739–3747. https://doi.org/10.1175/jas-d-16-0100.1

Fernández, A., Garcia, S., Herrera, F., & Chawla, N. V. (2018). SMOTE for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary. *Journal of Artificial Intelligence Research*, *61*, 863–905. https://doi.org/10.1613/jair.1.11192

Fischer, M. S., Tang, B. H., Corbosiero, K. L., & Rozoff, C. M. (2018). Normalized convective characteristics of tropical cyclone rapid intensification events in the North Atlantic and eastern North Pacific. *Monthly Weather Review*, *146*(4), 1133–1155. https://doi.org/10.1175/mwr-d-17-0239.1

Foltz, G. R., Balaguru, K., & Hagos, S. (2018). Interbasin differences in the relationship between SST and tropical cyclone intensification. *Monthly Weather Review*, *146*(3), 853–870. https://doi.org/10.1175/mwr-d-17-0155.1

Gall, R., Franklin, J., Marks, F., Rappaport, E. N., & Toepfer, F. (2013). The hurricane forecast improvement project. *Bulletin of the American Meteorological Society*, *94*(3), 329–343. https://doi.org/10.1175/bams-d-12-00071.1

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

Guan, X., & Burton, H. (2022). Bias-variance tradeoff in machine learning: Theoretical formulation and implications to structural engineering applications. *Structures*, *46*, 17–30. Paper presented at the Structures. https://doi.org/10.1016/j.istruc.2022.10.004

Harper, B., Kepert, J., & Ginger, J. (2008). Guidelines for converting between various wind averaging periods in tropical cyclone conditions. *World Meteorological Organization*.

Hartman, C. M., Chen, X., & Chan, M.-Y. (2023). Improving tropical cyclogenesis forecasts of Hurricane Irma (2017) through the assimilation of all-sky infrared brightness temperatures. *Monthly Weather Review*, *151*(4), 837–853. https://doi.org/10.1175/mwr-d-22-0196.1

Hobday, A. J., Alexander, L. V., Perkins, S. E., Smale, D. A., Straub, S. C., Oliver, E. C., et al. (2016). A hierarchical approach to defining marine heatwaves. *Progress in Oceanography*, *141*, 227–238. https://doi.org/10.1016/j.pocean.2015.12.014

Hsu, S. A. (2003). Coastal meteorology. In R. A. Meyers (Ed.), *Encyclopedia of physical science and technology* (3rd ed., pp. 155–173). Academic Press.

Huang, B., Liu, C., Banzon, V., Freeman, E., Graham, G., Hankins, B., et al. (2021). Improvements of the daily optimum interpolation sea surface temperature (DOISST) version 2.1 [Dataset]. *Journal of Climate*, *34*(8), 2923–2939. https://doi.org/10.1175/jcli-d-20-0166.1

Kaplan, J., & DeMaria, M. (2003). Large-scale characteristics of rapidly intensifying tropical cyclones in the North Atlantic basin. *Weather and Forecasting*, *18*(6), 1093–1108. https://doi.org/10.1175/1520-0434(2003)018<1093:lcorit>2.0.co;2

Kaplan, J., DeMaria, M., & Knaff, J. A. (2010). A revised tropical cyclone rapid intensification index for the Atlantic and eastern North Pacific basins. *Weather and Forecasting*, *25*(1), 220–241. https://doi.org/10.1175/2009waf2222280.1

Kaplan, J., Rozoff, C. M., DeMaria, M., Sampson, C. R., Kossin, J. P., Velden, C. S., et al. (2015). Evaluating environmental impacts on tropical cyclone rapid intensification predictability utilizing statistical models. *Weather and Forecasting*, *30*(5), 1374–1396. https://doi.org/10.1175/waf-d-15-0032.1

Kieu, C. Q., & Zhang, D. L. (2009). An analytical model for the rapid intensification of tropical cyclones. *Quarterly Journal of the Royal Meteorological Society: A Journal Of The Atmospheric Sciences, Applied Meteorology And Physical Oceanography*, *135*(642), 1336–1349. https://doi.org/10.1002/qj.433

Kim, S.-H., Lee, W., Kang, H.-W., & Kang, S. K. (2024). Predicting rapid intensification of tropical cyclones in the western North Pacific: A machine learning and net energy gain rate approach. *Frontiers in Marine Science*, *10*. Original Research. https://doi.org/10.3389/fmars.2023.1296274

Kimberlain, T. B. B., Eric, S., & Cangialosi, J. P. (2016). *Hurricane Patricia (EP202015). (Tropical Cyclone Report)*. National Hurricane Center.

Knaff, J. A., Sampson, C. R., Brammer, A., & Slocum, C. J. (2023). A rapid intensification deterministic ensemble (RIDE) for the joint typhoon warning center's area of responsibility. *Weather and Forecasting*, *38*(7), 1229–1238. https://doi.org/10.1175/waf-d-23-0012.1

Knapp, K. (2019). Technical documentation of international best track archive for climate stewardship (IBTrACS). [Dataset], Retrieved from https://www.ncei.noaa.gov/sites/default/files/2021-07/IBTrACS_version4_Technical_Details.pdf

Knapp, K. R., Diamond, H. J., Kossin, J. P., Kruk, M. C., & Schreck, C. J. (2018). International best track archive for climate stewardship (IBTrACS) project version 4. *NOAA National Centers for Environmental Information*, *10*

Knapp, K. R., Kruk, M. C., Levinson, D. H., Diamond, H. J., & Neumann, C. J. (2010). The international best track archive for climate stewardship (IBTrACS) unifying tropical cyclone data. *Bulletin of the American Meteorological Society*, *91*(3), 363–376. https://doi.org/10.1175/2009bams2755.1

Ko, M.-C., Chen, X., Kubat, M., & Gopalakrishnan, S. (2023). The development of a consensus machine learning model for hurricane rapid intensification forecasts with hurricane weather research and forecasting (HWRF) data. *Weather and Forecasting*, *38*(8), 1253–1270. https://doi.org/10.1175/waf-d-22-0217.1

Kowch, R., & Emanuel, K. (2015). Are special processes at work in the rapid intensification of tropical cyclones? *Monthly Weather Review*, *143*(3), 878–882. https://doi.org/10.1175/mwr-d-14-00360.1

Le Hénaff, M., Domingues, R., Halliwell, G., Zhang, J. A., Kim, H.-S., Aristizabal, M., et al. (2021). The role of the gulf of Mexico ocean conditions in the intensification of hurricane Michael (2018). *Journal of Geophysical Research: Oceans*, *126*(5), e2020JC016969. https://doi.org/10.1029/2020JC016969

Li, Y., Yang, R., Yang, C., Yu, M., Hu, F., & Jiang, Y. (2017). Leveraging LSTM for rapid intensifications prediction of tropical cyclones. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *4*, 101–105. https://doi.org/10.5194/isprs-annals-iv-4-w2-101-2017

Lok, C. C., Chan, J. C., & Toumi, R. (2021). Tropical cyclones near landfall can induce their own intensification through feedbacks on radiative forcing. *Communications Earth and Environment*, *2*(1), 184. https://doi.org/10.1038/s43247-021-00259-8

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*.

Mawren, D., Hermes, J., & Reason, C. (2022). Marine heat waves and tropical cyclones-Two devastating types of coastal hazard in South-eastern Africa. *Estuarine, Coastal and Shelf Science*, *277*, 108056. https://doi.org/10.1016/j.ecss.2022.108056

Mercer, A., & Grimes, A. (2015). Diagnosing tropical cyclone rapid intensification using kernel methods and reanalysis datasets. *Procedia Computer Science*, *61*, 422–427. https://doi.org/10.1016/j.procs.2015.09.179

Mercer, A., & Grimes, A. (2017). Atlantic tropical cyclone rapid intensification probabilistic forecasts from an ensemble of machine learning methods. *Procedia Computer Science*, *114*, 333–340. https://doi.org/10.1016/j.procs.2017.09.036

Mercer, A. E., Grimes, A. D., & Wood, K. M. (2021). Application of unsupervised learning techniques to identify atlantic tropical cyclone rapid intensification environments. *Journal of Applied Meteorology and Climatology*, *60*(1), 119–138. https://doi.org/10.1175/jamc-d-20-0105.1

Molnar, C. (2022). Interpretable machine learning: A guide for making black box models explainable (2nd ed.). Retrieved from https://christophm.github.io/interpretable-ml-book

Munsell, E. B., Zhang, F., Sippel, J. A., Braun, S. A., & Weng, Y. (2017). Dynamics and predictability of the intensification of hurricane edouard (2014). *Journal of the Atmospheric Sciences*, *74*(2), 573–595. https://doi.org/10.1175/jas-d-16-0018.1

Narayanan, A., Balaguru, K., Xu, W., & Leung, L. R. (2023). A new method for predicting hurricane rapid intensification based on co-occurring environmental parameters. *Natural Hazards*, *120*, 1–19. https://doi.org/10.1007/s11069-023-06100-z

Neetu, S., Lengaigne, M., Vialard, J., Mangeas, M., Menkès, C. E., Suresh, I., et al. (2020). Quantifying the benefits of nonlinear methods for global statistical hindcasts of tropical cyclones intensity. *Weather and Forecasting*, *35*(3), 807–820. https://doi.org/10.1175/waf-d-19-0163.1

Oliver, E. C., Burrows, M. T., Donat, M. G., Sen Gupta, A., Alexander, L. V., Perkins-Kirkpatrick, S. E., et al. (2019). Projected marine heatwaves in the 21st century and the potential for ecological impact. *Frontiers in Marine Science*, *6*, 734. https://doi.org/10.3389/fmars.2019.00734

Oliver, E. C., Donat, M. G., Burrows, M. T., Moore, P. J., Smale, D. A., Alexander, L. V., et al. (2018). Longer and more frequent marine heatwaves over the past century. *Nature Communications*, *9*(1), 1–12. https://doi.org/10.1038/s41467-018-03732-9

Pradipta, G. A., Wardoyo, R., Musdholifah, A., Sanjaya, I. N. H., & Ismail, M. (2021). SMOTE for handling imbalanced data problem: A review. In *Paper presented at the 2021 sixth international conference on informatics and computing (ICIC)*.

Pun, I.-F., Hsu, H.-H., Moon, I.-J., Lin, I. I., & Jeong, J.-Y. (2023). Marine heatwave as a supercharger for the strongest typhoon in the East China Sea. *NPJ Climate and Atmospheric Science*, *6*(1), 128. https://doi.org/10.1038/s41612-023-00449-5

Radfar, S., Foroumandi, E., Moftakhari, H., Moradkhani, H., Foltz, G., & Sen Gupta, A. (2024). Global predictability of marine heatwave induced rapid intensification of tropical cyclones (Predictive model codes and scripts, figure codes, and documentation) (version v1.0.0). *Zenodo*. [Software]. https://doi.org/10.5281/zenodo.13992639

Radfar, S., Moftakhari, H., & Moradkhani, H. (2024). Rapid intensification of tropical cyclones in the Gulf of Mexico is more likely during marine heatwaves. *Communications Earth and Environment*, *5*(421), 421. https://doi.org/10.1038/s43247-024-01578-2

Rathore, S., Goyal, R., Jangir, B., Ummenhofer, C. C., Feng, M., & Mishra, M. (2022). Interactions between a marine heatwave and tropical cyclone amphan in the Bay of bengal in 2020. *Frontiers in Climate*, *4*, 861477. https://doi.org/10.3389/fclim.2022.861477

Rozoff, C. M., & Kossin, J. P. (2011). New probabilistic forecast models for the prediction of tropical cyclone rapid intensification. *Weather and Forecasting*, *26*(5), 677–689. https://doi.org/10.1175/waf-d-10-05059.1

Rozoff, C. M., Velden, C. S., Kaplan, J., Kossin, J. P., & Wimmers, A. J. (2015). Improvements in the probabilistic prediction of tropical cyclone rapid intensification with passive microwave observations. *Weather and Forecasting*, *30*(4), 1016–1038. https://doi.org/10.1175/waf-d-14-00109.1

Schlegel, R. W., & Smit, A. J. (2018). heatwaveR: A central algorithm for the detection of heatwaves and cold-spells. *Journal of Open Source Software*, *3*(27), 821. [Software]. https://doi.org/10.21105/joss.00821

Sen Gupta, A., Thomsen, M., Benthuysen, J. A., Hobday, A. J., Oliver, E., Alexander, L. V., et al. (2020). Drivers and impacts of the most extreme marine heatwave events. *Scientific Reports*, *10*(1), 19359. https://doi.org/10.1038/s41598-020-75445-3

Shaiba, H., & Hahsler, M. (2016). Applying machine learning methods for predicting tropical cyclone rapid intensification events. *Research Journal of Applied Sciences, Engineering and Technology*, *13*(8), 638–651. https://doi.org/10.19026/rjaset.13.3050

Su, H., Wu, L., Jiang, J. H., Pai, R., Liu, A., Zhai, A. J., et al. (2020). Applying satellite observations of tropical cyclone internal structures to rapid intensification forecast with machine learning. *Geophysical Research Letters*, *47*(17), e2020GL089102. https://doi.org/10.1029/2020gl089102

Tallapragada, V., Bernardet, L., Biswas, M. K., Gopalakrishnan, S., Kwon, Y., Liu, Q., et al. (2014). Hurricane weather research and forecasting (HWRF) model: 2013 scientific documentation. *HWRF Development Testbed Center Tech. Rep*, *99*.

Tam, H.-f., Choy, C.-w., & Wong, W.-k. (2021). Development of objective forecast guidance on tropical cyclone rapid intensity change. *Meteorological Applications*, *28*(2), e1981. https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/met.1981

Tao, D., Van Leeuwen, P. J., Bell, M., & Ying, Y. (2022). Dynamics and predictability of tropical cyclone rapid intensification in ensemble simulations of Hurricane Patricia (2015). *Journal of Geophysical Research: Atmospheres*, *127*(8), e2021JD036079. https://doi.org/10.1029/2021jd036079

Trabing, B. C., & Bell, M. M. (2020). Understanding error distributions of hurricane intensity forecasts during rapid intensity changes. *Weather and Forecasting*, *35*(6), 2219–2234. https://doi.org/10.1175/waf-d-19-0253.1

Wang, L., Han, M., Li, X., Zhang, N., & Cheng, H. (2021). Review of classification methods on unbalanced data sets. *IEEE Access*, *9*, 64606–64628. https://doi.org/10.1109/access.2021.3074243

Wang, W., Zhang, Z., Cangialosi, J. P., Brennan, M., Cowan, L., Clegg, P., et al. (2023). A review of recent advances (2018-2021) on tropical cyclone intensity change from operational perspectives, Part 2: Forecasts by Operational Centers. *Tropical Cyclone Research and Review*, *12*(1), 50–63. https://doi.org/10.1016/j.tcrr.2023.05.003

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

Wei, Y., & Yang, R. (2021). An advanced artificial intelligence system for investigating tropical cyclone rapid intensification with the SHIPS database. *Atmosphere*, *12*(4), 484. https://doi.org/10.3390/atmos12040484

Wei, Y., Yang, R., & Sun, D. (2023). Investigating tropical cyclone rapid intensification with an advanced artificial intelligence system and gridded reanalysis data. *Atmosphere*, *14*(2), 195. https://doi.org/10.3390/atmos14020195

Yang, Q., Lee, C.-Y., & Tippett, M. K. (2020). A long short-term memory model for global rapid intensification prediction. *Weather and Forecasting*, *35*(4), 1203–1220. https://doi.org/10.1175/waf-d-19-0199.1

Zhang, C.-H., & Zhang, C.-J. (2023). Rapid intensification forecast for tropical cyclone by combining Multi-ConvLSTM model with infrared and microwave satellite images. In *Paper presented at the conference on infrared, millimeter, terahertz waves and applications (IMT2022)*.

Zhang, J. A., Cione, J. J., Kalina, E. A., Uhlhorn, E. W., Hock, T., & Smith, J. A. (2017). Observations of infrared Sea Surface temperature and air–sea interaction in hurricane edouard (2014) using GPS dropsondes. *Journal of Atmospheric and Oceanic Technology*, *34*(6), 1333–1349. https://doi.org/10.1175/jtech-d-16-0211.1

Zhang, L., & Oey, L. (2019). Young ocean waves favor the rapid intensification of tropical cyclones—A global observational analysis. *Monthly Weather Review*, *147*(1), 311–328. https://doi.org/10.1175/mwr-d-18-0214.1

Zhang, X., Zheng, F., Zhu, J., & Chen, X. (2022). Observed frequent occurrences of marine heatwaves in most ocean regions during the last two decades. *Advances in Atmospheric Sciences*, *39*(9), 1579–1587. https://doi.org/10.1007/s00376-022-1291-3

Zhang, Z., Wang, W., Doyle, J., Moskaitis, J., Komaromi, W., Heming, J., et al. (2023). A review of recent advances (2018-2021) on tropical cyclone intensity change from operational perspectives, Part 1: Dynamical model guidance. *Tropical Cyclone Research and Review*, *12*(1), 30–49. https://doi.org/10.1016/j.tcrr.2023.05.004

Zhou, G., Xu, J., Qian, Q., Xu, Y., & Xu, Y. (2022). Discriminating technique of typhoon rapid intensification trend based on artificial intelligence. *Atmosphere*, *13*(3), 448. https://doi.org/10.3390/atmos13030448

References From the Supporting Information

Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, *114*, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Foroumandi, E., Nourani, V., Jeanne Huang, J., & Moradkhani, H. (2023). Drought monitoring by downscaling GRACE-derived terrestrial water storage anomalies: A deep learning approach. *Journal of Hydrology*, *616*, 128838. https://doi.org/10.1016/j.jhydrol.2022.128838

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, *63*(1), 3–42. https://doi.org/10.1007/s10994-006-6226-1

Jin, Q., Fan, X., Liu, J., Xue, Z., & Jian, H. (2019). Using eXtreme gradient BOOSTing to predict changes in tropical cyclone intensity over the western North Pacific. *Atmosphere*, *10*(6), 341. https://doi.org/10.3390/atmos10060341

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., et al. (2017). LightGBM: A highly efficient gradient boosting decision tree. In *Neural information processing systems*. Paper presented at the.

Meng, F., Yao, Y., Wang, Z., Peng, S., Xu, D., & Song, T. (2023). Probabilistic forecasting of tropical cyclones intensity using machine learning model. *Environmental Research Letters*, *18*(4), 044042. https://doi.org/10.1088/1748-9326/acc8eb

NVIDIA. (2024). What is XGBoost? Retrieved from https://www.nvidia.com/en-us/glossary/xgboost/. accessed on 13 March 2024.

Zhou, Y., Zhao, J., Zhan, R., Chen, P., Wu, Z., & Wang, L. (2021). A logistic-growth-equation-based intensity prediction scheme for western North Pacific tropical cyclones. *Advances in Atmospheric Sciences*, *38*(10), 1750–1762. https://doi.org/10.1007/s00376-021-0435-1