North Indian Ocean Tropical Cyclone Intensity Prediction using Machine Learning Classifiers and Analysis

Manish Mawatwal [(1)], Saurabh Das [(1)]
(1) DAASE, IIT Indore, India

Abstract – Tropical cyclone (TC) intensity estimation of North Indian Ocean (NIO) is quite challenging and requires huge amount of computational power and data collected over open oceans to improve efficiency for meteorologists. Best Track data of TCs collected from Indian Meteorological Department (IMD) is used for intensity prediction. This research compares different Machine Learning (ML) classifiers based on different predictors received from Indian Meteorological Department (IMD). The five predictors which are used for estimating the intensity of TCs are Latitude, Longitude, Central Pressure, Pressure Drop and Maximum Sustained Wind Speed (MSW). Random Forest achieved the best accuracy in prediction.

Keywords – North Indian Ocean, Tropical Cyclone Intensity, Machine Learning Classifier, Best Track Data.

1. Introduction

Focused, rapidly rotating cloud storms that form in the tropics are called TCs. TC causes a huge number of casualties and consequences when landing in coastal areas [1]. A complete study of TC cloud strength is therefore essential to predict its destructive power and loss mitigation. Many measures are used to estimate TC intensity around the world.

Common methods for TC strength analysis are Dvorak method [2-3] and extended versions of Dvorak method [4]. TC intensity prediction by Numerical Weather Prediction (NWP) models are one of the most important models for TC intensity estimation [5].

However, NWP models remain a challenge for predicting most TC intensities [6]. The domain expert therefore applies these scales to satellite imagery by applying their data-driven NWP model [7] to TC intensity.

Other TC intensity estimation approaches include multi-layer perceptrons, deep learning architectures, and other machine learning approaches have gained popularity during recent years to improve classification results [9].

Other methods for TC study include TC eye-based intensity prediction [8], TC image study with elliptic Fourier descriptors [10], TC image texture study [11] and various ML techniques [13], image contour extraction [15]. Artificial Neural Network [14], multilayer perceptron [12], Convolution Neural Network (CNN) [15], K-nearest Neighbors (KNN) [16].

The contribution of this research can be summarized as follows:

1. This research takes the best track data from IMD, of 97 TCs during 2001- 2020 over NIO containing five different predictors. There are 2965 instances of 39 TCs. A relative analysis is done to measure accuracy using various ML classifier
2. An analysis approach to understand the cycles and occurrence of TCs.

The proposed work and some brief idea about the TC predictors along with a brief discussion about the different ML classifiers used in the research for comparative study are shown in section 2. The dataset and pre-processing used in this research is described in section 3. In section 4, results are discussed with the conclusions of the work.

2. Proposed Work

The best track data of the past 20 years comes from the Regional Specialized Meteorological Center (RSMC) New Delhi, India. This data consists of TC name, the hourly instance of each TC (Latitude and Longitude, Central Pressure, Pressure Drop and MSW). The research work is shown in figure 1. Additionally, a brief description of the above predictors is provided below. Status of the cyclones is used as a label.

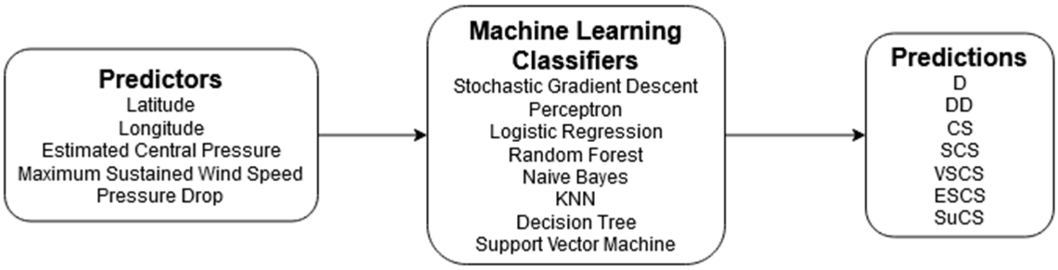

Figure 1. Diagram of the proposed work

This research is done to analyse the effect of predictors and used to estimate the TC intensity. The following predictors are collected from the best track data [24].

A. Latitude and Longitude of a TC
TC generally occur between 10 and 30 degrees of latitude, 60-to-70-degree longitude in the Arabian sea, 85-to-90-degree longitude in Bay of Bengal as shown in figure 7.

B. Estimated Central Pressure and Pressure Drop
It is calculated from wind speed [2]. Central pressure intensities help to estimate the destructive ability of a TC [17]. The wind and pressure relationship for a TC is calculated using equation 1.

$$S = C\sqrt{P_r - P_c} \qquad (1)$$

S is the sustained maximum wind speed,

C is the empirical parameter that varies between 0 and 1.

$P_r - P_c$ is the pressure drop

C. MSW
It can be calculated using central pressure, as seen in equation 1. Earlier the MSW is assessed depending on the cloud patterns of satellite imagery [18]. Finally, in the NIO basin equation 2 represents the wind and pressure drop relation [19].

$$S = 16\sqrt{P_r - P_c} \qquad (2)$$

Eventually, equation (2) was modified [20]:

$$S = 14.2\sqrt{P_r - P_c} \qquad (3)$$

The common ML classifiers used in this research is given below:

A. Naïve Bayes (NB)
NB algorithm is a classifier based on Bayes Theorem. It assumes that every pair of features being classified is independent of each other.

B. Logistic Regression (LR)
This algorithm computes the probabilities of class membership for each class. It is a supervised learning classification algorithm which is used to predict observations to a discrete set of classes, also called Logit Regression. The LR parameters used in the research are C: 10, solver: 'liblinear'.

C. Perceptron
An artificial neuron. It is the simplest possible neural network. It is used as an algorithm or a linear classifier to facilitate supervised learning of binary classifiers [21].

D. Random Forest (RF)
RF is a meta estimator with multiple decision tree which is commonly used for classification purpose.

E. Decision Trees (DT)
A type of Supervised ML where the data is continuously split according to a certain parameter [23]. The DT depth was kept 3 for this research.

F. KNN
It is a non-parametric method used for classification. It is also one of the best-known classification algorithms. The value of 'n' is taken as 3 in this research.

G. Support Vector Machine
Supervised Learning algorithm, which is used for classification as well as regression problems [22]. Support vectors are the data points that lie closest to the decision surface. SVM parameters used in the research are C: 1000, kernel: 'rbf'. Other kernels did not provide better results.

H. Stochastic Gradient Descent

Optimization algorithm often used in ML applications to predict the model parameters that correspond to the best fit between predicted and actual outputs.

3. Dataset and Processing

A temporal resolution of six/three-hour data of TCs from 2001 to 2020 are collected from the RSMC, India. Seven classes of TCs over NIO have been observed according to the IMD TC intensity scale as shown in the table 1 [25]. The training and testing dataset is divided into the ratio 8:2 using python package ‘sklearn’. ‘Standard Scaler’ is used to remove mean and scale all the features to unit variance. A ten-fold cross validation is used to ensure average and correct accuracy of the ML classifiers.

Table 1. IMD TC Intensity Scale

Depression (D)	17-27 knots
Deep Depression (DD)	28-33 knots
Cyclonic Storm (CS)	34-47 knots
Severe Cyclonic Storm (SCS)	48-63 knots
Very Severe Cyclonic Storm (VSCS)	64-89 knots
Extremely Severe Cyclonic Storm (ESCS)	90-119 knots
Super Cyclonic Storm (SuCS)	120+ knots

Here, 97 TC were extracted for the research. The distribution of count of these TC are given in figure 2. The average wind speeds of cyclones over the years in the NIO is given in figure 3. The scatterplot of cyclonic regions in the NIO is given in figure 4. Figure 5 shows the wind speed distribution over the complete dataset between 2001 to 2020. The outliers in the maximum wind speed are shown in figure 6. Figure 7 shows the distribution of cyclones over latitude and longitude in the NIO. Most number of cyclones pass between 12-to-15-degree latitude, 85-to-90-degree longitude. There are two maxima in the longitude which shows the cyclones occurring in the Arabian Sea on the left and Bay of Bengal on the right. Figure 8 shows the different cyclones occurring over different timestamps. Table 2 shows that most of TC occurrence over different timestamps.

Figure 2. Distribution of year wise cyclonic data over NIO

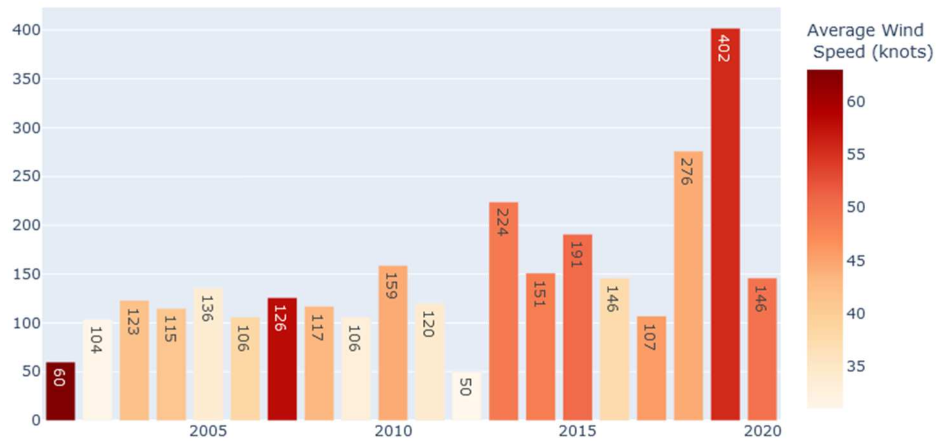

Figure 3. Average Wind Speed of Large storms in NIO (2001-2020)

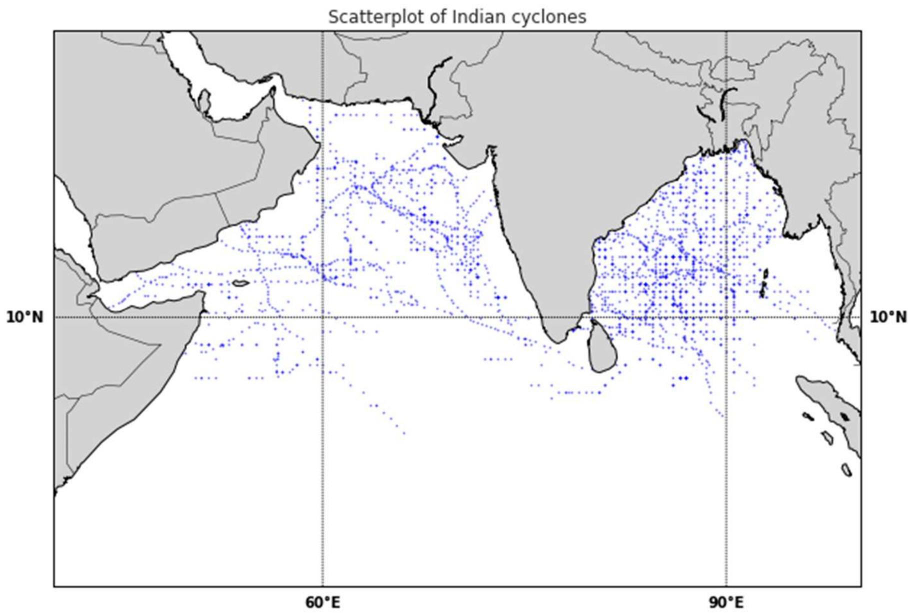

Figure 4. Scatterplot of NIO cyclones

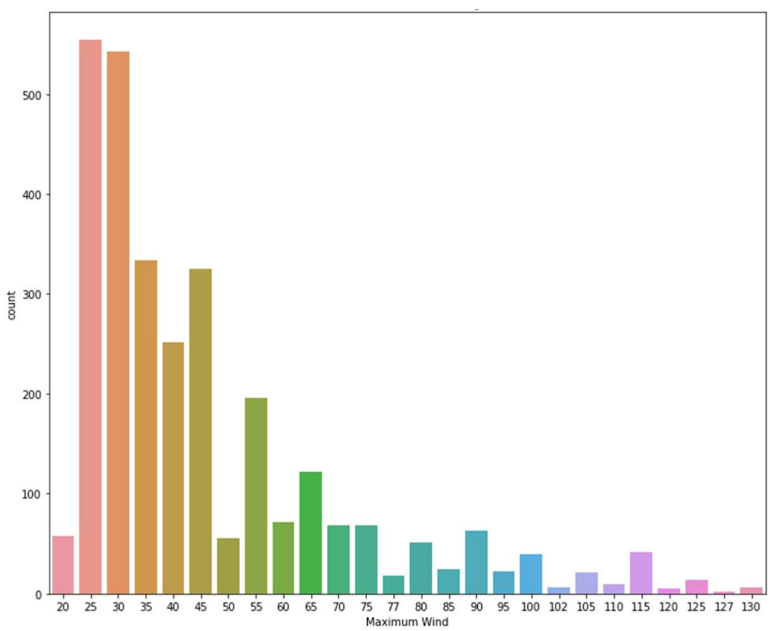

Figure 5. Maximum wind distribtuion

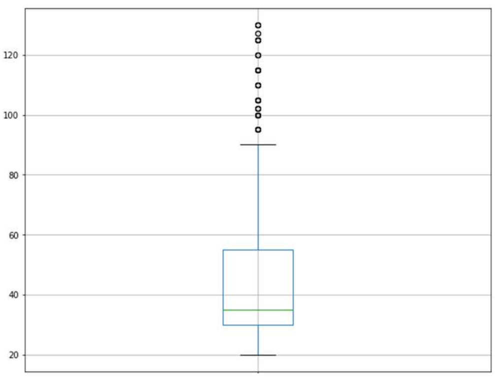

Figure 6. Boxplot of Maximum Wind Speed

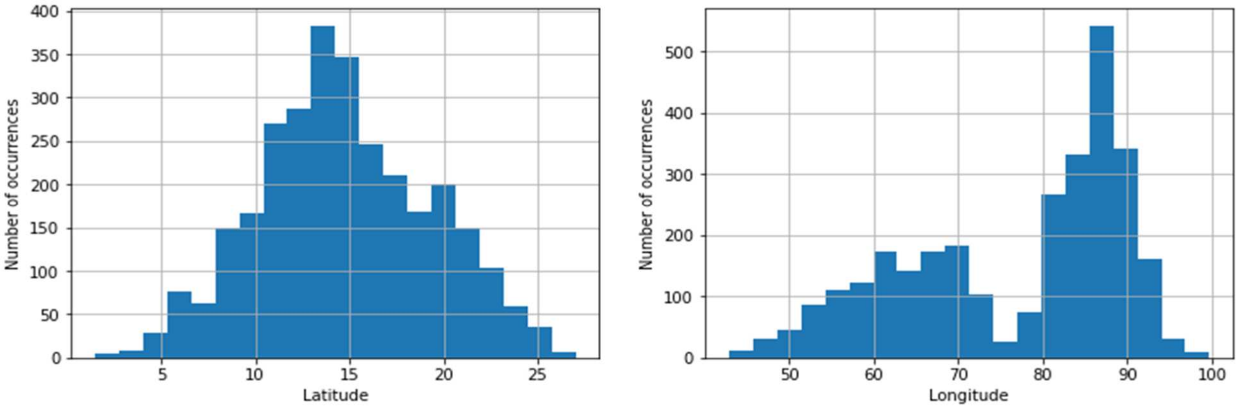

Figure 7. Number of occurrences of Latitude and Longitude

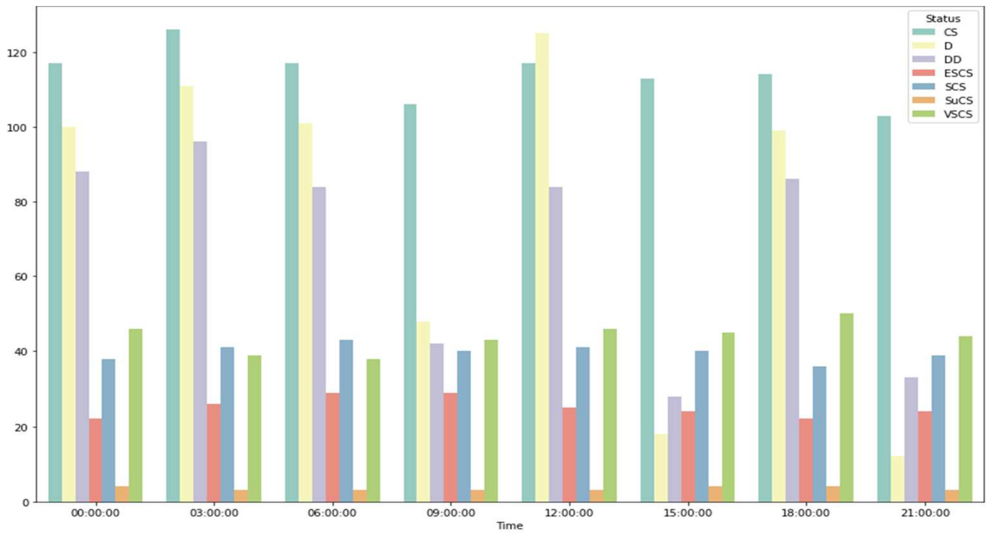

Figure 8. TC count over different time stamps

Table 2. Frequency distribution of data points across different timestamps

Time stamp	Number of data points of TC
03:00:00	442
12:00:00	441
00:00:00	415
06:00:00	415
18:00:00	411
09:00:00	311
15:00:00	272
21:00:00	258

4. Result and Discussion

Different ML classifiers are used for finding best performing classifier among them. The correlation matrix of all predictors with respect to maximum wind speed is shown in the figure 9. Predictors over the NIO are used for classification. Predictors like MSW etc over the NIO are used for TC classification. Performance of different ML classifiers is seen in figure 10. Both the testing and training accuracy is calculated to understand overfitting and underfitting.

This research was performed over NIO with the best track data available on IMD. The best track data of 97 individual storms with 2965 instances are taken for the research. The obtained result show that the TC predictors such as MSW and pressure drop strongly affect the accuracy of the classification. On the other hand, latitude, longitude do not affect the accuracy of the classification. One limitation of this research is that the predictive ML models presented here uses only the best track data. In the future, other predictors like sea surface temperature etc can be used to determine the robustness of the ML classifiers. Random Forest achieved the highest accuracy score of 99.16. The code for the same is published on GitHub https://github.com/manishmawatwal/conference_paper.

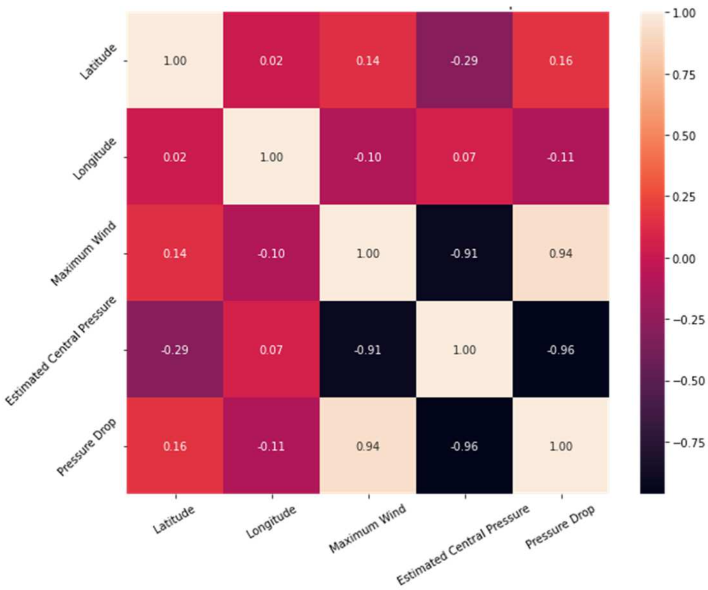

Figure 9. Correlation of predictors with Maximum Wind Speed

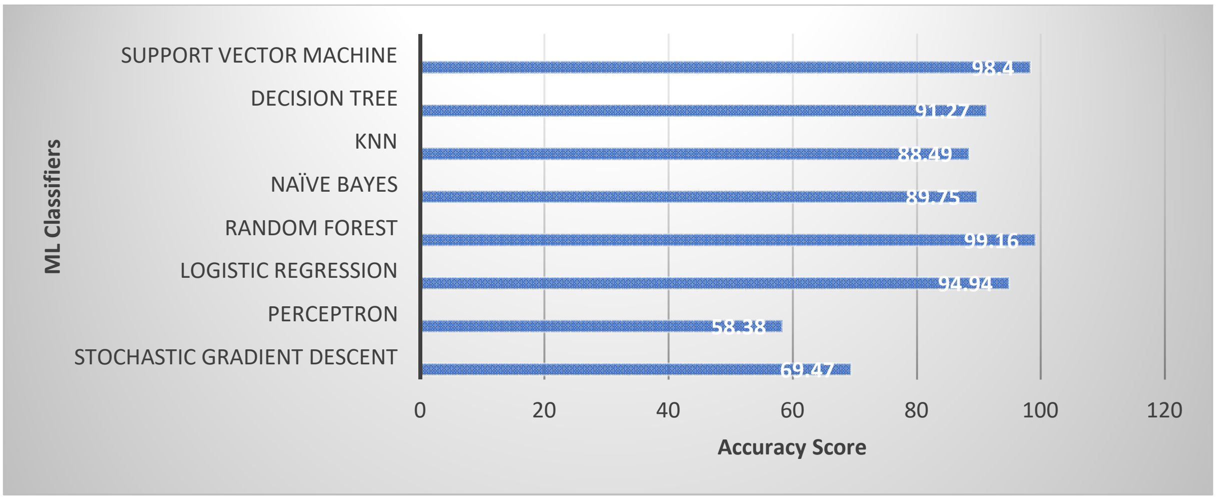

Figure 10. K fold cross validation score of different ML classifiers

5. Acknowledgment

The authors extend their thanks to IMD and RSMC for the best track data provided on their website.

6. References

[1] S. K. Dube, I. Jain, A.D. Rao, T.S. Murty, "Storm surge modeling for the Bay of Bengal and Arabian Sea", *Natural Hazards*, vol. 51, pp.3-27, 2009, https://doi.org/10.1007/s11069-009-9397-9.

[2] V. F. Dvorak, "Tropical Cyclone Intensity analysis and Forecasting from Satellite Imagery", *Mon. Weather. Re.*, vol. 103, pp. 420–430, 1975, https://doi.org/10.1175/1520-0493(1975)103.

[3] V. F. Dvorak, "Tropical cyclone intensity analysis using satellite data", US Department of Commerce, National Oceanic and Atmospheric Administration, National Environmental Satellite, Data, and Information Service: Washington, DC, USA, vol. 11, pp. 1–47, https://repository.library.noaa.gov/view/noaa/19322.

[4] T. L. Olander and C. S. Velden, “The Advanced Dvorak Technique (ADT) for Estimating Tropical Cyclone Intensity: Update and New Capabilities”, *Wea. Forecasting*, vol. 34, pp. 905–922, 2019, https://doi.org/10.1175/WAF-D-19-0007.1

[5] Y. L, Tsai, T. R Wu, C. Y. Lin, S.C. Lin, E. Yen, C. W. Lin "Discrepancies on Storm Surge Predictions by Parametric Wind Model and Numerical Weather Prediction Model in a Semi-Enclosed Bay: Case Study of Typhoon Haiyan", *Water*, vol. 12, no. 12, pp. 1-27, 2020, doi: 10.3390/w12123326.

[6] D. Burton, L. Bernardet, G. Faure, D. Herndon, J. Kna, Y. Li, J. Mayers, F. Radjab, C. Sampson, A. Waqaicelua, “Structure and intensity change: Operational guidance”, *In Proceedings of the 7th International Workshop on Tropical Cyclones*, La Réunion, France, 15–20 November 2010.

[7] M. Yamaguchi, J. Ishida, H. Sato, and M. Nakagawa, “WGNE Intercomparison of Tropical Cyclone Forecasts by Operational NWP Models: A Quarter Century and Beyond”, *Bulletin of the American Meteorological Society*, vol. 98, no. 11, pp. 2337-2349, 2017.

[8] J.S. Griffin, R.W. Burpee, F. D. Marks, J. L. Franklin, “Real Time airborne analysis of aircraft data supporting operational hurricane forecasting”, *Weather and Forecasting*, Vol. 7, pp. 480-490, 1992, doi: 10.1175/1520-0434(1992)007.

[9] R. S. T. Lee, J. N. K. Lin, “An elastic contour matching model for tropical cyclone pattern recognition," *IEEE T Syst Man Cy B*. vol. 31, no. 3, pp. 413-417, 2001, doi: 10.1109/3477.931532.

[10] I. Dutta, S. Banerjee, “Elliptic Fourier Descriptors in the Study of Cyclone Cloud Intensity Patterns”, *International Journal of Image Processing*, vol. 7, np. 4, pp. 402-417, 2013.

[11] C. Kar, S. Banerjee, “Intensity prediction of tropical cyclone using multilayer multi-block local binary pattern and tree-based classifiers over North Indian Ocean”, *Computers & Geosciences,* vol. 154, pp. 1- 14, 2021.

[12] C. Kar, A. Kumar, S. Banerjee, “Tropical cyclone intensity detection by geometric features of cyclone images and multilayer perceptron”, *SN Applied Sciences*, vol. 1, no. 1099, 2019, doi: 10.1007/s42452-019-1134-8.

[13] M. Kim, M. S. Park, J. Im, S. Park, M. I. Lee, “Machine Learning Approaches for Detecting Tropical Cyclone Formation Using Satellite Data”, *Remote Sensing*, vol. 11, no. 10, pp. 1-19, 2019, doi: 10.3390/rs11101195.

[14] A. Kulkarni, R. Bankert, M. Hadjimichael, "Tropical cyclone intensity estimation using neural networks." In *ASPRS Annual Conference*, San Diego, 2010.

[15] J. S. Combinido, J.R. Mendoza, J. A. Aborot, “Convolutional Neural Network Approach for Estimating Tropical Cyclone Intensity Using Satellite-based Infrared Images”, In Proceedings of the *2018 24th ICPR*, Beijing, 2018, pp. 20–24.

[16] R. L. Bankert, P.M. Tag, “An automated method to estimate tropical cyclone intensity using SSM/I imagery”, *J. Appl. Meteorol.*, vol. 45, no. 1, pp. 461, 2002, http://www.jstor.org/stable/26184990.

[17] J. A. Knaff, R. M. Zehr, “Reexamination of Tropical Cyclone Wind– Pressure Relationships”, *Weather and Forecasting*, vol. 22, no. 1, pp. 71-88.

[18] S. Fritz, L. F. Hubert, A. Timchalk, “Some inferences from satellite pictures of tropical disturbances”, *Mon Weather Rev*, vol. 94, pp. 231– 236, 1966.

[19] R. D. Fletcher, “Computation of maximum surface winds in hurricanes”, *Bull Am Meterol Soc* vol. 36, pp. 247–250, 1955, doi: 10.1175/1520-0477-36.6.247.

[20] D. K. Mishra, G. R. Gupta, “Estimation of maximum wind speed in tropical cyclones occurring in Indian Seas”, *Ind J Meteorol Hydrol Geophys*, vol. 27, pp. 285–290, 1976.

[21] S. Mitra, S. K. Pal, “Fuzzy multilayer perceptron, inferencing and rule generation”, *IEEE Transactions on Neural Networks*, vol. 6, no. 1, pp. 51-63, 1995, doi: 10.1109/72.363450.

[22] Cao, L. J., Keerthi, S. S., Ong, C. J., Zhang, J. Q., Periyathamby, U., Fu, X. J., & Lee, H. P. "Parallel sequential minimal optimization for the training of support vector machines", *IEEE Trans. Neural Networks*, vol. 17, no. 4, pp. 1039-1049, 2006, doi: 10.1109/TNN.2006.875989 .

[23] J. R. Quinlan, "C4.5: Programs for machine learning", In *Morgan Kaufmann*, San Francisco, 1993, https://doi.org/10.1007/BF00993309.

[24] Chinmoy Kar, Sreeparna Banerjee, "Tropical Cyclone Intensity Prediction Using Best Track Data Over North Indian Ocean by Machine Learning Classifiers", 2021 IEEE International India Geoscience and Remote Sensing Symposium (InGARSS), 2021, doi: 10.1109/InGARSS51564.2021.9792071.

[25] Rizwan Ahmed, M. Mohapatra, Ram Kumar Giri, Suneet Dwivedi, "An evaluation of the Advanced Dvorak Technique for the Tropical Cyclones over the North Indian Ocean", Tropical Cyclone Research and Review, 2021, doi: 10.1016/j.tcrr.2021.11.003.