Journal of the Indian Society of Remote Sensing (October 2024) 52(10):2165–2175
https://doi.org/10.1007/s12524-024-01929-8

RESEARCH ARTICLE

# An End-to-End Deep Learning Framework for Cyclone Intensity Estimation in North Indian Ocean Region Using Satellite Imagery

**Manish Kumar Mawatwal**[1] · **Saurabh Das**[1]

Received: 4 May 2023 / Accepted: 10 June 2024 / Published online: 5 July 2024


**Abstract**
Prediction of Tropical cyclones (TCs), particularly intensity prediction, has always been challenging for climate researchers due to the complicated physical mechanisms in TC dynamics and the way it interacts with upper-ocean and atmospheric circulation. Furthermore, the available data set over the North Indian Ocean (NIO) is also very limited for Machine Learning (ML) model development. Here, we demonstrated a simple yet robust hybrid architecture leveraging a Convolutional Neural Network for automated prediction of the intensity of the cyclone based on IR satellite imagery of 2000–2022. The model comprises a binary classifier, a multiclass classifier, a YOLOv3 based cyclone detector and a regression module. The paper also highlights the discrepancy between the results of independent testing wherein training is done on 2000 to 2019 dataset and tested on 2020 to 2022 dataset, as well as the outcomes of a stratified train-test split performed over the entire dataset using a 70:15:15 ratio for training, validation and testing, respectively. The model is tuned for the NIO region with a binary classification accuracy score of 98.4% (±0.003), multiclass classification accuracy of 63.83% (±1.3) and RMSE of 16.2 (±0.9) knots on stratified split. The results highlight the careful interpretation of the DL model's performance when applied to time series problems. Additionally, it discusses the limitations stemming from the dataset's small size and the challenges posed by the 5 kt resolution of the best track intensity estimation from the Indian Meteorological Department (IMD). The internal representations learned by the model through feature maps analysis were studied, shedding light on the model's decision-making process. The study underscores the need for further data accumulation and highlights avenues for enhancing model performance in the future.

**Keywords** Convolutional neural network (CNN) · Object detection · Deep learning (DL) · Remote sensing

## Introduction

An average of 90 Tropical Cyclones (TCs) form annually over the tropical ocean waters (Emanuel & Nolan., 2004). TC causes significant socio-economic damage to human life and assets, and the damage is directly correlated with the intensity of the cyclone (Zhang et al., 2009). Intensity predictions are, however, particularly very challenging due to the complicated physical mechanisms in tropical cyclone dynamics and the way they interact with upper-ocean and atmospheric circulation (Devaraj et al., 2021). As cyclones predominantly form over expansive water bodies where weather stations are scarce, meteorologists frequently face the challenge of approximating the wind speed of cyclones (Chandan et al., 2012). Understanding North Indian Ocean (NIO) cyclone occurrences is crucial due to the occurrence of extreme and rare events like tropical Super-cyclone "Shaheen", which took an unusual trajectory originating from cyclone "Gulab" remnants, making a rare landfall along the coast of Oman during the active southwest monsoon period in October 2021 (Meer et al., 2024).

Presently, data is sourced from varied types of observations and act as a boundary condition for the Numerical Weather Prediction (NWP) models. A limitation of the NWP model is that as the prediction period increases, the computation time increases to handle the interpolation of large amounts of data, and the errors continue to accumulate until new observations are made. Further, upper ocean feedback has important effects on TCs, but at present, only

✉ Manish Kumar Mawatwal
mawatwalmanish1997@gmail.com

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

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

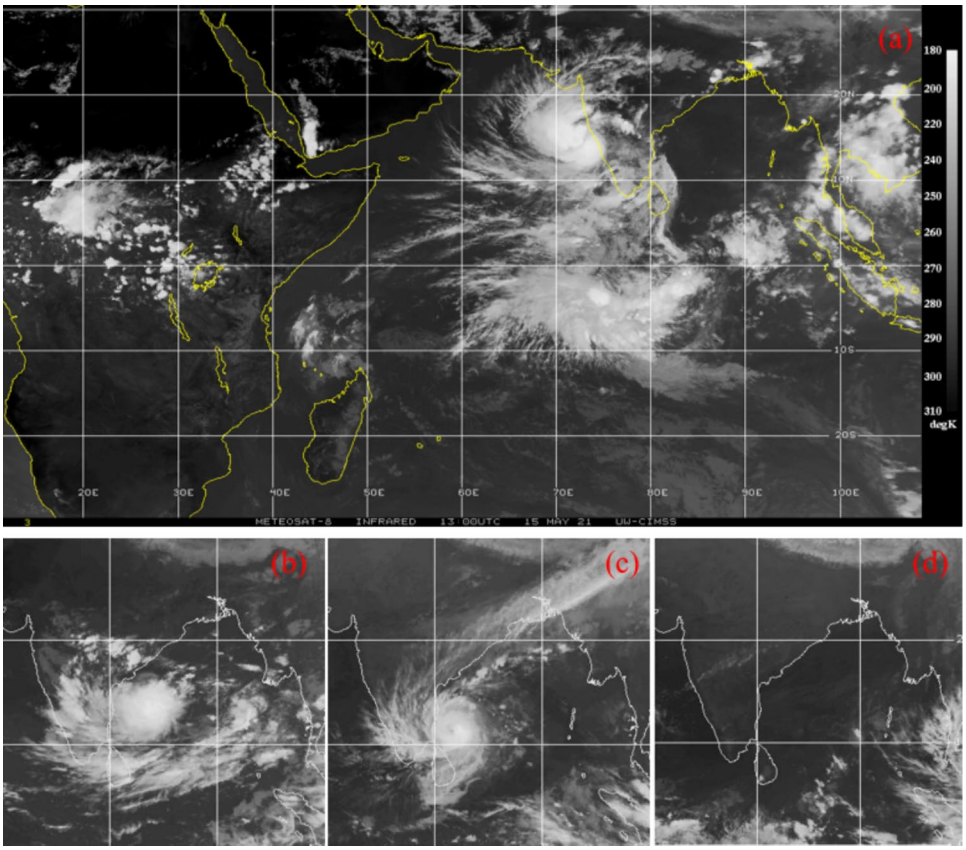


**Fig. 1** **a** Raw image from CIMSS dataset **b** Depression and **c** Very Severe Cyclone, and **d** shows an image with no cyclone

a few operational numerical forecast models take it into consideration (Moon et al., 2007).

As in-situ measurements over large spans of the ocean or remote areas of the land surface are difficult, data from satellite observations are also used as the alternate primary source of TC information. Satellites measure temperature, clouds, water vapor, and precipitation through active or passive mode and can be indirectly used for estimating TC intensity (Olander & Velden., 2016).

Dvorak technique estimated the intensity via human interpretation of the cyclone shape when direct measurements were not available. Techniques vary depending on the length and curvature of the storm bands; the strength is estimated by capturing the relationship between the features (Lee et al., 2021).

While the spatial patterns in infrared satellite imagery strongly relate to TC intensity, researchers found that parts of the Dvorak technique are subjective, leading to two different intensity estimates, decreasing the efficiency of the process. Advanced Dvorak Technique uses passive microwave data and measurements from aircraft to estimate the intensity, but the performance was worse for weaker storms (Olander & Velden., 2019; Ritchie & Valliere., 2012).

The deviation angle variation technique requires infrared satellite imagery and computes the gradient vector, the variance is used to estimate the intensity values. The constraint being the centre of TC images should be clearly marked making it difficult to tune parameters across multiple regions (Pineros et al., 2011).

Statistical models are also, found to perform poorly in handling complex and nonlinear relationships between TC-related predictors (Lee et al., 2015). In recent times,

**Table 1** Data split

| | Binary classification | Multiclass classification | Regression |
|---|---|---|---|
| Training | 1154 | 1316 | 1308 |
| Validation | 247 | 280 | 348 |
| Testing | 249 | 288 | 281 |
| Total | 1650 | 1884 | 1869 |

**Table 2** Category of cyclones (multiclass classification)

| Category | Wind speed | Actual images |
|---|---|---|
| Depression (D) | 20–25 knots | 352 |
| Deep depression (DD) | 30–35 knots | 534 |
| Cyclonic storm (CS) | 40–50 knots | 396 |
| Severe cyclonic storm (SevereCS) | 55–65 knots | 257 |
| Very severe cyclonic storm (VSCS) | 70 + knots | 305 |

Machine and Deep Learning based (ML/DL) models were also attempted to predict the cyclone intensity. (Sharma et al., 2013) has applied neural networks for this purpose to the cyclones over the Western North Pacific and (Chaudhari et al., 2013) on the NIO cyclones. An ensemble base Convolutional Neural Network (CNN) classifier on Outgoing Longwave Radiation (OLR) for classifying TC is done in (Matsuoka et al., 2018). Four state-of-the-art U-net models were developed in (Bonfanti et al., 2020) for the detection of Regions of Interest (ROI) for cyclones. There are attempts to estimate TC intensity from satellite data by using CNN in recent times (Tian et al., 2020), (Wang et al., 2021). Mask R-CNN and DenseNet architecture is also used for binary cyclone classification (Nair et al., 2022). The frequency of storms has decreased since 1960s, but their intensity has increased (Bandyopadhyay et al., 2021) which makes it more important to predict cyclones. (Kar et al., 2021) used feature extraction from cyclones to classify TC images. The emergence of DL and advancements in Computer Vision present an innovative and potent toolkit for addressing the challenging task of precisely determining the longitude and latitude of the cyclone centre within a 2D input image (Chen et al., 2019; Mooers et al., 2023).

Most of these early works were carried over the Atlantic basin, and there is a serious dearth of satisfactory models over the NIO region. The NIO, including the Arabian Sea (~24%) and Bay of Bengal (~76%), accounts for about 6% of the global TC and the cyclones affect a very large number of populations each year in surrounding regions. The major reason for limited performance of the DL models in this region is due to limited amounts of past data and complex tropical/equatorial weather phenomena.

The objective of this paper is to develop a simple yet effective end-to-end framework for estimating the intensity of TC directly from satellite images over the NIO region. For this purpose, the problem statement is divided into four steps:

- Binary classification: Does the satellite imagery contain any cyclone?
- Object detection: Identify the cyclone and extract.
- Multi-class classification: In which class (as per Dvorak) it belongs to?
- Regression: What is the expected intensity of the cyclone?

## Data and Pre-processing

### Data

Long wave infrared images are taken from CIMSS Tropical Data Archive (University of Wisconsin – CIMSS, https://tropic.ssec.wisc.edu/.) captured by Meteosat-5/7/8 for the period 2000–2022. The temporal resolution is 3 h/6 h and the image has a 10 km spatial resolution. Best track data of NIO region is taken from the IMD (IMD Cyclone Best Track Data, https://mausam.imd.gov.in/.) (RSMC Delhi) for labelling the intensity of the images. The best track cyclone intensity is provided at a resolution of 5 kt.

### Pre-processing

The images of cyclones in NIO are first cropped to an average dimension of 310×310 manually (average aspect ratio is maintained at 1) in a way that they don't contain cyclone eye in the centre, which makes the model more robust, as the region contains non-cyclonic parts as well. All the CNN models are fed with 310×310 image size. The reduction of image size is done only to reduce the processing time. The images are then converted into grayscale. The images are pre-processed such that all the pixels below 128 threshold value were converted to 0 to reduce noise from the images and improve cyclone classification. Data Augmentation by a rotation range of 10 degrees has been used to increase the number of samples for training. Data Augmentation was done only on the training dataset. Figure 1a shows the raw sample image before pre-processing, b Depression and c Very Severe Cyclone, and d shows an image with no cyclone.

**Table 3** Hyperparameter tuning

| | Binary classifier | Multiclass classifier | Regression |
|---|---|---|---|
| Epochs | 30 | 50 | 40 |
| Optimiser | Adam | Adam | Adam |
| Loss function | Binary cross-entropy | Categorical cross-entropy | Mean squared error (MSE) |
| Initial learning rate | 0.0001 | 0.0001 | 0.001 |
| Training/validation batch size | 2/1 | 2/1 | 2/1 |

**Table 4** Proposed CNN model

| Layer | Filters/dropout value/units/output shape | Kernel size/ pool size | Activation |
|---|---|---|---|
| Convolution | 32 (308, 308, 32) | 3,3 | ReLU |
| Max pooling | (154, 154, 32) | 2,2 | |
| Dropout | 0.4 | | |
| Convolution | 64 (152, 152, 64) | 3,3 | ReLU |
| Max pooling | (76, 76, 64) | 2,2 | |
| Dropout | 0.4 | | |
| Convolution | 128 (74, 74, 128) | 3,3 | ReLU |
| Max pooling | (37, 37, 128) | 2,2 | |
| Dropout | 0.4 | | |
| Convolution | 128 (35, 35, 128) | 3,3 | ReLU |
| Max pooling | (17, 17, 128) | 2,2 | |
| Dropout | 0.4 | | |
| Flatten | | | |
| Dropout | 0.4 | | |
| Dense | 512 | | ReLU |
| Dropout | 0.4 | | |
| Dense | 1(BC)[a] (R)[a], 5 (MC)[a] | | Sigmoid (BC)[a], SoftMax (MC)[a], Linear(R)[a] |

Terms—*BC* Binary Classification, *MC* Multiclass Classification, *R* Regression

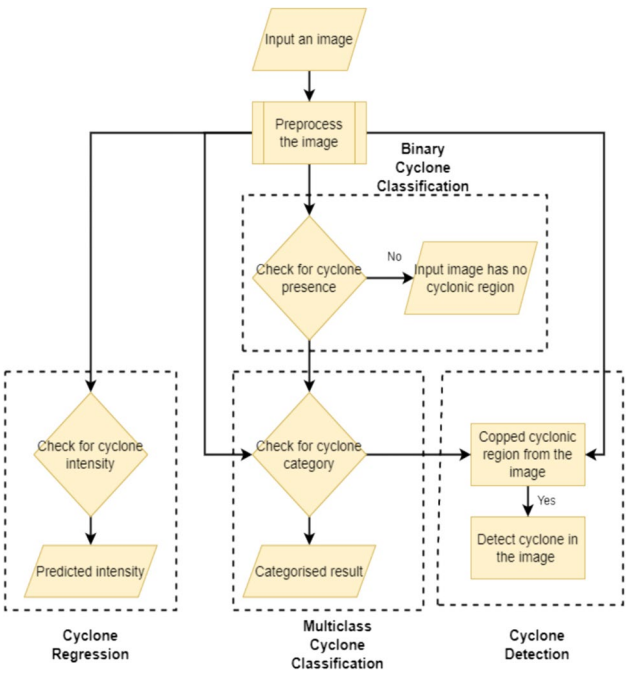


**Fig. 2** Flowchart of the implemented pipeline

## Training and Testing

Since we have a very limited dataset, the images have been split into training, validation and test datasets in the ratio 0.7:0.15:0.15 maintaining the ratios of different cyclones. However, it should be noted that as the cyclone images in a few hours doesn't change much, there is always a possibility of leakage of the data between training and testing samples. To avoid this, the evaluation is also carried separately with a testing dataset of 2020–22 while keeping 2000–19 as the training dataset. This, however, doesn't guarantee the proper representation of all the cyclone classes. Table 1 shows the data split of images for different models.

For multiclass classification and regression analysis, the images were taken from 2000 to 2022. Actual 1884 images are divided into 5 different classes with their corresponding wind speeds as shown in Table 2 for multiclass classification. Very high cyclone intensity ($\geq 120$) images are not used in the regression problem as the number of samples are very low (15 samples only).

For binary cyclone classification cyclone images from the year 2000 to 2022 were considered. For class category 'no', the images should not have any cyclones, the images are taken from different months across the years, carefully checking that the cyclone doesn't occur during that period. For class category 'yes', the higher intensity images (intensity $\geq 40$ knots) from multiclass classification dataset are used. Number of images used in class 'without cyclones' is 708 and 'with cyclones' is 942. The non-cyclonic images do have cloud patches.

## Methodology

Understanding the accuracy and reliability of satellite data is paramount in any remote sensing application, particularly in the context of cyclone detection and intensity prediction. While satellite technology offers invaluable insights into weather patterns and atmospheric conditions, it's essential to acknowledge and address potential sources of error that may affect the quality of the data. One significant consideration revolves around the inherent limitations and anomalies associated with satellite sensors, particularly those of the Meteosat first generation. These sensors, deployed aboard satellites like Meteosat-5 and Meteosat-7, are instrumental in capturing critical imagery used in cyclone monitoring. However, they are susceptible to various factors that can introduce errors into the data.

The errors related to Meteosat first generation of satellites include solar stray light effects in the water vapor (WV) and infrared (IR) channels of the Meteosat-5 and Meteosat-7 satellites. These effects occur when solar light enters the radiometer around local midnight and creates anomalies in the

**Table 5** Comparative study

| Refs. | Dataset, model used | Accuracy | Limitation |
|---|---|---|---|
| Nair et al., 2022 | 816 images from EUMETSAT, Deep Neural Net and Mask R-CNN | 86.55% accuracy on Cyclone Detection (Mask R-CNN)<br>Binary Classification: Recall score 76.14% | Only two classes were classified with small dataset |
| Shakya et al., 2020 | 995 images of cyclone from IMD, Deep Neural Net | 97% accuracy on Binary Classification, 84% confidence on Cyclone Detection (YOLO), 58% accuracy on Multiclass Classification | Small dataset, heavy pre-processing, only three categories of cyclone (SevereCS, CS, SuperCS) were classified |
| Kumawat & Jaiswal, 2021 | 931 images from IMD (Visual and Infrared), CNN and YOLO | 96% accuracy on Binary Classification, 85% confidence on Cyclone Detection (YOLO) | Binary classification with small dataset |
| Kar et al., 2019 | 379 images from IMD, Multilayer Perceptron | 84% accuracy on 3 classes of cyclone | Less dataset, Considered only three categories of cyclone (D, CS, ESCS) |
| Proposed | 1884 (additional 708 images for 'no' category) images from CIMSS, Best Track data from IMD | 98.4% accuracy score on Binary classification, 63.83% on multiclass classification (five categories), 16.2 knots on regression and YOLO cyclone detection | Manual cropping, imbalanced dataset |

**Table 6** Evaluation metrics for multiclass classification

| Category | Precision | Recall | F1-score |
|---|---|---|---|
| D (20–25) | 0.70 | 0.64 | 0.67 |
| DD (30–35) | 0.62 | 0.72 | 0.67 |
| CS (40–50) | 0.54 | 0.55 | 0.54 |
| SevereCS (55–65) | 0.59 | 0.59 | 0.59 |
| VSCS (70 +) | 0.76 | 0.60 | 0.67 |

radiances observed by the satellites. The anomalies include warm bows, intense warm spots, warm and cold horizontal stripes, and calibration anomalies (Liefhebber et al., 2020). These effects have been known since the early years of the Meteosat program. The amplitude of the anomalies can range from 2.4 K to over 10–25 K, depending on the time of year and proximity to the eclipse seasons. The anomalies get more pronounced due to the rapid decrease of radiated solar energy. The anomalies affect different time slots, with 0100 Coordinated Universal Time (UTC) being consistently affected throughout the year. Other time slots, such as 2300, 0000, 0200, and 0300 UTC, are affected to a different extent and at different times (Koepken, 2004).

In this research, high-quality images are taken for training the model and thresholding techniques are implemented to mitigate noise and improving the accuracy of cyclone detection. By carefully curating training data that minimize the presence of anomalies caused by solar stray light effects, the model can learn to better distinguish genuine cyclone features from spurious artifacts. Additionally, thresholding methods helped in identifying and removing noisy pixels, thus enhancing the clarity and reliability of the infrared imagery utilized by the model.

In this research, four different models are evaluated. As already mentioned, it involves four different steps: first, the input image is given to a binary classifier, second, the cyclone is detected and cropped using pre-trained weights on YOLOv3 architecture, third the image is given to a multiclass-classifier and fourth, an intensity predictor estimates the intensity of the image.

CNN is a network architecture for deep learning which learns directly from data, eliminating the need for manual feature extraction. Early stopping, Dropouts, Kernel Regularizers and Learning rate scheduler was used to prevent overfitting the model. Early stopping halts the training if the validation accuracy does not improve for a specified number of epochs (in our case, 10). Furthermore, a 'ReduceLROnPlateau' callback is utilized to dynamically adjust the learning rate by a factor of 0.1, if the validation loss plateau for a certain number of epochs (patience set to 5). The number of steps per epoch for both the training and validation sets is calculated based on the sample sizes of the respective generators, ensuring efficient training iterations. Table 3 shows the tuned hyperparameters used for different classifiers for the proposed model. Table 4 shows the proposed CNN model used for binary, multiclass and regression analysis. Figure 2 shows the pipeline flow of the cyclone classification, marking and intensity prediction algorithm. These callbacks and

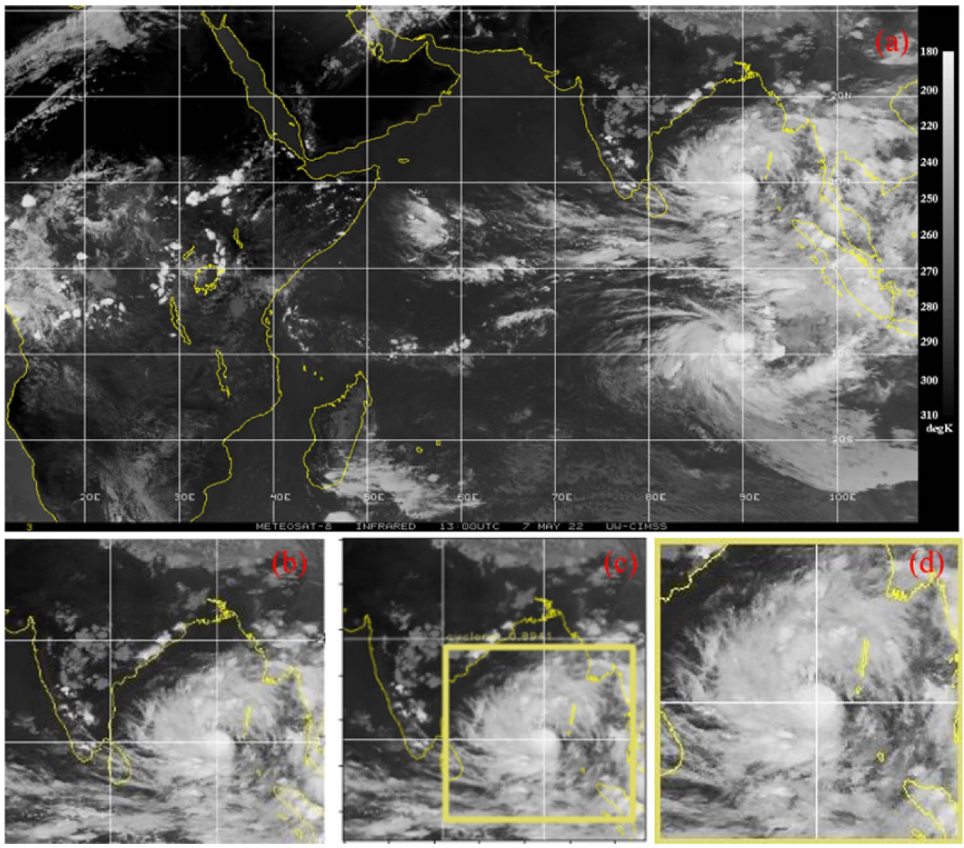


**Fig. 3** **a** shows original image (2022–05-07, 12:00 h), **b** shows the pre-processed image, **c** shows the detected cyclone and **d** shows the enlarged view of the detected cyclone 'Asani' testing on the proposed pipeline

**Table 7** Comparison of different models

| Model | Multiclass classification accuracy (%) | Binary classification accuracy (%) | Regression RMSE (knots) |
|---|---|---|---|
| Proposed Model | 63.83 ± 1.3 [31.2 ± 0.78] | 98.4 ± 0.003 | 16.2 ± 0.9 [26.69 ± 0.88] |
| VGG-16 | 39.83 ± 0.86 [30.3 ± 1.04] | 96.7 ± 0.007 | 19.79 ± 0.21 [20.67 ± 0.03] |
| Inceptionv3 | 51.51 ± 1.29 [36.1 ± 5.65] | 97.5 ± 0.01 | 12.4 ± 0.29 [17.28 ± 0.53] |
| XceptionNet | 40.19 ± 1.07 [30.9 ± 1.78] | 97.9 ± 0.0024 | 22.52 ± 0.8 [21.32 ± 1.43] |
| ResNet | 28.81 ± 0.005 [28.4 ± 0] | 98.3 ± 0.0015 | 13.88 ± 0.28 [22.35 ± 0.47] |

*Values in [] are based on Year-wise split (Training: 2000–2019, Testing: 2020–2022)

tuning aims to improve the models' generalization capability and optimize its performance during training.

YOLOv3 (You Only Look Once, Version 3) is a real-time object detection algorithm that identifies specific objects in videos, live feeds, or images. YOLO uses features learned by Darknet to detect an object. Darknet has 53 convolutional layers with consecutive 3 × 3 and 1 × 1 convolution layers followed by a skip connection to help the activations propagate through deeper layers without gradient diminishing (Redmon et al. 2018). Training for cyclone detection by YOLOv3 is done by annotating images with the cyclonic regions with a bounding box and the text files containing the class and coordinates of the bounding box in YOLO format. A custom weights file is generated using transfer learning with Darknet weights to train on our custom dataset. Our dataset had only one target, i.e., cyclones. Non maximum suppression is set to 0.3 to filter the best bounding box out of many overlapping boxes. Minimum probability to eliminate weak predictions is 0.3. The predicted bounding box size varies depending on the size of the cyclone.

## Results and Discussion

Compared to existing studies, our proposed pipeline simultaneously detects the cyclonic region, classifies and predicts the intensity. Table 5 shows the comparative study of the existing cyclone prediction models in the NIO region with the proposed model. The models were evaluated separately since most of the mis-classifications are

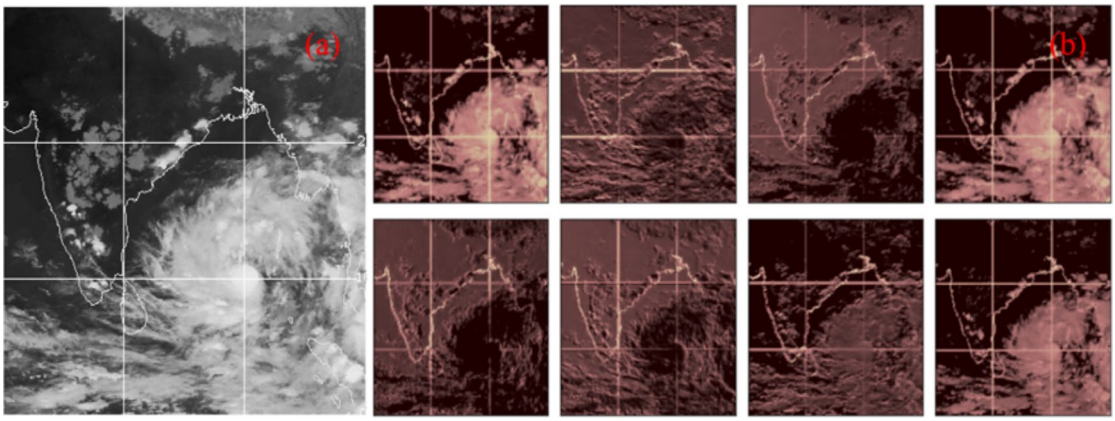


**Fig. 4** **a** shows original image and **b** shows the feature maps of the first layer of CNN

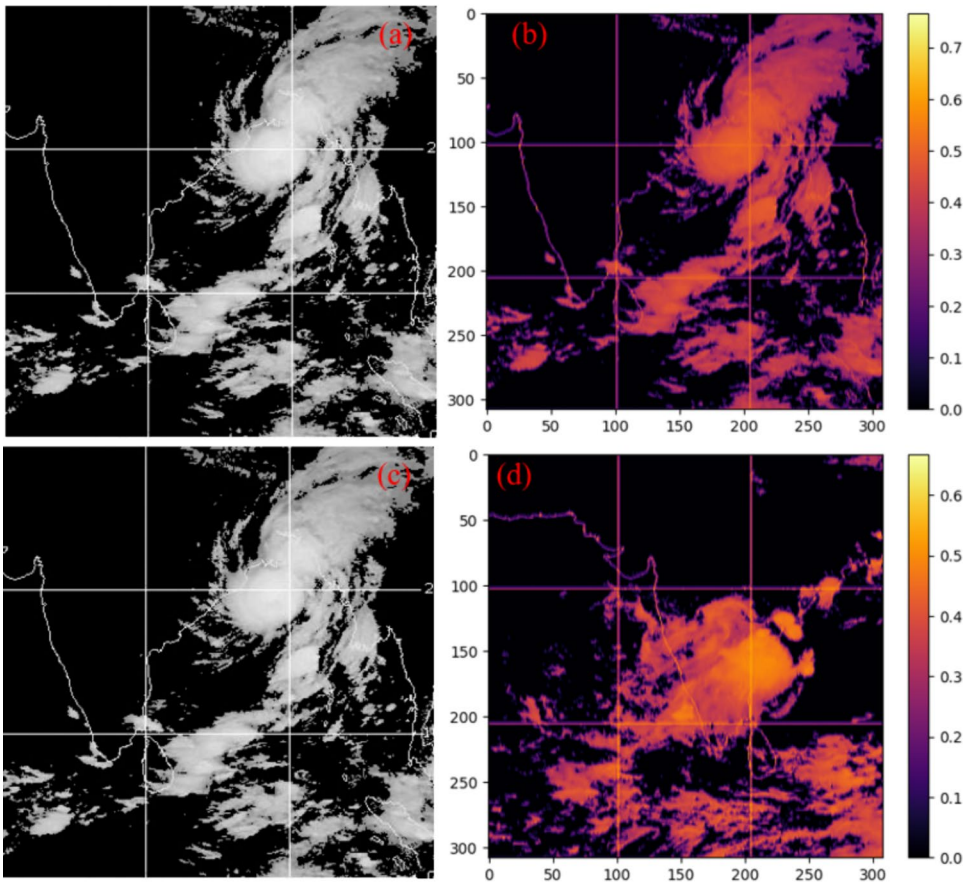


**Fig. 5** Heatmaps of cyclones, **a** and **c** are original image, **b** and **d** corresponding heatmaps

found to be within a single class difference and hence, the intensity value may be a better representation of the capabilities of the DL model. The proposed model is able to classify cyclones in five different classes as compared to most of the existing studies which classify only in one or two classes.

The False Positive Rate (FPR) represents the ratio of incorrectly predicted instances of a class to the total number of instances that do not belong to that class but were

**Table 8** Cyclone-wise metrics

| Category | Accuracy (%) | False positive rate (%) |
|---|---|---|
| D | 66.5 ± 2.17 | 6.77 ± 0.86 |
| DD | 70.54 ± 2.06 | 15.25 ± 0.82 |
| CS | 52.34 ± 0.96 | 9.63 ± 1.34 |
| SevereCS | 58.51 ± 1.57 | 5.22 ± 0.64 |
| VSCS | 71.37 ± 2.13 | 5.45 ± 0.63 |

predicted as belonging to it. In simpler terms, it measures how often the classifier incorrectly identifies instances from other classes as belonging to a specific class. False positives (FP) are instances where the model incorrectly predicted the class as belonging to a certain class when it actually doesn't. True negatives (TN) are instances where the model incorrectly predicted the class as belonging to a certain class when it actually doesn't.

The FPR is 0.7% for binary classification. The low FPR for binary classification underscores the reliability of our model in accurately identifying cyclones. Table 6 shows the multiclass testing dataset metrics for individual classes. It can be seen that the model performs best for 'VSCS' and worst for 'CS'. The well-defined structure in VSCS may be the reason for the high performance. However, it is important to note that the model exhibits comparatively lower performance in classifying 'CS', indicating potential areas for further refinement.

A case study of a specific cyclone 'ASANI' from the year 2022 (Timestamp: 2022–05-07, 12:00 h) has been done. Fig a, shows the original image, Fig. 3b shows the pre-processed image, Fig. 3c shows the detected cyclone and Fig. 3d shows the enlarged view of the detected cyclone. The pipeline predicted the intensity of the cyclone is 33 knots (Actual: 30 knots). The cyclonic region is also detected correctly.

Table 9 shows the results of early-stage detection. The results obtained from testing the early-stage detection using the proposed model are promising. During the critical early stages of these cyclones, ranging from Depression to Deep Depression, the model demonstrated a high level of accuracy in predicting the cyclone category. In the case of cyclone Asani, the model's predictions aligned closely with the actual cyclone category across various timestamps. Specifically, the model accurately predicted 'D' and 'DD' categories, with only one instance where a 'DD' category cyclone was misclassified as a 'D' category cyclone. The model's performance remained consistent and reliable throughout these early stages. These results are particularly significant considering the challenges associated with detecting weak cyclones in their nascent stages. Weak cyclones pose a unique challenge due to their subtlety and limited observable characteristics, making them difficult to identify using conventional techniques alone. However, the proposed model offers a promising solution by leveraging advanced algorithms to enhance early detection capabilities.

The development cycle of TC may be divided into three stages: formation and initial development stage, mature stage, and decay. As seen in Table 9, across the dataset, there's a notable accuracy in forecasting the majority of cyclone stages, particularly during critical periods of transition. This suggests a robust understanding of cyclone dynamics and an effective application of predictive algorithms. However, there are isolated instances where misclassifications occurred, such as predicting a 'CS' category as a 'SevereCS' category, indicating potential areas for improvement in capturing subtle variations or anomalies in cyclone behaviour.

At timestamp 2022–05-10 06:00 h, the model predicted 'CS' while the actual stage was 'SevereCS'. Similarly, at timestamp 2022–05-10 12:00 h, the model predicted 'SevereCS' while the actual stage was 'CS'. This could be due to various factors such as insufficient training data, feature representation, or complexity of the task. The incorporation of auxiliary data as well as other bands satellite images may be useful to overcome this issue and need further investigation.

Predefined CNN architectures provide general architectural recommendations for deep learning practitioners to handle a wide variety of problems and develop newer and data specific architectures when the training dataset is limited. Hence, these approaches were also tested with the same dataset. Table 7 shows the comparison of our proposed model with other predefined architectures on testing dataset, which shows that the proposed model is better than predefined architectures.

It is to be noted that the Advanced Dvorak Technique (ADT) has shown much better RMSE than the present model (Olander et al., 2021). It is, however, also to be noted that the performance of all the previous studies using Deep Learning techniques as well the pre-defined architectures evaluated in this work performed poor than the ADT. This is one of the major drawbacks of the Deep learning models as the performance directly depends on the number of training data available. As NIO region has a very small dataset at present, the performance may improve once more data is made available. Another reason for high RMSE is that the best track intensity estimation from IMD is only available with 5 kt resolution.

Feature maps provide insight into the internal representations for input in the CNN layers of the model. Figure 4 shows the feature maps of the first layer of CNN. It shows the learning pattern. The purpose of feature maps is to identify meaningful features in the input data, such as edges, corners, or texture patterns. These features are then used to classify or identify the input image.

Another case study to differentiate between correctly classified and incorrectly classified images is shown in Fig. 5. Figure 5a shows the original image of a cyclone (Timestamp: 2022-10-24, 00:00 h, actual class 'CS', predicted class 'DD'), which was wrongly classified, the heatmap of this cyclone is shown in Fig. 5b. Figure 5c shows the original image of a cyclone (Timestamp: 2022-05-11, 00:00 h, actual class 'CS', predicted class 'CS'), which was correctly classified. Figure 5d is the corresponding

**Table 9** Asani cyclone stage detections

| Timestamp | Actual | Predicted | |
|---|---|---|---|
| 2022–05-07 00:00 h | D | D | Early stage |
| 2022–05-07 06:00 h | D | D | |
| 2022–05-07 12:00 h | DD | D | |
| 2022–05-07 18:00 h | DD | DD | |
| 2022–05-08 00:00 h | DD | DD | |
| 2022–05-08 06:00 h | CS | CS | Mature Stage |
| 2022–05-08 12:00 h | CS | CS | |
| 2022–05-08 18:00 h | SevereCS | SevereCS | |
| 2022–05-09 00:00 h | SevereCS | SevereCS | |
| 2022–05-09 06:00 h | SevereCS | SevereCS | |
| 2022–05-09 12:00 h | SevereCS | SevereCS | |
| 2022–05-09 18:00 h | SevereCS | SevereCS | |
| 2022–05-10 00:00 h | SevereCS | SevereCS | |
| 2022–05-10 06:00 h | SevereCS | CS | |
| 2022–05-10 12:00 h | CS | SevereCS | |
| 2022–05-10 18:00 h | CS | CS | |
| 2022–05-11 00:00 h | CS | CS | |
| 2022–05-11 06:00 h | CS | CS | |
| 2022–05-11 12:00 h | DD | DD | Decay |
| 2022–05-11 18:00 h | DD | DD | |
| 2022–05-12 00:00 h | D | DD | |

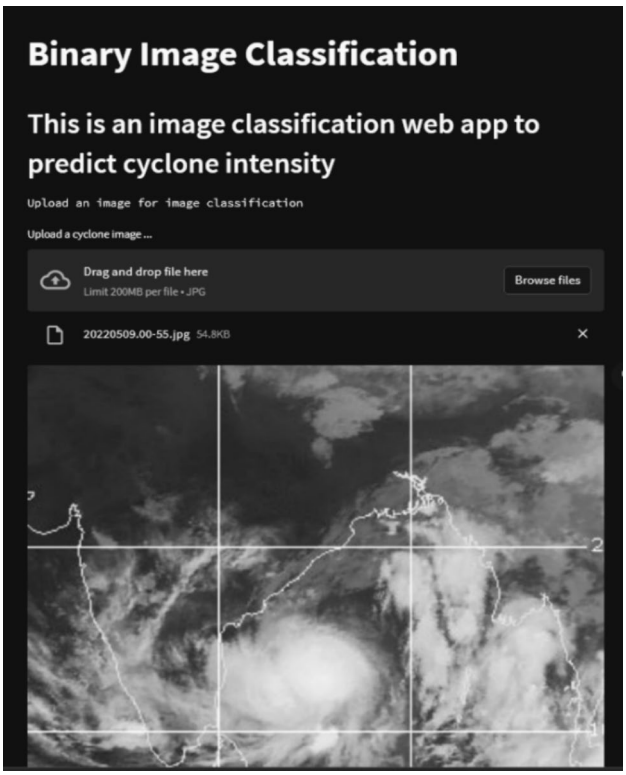


**Fig. 6** Streamlit (python library) app deployment of proposed model

heatmap of that cyclone. In heatmaps generated by CNNs, brighter areas typically indicate regions of the input image that have higher activation values in the corresponding feature map. These regions are considered more important or relevant for the model's prediction.

Another important observation from Table 7 is that the evaluation of all the DL models based on data-wise split shows higher accuracy than the year-wise split. As this data-wise split results were only available from previously published works in this domain, we keep these statistics as well for comparison. However, one should be careful in interpreting this result since data similar to training images may be present in testing samples as the cyclone images doesn't always change fast. Table 8 shows the specific cyclone wise accuracy based on the proposed model. These results highlight promising accuracy rates across categories, with 'DD' and 'VSCS' achieving above 70% accuracy. The accuracy for the 'CS' category is relatively low compared to the other categories. This could indicate a specific challenge in correctly identifying cyclones in this stage due to a similar cloud appearance to that of adjacent categories. This is also the reason of increased FPR in 'CS' category as the model sometimes wrongly detects 'DD' or 'SevereCS' category. Despite varying false positive rates, the overall performance suggests potential for continued refinement and optimization in classification techniques.

## Conclusion and Future Scope

A deep learning framework has been developed to estimate and classify TC from 10 km spatial resolution Meteosat satellite images over NIO regions for the period 2000–2022. The NIO region has a very less dataset compared to other regions, however, is one of the important areas of cyclone genesis.

Here we developed an architecture to detect, classify and predict the intensity of the cyclones. Here, for the first time, we tried to categorise the NIO cyclones in five groups. The architecture is statistically evaluated based on a data set spanning 2000–2022 both year-wise and data-wise. The performance of the architecture is found to be better or comparable with other existing studies in both the evaluation method. It is to be noted that most of the existing DL/ML based studies either used less data set for performance evaluation or developed for fewer cyclone classes. It is, however, also to be noted that the DL model performance is yet to surpass the Advanced Dvorak Technique for regression analysis.

The study also highlights the importance of careful interpretation of the results obtained from a data-wise split in the DL/ML approach. Based on the study, the application of deep learning to TC intensity analysis shows tremendous promise for further development with more advanced

methodologies and expanded training datasets. Detailed analysis of a particular storm to understand model performance with storm structural changes during rapid intensification is another future work that could be studied.

Early-stage cyclone detection is vital for meteorological agencies, governments, and communities, enabling timely warnings, evacuation planning, and infrastructure securing to minimize loss of life and property damage. Leveraging advanced forecasting techniques, including multi-spectral imagery and numerical weather model outputs, enhances early detection accuracy. By integrating data from multiple sources, such as infrared, visible, and microwave bands, alongside numerical weather models, predictive capabilities improve, allowing for more precise identification of cyclone formation and trajectory. This comprehensive approach strengthens disaster preparedness, reduces vulnerability, and fosters sustainable development in cyclone-prone regions.

To enhance multiclass classification accuracy and reduce RMSE, several methods can be employed. Integrating Generative Adversarial Networks (GANs) enriches the dataset, boosting diversity and quantity for training, thus improving model performance. Addressing class imbalance through techniques like oversampling or SMOTE ensures fair representation of all classes. Using interpretable models like SHAP aids in understanding prediction factors and feature relationships. Implementing continual learning techniques allows the model to adapt to new data, ensuring ongoing relevance and performance in evolving scenarios. By applying these strategies iteratively, both multiclass accuracy and RMSE can be significantly improved.

Incorporating multi-wavelength images into the classification pipeline can provide a wealth of additional information. These images capture data across different parts of the electromagnetic spectrum, offering insights into diverse physical properties of the objects being classified. By leveraging this comprehensive data, the model gains a more nuanced understanding of the underlying patterns, thereby enhancing its accuracy in distinguishing between classes. Furthermore, integrating physics-guided machine learning techniques can further refine the classification process. By incorporating domain-specific knowledge and principles from physics, such as light scattering behaviours or spectral signatures, into the machine learning algorithms, the model can make more informed predictions. This not only improves the accuracy of classification but also enhances the interpretability of the model's decisions, making it more transparent and trustworthy (Table 9).

The models can be deployed on online servers and can be used from anywhere in the world for quick response and analysis. Figure 6 shows the developed user interface on a local server. The research has also generated annotated dataset with cyclone location. The codes and results are publicly available at https://github.com/manishmawatwal/Cyclone

## Appendix

Precision is the measure of a positive predicted value and is defined as

$$Precision = \frac{TP}{TP + FP}$$

Recall or True Positive Rate is defined as

$$Recall = \frac{TP}{TP + FN}$$

True Positive (TP) is the outcome where the model correctly predicts the positive class. False Positive (FP) is the outcome where the model incorrectly predicts the positive class.

F1 score is the harmonic mean of precision and recall and is defined as

$$F1score = \frac{2 * (Precision * Recall)}{Precision + Recall}$$

False positive rate (FPR) is the ratio of negative events wrongly classified as positive and the total number of actual negative events.

$$FPR = \frac{FP}{FP + TN}$$

**Acknowledgements** The CIMSS Tropical Cyclone Archive for providing open access data. Author Saurabh Das also thankfully acknowledge 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: MM, SD. Data Collection and Analysis: MM. Interpretation of results: MM, SD. Draft Manuscript Preparation: MM. 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

Bandyopadhyay, S., Dasgupta, S., Khan, Z. H., & Wheeler, D. (2021). Correction to: spatiotemporal analysis of tropical cyclone landfalls in Northern Bay of Bengal, India and Bangladesh. *Asia-Pacific Journal of Atmospheric Sciences, 57*, 865–875. https://doi.org/10.1007/s13143-021-00227-4

Bonfanti, C., Stewart, J., Hall, D., & Govett, M. (2020). Tropical and extratropical cyclone detection using deep learning. *Journal of Applied Meteorology and Climatology, 59*(12), 1971–1985. https://doi.org/10.1175/JAMC-D-20-0117.1

Chandan, Roy Rita, Kovordányi (2012) Tropical cyclone track forecasting techniques —— A review Atmospheric Research 104-10540-69 10.1016/j.atmosres.2011.09.012

Chaudhari, S., Dutta, D., Goswami, S., & Middey, A. (2013). Intensity forecast of tropical cyclones over North Indian Ocean using multilayer perceptron model: Skill and performance verification. *Natural Hazards, 65*, 97–113. https://doi.org/10.1007/s11069-012-0346-7

Chen, B.-F., Chen, B., Lin, H.-T., & Elsberry, R. L. (2019). *Estimating tropical cyclone intensity by satellite imagery utilizing convolutional neural networks*, vol. 34, issue 2, pp. 447–465, 1 Apr 2019. https://doi.org/10.1175/WAF-D-18-0136.1.

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

Emanuel, K. A. & Nolan, D. S., (2004). Tropical cyclone activity and global climate. In Proceedings of the 26th Conference Hurricanes Tropical Meteorology, Miami, FL, USA, vol. 10, pp. 240–241. https://ams.confex.com/ams/26HURR/techprogram/paper_75463.htm .

Kar, C., Banerjee, S. (2021). Tropical cyclones intensity estimation by feature fusion and random forest classifier using satellite images. *Journal of the Indian Society of remote Sensing*, pp. 689–700. https://doi.org/10.1007/s12524-021-01477-5

Kar, C., Kumar, A., & Banerjee, S. (2019). Tropical cyclone intensity detection by geometric features of cyclone images and multilayer perceptron. *Springer Nature Appl. Sci., 1*, 1099. https://doi.org/10.1007/s42452-019-1134-8

Koepken, C. (2004). Solar stray light effects in meteosat radiances observed and quantified using operational data monitoring at ECMWF. *Journal of Applied Meteorology, 43*, 28–37. https://doi.org/10.1175/1520-0450(2004)043%3c0028:SSLEIM%3e2.0.CO;2

Kumawat, S., & Jaiswal, J. (2021). Cyclone detection and forecasting using deep neural networks through satellite data. *Proceedings of International Conference on Innovations in Software Architecture and Computational Systems*. https://doi.org/10.1007/978-981-16-4301-9_2

Lee, Y.-J. (2021). Interpretable tropical cyclone intensity estimation using Dvorak-inspired machine learning techniques. *Engineering Applications of Artificial Intelligence*, *101*, https://doi.org/10.1016/j.engappai.2021.104233.

Lee, C. Y., Tippett, M. K., Camargo, S. J., & Sobel, A. H. (2015). Probabilistic multiple linear regression modeling for tropical cyclone intensity. *Mon. Wea. Rev., 143*, 933–954. https://doi.org/10.1175/MWR-D-14-00171.1

Liefhebber, F., Lammens, S., Brussee, P. W. G., Bos, A., John, V. O., Rüthrich, F., Onderwaater, J., Grant, M. G., & Schulz, J. (2020). Automatic quality control of the Meteosat First Generation measurements. *Atmos. Meas. Tech., 13*, 1167–1179.

Matsuoka, D., Nakano, M., Sugiyama, D., & Uchida, S. (2018). Deep learning approach for detecting tropical cyclones and their precursors in the simulation by a cloud-resolving global nonhydrostatic atmospheric model. *Progress in Earth and Planetary Science, 5*(1), 1–16. https://doi.org/10.1186/s40645-018-0245-y

Meer, M. S., Mishra, A. K., & Nagaraju, V. (2024). Investigation of meteorological characteristics of tropical supercyclone shaheen insights from high-resolution satellite observations. *J Indian Soc Remote Sens*. https://doi.org/10.1007/s12524-024-01857-7

Mooers, G., Pritchard, M., Beucler, T., Srivastava, P., Mangipudi, H., Peng, L., Gentine, P., & Mandt, S. (2023). Comparing storm resolving models and climates via unsupervised machine learning. *Scientific Reports, 13*(1), 22365. https://doi.org/10.1038/s41598-023-49455-w

Moon, I. J., Ginis, I., Hara, T., & Thomas, B. (2007). A Physics-based parameterization of air-sea momentum flux at high wind speeds and its impact on hurricane intensity predictions. *Monthly Weather Review, 135*, 2869–2878. https://doi.org/10.1175/MWR3432.1

Nair, A., Srujan, K., Kulkarni, S., Alwadhi, K., Jain, N., Kodamana, H., Sandeep, S., John, V. (2022) A deep learning framework for the detection of tropical cyclones from satellite images. *IEEE Geoscience and Remote Sensing Letters*, *19*. https://doi.org/10.1109/LGRS.2021.3131638

Olander, T. L. & Velden, C. S. (2016). The current status of the UW-CIMSS Advanced Dvorak Technique (ADT). In: Confernece on Hurricanes and Tropical Meteorology.

Olander, T. L., & Velden, C. S. (2019). The advanced Dvorak technique (ADT) for estimating tropical cyclone intensity: Update and new capabilities. *Weather Forecasting, 34*(4), 905–922. https://doi.org/10.1175/WAF-D-19-0007.1

Olander, T., Wimmers, A., Velden, C., & Kossin, J. (2021). Investigation of machine learning using satellite-based advanced dvorak technique analysis parameters to estimate tropical cyclone intensity. *AMS Weather and Forecasting, 36*, 2161–2186. https://doi.org/10.1175/WAF-D-20-0234.1

Pineros, M. F., Ritchie, E. A., & Tyo, J. S. (2011). Estimating tropical cyclone intensity from infrared image data. *Weather Forecasting, 26*(5), 690–698. https://doi.org/10.1175/WAF-D-10-05062.1

Redmon, J., & Farhadi, A ., (2018). YOLOv3: an incremental improvement**,** April 2018, arXiv:1804.02767.

Ritchie, E. A., & Valliere, K. G. (2012). Tropical cyclone intensity estimation in the North Atlantic basin using an improved deviation angle variance technique. *Weather Forecasting, 27*(5), 1264–1277. https://doi.org/10.1175/WAF-D-11-00156.1

Shakya, S., Kumar, S., & Goswami, M. (2020). Deep learning algorithm for satellite imaging based cyclone detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol *13*, pp. 827–839. https://https://doi.org/10.1109/JSTARS.2020.2970253.

Sharma, N., Ali, M. M., Knaff, J. A., & Chand, P. (2013). A soft-computing cyclone intensity prediction scheme for the western North Pacific Ocean. *RMetS Atmospheric Science Letters*, *14*, pp. 187–192. https://doi.org/10.1002/asl2.438 .

Tian, W., Huang, W., Yi, L., Wu, L., & Wang, C. (2020). A CNN-based hybrid model for tropical cyclone intensity estimation in meteorological industry. *IEEE Access, 8*, 59158–59168. https://doi.org/10.1109/ACCESS.2020.2982772

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*., Early access, Mar 26, 202. https://doi.org/10.1109/TGRS.2021.3066299.

Zhang, Q., & Liguang, Q. (2009). Tropical cyclone damages in China 1983–2006. *Bulletin of the American Meteorological Society, 90*(4), 489. https://doi.org/10.1175/2008BAMS2631.1