Journal of the Indian Society of Remote Sensing (2025) 53:1965–1978
https://doi.org/10.1007/s12524-024-02101-y

RESEARCH ARTICLE

A Machine Learning Approach for Detecting Rapid Intensification in Tropical Cyclones

Tushar Sharma[1] · Manish Mawatwal[2] · Saurabh Das[2]

Received: 4 July 2024 / Accepted: 15 December 2024 / Published online: 17 January 2025

Abstract

Rapid intensification (RI) of Tropical Cyclones (TCs) is a significant threat to coastal regions worldwide. Although progress has been made in predicting TCs, RI events remain difficult to forecast accurately and in a timely manner. This study proposes a Machine Learning (ML)-based classification framework for RI prediction that utilizes Support Vector Machines (SVM) in conjunction with the Synthetic Minority Oversampling Technique (SMOTE) to handle the class imbalance of RI and non-RI cases. The Statistical Hurricane Intensity Prediction Scheme (SHIPS) data for the years 1982 to 2017 for the Atlantic Ocean basin and 1990 to 2010 for the Indian Ocean basin, respectively, are used to train and evaluate the proposed framework. Independent testing is conducted on operational data from 2010 to 2020 for the Atlantic basin and reanalysis data from 2011 to 2017 for the Indian basin. The framework demonstrates high skill, achieving a high Probability of Detection (POD) with a 24-h lead time in both basins, making it a potentially valuable tool for forecasters to issue timely warnings and prepare for the impacts of RI events. Overall, this study highlights the potential of ML-based approaches for improving RI prediction, and emphasizes the importance of ongoing efforts to develop more accurate and reliable forecasting methods for TCs.

Keywords Rapid Intensification · Machine Learning · Tropical cyclone · SVM · SMOTE

Introduction

Tropical cyclones are among the most destructive weather phenomena on earth, causing significant economic damage and loss of life in coastal regions worldwide. Although significant progress has been made in predicting TCs, RI events remain challenging to forecast in a timely manner. RI is defined as an event of a rapid increase in maximum sustained surface wind speed by at least 30 knots (approximately 35 miles per hour) over a 24-h period (DeMaria & Kaplan June, 1994). These events can occur suddenly, making it difficult for forecasters to issue timely warnings and prepare for their impacts. Deeper understanding and improvement in RI detection has been the top priority of the global agencies including National Hurricane Center (NHC) (Rappaport et al., 2009). RI is difficult to forecast, and seasons with more frequent RI cases tend to have large annual average forecast errors.

Researchers have explored the role of ocean characteristics, inner-core processes, and environmental interactions in TC intensity change (Zhu et al., 2004). Researchers have been exploring various techniques to improve TC intensity and RI prediction, including statistical models, ML, and Deep Learning methods (Li et al., 2017; Mawatwal & Das, 2024; Mercer & Grimes, 2017; Su et al., 2020; Wenwei et al., 2021). Traditional statistical analysis techniques have been used to find the associations between rapidly intensifying TCs and their environmental properties, but they are limited in their ability to capture the complex interactions between multiple non-linear factors for example, air-sea interaction. In recent years, ML based approaches have shown some promising results. Some works utilized SVM for the classification of RI cases (Li et al., 2017; Mercer & Grimes, 2017; Shaiba & Hahsler, 2016) used Long Short-Term Memory (LSTM) networks to predict RI of TCs.

✉ Tushar Sharma
tushar2016sharma@gmail.com

Manish Mawatwal
mawatwalmanish1997@gmail.com

Saurabh Das
saurabh.das@iiti.ac.in; das.saurabh01@gmail.com

[1] Department of Data Science, The George Washington University, Washington, D.C., USA

[2] Department of Astronomy, Astrophysics and Space Engineering, Indian Institute of Technology, Indore, India

LSTM is a type of recurrent neural network that can capture long-term dependencies in time series data (Cloud et al., 2019; Xu et al., 2021). Utilized multilayer perceptron (MLP) models to forecast the intensity of TC in the Atlantic and eastern Pacific regions. They used environmental predictors, which were derived from numerical model forecast data or from the statistical–dynamical SHIPS database. More recently, (Griffin et al., 2022) developed a convolutional neural network (CNN) for the prediction of RI in the North Atlantic and Eastern North Pacific basins. They found that a CNN approach is more skilful at predicting RI than the current operational RI models (Chen et al., 2023). Developed a deep learning ensemble model that achieved higher RI detection probabilities than the operational forecasts in the western Pacific basin (Wang et al., 2021). Developed a CNN with attention layers and noise reduction for TC intensity estimation using Himawari-8 satellite data.

This study builds on previous research by employing a ML-based classification approach enhanced with oversampling techniques to address the class imbalance inherent in RI prediction. A customized model evaluation strategy was utilized alongside effective feature selection methods to enhance model performance. This approach aims to provide a more comprehensive understanding of the factors contributing to RI events. The objectives of this study are: (1) To develop an interpretable ML model capable of detecting RI cases with high skill accuracy across different ocean basins, and (2) to comparatively implement various feature selection methods to identify the most relevant features for RI detection.

Methodology

Data

In this study, the SHIPS developmental data for RI detection was used (DeMaria & Kaplan June, 1994). SHIPS database is considered as the most complete and updated database for the study of TC intensity changes. SHIPS data is publicly available at the Regional and Mesoscale Meteorology Branch (RAMMB) website: (http://rammb.cira.colostate.edu/research/tropical_cyclones/ships/index.asp).

In the original SHIPS data, there are more than 500 weather, ocean, and climatological features. Many features in the data are time-dependent and are available in the original data in 5-day and 7-day window from the reporting time in 6-h intervals. This study utilized pre-processed data files taken from Xu et al. (2021). They retained a total of 121 SHIPS features with time-dependent features at the 24-h timestamp taken from the 5-day SHIPS data. The data has a total sample size of 56,223 taken from reanalysis data from 1982 to 2017 of all the basins. SHIPS data contains records from different ocean basins. This study analysed the Atlantic and the Indian basin (North Indian Ocean basin). The Atlantic basin was chosen because it is a widely studied basin and experiences frequent cyclone events. Its features are of high interest to the scientific community. In contrast, while the Indian basin experiences comparatively lesser cyclone events, the trend of RI suggests that there is an increasing trend in the last decade for this basin (Nadimpalli et al., 2021). The limited study of the rapid intensification phenomenon in this basin was another reason for studying it.

For the Atlantic basin, the reanalysis data from 1982 to 2017 was used for training and validating the model and the SHIPS operational data for the years 2010 to 2020 was used for testing the model. For the Indian basin, the reanalysis data from 1990 to 2010 formed the training and validation data, and reanalysis data from 2011 to 2017 formed the testing data, given the non-availability of operational data for this basin. The modeling approach that follows, is discussed in the subsequent section. Each record in the data indicates the value of the 121 SHIPS features at a 6-h interval. In the reanalysis data, there were a total of 496 cyclone events for the Atlantic basin with a total of 10,710 6-hourly data records from 1982 to 2017 and 146 cyclone events for the Indian basin with 2381 6-hourly data records covering the years 1990 to 2017. The threshold intensity for RI was taken as 30 kt over a period of 24 h.

Modeling Approach

In this study, a Leave-Three-Year-Out cross validation approach was employed, wherein three consecutive years constituted the validation data and all other years were used as training data. This iterative cross validation process continued until all the data was utilized, resulting in 12 cross validation folds (years 1982 to 2017) for the Atlantic basin and 7 for the Indian basin (years 1990 to 2010). This approach effectively addresses sample-size limitations and ensures a fair assessment of model performance. Given the limited number of RI cases, this cross-validation method allows for an effective distribution of data, ensuring that each fold has an adequate representation of both RI and non-RI samples for model validation. This method mitigates the risk of overfitting and ensures a comprehensive evaluation of model performance. The binary classification problem was then established by labelling all TC records as non-RI or RI based on the defined criteria, streamlining the analysis. The model was then tested on the operational data from 2010 to 2020 for the Atlantic basin and reanalysis data from 2011 to 2017 for the Indian basin. The SHIPS dataset exhibits a significant imbalance between RI and non-RI records, exemplified by the Atlantic basin's 10,710 6-hourly records, consisting of 10,134 non-RI and a mere 576 RI data points

(about 5.6% of the total instances). Similarly, in the Indian basin, there were a total of 138 RI data points in the total 2381 6-hourly records (5.8% of total).

The model's target variable dvs24, represents the intensity change over a 24-h period, calculated as the difference between the target time-step intensity minus the current intensity. Thus, with a 24-h lead time, the forecast made by this model can provide an up-to-24-h advance warning before the rapid intensification occurs. With a longer target lead time, say 48-h or 72-h, this prediction window will change accordingly.

SMOTE, is an oversampling technique commonly used in ML to address imbalances in class distribution, particularly in binary classification problems (Chawla et al., 2002). It involves selecting a minority class instance, identifying its k-nearest neighbors, randomly choosing one of them, and generating a synthetic instance along the line segment between the selected instance and the chosen neighbor. This process aids in creating diverse synthetic samples for the minority class, ultimately balancing the class distribution. To mitigate the imbalance in data among RI and non-RI classes, the SMOTE technique was employed exclusively on the training data to create synthetic data points. These synthetic data samples were not used in the validation and testing phase, ensuring that there was no information leakage.

For the choice of ML model, the SVM model was employed. SVM are a class of supervised ML algorithms used for classification and regression tasks (Cortes & Vapnik, 1995). Support Vector Classifier (SVC), a variant of SVM, finds a hyperplane to effectively separate two classes in the training data, maximizing the margin between them.

The optimal SVC parameters, namely gamma and cost, were determined using the GridsearchCV algorithm with the custom Leave-Three-Year-Out cross validation method instead of the default cross validation method. The optimal hyperparameters, gamma and cost for the Atlantic basin were 0.001 and 15, respectively and the non-RI to RI SMOTE sampling ratio was 0.60. For the Indian basin, they were 0.0005 and 5, respectively and the SMOTE sampling ratio was 0.60. The Radial Basis Function (RBF) kernel was the suitable kernel of choice for the SVC model. The true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) were computed from the classification results of each basin. Based on these values, the POD and False Alarm Ratio (FAR) were then evaluated which are commonly used metrics for binary classification problems. Another important metric used in meteorological forecasting called Heidke Skill Score (HSS)/Cohen's Kappa was calculated which evaluates a model's accuracy against random chance. The Pierce Skill Score (PSS), assesses a model's ability to correctly distinguish between RI events and non-RI events. The POD, FAR, HSS, and PSS are given as:

$$POD = \frac{TP}{TP + FN} \tag{1}$$

$$FAR = \frac{FP}{TP + FP} \tag{2}$$

$$HSS = \frac{2[(TP \times TN) - (FN \times FP)]}{(TP + FN)(FN + TN) + (TP + FP)(FP + TN)} \tag{3}$$

$$PSS = POD - POFD \tag{4}$$

where POFD (Probability of False Detection) is the proportion of non-RI events that are incorrectly forecasted as RI events.

Feature Selection

There were a total of 121 features in the dataset to start with. As an attempt to reduce the feature set size to a more relevant and reasonable one, two methods- random forest ranking and permutation feature importance were used to select the most significant features. These techniques are outlined below:

(a) *Random Forest (RF) Ranking*- It is an ensemble learning method that builds multiple decision trees during training and outputs the mode of their class distributions. It inherently provides feature ranking based on their importance in reducing impurity or error in tree nodes. In this study, an RF Classifier with 200 trees was used to rank the 121 features.

(b) *Permutation Importance (PI)*- It is a technique used to evaluate the significance of features in a ML model. Instead of relying on the inherent structure of the algorithm, PI assesses the impact of shuffling or permuting the values of a single feature on the model's performance. The importance score for that feature is high if there is a decrease in model performance when the feature is permuted.

These feature selection methods were applied independently in all the Leave-Three-Year-Out cross validation folds of data to ensure that there is no data leakage. While dimensionality reduction method for example Principal Component Analysis (PCA) is effective for reducing feature redundancy and interdependence, it is an unsupervised method that does not consider the target variable and only captures linear relationships. This can lead to the retention of components that explain variance but may not be predictive. However, a preliminary feature correlation analysis prior to modeling, was conducted based on the Pearson correlation

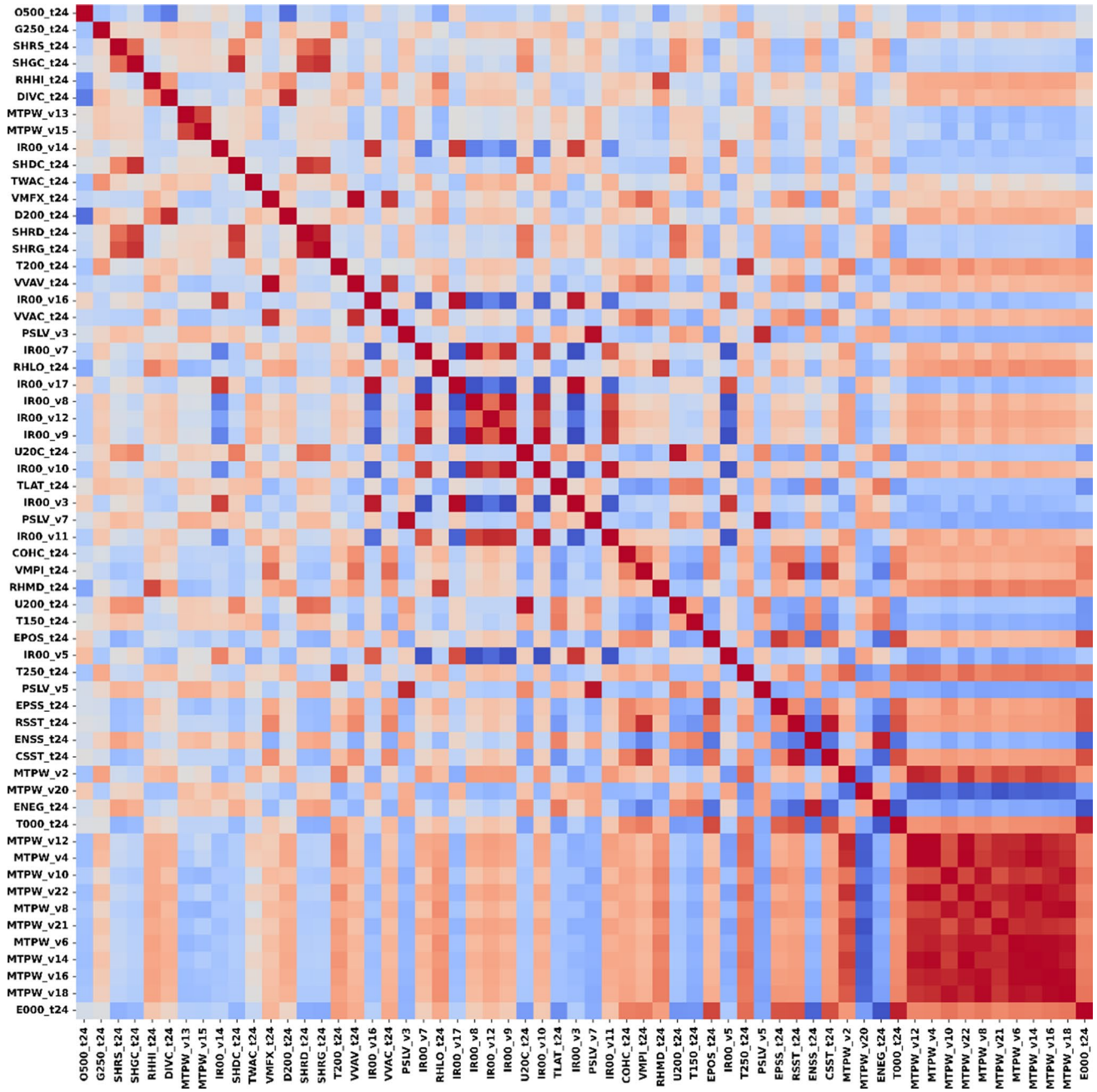

Fig. 1 Correlation matrix based on Pearson correlation of the top 60 features for the Atlantic basin. Blue and red colour represents low and high correlation, respectively.

coefficient. Figures 1 and 2 show the correlation matrix of the top 60 Atlantic and Indian basin features, respectively.

Both RF ranking and PI method are inherently non-deterministic. Additionally, the SMOTE oversampling technique used to balance the dataset introduces further variability, potentially resulting in slight variations in the top-ranked feature sets across different model executions. This may cause features to appear in different orders, particularly around the ranking threshold. To address this, both feature selection methods were repeated 10 times, and the importance scores for each feature were averaged to prioritize consistently important features for further analysis. Prior to this, importance scores were already averaged across all Leave-Three-Year-Out cross validation folds.

After evaluating the SVM models in each cross-validation fold, the feature ranking methods as discussed above were employed which ranked the 121 features in order of their importance in detecting the target. Thereafter, three

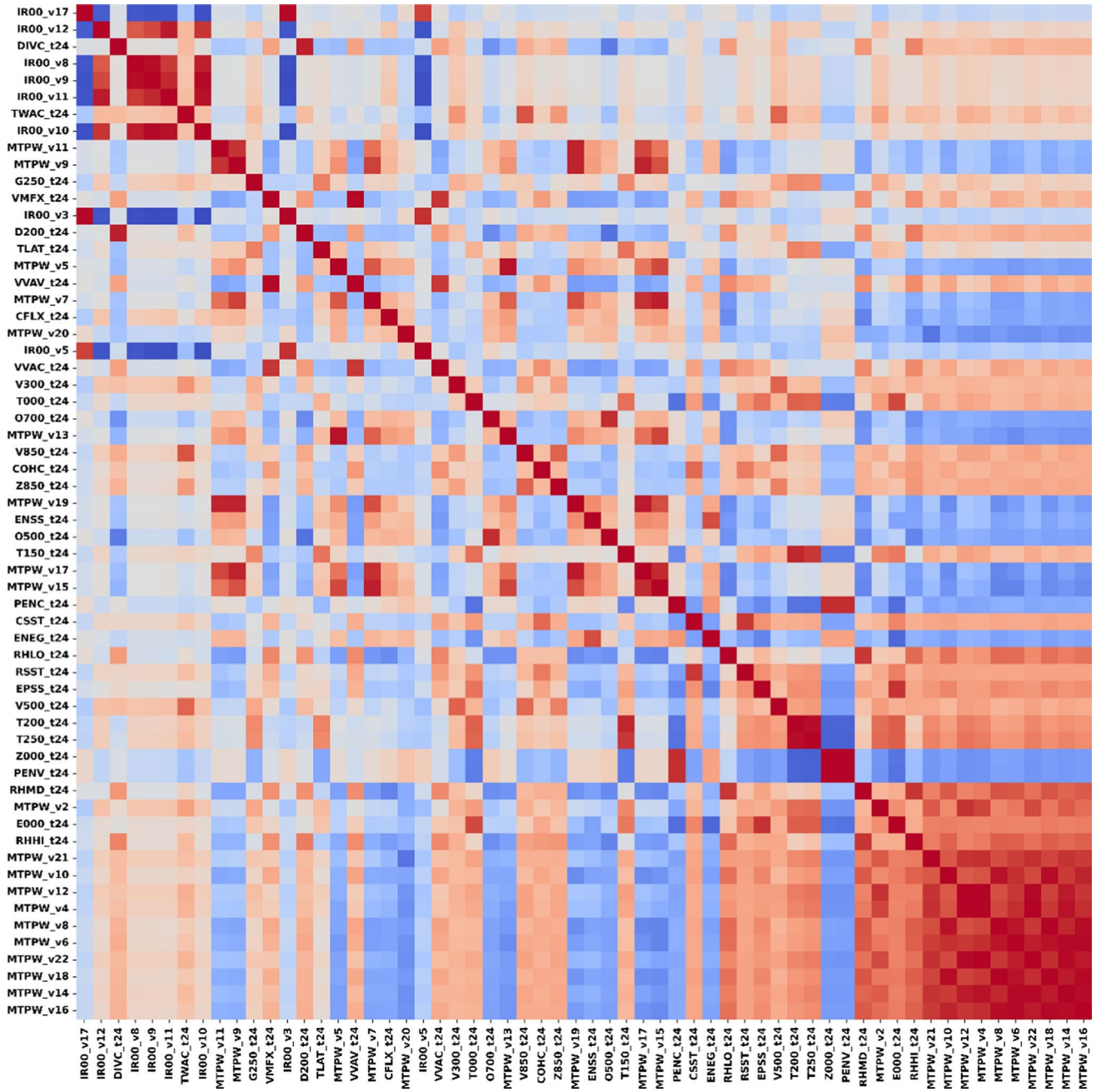

Fig. 2 Correlation matrix based on Pearson correlation of the top 60 features for the Indian basin. Blue and red colour represents low and high correlation, respectively.

thresholds of 20, 40, and 60 best features were set so as to reduce the feature set size. The model was then retrained again and then validated in each fold, giving the validation metrics. The flowchart for the entire RI prediction framework is shown in Fig. 3.

Results and Discussion

Proposed Framework Results

In accordance with the outlined methodology, the evaluation metrics were derived through averaging the results across all folds within the Leave-Three-Year-Out cross validation, conducted in the Atlantic and Indian basin. The objective was to assess model performance using the best 20, 40, and

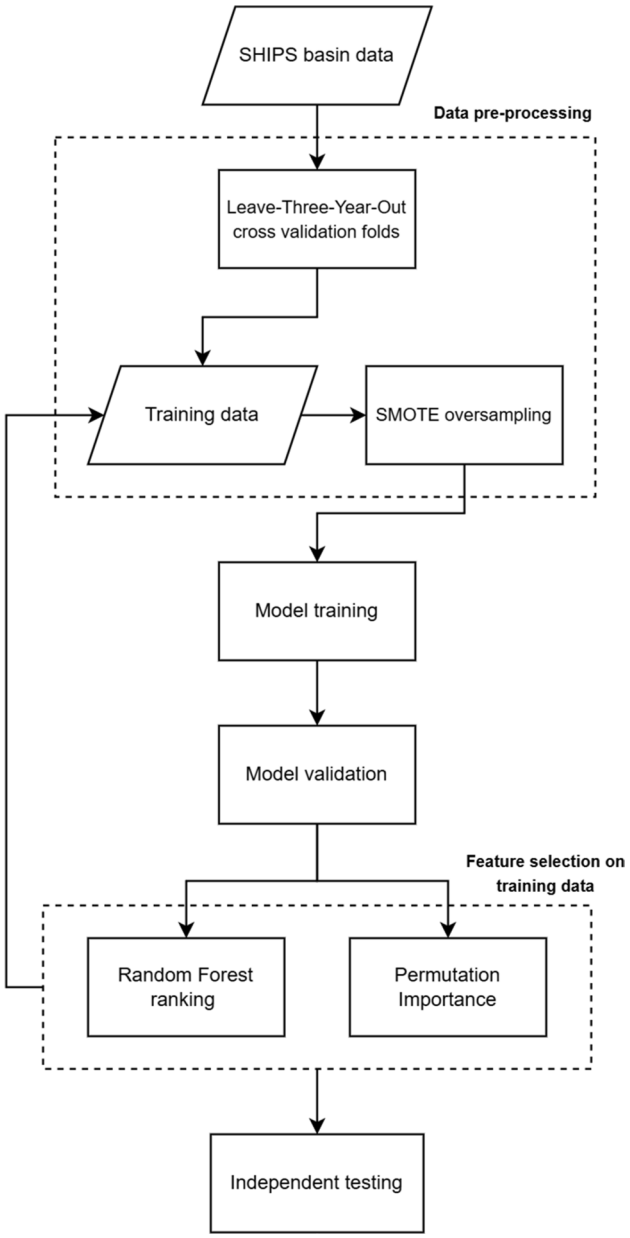

Fig. 3 Flowchart for the RI prediction framework.

60 features identified by the two distinct feature ranking methods, subsequently determining the more effective one.

With a threshold of top 20 features, there were three common features identified by both the ranking methods and for both the basins. These were DELV-12 (12-h intensity change of TC), vs0 (current intensity or maximum 1-min sustained wind speed of TC), and TWXC_t24 (maximum 850 hPa symmetric tangential wind at 850 hPa). This underscores the importance of these predictors in the prediction of RI of TCs. Other important features identified were DTL_t24 (distance to nearest landmass), VMPI_t24 (maximum potential intensity), and SHGC_t24 (generalized 850–200-hPa shear magnitude vs time with vortex removed and averaged from 0 to 500 km relative to 850-hPa vortex center).

Certain features considered important in the literature, such as the Climatological Ocean Heat Content (COHC_t24) and the dry air predictor CFLX_t24 (refer Appendix Table IV), did not appear in the top 20 features but were consistently ranked within the top 20–30 features. This positioning suggests that while such features may not be among the most dominant features, their consistent presence within this range highlights their relevance and potential contribution to the model's performance, likely influenced by specific data characteristics and interactions with other features.

Table I shows the cross validation results for both basins, showcasing results from both feature ranking methods and the different number of features selected. The results were evaluated in terms of the 95% confidence intervals based on 20 executions of the model in each case. The variability in the metrics arises due to random oversampling of the minority class by SMOTE in every cross-validation fold. A comparative analysis of metrics across the two basins indicates that RI cases are classified with greater accuracy in the Indian basin compared to the Atlantic basin. A plausible explanation for this observation is the occurrence of more RI events relative to the total cyclone events in the Indian basin compared to the Atlantic basin. In some years, there were more than 10% RI cases from the total cases which when coupled with the Leave-Three-Year-Out cross validation,

Table I RI detection leave-three-year-out cross validation results for the Atlantic and Indian basin in terms of the 95% confidence intervals. The model was re-trained 20 times for the evaluation and the selected number of features varied from 20, 40, 60, and 121 features

Ranking method	Metric	Atlantic				Indian			
		Best 20	Best 40	Best 60	All 121	Best 20	Best 40	Best 60	All 121
Random forest	POD	0.662 ± 0.003	0.638 ± 0.004	0.613 ± 0.003	0.430 ± 0.004	0.727 ± 0.003	0.692 ± 0.004	0.691 ± 0.005	0.582 ± 0.006
	FAR	0.819 ± 0.001	0.812 ± 0.001	0.800 ± 0.001	0.791 ± 0.002	0.731 ± 0.001	0.688 ± 0.003	0.693 ± 0.002	0.689 ± 0.006
	HSS	0.216 ± 0.001	0.225 ± 0.001	0.238 ± 0.002	0.222 ± 0.002	0.323 ± 0.002	0.365 ± 0.004	0.360 ± 0.003	0.339 ± 0.005
	PSS	0.310 ± 0.002	0.298 ± 0.002	0.285 ± 0.002	0.250 ± 0.002	0.354 ± 0.003	0.340 ± 0.003	0.335 ± 0.004	0.320 ± 0.004
Permutation importance	POD	0.629 ± 0.002	0.611 ± 0.004	0.585 ± 0.003	0.430 ± 0.004	0.803 ± 0.002	0.760 ± 0.006	0.725 ± 0.006	0.582 ± 0.006
	FAR	0.823 ± 0.001	0.803 ± 0.002	0.790 ± 0.001	0.791 ± 0.002	0.676 ± 0.001	0.679 ± 0.003	0.668 ± 0.002	0.689 ± 0.006
	HSS	0.208 ± 0.001	0.234 ± 0.001	0.246 ± 0.001	0.222 ± 0.002	0.400 ± 0.001	0.392 ± 0.004	0.397 ± 0.003	0.339 ± 0.005
	PSS	0.298 ± 0.002	0.285 ± 0.002	0.270 ± 0.002	0.250 ± 0.002	0.370 ± 0.002	0.355 ± 0.003	0.340 ± 0.003	0.320 ± 0.004

Table II Test results for the Atlantic and Indian basin

Metric	Atlantic basin	Indian basin
POD	0.510	0.880
FAR	0.711	0.627
HSS	0.328	0.492
PSS	0.447	0.811

significantly mitigates the class imbalance issue in the Indian basin. Consequently, the average metrics were improved.

From Table I, the optimal RI forecasting in the Atlantic basin is achieved using the top 60 features obtained from the PI method. These results align closely with those obtained from RF ranking using the top 60 features. The POD and FAR are conflicting metrics and therefore, the HSS or PSS can provide a balanced assessment of the model performance among the different configurations. The highest HSS at 0.238 (mean value of the confidence interval) is of the configuration with 60 features. Overall, both RF ranking and the PI techniques retained feature sets with comparable skill scores for the Atlantic basin. In the Indian basin, models with the top 20, 40, and 60 features from PI method, showed skilful RI forecasts, with the model with 20 features performing the best in terms of HSS (0.400) and PSS (0.370). However, PI yielded feature sets with only marginally better skill scores compared to RF ranking.

Overall, the obtained metrics demonstrate that the models incorporating features up to fixed thresholds, outperform those with a larger number or all of the 121 features. This finding simplifies the operational implementation by focusing on a concise set of variables that consistently yield high skill scores. The ability to achieve reliable forecasts with fewer features reduces computational complexity and facilitates real-time applications. The results also demonstrate the usefulness of the feature selection methods being employed.

Table II presents the testing results for the Atlantic and Indian basins using independent datasets: the SHIPS operational data from 2010 to 2020 for the Atlantic basin and reanalysis data from 2011 to 2017 for the Indian basin. The framework demonstrates high skill in detecting RI events across both regions. In the Atlantic basin, the model achieves a POD of 0.510 and a FAR of 0.711, with an HSS of 0.328 and a PSS of 0.447. These metrics highlight the framework's high skill in the Atlantic basin. In the Indian basin, the model achieves even higher skill, with a POD of 0.880 and a slightly lower FAR of 0.627, along with an HSS of 0.492 and a PSS of 0.811. These results highlight the model's consistent ability to detect RI events in both basins, though specific characteristics of the Indian basin's data facilitate even greater skill in this basin. Another useful metric, the False Alarm Rate (FARate), equivalent to POFD in the PSS equation, was calculated for the obtained results. The FARate of the Atlantic and Indian basin test results was 0.063 and 0.069, respectively.

Model Interpretation

SHapley Additive exPlanations (SHAP) is a game-theoretic approach to explain the output of any ML model by computing the Shapley values of each feature (Lundberg & Lee, 2017). These values provide a measure of the contribution of each feature in the model's prediction for a particular input. Figure 4 shows the SHAP summary plot for the top 20 features obtained from RF ranking for both basins. SHAP computation re-ranks the features based on their contribution to the model's outcome in terms of SHAP values. Each dot represents a cyclone event prediction and the colour represents the magnitude of a feature.

From the figure, it can be noted that the impact on the model predictions toward the positive class (RI) increases with an increase in the value of DELV-12 in both basins. This is likely because DELV-12 is the 12-h intensity change in the wind speed of the cyclone which is directly correlated with the target. This finding is in agreement with existing studies (Griffin et al., 2022; Meng et al., 2023). On the other hand, a decrease in SHGC_t24 resulted in an increase in its SHAP value and consequently, a positive impact in the model predictions toward RI. Studies have shown that the vertical wind shear is negatively correlated with the TC intensity change (Meng et al., 2023; Zeng et al., 2010). It is interesting to note the impact of vs0 on RI predictions in the two basins. For the Atlantic basin, vs0 had a high impact on the SHAP value and its spread across the SHAP value range is large. This suggests that TCs with a low value of maximum 1-min sustained wind speed (or current intensity) are more likely to undergo RI, which is consistent with (Meng et al., 2023). Conversely, for the Indian basin, the correlation between vs0 and SHAP value itself is low. The distance to the nearest major landmass (DTL_t24) is positively correlated with the SHAP value and consequently, with the probability of RI. The further away a cyclone is from land, the more intense it can become. Conversely, it weakens as it makes a landfall. This finding also aligns with existing studies (Meng et al., 2023; Schade & Emanuel, 1999).

To understand the relationships between the value of a feature and its impact on the model's prediction, SHAP dependence plots were generated for the years 2015 to 2017 and are shown in Fig. 5. Among the 20 features retained from the RF ranking method, most features for example DELV-12 and TWXC_t24 exhibited linear relationship with their SHAP values. There appears clear transition from negative to positive SHAP values for both these plots. Three features- SHRG_t24 (generalized 850–200-hPa

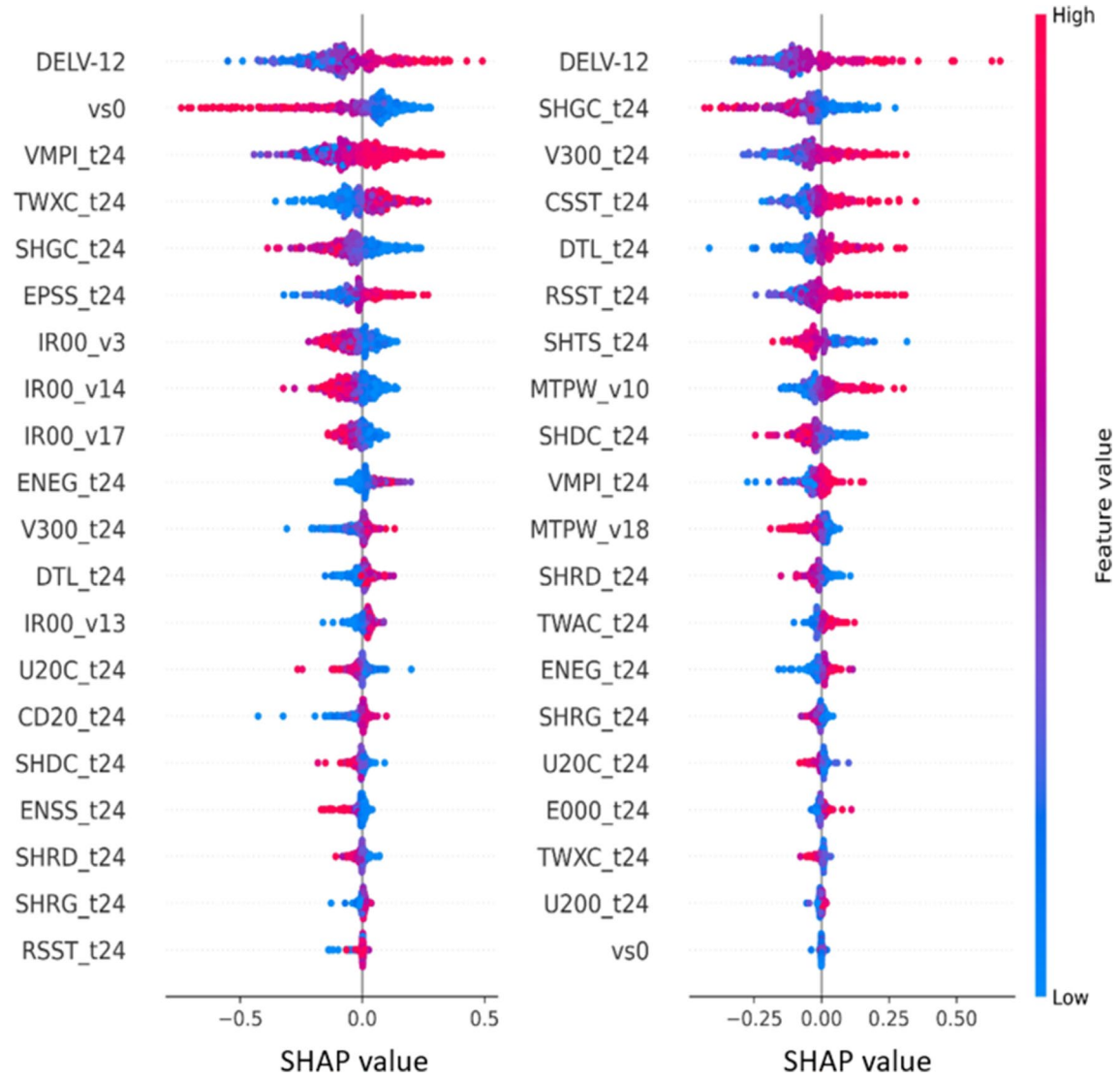

Fig. 4 SHAP summary plots for the top 20 features in the Atlantic (left) and Indian (right) basin obtained from RF ranking. The SHAP values are computed for the last cross validation fold which are the years 2015–2017.

shear magnitude (kt × 10) vs time), RSST_t24 (Reynolds Sea Surface Temperature), and V300_t24 (tangential wind (m/s × 10) azimuthally averaged at r = 500 km from (TLAT, TLON) at 300 hPa)- exhibited notable nonlinear trends in the Atlantic basin. The plot for SHRG_t24 indicates that shows a more scattered pattern, indicating variability in the SHAP values for different SHRG_t24 values. RSST_t24 has a negative impact toward RI favour-ability initially, but this effect stabilizes after a certain point, indicating a threshold beyond which RSST_t24 does not significantly affect the model's output. The statistical distribution of these features for other years are similar, and therefore, the SHAP dependence plots for them are also expected to be similar.

Model Performance Comparison

Probabilistic models gain operational priority over deterministic models at NHC. The SVM is a deterministic model by nature. However, the probabilities of SVM predictions were obtained which enabled a comparison with probabilistic models. The best model of the Atlantic basin, which had 60 features retained from permutation importance, was compared with the Operational SHIPS-RII Consensus model (SHIPS RII-C) (DeMaria et al., 2021). The SHIPS-RII-C model essentially combines the forecast from multiple different models and is shown to contribute to an improved forecast skill (Simon et al., 2018). The Brier Skill Score (BSS) is a suitable metric for evaluating probabilistic models. It is used to assess the accuracy of probabilistic forecasts by comparing them to a

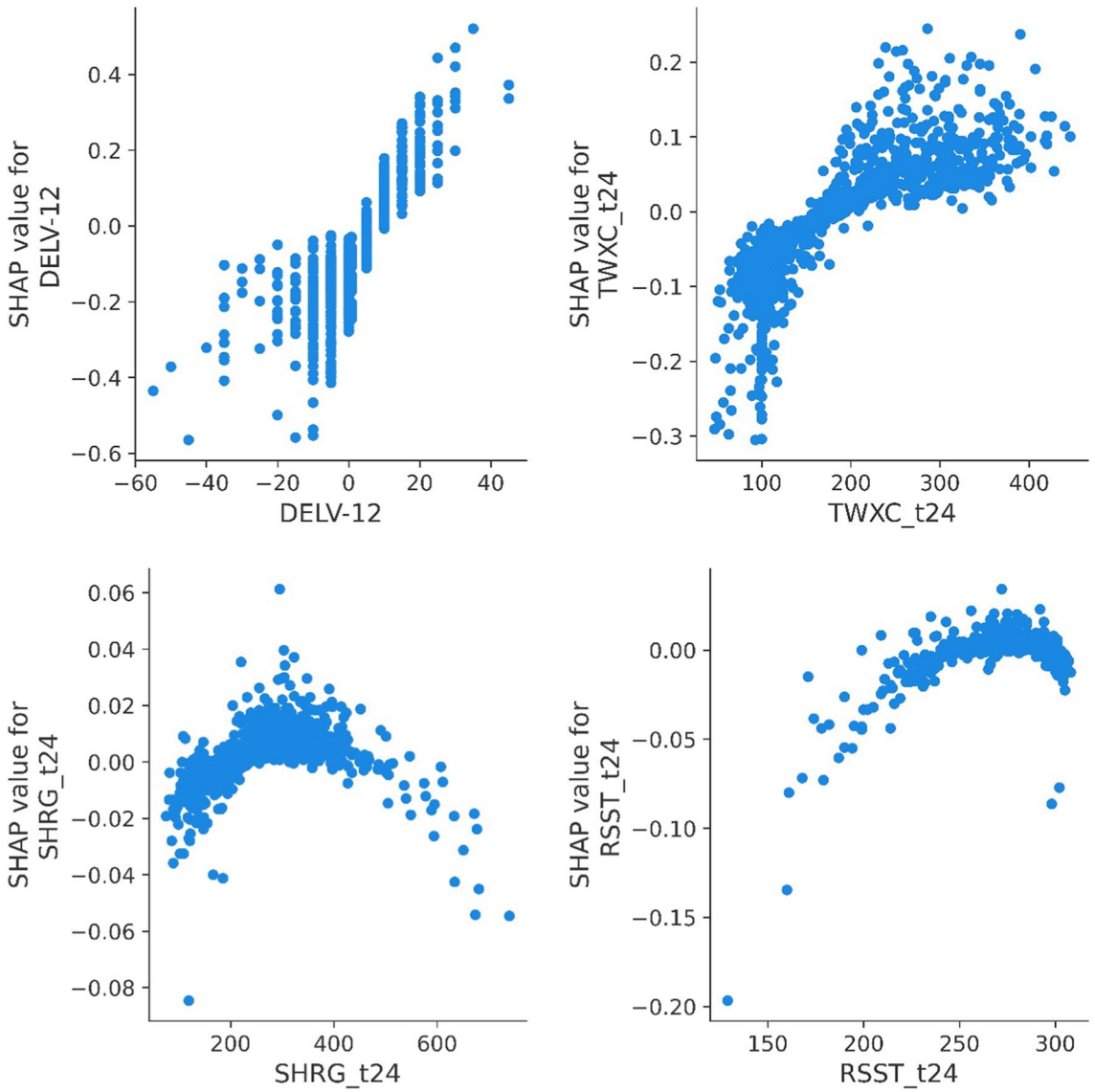

Fig. 5 SHAP dependence plots highlighting linearly (top) and non-linearly (bottom) correlated features with their SHAP values for the Atlantic basin in the last cross validation fold (years 2015–17).

reference forecast. To calculate BSS, the Brier Score (BS) of a reference model is required to be calculated first. BS measures the magnitude of the probability forecast errors. Thus, the lower the BS, the better the model performance. The BS and BSS are given as:

$$BS = \frac{1}{N}\sum_{i=1}^{N}\left(F_i - O_i\right)^2 \tag{5}$$

$$BSS = \frac{BS_{ref} - BS_{mod}}{BS_{ref}} \tag{6}$$

where N is the number of cyclone cases, F_i is the forecast probability of an event, and O_i is the observed probability which is 1 or 0 if the event is observed or not, respectively. BS_{ref} is the BS of the reference model and BS_{mod} is the BS of the model under consideration.

The reference model for this study is the SHIPS-RII-C model, the BSS of which is 0.19 (DeMaria et al., 2021). On the 2010–2020 operational data, the best Atlantic basin model in this study showed a BSS of 0.68, which is significantly improved over the BSS of the SHIPS-RII-C model. The best Indian basin model on the 2011–2017 reanalysis data, showed a BSS of 0.71, suggesting the model's high skill in operational forecasting.

Conclusion and Future Work

This study aimed at detecting RI events in TC using a ML approach. It demonstrated the potential of ML algorithms in tackling problems related to weather and climate. The SVM model was chosen as the ML model, and SMOTE was used for oversampling the minority classes (RI events) to balance the class distribution. The feature set was reduced with the help of a comparative implementation of two different approaches, namely RF ranking and PI. DELV-12, vs0, TWXC_t24, etc., were identified as some of the most important features relevant for skilful RI prediction. A highlight of the study was that minor adjustments to the SVM model parameters, while maintaining the established pipeline, yielded superior results in the two different ocean basins, suggesting the framework's adaptability and broader applicability. The Leave-Three-Year Out cross validation strategy helped overcome overfitting, which has been a challenging aspect for researchers. The model is limited to predicting the RI up to 24 h in advance. Extending predictions beyond 24 h could be taken in future work.

The validation results indicated that RI detection for the Indian basin exhibited better HSS than the Atlantic basin, which was likely due to the inherent nature of the data from that basin. The results in the Atlantic basin showed that using the top 60 features from PI yielded the best skill scores. Conversely, for the Indian basin, the top performing models had 20 and 40 top features from PI. Between the two feature ranking methods, PI resulted in better models when compared to RF for the Atlantic basin. For the Indian basin, PI resulted in marginally better models than RF. In independent testing, the best model for both basins demonstrated improved forecast skill compared to the SHIPS-RII-C operational model, as indicated by a higher BSS, underscoring the model's potential value in operational settings. The model demonstrated a favourable high POD, but also a high FAR, which is an area for improvement in the current study.

In summary, this study's outcomes can contribute to improving the detection and preparation for RI of TCs, particularly in regions susceptible to such weather phenomena. Future research avenues could explore alternative ML algorithms, advanced feature selection methods, and robustness tests to further enhance the framework's performance.

Table III Imbalanced dataset results for Atlantic basin

Metric	Best 20 RF features	Best 20 PI features
POD	0.033	0.087
FAR	0.785	0.702
HSS	0.046	0.11
PSS	0.029	0.075

Appendix

The results of the imbalanced dataset were initially not included due to their underwhelming performance. After thorough consideration and analysis, it was determined that the inclusion of these results would not contribute meaningful in sights or enhance the discussion. The results of the imbalanced dataset are as shown in Table III.The complete list of all the 121 SHIPS features is given in the Table IV.

Table IV Complete list of all 121 engineered features used in the 24-h intensity model for RI prediction

Predictors	Long name
vs0	Initialmaximum1-minsustainedwindspeedat 10 m(kt)
PSLV_v2	Pressure of the centre of mass (hPa)of the layer where storm motion best matches environmental flow
PSLV_v3	The observed zonal storm motion component ($ms^{-1} \times 10$)
PSLV_v4	The observed meridional storm motion component ($ms^{-1} \times 10$)
PSLV_v5	As in PSLV_v2, but for the 1000–100-hPamass weighted deep layer environmental wind ($ms^{-1} \times 10$)
PSLV_v6	As in PSLV_v3, but for the 1000–100-hPamass weighted deep layer environmental wind ($ms^{-1} \times 10$)
PSLV_v7	As in PSLV_v2, but for the optimally weighted deep layer mean flow ($ms^{-1} \times 10$)
PSLV_v8	As in PSLV_v3, but for the optimally weighted deep layer mean flow($ms^{-1} \times 10$)
PSLV_v9	The parameter alpha that controls the constraint on the weights from being not too "far" from the deep layer mean weights (non-dimensional × 100)
PSLV_v10	The optimal vertical weights for $p = 100$ hPa (non-dimensional × 1000)
PSLV_v11	The optimal vertical weights for $p = 150$ hPa (non-dimensional × 1000)
PSLV_v12	The optimal vertical weights for $p = 200$ hPa (non-dimensional × 1000)
PSLV_v13	The optimal vertical weights for $p = 250$ hPa (non-dimensional × 1000)
PSLV_v14	The optimal vertical weights for $p = 300$ hPa (non-dimensional × 1000)

Table IV (continued)

Predictors	Long name
PSLV_v15	The optimal vertical weights for $p = 400$ hPa (non-dimensional × 1000)
PSLV_v16	The optimal vertical weights for $p = 500$ hPa (non-dimensional × 1000)
PSLV_v17	The optimal vertical weights for $p = 700$ hPa (non-dimensional × 1000)
PSLV_v18	The optimal vertical weights for $p = 850$ hPa (non-dimensional × 1000)
PSLV_v19	The optimal vertical weights for $p = 1000$ hPa (non-dimensional × 1000)
MTPW_v2	0–200-km average total precipitable water (TPW) at t = 0 from the GFS analysis (mm × 10)
MTPW_v3	0–200-km TPW standard deviation (mm × 10)
MTPW_v4	200–400-km average TPW (mm × 10)
MTPW_v5	200–400-km TPW standard deviation (mm × 10)
MTPW_v6	400–600-km average TPW (mm × 10)
MTPW_v7	400–600-km TPW standard deviation (mm × 10)
MTPW_v8	600–800-km average TPW (mm × 10)
MTPW_v9	600–800-km TPW standard deviation (mm × 10)
MTPW_v10	800–1000-km average TPW (mm × 10)
MTPW_v11	800–1000-km TPW standard deviation (mm × 10)
MTPW_v12	0–400-km average TPW (mm × 10)
MTPW_v13	0–400-km TPW standard deviation (mm × 10)
MTPW_v14	0–600-km average TPW (mm × 10)
MTPW_v15	0–600-km TPW standard deviation (mm × 10)
MTPW_v16	0–800-km average TPW (mm × 10)
MTPW_v17	0–800-km TPW standard deviation (mm × 10)
MTPW_v18	0–1000-km average TPW (mm × 10)
MTPW_v19	0–1000-km TPW standard deviation (mm × 10)
MTPW_v20	Percent TPW less than 45 mm, r = 0–500 km in 908 azimuthal quadrants centred on up-shear direction
MTPW_v21	0–500-km averaged TPW (mm × 10) in 908 up-shear quadrant
MTPW_v22	0–500-km average TPW (mm × 10)
IR00_v2	Time (hhmm) of the GOES image
IR00_v3	Average GOES channel-4 brightness temperature (BT) (°C × 10), r = 0–200 km
IR00_v4	Std dev of GOES BT (°C × 10), r = 0–200 km
TGRD_t24	The magnitude of the temperature gradient between 850 and 700 hPa averaged from 0 to500 km estimated from the geostrophic thermal wind
TADV_t24	The temperature advection between 850 and 700 hPa averaged from 0 to 500 km from the geostrophic thermal wind
PENC_t24	Azimuthally averaged surface pressure at outer edge of vorte × [(hPa-1000) × 10]
SHRD_t24	850–200-hPa shear magnitude (kt × 10) vs time (200–800 km)
SHTD_t24	Heading of above shear vector
SHRS_t24	850–500-hPashearmagnitude(kt × 10)
SHTS_t24	Heading of above shear vector
SHRG_t24	Generalized 850–200-hPa shear magnitude (kt × 10) vs time (takes into account all levels from1000 to 100 hPa)
SHDC_t24	As in SHRD, but with vortex removed and averaged from 0 to 500 km relative to 850-hPa vortex center
SDDC_t24	Heading of above shear vector
SHGC_t24	As in SHRG, but with vortex removed and averaged from 0to500 km relative to 850-hPa vortex center
DIVC_t24	As in D200, but centered at 850-hPa vortex location
T150_t24	200–800-km area average150-hPa temperature
T200_t24	As above, but for 200-hPa temperature
T250_t24	As above, but for 250-hPa temperature
PENV_t24	200–800-km average surface pressure
VMPI_t24	Maximum potential intensity from K. Emanuel equation (kt)
VVAV_t24	Average (0–15 km) vertical velocity of a parcel lifted from the surface where entrainment, the ice phase and the condensate weight are accounted for
VMFX_t24	As in VVAV, but a density weighted vertical average

Table IV (continued)

Predictors	Long name
VVAC_t24	As in VVAV, but with soundings from 0 to 500 km with GFS vortex removed
HE07_t24	Storm motion relative helicity for $p = 1000–700$ hPa, r = 200–800 km
HE05_t24	As in HE05, but for $P = 1000–500$ hPa
O500_t24	Pressure vertical velocity at 500 hPa, averaged from r = 0 to1000km
O700_t24	As in O500, but at 700 hPa
CFLX_t24	Dry air predictor based on the difference in surface moisture flux between air with the observed (GFS) RH value, andwithRHo-fairmixedfrom500hPa to the surface
DELV-12	Last 12-h intensity change (kt)
IR00_v5	Average GOES channel-4 brightness temperature (BT) (°C × 10), r = 100–300 km
IR00_v6	Std dev of GOES BT (°C × 10), r = 100–300 km
IR00_v7	Percent area r = 50–200 km of GOES channel 4 BT < − 10 °C
IR00_v8	Percent area r = 50–200 km of GOES channel 4 BT < − 20 °C
IR00_v9	Percent area r = 50–200 km of GOES channel4BT < − 30 °C
IR00_v10	Percent area r = 50–200 km of GOES channel4BT < − 40 °C
IR00_v11	Percent area r = 50–200 km of GOES channel4BT < − 50 °C
IR00_v12	Percent area r = 50–200 km of GOES channel4BT < − 60 °C
IR00_v13	Max BT from 0 to 30-km radius (°C × 10)
IR00_v14	Avg. BT from 0 to 30-km radius (°C × 10)
IR00_v15	Radius of max BT (km)
IR00_v16	Minimum GOES brightness temperature from 20 to 120-km radius (°C × 10)
IR00_v17	Avg. BT from 20 to 120-km radius (°C × 10)
IR00_v18	Radius of min BT (km)
IR00_v19	Variables No.1 need for storm size estimation
IR00_v20	Variables No.2 need for storm size estimation
IR00_v21	Variables No.3 need for storm size estimation
CSST_t24	Climatological SST (°C × 10)
CD20_t24	Climatological depth (m) of 20 °C isotherm from 2005 to 2010 NCODA analyses
CD26_t24	As for CD20_t24, but for the 26 °C isotherm
COHC_t24	As above, but for ocean heat content (kJ cm^{-2})
DTL_t24	Distance to nearest major landmass (km)
RSST_t24	Reynolds SST (°C × 10)
U200_t24	200-hPa zonal wind (kt × 10) (r = 200–800 km)
U20C_t24	As in U200_t24, but for r = 0–500 km
V20C_t24	As in U20C_t24, but for the v component of the wind
E000_t24	100 hPa theta_e (r = 200–800 km) vs time (K × 10)
EPOS_t24	The average theta_e difference between a parcel lifted from the surface and its environment vs time
ENEG_t24	As in EPOS, but only negative differences are included
EPSS_t24	As in EPOS, but the parcel theta_e is compared with the saturated theta_e of the environment
ENSS_t24	As in ENEG, but the parcel theta_e is compared with the saturated theta_e of the environment
RHLO_t24	850–700-hPa relative humidity (%) vs time (200–800 km)
RHMD_t24	As in RHLO, but for700–500 hPa
RHHI_t24	As in RHLO, but for500–300 hPa
Z850_t24	850-hPa vorticity vs time (r50–1000 km)
D200_t24	200-hPa divergence vs time (r50–1000 km)
REFC_t24	Relative eddy momentum flux convergence (100–600-km avg)
PEFC_t24	Planetary eddy momentum flux convergence (100–600-km avg)
T000_t24	1000-hPa temperature (200–800-kmaverage)
R000_t24	1000-hPa relative humidity (200–800-kmaverage)
Z000_t24	1000-hPa height deviation (m) from the U.S. standard atmosphere

Table IV (continued)

Predictors	Long name
TLAT_t24	Latitude of 850-hPa vortex center in NCEP analysis
TLON_t24	Longitude of 850-hPa vortex center in NCEP analysis
TWAC_t24	0–600 km average symmetric tangential wind at 850 hPa from NCEP analysis
TWXC_t24	Maximum 850-hPa symmetric tangential wind at 850 hPa from NCEP analysis
G150_t24	Temperature perturbation at 150 hPa due to the symmetric vortex calculated from the gradient thermal wind
G200_t24	As in G150, but at 200 hPa
G250_t24	As in G150, but at 250 hPa
V000_t24	The tangential wind azimuthally averaged at r = 5500 km from (TLAT, TLON)
V850_t24	As inV000, but at 850 hPa
V500_t24	As inV000, but at 300 hPa
V300_t24	As inV000, but at 300 hPa

Acknowledgements The authors thank RAMMB/CIRA for providing the SHIPS developmental data and Wenwei Xu and co-authors for providing the open access data files. Author Saurabh Das also thankfully acknowledges the financial support received under SERB project MTR/2019/001581.

Author's Contribution The authors confirm contribution to the paper as follows: Study conception and design: T. Sharma, M. Mawatwal; Data Collection and Analysis: T. Sharma; Interpretation of results: T. Sharma, M. Mawatwal, S. Das; Draft Manuscript Preparation: T. Sharma, M. Mawatwal. All authors reviewed the results and approved the final version of the manuscript.

Funding The research is funded by SERB project MTR/2019/001581.

Declarations

Conflict of interest The authors declared that they have no conflict of interest.

References

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

Chen, B.-F., Kuo, Y.-T., & Huang, T.-S. (2023). A deep learning ensemble approach for predicting tropical cyclone rapid intensification. *Atmospheric Science Letters, 24*(5), e1151. https://doi.org/10.1002/asl.1151

Cloud, K. A., Reich, B. J., Rozoff, C. M., Alessandrini, S., Lewis, W. E., & DelleMonache, L. (2019). A feed forward neural network based on model output statistics for short-term hurricane intensity prediction. *Weather and Forecasting, 34*(4), 985–997. https://doi.org/10.1175/waf-d-18-0173.1

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*, 273–297.

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

DeMaria, M., & Kaplan, J. (1994). A statistical hurricane intensity prediction scheme (SHIPS) for the Atlantic basin. *Weather and Forecasting, 9*(2), 209–220. https://doi.org/10.1175/1520-0434(1994)009

Griffin, S. M., Wimmers, A., & Velden, C. S. (2022). Predicting rapid intensification in North Atlantic and Eastern North Pacific tropical cyclones using a convolutional neural network. *Weather and Forecasting., 37*(8), 1333–1355. https://doi.org/10.1175/waf-d-21-0194.1

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, IV-4/W2*, 101–105. https://doi.org/10.5194/isprsannals-iv-4-w2-101-2017

Lundberg, S. M., & Lee, S. -I. (2017). A unified approach to interpreting model predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). pp. 4768–4777. Curran Associates Inc. https://dl.acm.org/doi/10.5555/3295222.3295230

Mawatwal, M., & Das, S. (2024). An end-to-end deep learning framework for cyclone intensity estimation in North Indian Ocean region using satellite imagery. *Journal of the Indian Society of Remote Sensing, 52*(10), 2165–2175. https://doi.org/10.1007/s12524-024-01929-8

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

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

Nadimpalli, R., Mohanty, S., Pathak, N., et al. (2021). Understanding the characteristics of rapid intensity changes of tropical cyclones over North Indian Ocean. *SN Applied Sciences, 3*(1), 68. https://doi.org/10.1007/s42452-020-03995-2

Rappaport, E. N., et al. (2009). Advances and challenges at the National Hurricane Center. *Weather and Forecasting, 24*(2), 395–419. https://doi.org/10.1175/2008waf2222128.1

Schade, L. R., & Emanuel, K. A. (1999). The ocean's effect on the intensity of tropical cyclones: Results from a simple coupled atmosphere – ocean model. *Journal of the Atmospheric Sciences, 56*(4), 642–651. https://doi.org/10.1175/1520-0469(1999)056

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

Simon, A., Penny, A. B., DeMaria, M., Franklin, J. L., Pasch, R. J., Rappaport, E. N., & Zelinsky, D. A. (2018). A description of the real-time HFIP corrected consensus approach (HCCA) for tropical cyclone track and intensity guidance. *Weather and Forecasting, 33*(1), 37–57. https://doi.org/10.1175/waf-d-17-0068.1

Su, H., et al. (2020). Applying satellite observations of tropical cyclone internal structures to rapid intensification forecast with machine learning. *Geophysical Research Letters, 47*(17). https://doi.org/10.1029/2020gl089102

Wang, C., Zheng, G., Li, X., Xu, Q., Liu, B., & Zhang, J. (2021). Tropical cyclone intensity estimation from geostationary satellite imagery using deep convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing, 60*, 1–16. https://doi.org/10.1109/tgrs.2021.3066299

Wenwei, X., Karthik, B., Andrew, A., Nicholas, L., Nathan, H., Mark, D., & David, J. (2021). Deep learning experiments for tropical cyclone intensity forecasts. *Weather and Forecasting*. https://doi.org/10.1175/waf-d-20-0104.1

Xu, W., et al. (2021). Deep learning experiments for tropical cyclone intensity forecasts. *Weather and Forecasting, 36*, 1453.

Zeng, Z., Wang, Y., & Chen, L. (2010). A statistical analysis of vertical shear effect on tropical cyclone intensity change in the North Atlantic. *Geophysical Research Letters*. https://doi.org/10.1029/2009GL041788

Zhu, T., Zhang, D.-L., & Weng, F. (2004). Numerical simulation of Hurricane Bonnie Part I: Eyewall evolution and intensity changes". *Monthly Weather Review, 132*(1), 225–241. https://doi.org/10.1029/2009gl041788