Check for updates

# Improved forecasting of extreme hurricane events by integrating spatio-temporal CNN-RF learning of tropical cyclone characteristics

Javier Martinez-Amaya[1]*, Nicolas Longépé[2], Veronica Nieves[1] and Jordi Muñoz-Marí[3]

[1]Image Processing Laboratory, University of Valencia, Valencia, Spain, [2]Φ-lab Explore Office, European Space Agency, Frascaty, Italy, [3]School of Engineering, University of Valencia, Valencia, Spain

Assessing hurricane predictions in a changing climate is one of the most challenging weather forecast problems today. Furthermore, effectively integrating information-rich features that are specific to the growth of hurricanes proves to be a difficult task due to the anticipated nonlinear interactions during the spatio-temporal evolution of the tropical cyclone system. Consequently, the need arises for complex and nonlinear models to address this formidable scenario. In light of this, we introduce a novel framework that combines a Convolutional Neural Network with a Random Forest classification configuration. This innovative approach aims to identify the critical spatial and temporal characteristics associated with the formation of major hurricanes within the hurricane and surrounding regions of the Atlantic and Pacific oceans. Here, we demonstrate that the inclusion of these unprecedented spatio-temporal features extracted from brightness temperature data, along with the temperature and anatomical cloud properties of the system, results in an average improvement of 12% in the prediction of severe hurricanes, using the previous model version as a benchmark. This enhancement in the prediction accuracy extends up to 3 days in advance, considering both regions collectively. Although these innovative attributes may be relatively more costly to generate, it allows us to gain a more refined understanding of the intricate relationships between different spatial locations and temporal dynamics, leading to more efficient and effective solutions. This hybrid machine learning approach also offers adaptability, enabling the exploration of other suitable hurricane or environmental-related conditions, making it suitable for potential future applications.

## 1 Introduction

The diagnosis of extreme hurricane events is hampered by the need for new ways to analyze the multiple factors involved in the development of tropical cyclones (TCs), both directly and indirectly. This challenge becomes particularly significant in the context of climate change, which may alter the pattern and features of TCs (Reed et al., 2018; Bhatia et al., 2019; Knutson et al., 2019; IPCC, 2021). Hurricanes are complex and costly

phenomena that are influenced by a wide range of non-linearly interrelated physical factors. This complexity poses a significant obstacle to accurate predictions, particularly in the days leading up to their maximum development (Asthana et al., 2021; Jiang et al., 2022).

Present-day models exhibit significant variation in their complexity and functional forms. For instance, they differ in their ability to capture the dynamics of atmospheric physical processes, as seen in the case of dynamical or statistical-dynamical forecasting techniques. In contrast, when employing machine learning methods, models must determine tree structures and make parameter/category assignments (refer to Chen et al. (2020) for a more extensive discussion on this topic). However, it is important to note that recent advancements in hybrid machine learning models have made notable progress. These models have the capacity to 1) extract essential information associated with the intensification of tropical storms; 2) categorize and learn complex nonlinear patterns that change over time; 3) improve predictions, especially in cases of rapid intensification (RI)– all while maintaining a low computational cost compared to traditional methods (Martinez-Amaya et al., 2022; Chen et al., 2023; Wei et al., 2023). Among these efforts, some focus on analyzing the structure (temperature and anatomy) of the cloud system using Geostationary Operational Environmental Satellites (GOES) imagery, converted to brightness temperature for analysis. The cloud system strengthens as the storm matures, and it can assist in identifying the potential for achieving "major hurricane" status using a Random Forest (RF) classification framework (Martinez-Amaya et al., 2022). RF effectively classifies and categorizes TCs patterns into groups that encompass a range of intensities. For example, it can distinguish between tropical storms (TSs) that do not reach hurricane status and TSs that intensify into major hurricanes (MH) with category 3, 4 or 5 strength. This classification framework employs a majority-voting ensemble for prediction refinement (Martinez-Amaya et al., 2022). Consequently, the goal of this classification approach is not to provide specific wind intensities but rather to offer an informed assessment of whether a TS is likely to develop into a MH based on key predictors associated with the cloud system. Considering that the previous analysis of hurricane-related features by Martinez-Amaya et al. (2022) was conducted using a simplified RF model that did not account for the cumulative knowledge of these features over time, nor the spatial information surrounding and within the cloud structure, our present study establishes a more robust prediction model. This updated model incorporates new input variables that capture spatio-temporal changes. Specifically, we investigate the time evolution of TC structural parameters across various temporal intervals and incorporate high-level spatial features extracted using a Convolutional Neural Network (CNN) model. CNNs are a type of artificial neural network widely used for recognizing patterns in images, and their applicability to tropical cyclones has been proven (Maskey et al., 2020; Lee et al., 2019; Carmo et al., 2021).

To the best of our knowledge, a joint CNN-RF model for TCs has not been attempted in previous reported studies on TCs. The innovative hybrid approach we propose, combining the power of CNN with RF, provides a solid foundation for the ongoing exploration of TC prediction. By incorporating spatio-temporal properties into the RF system, the model is able to recognize
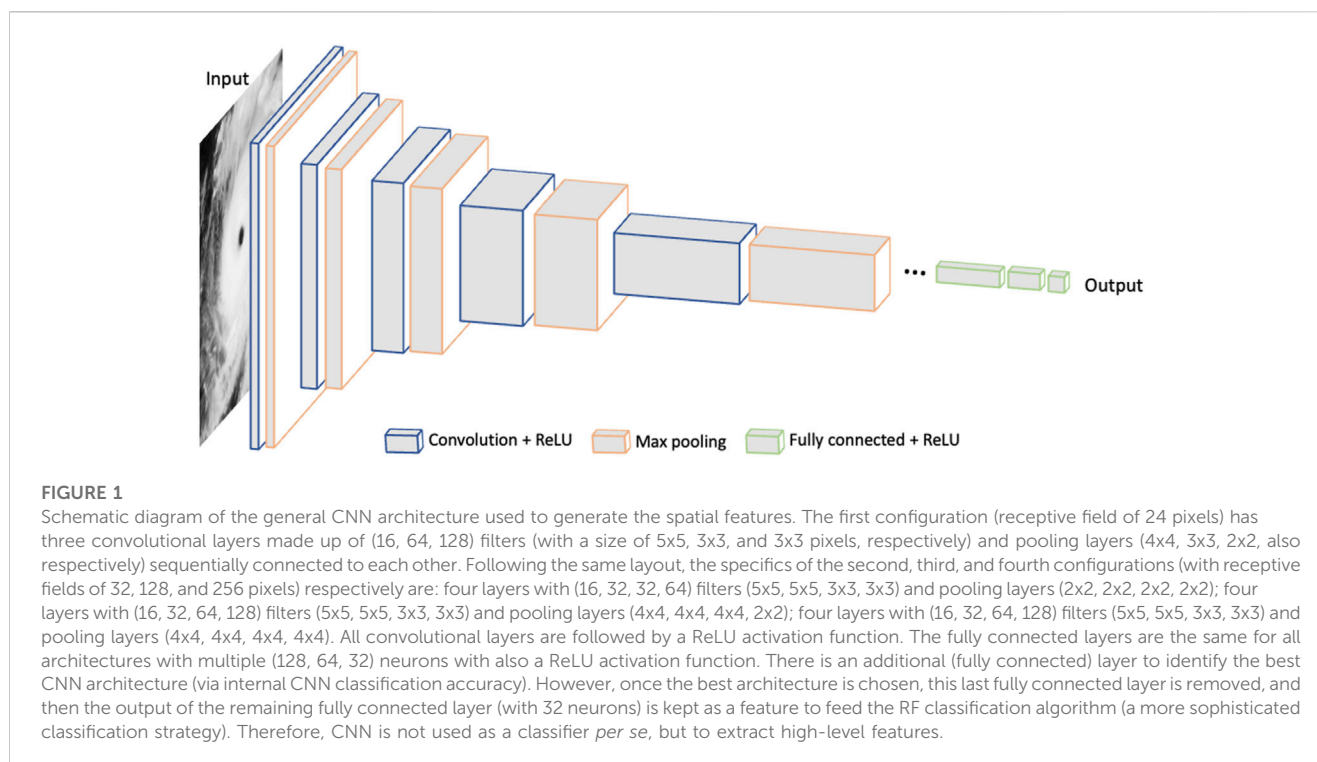
crucial indicators associated with the development and intensification of TCs, enhancing the prediction performance of MH events. This integration allows for a more comprehensive understanding of the complex relationships between TC characteristics and the occurrence of MH events. The highly adjustable nature of the model components suggests that further advancements can be achieved by incorporating new characteristics or factors that might contribute to forecasting the TC behavior.

# 2 Methodology

We start with two datasets, one for each group of hurricanes, TS and MH, according to a range of intensity values: a) the TS dataset is the collection of tropical storms with wind speeds between 63 and 118 km/h; b) the MH dataset includes extreme events where winds exceed 178 km/h. The location and intensity information of the TCs were obtained from the International Best Track Archive for Climate Stewardship (IBTrACS) at 3-hourly resolution (Knapp et al., 2010), interpolated to 15- minutes time resolution (which is the GOES resolution). For consistency with the previous study, the same number of events was considered (from 1995 to 2019), although data augmentation was also tested in order to improve the system's performance (please refer to the table in Figure 4 for the sample size). The areas of study encompass the Atlantic ([−110° 0° E] [5° 50° N]) and northeast (NE) Pacific ([−180° −75° E] [5° 50° N]) basins, which represent the typical formation regions. Predictions generated by our CNN-RF model (discussed in the subsequent sections) are given for several different forecast (or lead-) times (from 6 to 54-h). The lead-times define the specific time considered for prediction (made at regular 6-h intervals) leading up to the maximum development of the event. To provide an example, if we consider a 18-h lead-time, the MH prediction relies on the TC attributes gathered (from GOES images) 18 h prior to the identified peak (unless otherwise specified). These lead-times allow for monitoring and analysis of the phenomenon's progression and behavior leading up to its peak (Martinez-Amaya et al., 2022). All images were previously converted to brightness temperature (Yang et al., 2017; NOAA, 2019).

## 2.1 Prominent time evolution features

To uncover the rapidly evolving features, the most relevant cloud structural parameters (the area, the temperature difference between the inner core and outer part of the storm cloud, and the morphology) were repeatedly extracted from the GOES images every hour following Martinez-Amaya et al. (2022). Furthermore, here, we explore the space-time evolution of these features over several periods of time (from 2 to 24 h) prior to each lead-time. Then, for each period of time, the mean and standard deviation for each above-listed structural parameter (such as the area, *etc.*) were calculated and used as inputs for the model (explained later). In a different experiment, the same features were extracted but for a single period of time covering the full range of information, from the formation of the storm to the lead-time, which provides comparable results (also discussed and tested).

**FIGURE 1**
Schematic diagram of the general CNN architecture used to generate the spatial features. The first configuration (receptive field of 24 pixels) has three convolutional layers made up of (16, 64, 128) filters (with a size of 5x5, 3x3, and 3x3 pixels, respectively) and pooling layers (4x4, 3x3, 2x2, also respectively) sequentially connected to each other. Following the same layout, the specifics of the second, third, and fourth configurations (with receptive fields of 32, 128, and 256 pixels) respectively are: four layers with (16, 32, 32, 64) filters (5x5, 5x5, 3x3, 3x3) and pooling layers (2x2, 2x2, 2x2, 2x2); four layers with (16, 32, 64, 128) filters (5x5, 5x5, 3x3, 3x3) and pooling layers (4x4, 4x4, 4x4, 2x2); four layers with (16, 32, 64, 128) filters (5x5, 5x5, 3x3, 3x3) and pooling layers (4x4, 4x4, 4x4, 4x4). All convolutional layers are followed by a ReLU activation function. The fully connected layers are the same for all architectures with multiple (128, 64, 32) neurons with also a ReLU activation function. There is an additional (fully connected) layer to identify the best CNN architecture (via internal CNN classification accuracy). However, once the best architecture is chosen, this last fully connected layer is removed, and then the output of the remaining fully connected layer (with 32 neurons) is kept as a feature to feed the RF classification algorithm (a more sophisticated classification strategy). Therefore, CNN is not used as a classifier *per se*, but to extract high-level features.

## 2.2 CNN parameters estimation

A 2D CNN algorithm was also developed to extract the high-level spatial features in brightness temperature data for each lead-time using a sliding window of 10° x 10° latitude/longitude, centered on the storm's location given by IBTrACS (Knapp et al., 2010). For this, all images were standardized, i.e., resized to 256x256 pixels using bilinear interpolation and normalized by Z-score feature distribution (Jiang and Tao, 2022).

Our CNN architecture consists of a set of interconnected layers (convolutional, pooling and fully-connected) with multiple configurations and receptive fields of various sizes to extract features at different local regions (portions of the input image covered by the CNN) (Luo et al., 2016) (see Figure 1). The convolutional layer is the primary building block of a CNN, which captures local information and learns features from the datasets in filter/kernel size. This hyperparameter, the kernel size, refers to the size of the matrix that contains the weights, which are learned using the Adam optimizer (with a learning rate of 0.001) and determines the size of these features locally (Yamashita et al., 2018; Sun, 2020). The pooling layer is used to reduce the number of input features and increase computational efficiency by selecting the maximum of the window region (max-pooling layer) (Boureau et al., 2010). We defined a pooling layer of different sizes to capture more detailed information. The fully connected layer, which connects all the neurons of a layer with all the neurons of the preceding layer, finally outputs the information of the last feature map (Maskey et al., 2020; Kwak et al., 2021). We adjusted the batch size (the number of training images in one iteration), which was set to 16, and used batch normalization after each max-pooling layer (Goodfellow et al., 2017; Devaraj et al., 2021). Standard

regularization terms (a dropout value of 0.65, L2 norm of 0.01 and early stopping after 10 epochs without improvement) were also applied to optimize the algorithm and prevent overfitting (Kwak et al., 2021; Devaraj et al., 2021; Calton and Wei, 2022). Besides, we used the Rectified Linear Unit Activation Function (ReLU), a computationally efficient activation function that introduces the property of nonlinearity to a deep learning model and allows the network to focus on the more relevant features in the data. It was implemented after each convolutional and fully connected layer (except for the last one) to train the network faster (Ding et al., 2018; Santosh et al., 2022). A sigmoid activation function was used in the last (fully connected) layer of the CNN model to predict the classification performance. Finally, CNN network depths of up to 13 layers were used to examine the right balance between complexity and accuracy (Kwak et al., 2021). To select the best CNN architecture configurations, we performed hyperparameter tuning using Grid Search to identify the optimal set of parameters that achieve the highest accuracy (Elgeldawi et al., 2021). We used the (binary cross entropy) loss function metric produced by the CNN classifier (Kwak et al., 2021). Note, however, that, as we want to keep the spatial features, we retain the output of the second-to-last fully connected layer of the CNN. These features are treated as another input variable of the RF algorithm to perform the classification into various categories. RF is preferred over CNN for binary classification tasks as it is more robust to overcome undersampling and overfitting issues (Kwak et al., 2021), which was also tested by us in one experiment (not included). A schematic overview detailing all architectures can be found in Figure 1. The CNN and RF models were developed using the Tensor Flow 2.6.0 and Scikit-Learn 1.2.1 libraries on Python 3.8.16, respectively.
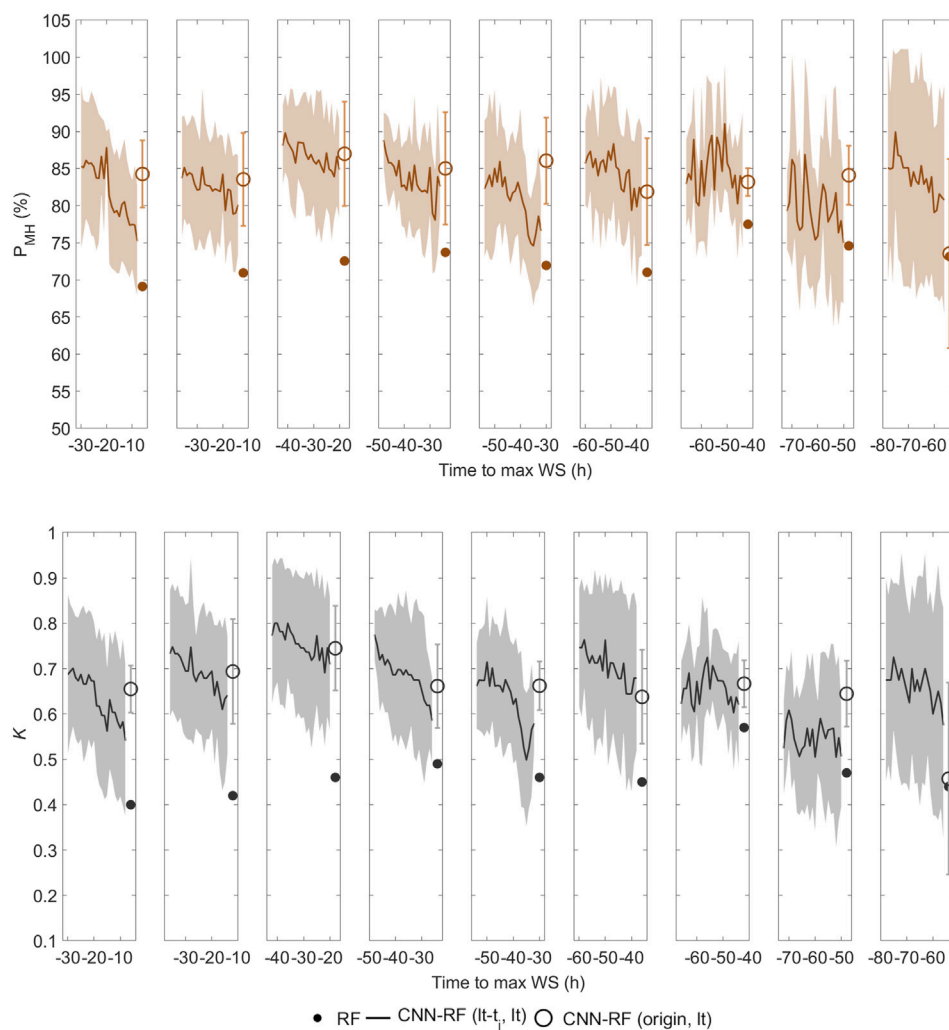
**FIGURE 2**
Metrics of our predictive models for the different proposed approaches, which include information on the: 1) TC structural parameters for each lead-time (our former study (Martinez-Amaya et al., 2022), shown in dots); 2) temporal evolution of the TC structural parameters for a time window of several hours prior to a given lead-time (lt), from 2 to 24-h ($t_i$), in combination with the CNN features obtained from a receptive field of 256 pixels (that draws a solid line); 3) as in 2) but considering the time-average (of the temporal evolution of the TC structural parameters) over the entire time range, from when the TC forms (origin) to a given lead-time (lt, in circles). The last two approaches also incorporate the standard deviation from the 5-fold cross-validation models (see Section 2.4). Confidence intervals correspond to one standard deviation (represented with a shaded area in the case of the second approach and depicted in bars for the third one). The predictive ability of our models was tested for lead-times between 6 and 54-h every 6 h (from left to right).

## 2.3 Design of the hybrid CNN-RF model

Once the most important spatial and temporal hurricane-related characteristics were determined for each lead-time, the events were classified based on these features according to two main hurricane categories (TS *versus* MH) using a RF algorithm to predict whether a storm will intensify to major hurricane status (as in Martinez-Amaya et al., 2022). Like in the previous study, several recommended steps were taken to achieve better results (Martinez-Amaya et al., 2022): 1) samples from the majority class were randomly removed to build a balanced dataset; 2) an 80/20 train/test split was applied on the dataset; 3) a 5-fold cross-validation process was conducted; 4) the RF was constructed with a maximum of 100 trees. The RF algorithm tends to reach saturation typically after using 100 trees (Martinez-Amaya et al., 2022). Please

also note that by employing random undersampling, we strike a balance between capturing rare events and avoiding misclassification problems caused by imbalanced datasets. This approach helps to improve the model's ability to correctly classify instances from both the majority and minority classes, leading to more accurate predictions overall (Chawla, 2003; Oh et al., 2022). Furthermore, here, we implemented a data augmentation strategy, including rotations (90°, 180°, 270°) and flips (horizontally and vertically) to generate new samples (Pradhan et al., 2017; Jiang and Tao, 2022). In this case, we ran one additional experiment with 20% of the training dataset that was set aside for additional validation purposes. Note that the models are either way trained and internally tested against data in the classification task (with the portion of the dataset held for testing).
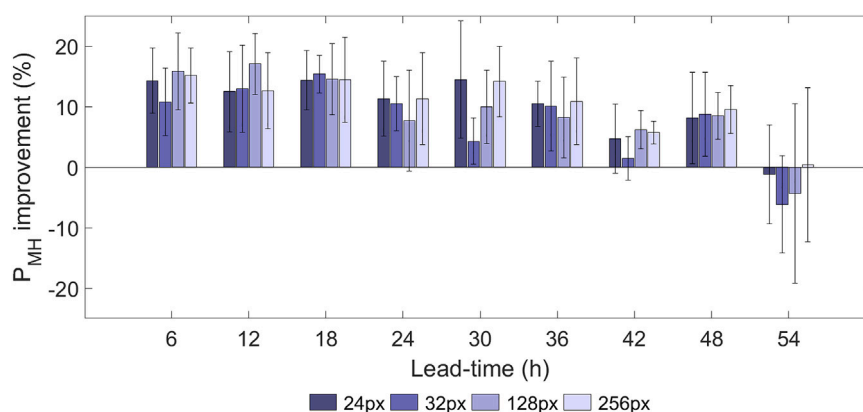
**FIGURE 3**
Model $P_{MH}$ improvement (in %) with respect to the reference method (Martinez-Amaya et al., 2022) for each lead-time and receptive field size (in pixels, px) with both regions combined. The new version of the model constructed using a RF approach, which contemplates the temporal variations of the TC structural parameters (over the entire period of each event in this case) and the CNN spatial features nearby and within the cloud system, always provides the best results up to 48-h (or 54-h using a receptive field of 256 pixels).

## 2.4 Model evaluation

As in Martinez-Amaya et al. (2022), both Cohen's Kappa $k$) and the precision ($p_{MH}$) were computed to evaluate the performance of the model on all experiments (Landis and Koch, 1977; Wei and Yang, 2021). The $p_{MH}$ of our classification represents the probability that an event predicted by the algorithm as a major hurricane did actually occur. The $k$ measures the reliability of the classification model in predicting the two classes. We report the mean and standard deviation of the accuracy ($p_{MH}$ and k) across all folds. For the RI events, the evaluation was done beyond the study period (to October 2022) using the hit ratio (HR) metric (Martinez-Amaya et al., 2022). The HR indicates the number of times that a correct prediction was made (out of the total number of predictions) with our random forest classifier.

## 3 Results and discussion

### 3.1 Assessment of the spatio-temporal CNN-RF predictions

In this section, the potential of a hybrid CNN-RF approach in the context of TC is assessed. The approach integrates the automatic feature extraction capability of CNN and also accounts for the temporal variations in key hurricane cloud parameters, for all the lead-times considered.

Our results show that when considering information from 2 to 24-h time period (prior to the lead-time) (see Section 2.1), the $p_{MH}$ $k$) ranges between 72% and 90% (0.45 and 0.70). By analyzing a longer period of time, from the early stage of the cyclone, predictions showcased similar accuracy values than that of 24-h. It can therefore be argued that it is not necessary to consider the information beyond 24-h to have reasonably accurate results.

Adding the final spatial features produced by the CNN method into the RF classification process (Sections 2.2 and 2.3) most often helps the overall skill of the model, yet it is sensitive to the receptive

field and depth in the CNN network architecture. The best CNN architecture was achieved by setting the number and type of layers as shown Figure 1. Having a smaller receptive field (of 24 or 32 pixels), makes the localization of distinctive features difficult. In contrast, large-scale complex features can be extracted with a larger receptive field (128 or 256 pixels), yielding to better results as evidenced by higher $p_{MH}$ (and $k$) values (see Figure 2 for a receptive field size of 256 pixels; Supplementary Figures S1–S3 for the remaining architectures). A receptive field with a size of 256 pixels offers the optimal outcome in test statistics ($p_{MH}$ =75–91% and $k$ = 0.50–0.80 for a moving temporal window that spans back from 2- to 24-h before a given lead-time, or $p_{MH}$ = 77–86% and $k$ = 0.47–0.72 for the entire duration of the event from when the TC forms to the lead-time) (see Figure 2). Either way, as expected, results are less consistent for lead-times that are far apart or too close to the maximum intensity of the cyclone due to the difficulties in sampling and in capturing contrasting patterns (Martinez-Amaya et al., 2022).

Considering the above limitations, an independent test was adopted to analyze the effect of data augmentation for the best architecture (this is the case of a receptive field of dimension 256 x 256 pixels), it made a positive impact on the performance that increased to 3% ($p_{MH}$) across all lead-times and time intervals. Splitting the data into training and validation test, however, did not seem to improve the results.

### 3.2 Comparison with the former RF-based model

Here, we compare the performance of the Martinez-Amaya et al. (2022) model and the different proposed predictive models based on the use of spatio-temporal features derived from brightness temperature data associated with TC intensification.

By only adding the temporal information of up to 24-h (or the entire period of time involved) prior to each lead-time, we obtain a
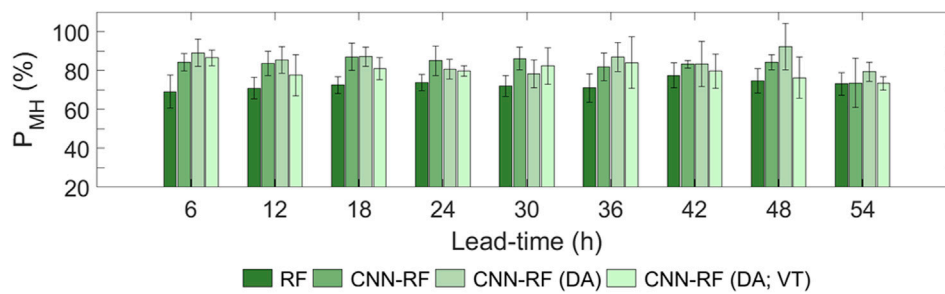
**FIGURE 4**
Accuracy comparison of all experiments, from the simpler (left) to the more sophisticated (right) test, with a receptive field size of 256 pixels: 1) RF as in Martinez-Amaya et al. (2022); 2) spatio-temporal CNN-RF, where the temporal evolution here is over the entire period of each event; 3) as in 2) but with data augmentation and 4) an additional validation test was performed. DA and VT stand for data augmentation and validation tests, respectively. The following numbers represent the cases that have been excluded from the dataset after applying random undersampling (before DA), aiming to achieve a balanced distribution between the classes, for each lead-time, ranging from 6 to 54-h: 125 T, 95 T, 79 T, 52 T, 24 T, 2 T, 5 MH, 25 MH, and 50 MH. The number of samples (# Samples) refers to the total count of events considered for the analysis with an equal representation of TSs and MHs.

performance boost of 8.5% (7.5%) in terms of $p_{MH}$. The accumulated temporal integration is, hence, fundamental to improve the accuracy in predicting MH. In addition, integrating the high-level spatial features (extracted from CNN using an effective architecture such as with a receptive field of 256 pixels) on the RF classification scheme, the improvement in $p_{MH}$ can reach 14%, depending on the lead-time (see Figure 3). Generally, including the complex spatial patterns (around and within the cloud region) learned by the CNN method is advantageous. However, for smaller sample sizes and close to the maximum development (such as with shorter lead-times), when the TC exhibits less identifiable spatial attributes, the model enhancement is questionable. Therefore, data augmentation seems crucial to adequately train the CNN model. This procedure allows for an increase in the short lead-time and the overall (all lead-times) classification prediction accuracy by up to 6% and 20%, respectively (see Figure 4). More importantly, the forecasting range beyond 36-h also improved by 10%. It is pivotal to make accurate forecasts several days in advance. Analysis with the $k$ statistic led to the same conclusions (not shown).

As a final remark, we note that even if the computational training cost scales up with the CNN-RF approach, once the model is trained and the new data processed, the time for each prediction is almost instantaneous (0.5 s).

## 3.3 Individual cases: Forecasting TC rapid intensification

Our best CNN-RF model configuration (see experiment three) in Figure 4) was tested for four relatively recent RI cases, two for each region, which were not included in the model training: category 4 Ida (2021) and category 4 Sam (2021) for the N AO, and category 4 Oren (2022) and category 4 Roslyn (2022) for the NE PO. Ida

underwent two periods of RI in just over a few hours, each with a different level of complexity. Sam was one of the most long-lived storms as it remained in category 4 status for more than 7 days. Oren exhibited a 24-h period of rapid intensification with a well-defined eye-core region and surrounding rings at the time of peak intensity (213 km/h). Roslyn rapidly intensified to a category 4 hurricane within 18 h, carrying sustained winds of 215 km/h. In all cases we always observed a HR higher than 60% for all lead-times. In particular, results for 12–18-h and 36–54-h lead-times showed better HR of at least 80% on average. Intermediate values are found for 24–30-h lead-times. By comparison, the Martinez-Amaya et al. (2022) model sometimes failed to outperform the classification results in the studied cases at different lead-times. Detailed results are given in the tables included in Supplementary Figure S4.

## 4 Conclusion

Our advanced parameter extraction and classification of the storm cloud system demonstrates that the patterns at multiple spatial and temporal scales displayed improved performance in predicting major hurricanes when compared to the Martinez-Amaya et al. (2022) model. The final improvement (relative to this previous model version) is 12% with data augmentation applied. Our model also enables us to make more accurate and consistent predictions of specific extreme cases of RI in recent years up to 54-h in anticipation. This advancement can have significant practical implications in terms of early warning systems. Yet, it is important to consider that the performance of the proposed approach may vary when applied to different regions or different types of storms for coastal communities prone to hurricanes. As the climate continues to change, future research could also explore the generalizability and

performance of the model under different climate scenarios. In any case, the model's ability to make predictions will be influenced by the availability of the training dataset that applies to different situations. As a whole, our cost-effective and scalable technique suggests a promising future impact on the weather forecasting paradigm through the investigation of further relevant indicators and diverse contexts for enhancing hurricane prediction, thus mitigating the impact of these destructive storms. However, expanding the range of input data may come at a higher cost.

## Data availability statement

The datasets presented in this study can be found in the AI4OCEANS repository (https://github.com/AI4OCEANS/).

## Author contributions

VN provided the resources, conceived and designed this study, coordinated the analysis of data and wrote the paper. JM-A analyzed the data, conducted the ML experiments, and wrote the paper. NL and JM-M provided expertise on ML techniques. All authors contributed to the article and approved the submitted version.

## Funding

## Acknowledgments

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feart.2023.1223154/full#supplementary-material

## References

Asthana, T., Krim, H., Sun, X., Roheda, S., and Xie, L. (2021). Atlantic hurricane activity prediction: A machine learning approach. *Atmosphere* 12, 455. doi:10.3390/atmos12040455

Bhatia, K. T., Vecchi, G. A., Knutson, T. R., Murakami, H., Kossin, J., Dixon, H. W., et al. (2019). Recent increases in tropical cyclone intensification rates. *Nat. Commun.* 10, 635–639. doi:10.1038/s41467-019-08471-z

Boureau, Y.-L., Ponce, J., and LeCun, Y. (2010). "A theoretical analysis of feature pooling in visual recognition," in Proceedings of the 27th International Conference on International Conference on Machine Learning, Haifa, Israel, 21–24 June 2010, 111–118.

Calton, L., and Wei, Z. (2022). Using artificial neural network models to assess hurricane damage through transfer learning. *Appl. Sci.* 12, 1466. doi:10.3390/app12031466

Carmo, A. R., Longépé, N., Mouche, A., Amorosi, D., and Cremer, N. (2021). "Deep learning approach for tropical cyclones classification based on C-band sentinel-1 SAR images," in 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11-16 July 2021, 3010–3013. doi:10.1109/IGARSS47720.2021.9554111

Chawla, N. V. (2003). "C4.5 and imbalanced datasets: Investigating the effect of sampling method probabilistic estimate and decision tree structure," in Proceedings of International Conference on Machine Learning Machine Learning, Washington, DC, USA, 21 August 2003, 1–8.

Chen, B. F., Kuo, Y. T., and Huang, T. S. (2023). A deep learning ensemble approach for predicting tropical cyclone rapid intensification. *Atmos. Sci. Lett.* 24, e1151. doi:10.1002/asl.1151

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

Devaraj, J., Ganesan, S., Elavarasan, R. M., and Subramaniam, U. (2021). A novel deep learning based model for tropical intensity estimation and post-disaster management of hurricanes. *Appl. Sci.* 11, 4129. doi:10.3390/app11094129

Ding, B., Qian, H., and Zhou, J. (2018). "Activation functions and their characteristics in deep neural networks," in Chinese Control and Decision Conference (CCDC), Shenyang, China, 9-11 June 2018, 1836–1841. doi:10.1109/CCDC.2018.8407425

Elgeldawi, E., Sayed, A., Galal, A. R., and Zaki, A. M. (2021). Hyperparameter tuning for machine learning algorithms used for Arabic sentiment analysis. *Informatics* 8, 79. doi:10.3390/informatics8040079

Goodfellow, I., Bengio, Y., and Courville, A. (2017). *Deep learning*. Cambridge: MIT press, 321–359.

Jiang, S., Fan, H., and Wang, C. (2022). Improvement of typhoon intensity forecasting by using a novel spatio-temporal deep learning model. *Remote Sens.* 14, 5205. doi:10.3390/rs14205205

Jiang, S., and Tao, L. (2022). Classification and estimation of typhoon intensity from geostationary meteorological satellite images based on deep learning. *Atmosphere* 13, 1113. doi:10.3390/atmos13071113

Knapp, K. R., Kruk, M. C., Levinson, D. H., Diamond, H. J., and Neumann, C. J. (2010). The international best Track archive for climate stewardship (IBTrACS): Unifying tropical cyclone best track data. *Bull. Amer. Meteor. Soc.* 91, 363–376. doi:10.1175/2009BAMS2755.1

Knutson, T., Camargo, S. J., Chan, J. C. L., Emanuel, K., Ho, C. H., Kossin, J., et al. (2019). Tropical cyclones and climate change assessment: Part I: Detection and attribution. *Bull. Amer. Meteor. Soc.* 100, 1987–2007. doi:10.1175/BAMS-D-18-0189.1

Kwak, G. H., Park, C., Lee, K., Na, S., Ahn, H., and Park, N. W. (2021). Potential of hybrid CNN-RF model for early crop mapping with limited input data. *Remote Sens.* 13, 1629. doi:10.3390/rs13091629

Landis, J., and Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174. doi:10.2307/2529310

Lee, J., Im, J., Cha, D. H., Park, H., and Sim, S. (2019). Tropical cyclone intensity estimation using multi-dimensional convolutional neural networks from geostationary satellite data. *Remote Sens.* 12, 108. doi:10.3390/rs12010108

Luo, W., Li, Y., Urtasun, R., and Zemel, R. (2016). "Understanding the effective receptive field in deep convolutional neural networks," in 30th Conference on NIPS, Barcelona, Spain, Dec. 5-10, 2016, 4905–4913.

Martinez-Amaya, J., Radin, C., and Nieves, V. (2022). Advanced machine learning methods for major hurricane forecasting. *Remote Sens.* 15, 119. doi:10.3390/rs15010119

Maskey, M., Ramachandran, R., Ramasubramanian, M., Gurung, I., Freitag, B., Kaulfus, A., et al. (2020). Deepti: Deep-Learning-Based tropical cyclone intensity estimation system. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 13, 4271–4281. doi:10.1109/JSTARS.2020.3011907

IPCC (2021). "Weather and climate extreme events in a changing climate," in *Climate change 2021: The physical science basis. Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change.* V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, and S. Berger Editors (Geneva, Switzerland: Intergovernmental Panel on Climate Change).

NOAA (2019). Product definition and user's guide (PUG), volume 5: Level 2+ products, DCN 7035538, GOES-R/code 416 (NOAA, 2019). Available at: https://www.goes-r.gov/products/docs/PUG-L2+-vol5.pdf (Accessed November 16, 2022).

Oh, S., Jufri, F. H., Choi, M.-H., and Jung, J. (2022). A study of tropical cyclone impact on the power distribution grid in South Korea for estimating damage. *Renew. Sustain. Energy Rev.* 156, 112010. doi:10.1016/j.rser.2021.112010

Pradhan, R., Aygun, R. S., Maskey, M., Ramachandran, R., and Cecil, D. J. (2017). Tropical cyclone intensity estimation using a deep convolutional neural network. *IEEE Trans. Image Process.* 27, 692–702. doi:10.1109/TIP.2017.2766358

Reed, K. A., Stansfield, A. M., Wehner, M. F., and Zarzycki, C. M. (2018). Forecasted attribution of the human influence on Hurricane Florence. *Sci. Adv.* 6, eaaw9253. doi:10.1126/sciadv.aaw9253

Santosh, K. C., Das, N., and Ghosh, S. (2022). "Chapter 2 - deep learning: A review," in *Deep learning models for medical imaging*. Editors K. C. Santosh, N. Das, and S. Ghosh (Cambridge: Primers in Biomedical Imaging Devices and Systems, Academic Press), 29–63.

Sun, R. Y. (2020). Optimization for deep learning: An overview. *J. Oper. Res. Soc. China* 8, 249–294. doi:10.1007/s40305-020-00309-6

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

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

Yamashita, R., Nishio, M., Do, R. K. G., and Togashi, K. (2018). Convolutional neural networks: An overview and application in radiology. *Insights Imaging* 9, 611–629. doi:10.1007/s13244-018-0639-9

Yang, C., Liu, Z., Gao, F., Childs, P. P., and Min, J. (2017). Impact of assimilating GOES imager clear-sky radiance with a rapid refresh assimilation system for convection-permitting forecast over Mexico. *J. Geophys. Res. Atmos.* 122, 5472–5490. doi:10.1002/2016JD026436