Cyclone Intensity Prediction using ERA5

Aman Goel
Department of Computing Technologies
SRM Institute of Science and Technology
Kattankulathur-603203, India
ag2351@srmist.edu.in

Rishaan Yadav
Department of Computing Technologies
SRM Institute of Science and Technology
Kattankulathur-603203 , India
ry1178@srmist.edu.in

Priya S
Department of Computing Technologies
SRM Institute of Science and Technology
Kattankulathur-603203, India
priyas3@srmist.edu.in

***Abstract*— Cyclone intensity prediction poses a significant challenge in meteorology due to the highly dynamic and nonlinear nature of atmospheric systems. This research work introduces a machine learning framework designed to predict cyclone intensity using the ERA5 reanalysis dataset, which consists of over one million data points. Surface pressure is used as the primary target variable, representing the intensity of cyclonic systems. A specialized three-tier feature selection technique has been developed for this work, combining statistical filtering, correlation-based analysis, and model-driven selection. Multiple machine learning models were trained and evaluated using standard performance metrics, including R^2 score, Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). The results indicate that the proposed framework achieves higher predictive accuracy compared to existing models. The improvement is attributed to the tailored feature engineering strategy and meticulous data preparation methods. By enabling more accurate forecasting of cyclone intensity, the proposed approach has the potential to support timely evacuations, reduce economic losses, and save lives during severe weather events.**

Keywords—cyclone intensity prediction, ERA5 reanalysis data, machine learning, feature engineering, surface pressure, comparative study

I. Introduction

Cyclone intensity prediction is still a challenging task in meteorological studies, particularly in areas with high susceptibility to extreme weather phenomena such as the Indian subcontinent. In this research, Proposed an ML-based approach to predict cyclone intensity using the ERA5 reanalysis data set. A data set of more than a million points is used, with surface pressure as target variable and a focus region being the Indian region. This region was selected manually by defining our custom latitude and longitude boundaries to get appropriate data for research. From the ERA5 dataset with more than 100 meteorological features, an initial narrowing identified 40 significant ones. Employing a sophisticated three-tier feature engineering technique, specifically developed for this project, further reduced the feature set to 22 significant predictors.

Each feature was personally reviewed and grouped into three categories: Useful, Maybe Useful, and Not Useful. Following the reduction to 40 relevant features, a sophisticated three-tier feature selection technique was used, and eventually, 22 of the most relevant predictors for cyclone intensity estimation were found.

For the choice of the output variable, a vast amount of research on cyclones was carried out. The most important factor for estimating cyclone intensity was determined to be the surface pressure caused by the core of the cyclone, expressed in kPa. Surface pressure was selected as the main target variable, providing precise estimation of cyclone intensity.

A number of ML models were trained and tested to compare their predictive accuracy. The approach presented is characterized by its expert-based data selection and feature engineering techniques, leading to improved accuracy over other models. The results of this study contribute to the improvement of early warning systems, ultimately leading to better disaster preparedness and mitigation in India.

II. Literature Survey

Accarino et al. [1] suggested an ensemble machine learning technique for tropical cyclone detection based on ERA5 reanalysis data. Their technique was superior to single-model methods by using a combination of classifiers, enhancing detection accuracy and robustness. This research illustrates the promise of ensemble learning in improving climate-related predictive models.

An analogous approach using an ensemble ML framework of a few VGG-like neural networks was proposed by [2] for improving cyclone localization and detection. The model successfully reduced localization errors and increased classification reliability but heavily diverged in performance when evaluated on unbalanced datasets, thereby failing for extreme cases of cyclones.

Chen, R [3] proposed cyclone intensity forecasting using ResNet-18. The approach improved generalizability to new storms, reducing RMSE to 67.98 knots, though sparse prediction errors indicated the need for further optimization. Another study in [4] examined inter- and intra-pattern fusion techniques for intensity prediction, merging multi-source data to increase forecasts.

Giffard-Roisin S et al. [5] had built a combined deep learning model of wind CNN, pressure CNN, and past cyclone paths, which was better than traditional statistical models like CLP5. Although this model performed extremely well in predicting global cyclones, it was inaccurate near land and in weaker storm development like depressions. The role of large-scale environmental conditions on cyclone intensity was investigated in [6], where scientists processed reanalysis data to understand the effect of external climatic conditions on cyclone behaviour.

A comprehensive literature survey was done before the project. Sandeep et al. [7] combined Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks to forecast cyclone formation up to 60 hours in advance from ERA5 and IBTrACS data. The approach was extremely accurate in forecasting formation in six ocean basins, showcasing the potential of deep learning in cyclone forecasting.

Another study in [8] was centered on landfall location and time prediction with CNN-LSTM networks that were trained on NOAA IBTrACS datasets. Although this model had high accuracy for landfall prediction, it did not generalize to larger cyclone tracking or intensity estimation.

Also, Koteshwaramma et al. [9] employed a high-resolution Weather Research and Forecasting (WRF) model using the IMDAA regional reanalysis dataset for the simulation of severe cyclones over the Bay of Bengal. Highly detailed regional simulations were generated by this research, but computational expenses involved in running such models were difficult to accommodate on large scales. A physics-informed solution was proposed in [10], which developed an open-source downscaling model of tropical cyclones with intensity-dependent steering.

Chibueze N [11] examined the effect of sea surface temperature on cyclone development using a regional climate model (UEMS-WRF). While this study produced valuable results on cyclone intensification, it required massive computation and was limited to specific oceanic conditions. A work by [12] assessed whether ERA5 marked a new era in tropical cyclone inner-core structure resolution. Even though this research analysed the ERA5 data quality, it did not develop a forecasting model, thus limiting direct application in intensity prediction.

A more sophisticated ML model was proposed in [13], where a causal autoregressive model was combined using multi-modal and multi-scale information via a causal cross-attention mechanism and LSTM-based autoregressive decoder. This model had better forecasting accuracy but was computationally intensive and thus not suitable for most applications. North Atlantic synthetic dataset [14] explores the cyclone track, intensity, and rainfall forecasting with ERA5-based statistical-dynamical models. The dataset was helpful in verifying predictive models, but it did not capture the actual-world intricacies of true cyclone occurrences.

A number of studies highlighting the reconstruction of tropical cyclone attributes using machine learning techniques. A dataset reconstruction technique in [15] used Random Forest models trained on ERA5 and IBTrACS data to compute maximum wind speed and cyclone size parameters. The approach improved reanalysis data quality and offered improved climatological insights, although the estimates propagated biases in the datasets, resulting in occasional discrepancies.

Overall, the literature for cyclone intensity forecasting has surveyed various methodologies ranging from deep learning and ensemble ML models to physics-based and hybrid approaches. Although ML-based approaches like CNN-LSTM and ResNet-18 have shown high predictive strength, they are still facing difficulties in dealing with imbalanced data, computational expense, and generalizability to extreme storm occurrences. Physics-based models offer a more transparent approach but tend to be computationally costly. An integration of sophisticated feature engineering, as investigated in the present work, with ML methods may provide a very effective avenue towards enhanced cyclone intensity prediction.

III. Proposed Methodology

Cyclones are prevalent in coastal areas. In the Indian subcontinent, primarily West Bengal, Odisha, Andhra Pradesh, and Tamil Nadu are hit by cyclones. There is no machine or model that can tell us what is going to happen in the next few days.

The research work proposes a machine learning model that assists in predicting storm intensity by considering 22 independent features and providing the output as surface pressure which help in predicting the intensity of cyclone.

The predictive pipeline consists of the following models:

A. Linear Regression

Model provides a baseline to compare against. It works well if there is a linear correlation between cyclone intensity and meteorological parameters (e.g., wind speed, mean sea level). There are some drawbacks: it is unable to capture complex and non-linear interactions since cyclone intensity is influenced by them, high multicollinearity between meteorological parameters impacts model stability; and it is outlier-sensitive.

B. Lasso Regression

Model achieves automatic feature selection by setting irrelevant variable coefficients to zero, and it is useful in high-dimensional data such as ERA5, where most features can have minimal importance. It is, however, subject to some limitations: it deletes significant variables, resulting in an underfitted model or incorrect prediction, and it does not perform well with highly correlated features.

C. Ridge Regression

Model handles multicollinearity by scaling down coefficient values, and it is useful when several features contribute to cyclone intensity in a dependent relationship. Its limitations are that it does not perform feature selection and makes the assumption of a linear relationship, which is not a realistic assumption for cyclones.

D. Bayesian Ridge Regression

Model incorporates probabilistic reasoning and offers confidence intervals, which can be helpful in uncertainty estimates. Its drawback is that it has the assumption of data being distributed as a Gaussian distribution and is computationally heavier than regular Ridge Regression. Further, the probabilistic nature of the model supports enhanced risk calculation.

E. K-Nearest Neighbors (KNN) Regression

Model operates by identifying previous cyclones with comparable meteorological conditions and employing their intensity values to make predictions. It is able to model non-linear relationships without assumptions regarding data distribution. Its drawbacks, however, are that it does not perform well with high-dimensional data (curse of dimensionality) and is computationally intensive for large datasets such as ERA5.

F. *Decision Tree Regressor*

Model is capable of capturing non-linear relationships and interactions between features and automatically identifies significant meteorological variables that affect cyclone intensity. But its drawbacks are that it has high variance and is susceptible to overfitting, and it can produce varied tree structures.

G. Random Forest Regressor

Model ensembles several decision trees, preventing overfitting and still preserving the ability to model non-linearity. It performs well on high-dimensional data and is less sensitive to outliers and noise. But it is limited in that it is computationally costly and less interpretable.

H. Adaboost Regressor

Model increases poor learners for the hard-to-predict intensities of cyclones and is good when a specific intensity is consistently underestimated. Nevertheless, limitations of this approach are that it is noise sensitive and computationally intensive.

I. Support Vector Regression (SVR)

Model employs kernel tricks to describe complicated, non-linear relationships between meteorological parameters and cyclone intensity and performs well when the relationship is not clearly known. Its shortcomings, however, are that it is extremely slow with large datasets and that the correct kernel function to use is difficult to determine.

J. *XGBoost Regressor*

Model deals with difficult, non-linear relationships effectively through gradient boosting, is less overfitting-prone, and performs well on highly correlated variables. Its shortcomings are that it needs to be carefully tuned for hyperparameters in order to be optimally performing and is computationally expensive.

The objective function for XGBoost was defined as:

$$\mathbb{Q} = \sum_{i=1}^{n} L(y_i, \hat{y}_i) + \sum_{k=1}^{k} \Omega(f_k) \text{ --------- (1)}$$

y_i is the actual cyclone intensity. y_i is the predicted intensity and $\Omega(f_k)$ is the regularization term to prevent overfitting

Taylor Expansion Approximation for Cyclone Intensity Prediction:

$$L(y_i, \hat{y}_i) \approx L\big(y_i, \hat{y}_i^{(t)}\big) + g_i\big(f(x_i)\big) + \frac{1}{2} h_i\big(f(x_i)\big)^2 \text{----(2)}$$

where g_i is first derivative of loss, h_i is the second derivative of loss and $\hat{y}_i^{(t)}$ is the previous iteration prediction.

Predicted Surface Pressure (Cyclone Intensity) using XGBoost Regressor:

$$y = \sum_{k=1}^{k} f_k(x_i) \text{ -------------(3)}$$

Equation 3, estimates cyclone intensity by considering meteorological features such as wind speed, temperature, and pressure variations

IV. Implementation & Discussion

The training data comprises of first 15 days of each month. Similarly, the testing data comprises of the next 10 days of each month and validation consist of remaining 5-6 days of each month ensuring the time-series aspect of dataset is respected. This methodology prevents the machine learning model from training on the future data before getting inferences from the past.

An advanced three-tier feature selection for feature selection.

A. Three-tier Feature Selection

Firstly, perform correlation coefficient with threshold of 0.1 to remove the weak correlated features among 40 features. Remove weak correlated features because it doesn't have any impact on cyclone intensity prediction.

Following correlation coefficient, SelectKBest is applied that selects the top K features from the dataset based on their statistical importance in relation to the target variable and for scoring function we use f_regression as it handles continuous targets, fast and computationally efficient. SelectKBest helps to remove irrelevant or redundant features and reduces overfitting by eliminating noise.

Finally, Recursive Feature Elimination (RFE) was performed to rank the features based on their importance and it eliminates the least important feature and it repeat this recursively until only the selected number of feature remains. 22 features are selected, as experimental studies show that less than 20 features result in underfitting, and 20 to 25 features give the best performance.

After applying three-tier feature selection technique the dataset is filtered with only the selected 22 features and the target variable. It is then transformed into Parquet file format that is more efficient than CSV. Speed is improved by the following: data is processed chunk-by-chunk, allowing memory-efficient processing of the large ERA5 dataset.

The primary performance is improved by using an Advanced Three-Tier Feature Selection Technique that applies Correlation Coefficient analysis, SelectKBest, and Recursive Feature Elimination (RFE) one after the other. This filters out irrelevant and redundant features and improves the machine learning model performance. The focus input results in improved model generalization. Data are divided into training (50%), test (30%), and validation (20%) sets without random shuffling such that the time-dependent nature of the data is preserved.

In cyclone intensity prediction, a wide array of meteorological and environmental factors is used to identify important cyclone dynamics. Latitude and longitude are among such features that determine the spatial coordinates of cyclone observations. Surface Latent Heat Flux (slhf) and Surface Sensible Heat Flux (sshf) measure the atmospheric-oceanic energy exchange responsible for cyclone generation and intensification. Top-of-Atmosphere Incident Solar Radiation (tisr) and Surface Solar Radiation (SSR) impact sea surface temperature variability, modulating cyclone energy availability. Evaporation (e) and Potential Evaporation (pev) reflect moisture availability, which in turn influences convective activity and precipitation intensity. The Large-Scale Precipitation Fraction (lspf) also estimates precipitation spatial coverage, critical for cyclone impact assessment.

A number of averaged parameters are also provided: Average Instantaneous Sensible Heat Flux (avg_ishf), Average Instantaneous Evaporation (avg_ie), Average Instantaneous Boundary Layer Dissipation (abg_ibld), Average Instantaneous Eastward Turbulent Surface Stress (avg_iews), and Average Instantaneous Northward Turbulent Surface Stress (avg_inss). These parameters control wind intensities and cyclone structure by modulating surface energy transfer and turbulent stress. Wind parameters are the 10m U-Component of Wind (u10), the 100m V-Component of Wind (v100), and the 10m V-Component of Neutral Wind (v10n), which represent the speed and direction of near-surface and upper-level winds. Atmospheric and cyclone intensity parameters are Dewpoint Temperature at 2m (d2m), Air Temperature at 2m (t2m), and Mean Sea Level Pressure (msl). Also, Total Precipitation (tp) and Skin Temperature (skt) are important in determining precipitation trends and surface warmth, both of which directly affect cyclone intensity and development.

Performed standard scaling on all three set of data thereby making the mean as 0 and standard deviation as 1. The training and validation datasets are used to train the machine learning model, and performance evaluation is done using the test dataset.

The prediction process involves a few main steps. It starts with the extraction of cyclone parameters from the ERA5 database and then preprocessing of time series data for analysis suitability. A sophisticated three-tier feature selection method is then used to discover the most appropriate features. The time-consistent data is then separated into training, testing, and validation sets in order to preserve the integrity of the time series format. Baseline machine learning models such as Random Forest, Support Vector Machine, and Decision Tree are subsequently trained on the pre-processed data. Throughout the process, the major concern is always to correctly predict surface pressure values while maintaining the temporal coherence of the data pipeline.

Table 4.1 – Comparative Model Analysis

Model	Train/Test	R^2 Score	MAE	MSE	RMSE
Linear	Train Validation	0.9089 0.8866	3.0877 3.3813	21.9219 27.2913	4.6821 5.2241
Lasso	Train Validation	0.8773 0.8264	3.4091 4.1669	29.5354 41.7841	5.4346 6.4641
Ridge	Train Validation	0.9089 0.8866	3.0879 3.3814	21.9220 27.2906	4.6821 5.2240
Bayesian Ridge	Train Validation	0.9089 0.8866	3.0877 3.3813	21.9219 27.2912	4.6821 5.2241
KNN	Train Validation	0.9980 0.9782	0.2785 1.0229	0.4896 5.2384	0.6997 2.2287
Decision Tree	Train Validation	1.0000 0.9890	0.0000 0.4247	0.0000 2.6461	0.0000 1.6267
Random Forest	Train Validation	0.9999 0.9926	0.0423 0.3620	0.0225 1.7891	0.1500 1.3375
AdaBoost	Train Validation	0.8969 0.8851	4.5715 4.7417	24.8178 27.6484	4.9817 5.2582
XGBoost	Train Validation Test	0.9951 0.9905 0.9931	0.5580 0.7482 0.6658	1.1740 2.2875 1.6521	1.0835 1.5124 1.2853
SVR	Train Validation	0.9953 0.9532	0.5551 3.6314	1.1471 4.7719	1.0710 2.1845

In Table 4.1, the evaluation metrics like R2 score, MAE, MSE, and RMSE for Linear, Lasso, Ridge and Bayesian Ridge were not giving expected accuracy compared to the rest. Whereas KNN, Adaboost, SVR and tree-based structure model like Decision Tree and Random Forest were overfitting because there were not generalized at all. While XGBoost Regressor performed best with the best generalization on train validation and unseen test data.

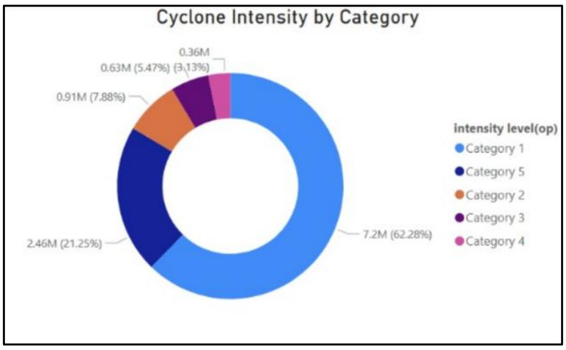

Figure 4.1 – Cyclone Intensity by Category

From Figure 4.1, there are 5 categories. Category 1 is most frequent (62.28%), indicating weaker cyclones are common, and Category 5 is 21.25%, indicating strong cyclones do occur with great frequency. Low-middle Categories 2 and 3 are less common, at 7.88% and 5.47%, respectively. Category 4 is least frequent at 3.13%, indicating strongest cyclones are least common.

This pattern shows that as weaker cyclones are prevalent, a significant percentage of the strong cyclones still remain dangerous. The unevenness of cyclone classes emphasizes the need for proper forecasting in order to reduce the risks related to intense storms. It also helps in developing improved early warning systems and response measures by understanding the evolution of cyclones through these classes.

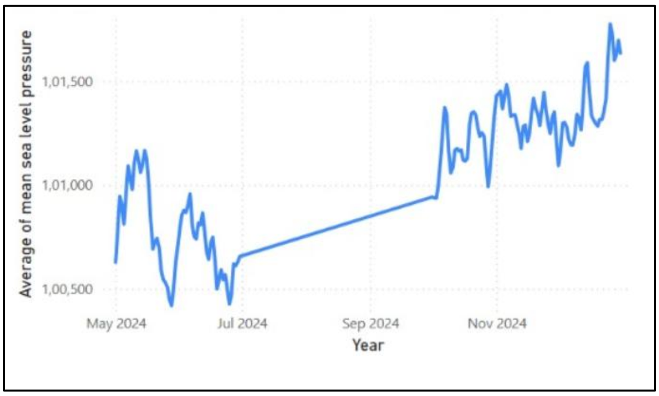

Figure 4.2 – Mean sea level pressure over the time

From Figure 4.2, we state that from month May to July and from October to December there is peak in mean sea level pressure which help to classify cyclones into categories.This seasonal fluctuation in mean sea level pressure is consistent with the principal cyclone seasons in the Indian Ocean, wherein pre- and post-monsoon seasons experience heightened storm activity. The pressure increase in these months reflects atmospheric conditions conducive to cyclone development and strengthening. Recognizing these peaks permits improved categorization of cyclones on the basis of their pressure characteristics. The trends witnessed in these months help deliver important details regarding cyclone activity, and these predictions can prove to be more accurate. These trends support the significance of monitoring pressure fluctuations to develop successful early warning systems.

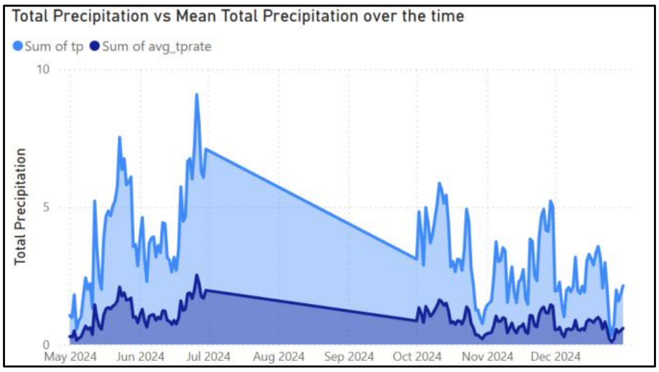

Figure 4.3 – Total Precipitation vs Mean Total Precipitation

Figure 4.3, depicts that spikes in precipitation indicate cyclone landfalls. Higher precipitation from May to July suggests an active cyclone or monsoon season, followed by fluctuation later in the year.These spates of precipitation coincide with seasonal monsoon patterns and cyclone activity near the coastlines. The augmentation of rainfall for these months demonstrates the influence of cyclones on interior flooding as well as storm surges. These fluctuations experienced toward the later months reflect fluctuating cyclone strengths and the dissipation of storms. Such monitoring of the trends in precipitation is important in order to refine disaster preparedness and response programs.

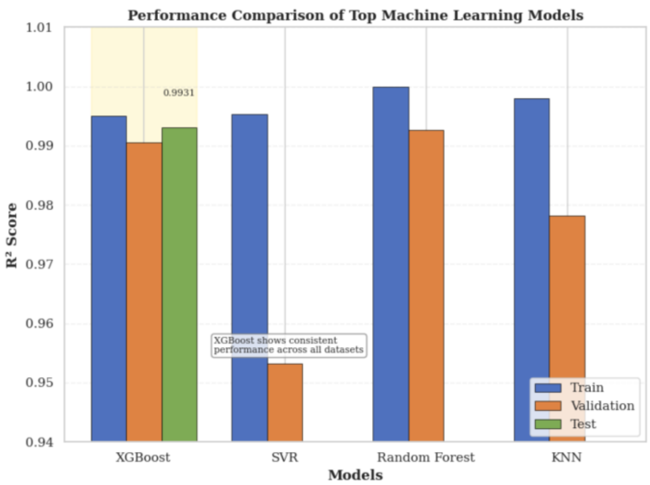

Figure 4.4 – Performance Comparison of Top Machine Learning Models

In Fig. 4.4, XGBoost beats the rest of the machine learning models as far as R^2 score is concerned for all datasets—training, validation, and test. It has a high and stable performance with the test score being around 0.9931, meaning that it has very good predictive accuracy and generalization ability. Unlike SVR, which has a large deterioration in validation performance, and KNN, which performs poorly on the test set, XGBoost has a well-balanced and stable performance. Random Forest works well too but just a bit behind XGBoost in consistency with datasets. Based on these findings, XGBoost was chosen for deployment owing to its high accuracy and high generalization over unseen samples, ranking it as the best and most reliable model among those tested.

V. Conclusion & Future Work

This study created several machine learning models to forecast surface pressure, which is used to predict cyclone intensity based on the ERA5 reanalysis dataset. A significant aspect of this approach was the use of a proprietary three-tier feature selection process, specifically fine-tuned to meet certain demands, which reduced over 40 features to the 22 most effective predictors. This step effectively eliminated redundant and lesser features, yielding very accurate forecasts.

Out of the different ML models that were experimented with, XGBoost Regressor performed the best with excellent generalization. The main aim was to create a model that not only has high predictive accuracy but also generalizes well, which would make it appropriate for sophisticated cyclone tracking and intensity estimation. Tree-based models showed overfitting characteristics, which implies poor generalization. The feature selection pipeline is still very robust, and repeated usage can be employed to identify the best features to extract for the dataset.

Future work involves extrapolating the ERA5 hourly dataset outside the Indian region to include global data. Moreover, implementing the models on cloud computing platforms and interfacing them into a web portal might facilitate researchers to engage with the study and compute surface pressure to help estimate cyclone intensity. Improvement could also be through researching deep learning models for learning long-term dependencies in cyclone dynamics.

References

[1] Accarino, Gabriele, Davide Donno, Francesco Immorlano, Donatello Elia, and Giovanni Aloisio. 2023. An Ensemble Machine Learning Approach for Tropical Cyclone Detection Using ERA5 Reanalysis Data. arXiv preprint arXiv:2306.07291. https://arxiv.org/abs/2306.07291.

[2] Chen, Rui, Ralf Toumi, Xinjie Shi, Xiang Wang, Yao Duan, and Weimin Zhang. 2023. "An Adaptive Learning Approach for Tropical Cyclone Intensity Correction." Remote Sensing 15 (22): 5341. https://doi.org/10.3390/rs15225341.

[3] Chen, Rui, Ralf Toumi, Xinjie Shi, Xiang Wang, Yao Duan, and Weimin Zhang. 2023. "An Adaptive Learning Approach for Tropical Cyclone Intensity Correction." Remote Sens. 15: 5341. https://doi.org/10.3390/rs15225341.

[4] Dongfang, Ma, et al. 2023. "Study on [title]." Environmental Research Letters 18: 014020. https://doi.org/10.1088/1748-9326/aca9e2.

[5] Giffard-Roisin, Samuel, Mingkai Yang, Guillaume Charpiat, Christine Kumler Bonfanti, Balázs Kégl, and Claire Monteleoni. 2020. "Tropical Cyclone Track Forecasting Using Fused Deep Learning From Aligned Reanalysis Data." Frontiers in Big Data 3: 1. https://doi.org/10.3389/fdata.2020.00001.

[6] Gorja, M. M. K., V. S. Gulakaram, N. K. Vissa, Y. Viswanadhapalli, and B. Tyagi. 2023. "Analysis of Large-Scale Environmental Features during Maximum Intensity of Tropical Cyclones Using Reanalysis Data." Atmosphere 14 (2): 333. https://doi.org/10.3390/atmos14020333.

[7] Kumar, Sandeep, Koushik Biswas, and Ashish Kumar Pandey. 2022. "Forecasting Formation of a Tropical Cyclone Using Reanalysis Data." arXiv preprint arXiv:2212.06149. https://arxiv.org/abs/2212.06149.

[8] Kumar, Sandeep, Koushik Biswas, and Ashish Kumar Pandey. 2021. "Predicting Landfall's Location and Time of a Tropical Cyclone Using

Reanalysis Data." In Artificial Neural Networks and Machine Learning – ICANN 2021: 30th International Conference on Artificial Neural Networks, Bratislava, Slovakia, September 14–17, 2021, Proceedings, Part IV 30, Springer International Publishing.

[9] Koteshwaramma, T., K. S. Singh, and S. Nayak. 2025. "The Performance of a High-Resolution WRF Modelling System in the Simulation of Severe Tropical Cyclones over the Bay of Bengal Using the IMDAA Regional Reanalysis Dataset." Climate 13 (1): 17. https://doi.org/10.3390/cli13010017.

[10] Lin, Jonathan, et al. 2023. "An Open‐Source, Physics‐Based, Tropical Cyclone Downscaling Model With Intensity‐Dependent Steering." Journal of Advances in Modeling Earth Systems 15 (11): e2023MS003686.

[11] Oguejiofor, Chibueze N., and Babatunde J. Abiodun. 2019. "Simulating the Influence of Sea-Surface-Temperature (SST) on Tropical Cyclones over South-West Indian Ocean, Using the UEMS-WRF Regional Climate Model." arXiv preprint arXiv:1906.08298. https://arxiv.org/abs/1906.08298.

[12] Slocum, Christopher J., Muhammad Naufal Razin, John A. Knaff, and Justin P. Stow. 2022. "Does ERA5 Mark a New Era for Resolving the Tropical Cyclone Environment?" Journal of Climate 35 (21): 7147–7164. https://doi.org/10.1175/JCLI-D-22-0127.1.

[13] Wang, Xinyu, et al. 2025. "Global Tropical Cyclone Intensity Forecasting with Multi-Modal Multi-Scale Causal Autoregressive Model." In ICASSP 2025–2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE.

[14] Xu, W., K. Balaguru, D. R. Judi, et al. 2024. "A North Atlantic Synthetic Tropical Cyclone Track, Intensity, and Rainfall Dataset." Scientific Data 11: 130. https://doi.org/10.1038/s41597-024-02952-7.

[15] Zhang, Y., et al. 2024. "Global Tropical Cyclone Size and Intensity Reconstruction Dataset for 1959–2022 Based on IBTrACS and ERA5 Data." Earth System Science Data 16: 5753–5780. https://essd.copernicus.org/articles/16/5753/2024/.