International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 4, 2025
https://www.theaspd.com/ijes.php

Tropical Cyclone Intensity Estimation: A Study Of Neural Architecture Search And Transfer Learning In Cnns

Kancharagunta Kishan Babu[1], Ajay Dilip Kumar Marapatla[2], Harikesh Manchala[3], Uday Kiran Pillalamarri[4], Danush Kanthala[5], Kamsani Charan Sai[6], and Musku Sai Ritish Reddy[7] 1,

[1,3,4,5,6&7]*Department of Computer Science and Engineering – CSE(AIML&IoT), Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering and Technology, Hyderabad-500090, Telangana, India*
[2] *Department of Computer Science and Engineering, SRM University-AP, Neerukonda-522502, Andhra Pradesh, India*
Email: [1]*kishan.kancharagunta@gmail.com* ; [2]*ajaydilipkumar.m@srmap.edu.in*; [3]*manchalaharikesh@gmail.com* ; [4]*udaykiran6463@gmail.com* ; [5]*kanthaladanush1234@gmail.com* ; [6]*csai0577@gmail.com* ;[7] *sritish.10@gmail.com*

Abstract: *Tropical cyclones (TCs) are severe weather phenomena that can significantly affect human lives. These events can lead to calamities characterized by strong sustainable winds and enormous waves. We proposed an architecture based on convolutional neural networks (CNNs) to tackle this problem. This method makes use of cyclone infrared images. Using customized FCL architectures, we used transfer learning and fine-tuning on CNN architectures such as VGG16, VGG19, and ResNet50, both with and without data augmentation. Fine-tuning involved 4 layers of VGG16, 8 layers of VGG19, and 12 layers of ResNet50 to capture cyclone features effectively. The CNN models were used to these architectures to extract features, and the resulting feature maps were fed to various combinations of Fully Connected Networks (FCL). The most optimistic results were achieved with the VGG16 + FCL (128 x 64 x 1) architecture through transfer learning, producing a Mean Absolute Error (MAE) of 7.51 kts, Root Mean Square Error (RMSE) of 9.63 kts, and an R2 Score of 0.92. Consequently, we identified this model as the foundational basis for Neural Architecture Search (NAS) to enhance the FCL architecture. The NAS process generated various architectures, among which the VGG16 + FCL (128 x 128 x 1) architecture stood out with notable performance, featuring a Mean Squared Error (MSE) of 6.77 kts, RMSE of 8.88kts, and an impressive R2-Score of 0.945.*
Keywords: *Cyclone Intensity Estimation, Transfer Learning, Neural Architecture Search, Infrared Images, CNNs*

INTRODUCTION

Tropical cyclones are among extreme weather events for their profound impact on human life , capable of causing disasters such as intense winds and substantial waves. Accurate intensity forecasting of TCs is of paramount importance for social and economic reasons. The Dvorak Technique, which is a traditional method to estimate cyclone intensity used since 1970 [1], is very subjective and includes assumptions. Dvorak Technique is the popular technique that uses the satellite to estimate the sustained maximum wind speed of a tropical cyclone. It uses IR(Infrared) imagery and visible(VIS) imagery and also the cloud patterns to estimate the intensity of the tropical cyclone. Latitude, translation rate, intensity, size, and the 12-hour intensity trend as indicated by the radius of the outer closed isobar are among the variables affecting the biases in DVT-based intensity assessments [2]. Due to complex conditions like the western North Pacific, the Dvorak Technique was modified by including two distinct calculation radii, leading to a parametric surface that connects intensity and TC axis symmetry [3]. Better alternatives like the advanced Dvorak techniques come into existence due to the very subjective character of the traditional Dvorak technique. advancements in the Advanced Dvorak Techniques 1)Intensity estimation using the microwave information from the satellite. 2)Advanced automated tropical cyclone centre fixing methods. 3)using aircraft for the tropical cyclone intensity estimation. 4)modifications to intensity projections for TCs going through an extratropical transition and subtropical systems. 5)surface wind radii protection procedures [4]. This advanced Dvorak technique is been applied to the Indian Ocean region [5]. It performs very decently in this region for the T greater than 4.0 (VSCS to SuCS) cyclones [5]. We also have the statistical methods which consider various parameters for the cyclone intensity predictions [6]. The sea surface temperature, vertical wind shear, storm movement velocity, vorticity at 850 hPa, initial storm intensity, and divergence at 200 hPa are some of the variables that affect storm formation. The models like SHIPS(Statistical Hurricane Intensity Prediction System) [7] in the Atlantic basin the proposed system is also called the statistical–dynamical model. The constantly shifting conditions make

it challenging to forecast cyclone intensity, even with advances in modelling. To improve accuracy and reduce variability in intensity estimation, limitations still exist and continual advances in observational capabilities and Innovative model development are required. Several machine Learning techniques also evolved to estimate the cyclone intensity estimation with less time and fewer resources like the [8] use of satellite information like spatial and temporal relations data, essentially a Dvorak-based machine learning model. The alternative machine learning model proposed by [9] support vector regression model is built on the specific statistical factors taken from IR(Infrared) imagery of hurricanes. The findings indicate that a cyclone is stable over a wide temperature range. Another major development in the ML field is eXtreme Gradient BOOSTing for cyclone intensity in the northwestern Pacific region [10]. These ML approaches also have limitations [11] such as limited data for complex physical processes in cyclone intensity estimations, inaccurate voted initialization, and poor resolution limiting the models' performance. The tropical cyclone strength as its output by taking into consideration the climatology and persistence factors, environmental situations, meteorological characteristics, intensity categories, and TC months as inputs. Another model that uses a multiple linear regression model [12] which considers seven major parameters that change over 12 hours of the forecast. Even the analysis contributes to the continuous efforts to enhance the TC's (Tropical Cyclone) precision and accuracy. There are Machine Learning forecasting models for cyclone intensity estimation that use a TCP-NGBoost framework [13] it shows superior performance compared to the state-of-the-art statistic-dynamical model. These machine learning models excel in collecting forecast intervals, enabling the generation of dependable probabilistic forecasts—an important element in disaster warnings There are major Advancements in the deep-learning(DL) field for intensity estimations which uses Deep Convolutional Neural Network architecture [14] which uses Infrared Images, interpolating hurricane data, Augmenting additional images of the HURDAT2 dataset. Using DL (deep learning) and IR (infrared imagery) satellite imaging models, an analytical evaluation of TC (Tropical Storm) Strom intensity estimation was proposed [15]. Convolutional networks (CNNs) are used when combined with deep learning and multi-platform remote sensing utilizing environmental field data on the INSAT-3D IR dataset [16]. The deep Learning approach has evolved into the Hybrid approach, which The deep learning-based hybrid model proposes merging the CNN (convolution neural network) with satellite remote sensing [17]. Additional approaches were for estimating tropical cyclone intensity utilizing DCNN + ViT [18], estimating strength using CNN + LSTM [19], and detecting hurricanes utilizing imagery from satellites and YOLOV5 [20]. Furthermore, a hybrid deep learning method based on AlexNet was under consideration [21].

METHOD

The research's suggested methodology is divided into different stages, as shown. in the flowchart below. The entire procedure is provided in a very understandable manner. The MOSDAC ISRO database was the source of the dataset. Next, the pre-processing of the data involves interpolating missing intensity values, resizing. im-ages, ROI extraction, scaling pixel values and intensities, and data augmentation. utilizing model training techniques like neural architecture search, transfer learning, and layer-by-layer optimization. Mean Square Error (MSE), Root Mean Square Error (RMSE), and R2 Score are used in the evaluation of models. These evaluation metrics help us to select the best-performing model in cyclone intensity estimation.

Figure 1. The entire methodology is visually represented in the flow chart.

1.1 Data Acquisition

The cyclone image data has been collected from MOSDAC, ISRO's oceanographic data archive center. The data was collected from 2013 to 2023. The images gathered are from the INSAT-3D satellite. The images are 3 channeled IR images. The images are available with a duration of 30 mins intervals. The intensity values are provided by IMD (Indian Meteorological Department). The total images available are around 17112. The images captured enclosed the Indian-subcontinent region

along with Arabian sea, Indian ocean, and Bay of Bengal. The images are captured with a spatial resolution of 4 km × 4 km.

Figure 2. Sample Images included in the dataset.

The Indian Ocean basin Imagery will be updated regularly in the MOSDAC database every thirty minutes. The Indian Meteorological Department, or IMD, provides the intensities. The dataset used in this study almost comprises 17112 images. The region of the Indian subcontinent, the Arabian Sea, the Indian Ocean, and the Bay of Bengal were all included in the imagery. A (4 km x 4 km) spatial resolution is used to capture the photos.

Data Pre-Processing

Preprocessing involves several procedures, such as scaling pixel values and intensities, ROI extraction, image resizing, data augmentation, and interpolating missing intensity values. The cyclone intensity best-track data is available at 6-hour intervals. As the images are available at 30-minute intervals, The missing intensity values are interpolated using time-series-based linear interpolation after the data has been preprocessed. The imagery includes the whole Indian basin, comprising the Bay of Bengal, the Arabian Sea, and the Indian Ocean. To estimate the intensity of a cyclone, we must extract the ROI (Region of Interest), which is a cyclone throughout the whole image.

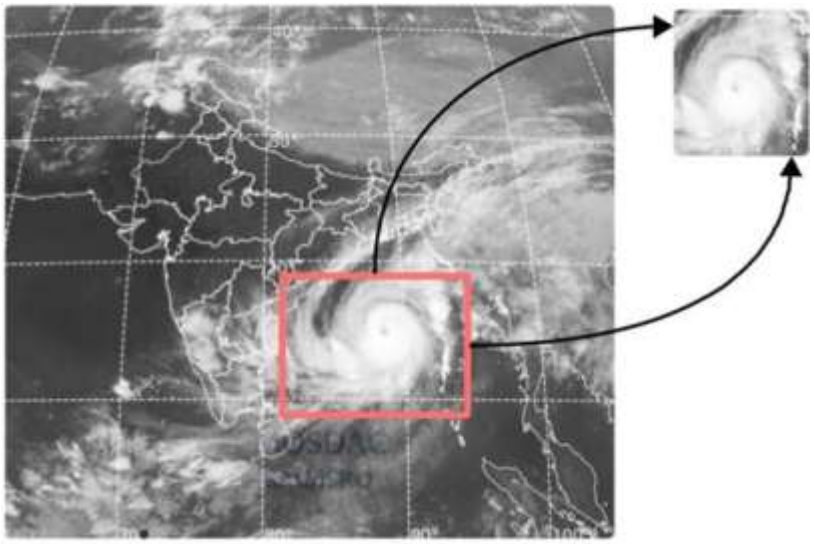

Figure 3. Region of Interest Extraction from the image**.**

The images are adjusted to match the CNN model architecture's input dimensions. The models are designed for 224 X 224-pixel dimensions. In addition, by dividing the pixel intensities by 255, the maximum intensity value, the intensities are normalized to a range of 0 to 1. The cyclone intensity values have also been scaled using the Min-Max scaler to make training easier and ensure their values fall between 0 and 1. After processing the cyclone image dataset, the total dataset contains approximately 17,112 images. These images are distributed among training, validation, and test sets, with the training set containing 80%, the validation set containing 10%, and the test set comprising the remaining 10% of the data items. Different data augmentations were used, including translation, rotation (approximately 20%), horizontal flips, zoom (around 20%), and a shear range of about 20%. Moreover, to enhance model training, the pixel intensities of the cyclone images and the cyclone's Maximum Sustained Wind (MSW) intensity values were normalized to a range of 0 to 1. The pixel values were divided by 255 (the maximum intensity value) to accomplish this scaling, while the MSW values underwent scaling using the Min-Max Scaler.

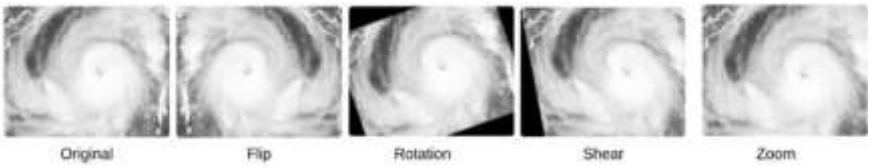

Figure 4. The sample images after data augmentation.

1.2 CNN Model Design and Training

Due to the technological advancement in Deep Learning and Machine Learning fields, the performance of Deep Learning models, particularly hybrid models, out-performs that of traditional Dvorak techniques presented in [15][18][32] and statis-tical models depending on the analysis of a limited set of parameters in evaluating Cyclone intensity in study [6] [7] [14]. This improvement allows for more accurate estimations in less time and at a lower cost. The CNN models have vast applications in classification, object detection, image analysis, etc. The present task falls under regression, given that the target variable is the Maximum Sustained Wind (MSW) of the cyclone, representing continuous numerical data. The proposed approach involves utilizing established CNN architectures trained on large datasets, such as ImageNet. Our research introduces three models based on Transfer Learning, Fine-Tuning, and Neural Architecture Search applied to transfer learning. Our goal is to leverage existing classification task architectures, incorporating custom fully connected (dense) layers and an output layer (single neuron) adjusted especially for the task of cyclone intensity.

Transfer Learning

This study aims to effectively utilize widely recognized CNN architectures developed for image classification tasks, such as VGG16, VGG19, and ResNet50 [21]. These networks have trained on vast ImageNet datasets, providing them with the knowledge to extract information about edges, texture, shape, and size of objects. This knowledge is later applied in the conventional layers of pre-existing models, transferring the extracted features to the custom-selected fully connected layers and output layer. We particularly train the weights of the Fully Connected Layers (FCL) and output layer for transfer learning in this study. The various setups for FCL set-ups used along with the above-mentioned architectures include 128 X 1, 128 X 64 X 1, 128 X 64 X 32 X 1. Each Fully Connected Layer (FCL) comprises the Rectified Linear Unit (ReLU) activation function, while the output layer utilizes a linear activation function. A dropout layer of 10% is added to the output layer to prevent over-fitting. These designs are consistently implemented across all classification architectures. The various models were trained separately on both augmented and non-augmented datasets.

ReLU Activation: f(x) = max (0, x)Linear Activation: f(x) = x

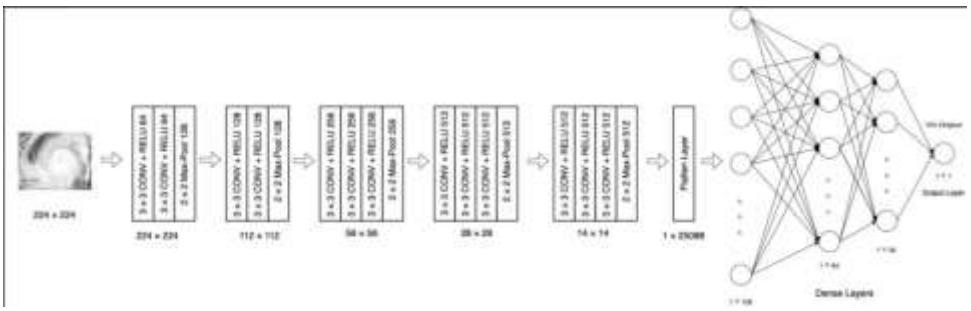

Figure 5. VGG16-128×64×32×1 architecture for Transfer Learning.

Similarly, transfer learning is implemented for VGG19 and ResNet50 architectures with mentioned combinations of FCL setups.This study outlines the transfer learning process involving fully connected layers, convolution functions, and regression functions. The entire process is described, including the extraction of features, prediction, loss function, and optimization using the Adam optimizer.

Transfer Learning Process

Let θ_{fc} be the weights of the fully connected layers, trainable by backpropagation. Let ϕ be the convolution function that extracts features from image I, freg be the regression function implemented by fully connected layers, and F be the features extracted from convolution. The overall transfer learning process is as follows:

$F = \phi(I)$

$Y_{pred} = F_{reg}(F, \theta_{fc})$

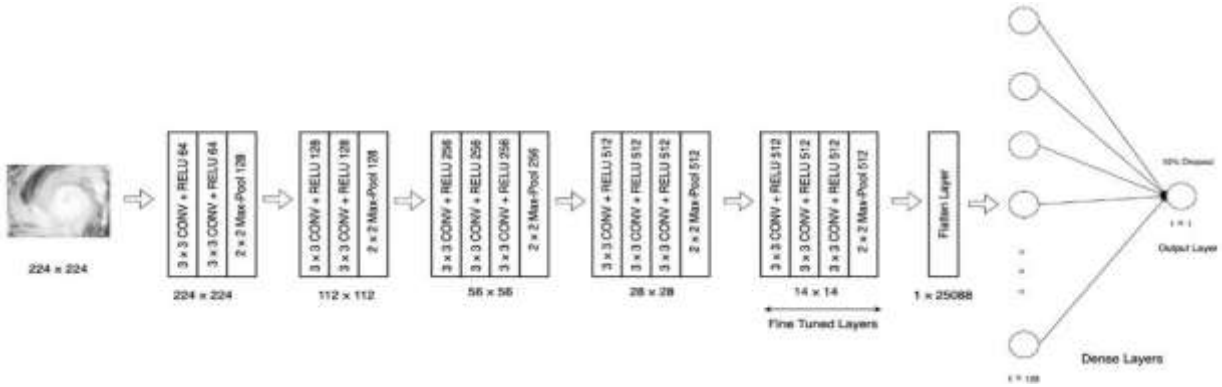

Figure 6. VGG16-128×1 architecture for Fine-Tuning (4 trainable layers).

Let Y_{actual} be the actual intensity and Y_{pred} be the predicted intensity. The loss function is given as:

$L_{reg} = \text{Loss}(Y_{pred}, Y_{actual})$

The loss function is the Mean Squared Error. The optimization is done using the Adam optimizer with a learning rate of 0.0001, as follows for the transfer learning task:

$$\theta^*_{fc} = \text{argmin}_{\theta fc} L_{reg}$$

θ^*_{fc} represents the updated weights of the fully connected layers.

The training has been carried out using early stopping with a patience of 5, monitoring the validation loss.

Fine Tuning

Within transfer learning techniques, the convolution layers of CNN architectures remain consistent, with only the Fully Connected Layers (FCL) being trained on specific data. In fine-tuning, selective layers of the convolution network are trained on specific data for a given task, focusing particularly on the last few layers. The motivation behind fine-tuning the last layers lies in their ability to extract high-level and task-specific features. Limiting the fine-tuning to these last layers helps in showing the early layers to extract features based on pre-trained data, thereby preventing overfitting due to the limited size of task-specific data. For VGG16, the last 4 layers, VGG19, the last 8 layers and ResNet50, the last 12 layers were fine-tuned.

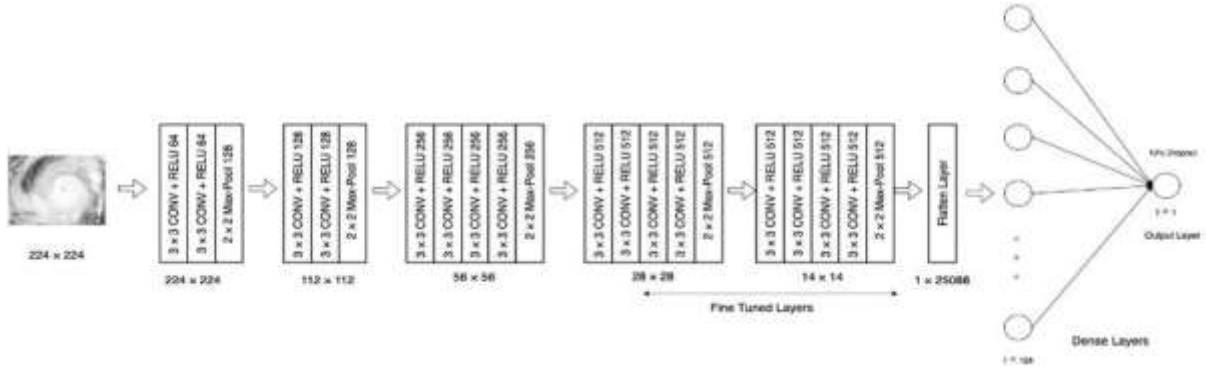

Figure 7. VGG19-128×1 architecture for Fine-Tuning (8 trainable layers).

The FCL layers utilize the same Rectified Linear Unit (ReLU) activation function, while the output layer uses a linear activation function. This pattern is maintained. across other architectures, and models with various FCL designs are trained both with data augmentation and without augmented data.Consider θ_{fc}

as the trainable weights for the fully connected layers, which can be adjustable through backpropagation. Let θ represent the weights for the convolution layers, which can be fine-tuned. Further, let ϕ denote the convolution function responsible for extracting features from image I, f_{reg} represents the regression function implemented by the fully connected layers, and F represents the features extracted through convolution.

The complete process of transfer learning is as follows:

$F_1 = \phi(I)$

$F_2 = \phi(\theta, F_1)$

$Y_{pred} = f_{reg}(F_2, \theta_{fc})$

Let Y_{actual} be the actual intensity, and Ypred be the predicted intensity. The loss function is given as:

$L_{reg} = \text{Loss}(Y_{pred}, Y_{actual})$

The loss function is the Mean Squared Error.

The optimization is done using the Adam optimizer with a learning rate of 0.0001, which is as follows for the transfer learning task:

$\theta^*_{fc}, \theta^* = \text{argmin}_{\theta fc, \theta} L_{reg}$

- θ^*_{fc} are the updated weights of the fully connected layers
- θ^* are the updated weights of the trainable convolution layers.

The training has been carried out using the early stopping with a patience of 5 with a monitor on validation loss.

Transfer Learning with Neural Architecture Search

A Neural Architecture Search (NAS) based on Hyperband and successive halving is achieved using the VGG16 transfer learning model. NAS is especially applied to VGG16 as the base model in the context of transfer learning, without including any. data augmentation. The reason behind the decision lies in the findings presented in the subsequent section, indicating the better performance of VGG16 compared to others. architectures, particularly in the absence of data augmentation.In this Neural Architecture Search (NAS) technique, VGG16 helps as the base model, and within the search space, no modifications were made to the convolution layers, they maintained the pre-trained weights from the ImageNet dataset. NAS is exclusively employed to identify the optimal Fully Connected Layer (FCL) arrangement for the VGG16 architecture. Among various available search strategies, we selected for the hyperband-based successive halving approach due to its efficient distribution of resources to various sampled architectures. This technique facilitates the early termination of poorly performing architectures and reallocates resources to more successful architectures. Such a technique promotes effective resource utilization.

The search space contains:

- The number of Fully Connected Layer (FCL) layers ranges from a minimum of 1 to a maximum of 3.
- Each FCL layer is sampled with neurons ranging from a minimum of 32 to a maximum of 256, with a step of 32 in each iteration.
- A dropout layer at the output is also sampled, varying from 1.
- The search is performed with a maximum allowed number of trials set at 20 and a maximum of 10 epochs for each architecture.
- All FCL neurons utilize ReLU as the activation function in each sampling.
- The activation function at the output neuron is always linear in each sampling.
- The primary goal is to minimize the loss function, specifically the mean-squared error loss.

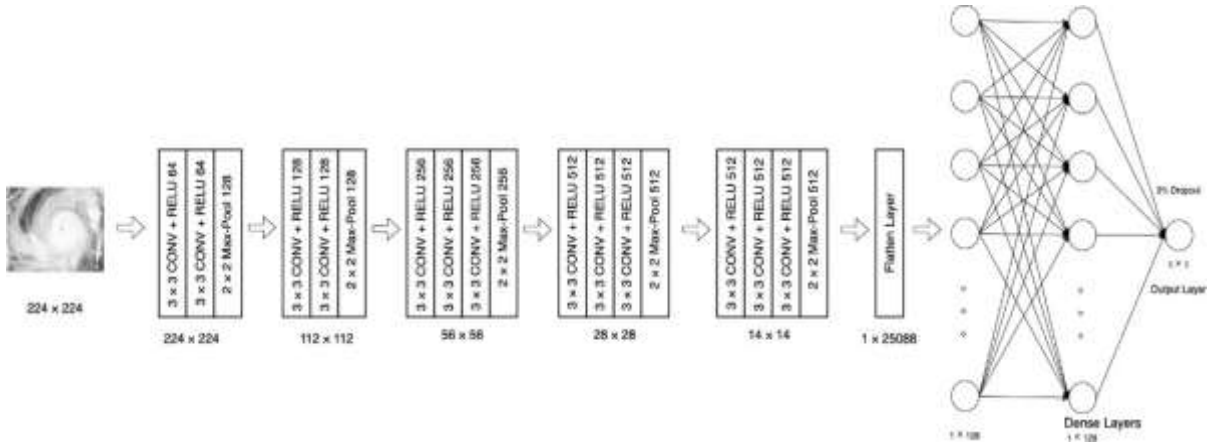

Figure 8. The best-performing architecture was NAS with VGG16 and no data augmentation. (VGG16-128×128×1).

Based on the above-mentioned search, the top 5 performing architectures were chosen for additional training with an extended number of epochs using a transfer learning approach without data augmentation, as detailed in the previous section. These architectures experienced training with an early stopping measure, where the patience was set at 5, and the validation loss function monitored was the mean-squared error.The best-preforming architecture derived from the previously mentioned Neural Architecture Search (NAS) involving VGG16, conducted without data augmentation and thereafter trained for extra epochs, is the VGG19-128×128×1 configuration, featuring a 3% dropout at the output layer.

FINDINGS AND DISCUSSIONS

Evaluation Metrics

We used a wide range of evaluation metrics to evaluate our deep learning models' performance. The selection of evaluation measures is essential in evaluating the precision and effectiveness of our models in forecasting the intensity levels of cyclones. We used Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R2 Square (R-squared) as the metrics in our study.

3.1.1 Mean Absolute Error (MAE)

Mean Absolute Error (MAE) is a measurement of the average of absolute differences between forecasted and actual values. It is calculated by taking the average of the absolute differences between the predicted and observed values.

The formula for MAE is as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |Y_i - \hat{Y}_i|$$

3.1.2 RMSE – Root Mean Square Error

RMSE is another parameter that accounts for both the magnitude and direction of errors. It is calculated by carrying the square root of the average of the square differences between predicted and observed values. The formula for RMSE is as follows:

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$$

$\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n$ are predicted values
$y_1, y_2, \ldots, y_n$ are observed values
n is the number of observations

3.1.3 R2-Score

The coefficient of determination, or R2 Square, measures the percentage of the variance in the predicted intensity values that of our deep learning model. It pro-vides insight into the quality of fit, or how effectively the variability in the observed intensity levels is reflected by the models. An increased R2 Square value indicates that our models have a better fit and more accurate predictions for estimating cy-clone intensity The R-squared (coefficient of determination) is calculated using the formula:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

3.2 RESULTS

3.2.1 Transfer Learning and Fine Tuning without Data Augmentation

Table 1: Results of the Transfer Learning Architecture sets that were trained without data augmentation.

Transfer Learning without Data-Augmentation								
		MAE			**RMSE**			
CNN Model	**FCL**	**Train**	**Val**	**Test**	**Train**	**Val**	**Test**	**R2-Score**
VGG-16	128×1	8.88	9.32	8.58	19.01	12.31	11.12	0.89
VGG-16	**128×64×1**	**7.37**	**7.51**	**7.56**	**9.83**	**9.63**	**10.02**	**0.92**
VGG-16	128×64×32×1	8.98	8.976	7.19	8.98	11.71	9.68	0.92
VGG-19	128×1	9.293	12.12	7.913	9.293	14.814	10.087	0.915
VGG-19	128×64×1	9.0387	9.162	7.805	11.455	11.726	9.998	0.916
VGG-19	128×64×32×1	10.109	8.106	8.455	13.778	10.188	10.881	0.9
ResNet50	128×1	28.71	29.46	28.89	28.71	37.54	36.52	-0.2114
ResNet50	128×64×1	13.003	14.083	12.604	13.003	17.762	16.573	0.76
ResNet50	128×64×32×1	16.753	20.937	16.257	16.753	24.796	20.558	0.62555

From the above-mentioned results of the transfer learning technique, it is obvious that VGG16 has exceeded the performance in all other architecture designs. The most performed architecture is VGG16-128×64×1. This observation implies that, for our dataset, less complex feature extraction is more effective, given that VGG16's feature extraction is less complex compared to that of VGG19 and ResNet50. In transfer learning, pre-trained models are employed, having been trained on vast and various datasets to extract generic features useful to a wide collection of tasks. The generalized features extracted from our cyclone data played a crucial role in the efficient training of the Fully Connected Layer (FCL) network.As previously described, the models trained with an early stopping criterion, monitoring the validation loss. This resulted in the VGG16-128×64×1 model being trained for 24 epochs. The maximum Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) values achieved through transfer learning, generalized to the test set, are 7.56 knots and 10.02 knots, respectively.

The selection of the best-performing models was based on two criteria:

- Their capability to manage overfitting (indicated by a small difference be-tween training and validation metrics)
- Their capability to strike a balance between lower and higher intensity values.

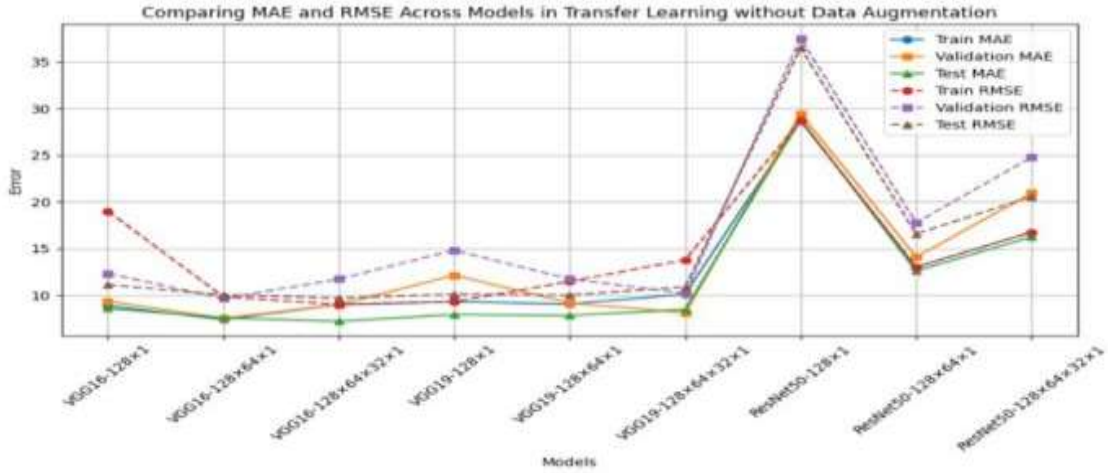

Figure 9. Comparison between MAE and RMSE of models in Transfer Learning without data augmentation.

This balance was evaluated by comparing the differences in Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) values over the training, validation, and test sets. For a more in-depth understanding, refer to the following graphs for detailed insights:

In the above model, the VGG16-128×64×1 points are closely clustered, implying its compliance with the selected requirements for being the best model. The calculation of the R2 score is performed on the test data, it is calculated only for the test data as it is used to represent the ability of the model towards generalization.

Table 2: Results of the Fine-tuning architecture sets trained without data augmentation.

Fine Tuning without Data-Augmentation								
		MAE			**RMSE**			
CNN Model	**FCL**	**Train**	**Val**	**Test**	**Train**	**Val**	**Test**	**R2-Score**
VGG-16	128×1	6.681	11.619	10.693	8.543	14.902	14.038	0.83
VGG-16	128×64×1	6.03	11.073	9.766	7.853	14.591	13.297	0.85
VGG-16	128×64×32×1	7.869	11.037	10.774	10.477	14.703	14.146	0.82
VGG-19	128×1	28.72	26.67	28.13	33.19	33.07	33.2	0.06
VGG-19	128×64×1	29.67	29.099	29.21	34.569	33.146	33.305	-0.00436
VGG-19	128×64×32×1	25.993	23.213	23.752	31.562	29.38	29.459	0.217
ResNet50	128×1	20.25	56.19	26.47	27.29	83.05	31.14	0.12
ResNet50	128×64×1	22.67	29.13	25.41	27.71	34.39	31.11	0.13
ResNet50	128×64×32×1	15.3	21.14	17.85	18.95	24.9	22.11	0.57

The poor performance of fine-tuning becomes obvious, even when attempting to train the high-level feature extraction layers of the model. This can be attributed to the fact that the chosen architectures, the reason for this is, were initially trained on extensive and various datasets. However, our cyclone task-specific dataset, despite comprising over 17,000 data items, proves to be relatively small in comparison to the scale of these extensive architectures. Additionally, our observations indicate that as network complexity increases, performance tends to decline in transfer learning designs.This indicates that our data is most consistent with simpler network architectures. In the process of fine-tuning, the models tend to display overfitting due to the limited qualities of the task-specific data. This overfitting becomes

apparent when comparing the metrics of the training sets and validation sets. The significant difference in metrics between these two sets highlights the drawbacks associated with utilizing a small dataset in the fine-tuning process.

3.2.2 Transfer Learning and Fine Tuning with Data Augmentation

The delivered results highlight that data augmentation did not produce any improvement in the overall model performance across both methods and all architectures. This highlights a challenge previously discussed in the study by [18], which correlates to issues associated with data imbalance. The phenomena connected to cyclone life and its intensity may clarify these results. The data makes it obvious that cyclones had a lengthy lifespan in these lower intensity values and that their levels of intensity were low in the initial phases of the cyclone. The cyclone's lifetime reduces when it reaches higher intensity levels because it reaches the coast sooner. This suggests that any data we collect for cyclones is highly skewed to the lower intensity values. This can also be seen in the difference between MAE and RMSE values. As there are more low intensity values in the data the RMSE values are quite high compared to MAE values as models tend to fit towards skewed data. As RMSE by squaring gives weightage to high errors while MAE averages errors with equal preference, high RMSE shows the existence of these high-intensity values as outliers in data distribution. Since even the instances of higher intensities are fair data points, they cannot be discarded similarly to the handling of outliers in other scenarios. The application of data augmentation to this skewed data distribution adds complexity, making it challenging for the models to effectively learn from the fine data.

Table 3: Results of the Transfer Learning architecture sets trained with data augmentation.

Transfer Learning with Data-Augmentation								
		MAE			**RMSE**			
CNN Model	**FCL**	**Train**	**Val**	**Test**	**Train**	**Val**	**Test**	**R2-Score**
VGG-16	128×1	14.87	13.39	11.69	19.26	17.99	14.36	0.82
VGG-16	128×64×1	15.44	14.98	11.77	19.53	19.15	14.96	0.81
VGG-16	128×64×32×1	15.31	15.31	12.29	15.32	19.27	15.64	0.79
VGG-19	128×1	21.45	20.93	20.502	21.45	25.06	25.77	0.4
VGG-19	128×64×1	16.07	15.74	15.63	16.07	20.037	19.124	0.68
VGG-19	128×64×32×1	16.945	16.85	15.71	16.94	20.44	19.39	0.67
ResNet50	128×1	30.677	30.677	26.33	30.677	36.702	30.601	0.155
ResNet50	128×64×1	30.915	30.915	25.34	30.92	30.63	28.76	0.26
ResNet50	128×64×32×1	28.44	26.77	27.92	28.44	31.04	31.6	0.097

Table 4: Results of the Fine-Tuning architecture sets trained with data augmentation.

Fine Tuning with Data-Augmentation								
		MAE			**RMSE**			
CNN Model	**FCL**	**Train**	**Val**	**Test**	**Train**	**Val**	**Test**	**R2-Score**
VGG-16	128×1	25.886	23.578	25.785	31.284	29.692	30.633	0.233
VGG-16	128×64×1	26.811	26.2415	27.283	32.37	29.696	32.785	0.119
VGG-16	128×64×32×1	26.241	25.342	26.078	31.666	38.573	32.805	0.119
VGG-19	128×1	23.892	19.948	22.72	29.087	24.604	28.398	0.343
VGG-19	128×64×1	22.172	20.47	24.947	26.828	25.038	30.506	0.239

VGG-19	128×64×32×1	23.375	25.226	21.972	28.761	29.261	27.41	0.39
ResNet50	128×1	29.73	29.34	31.37	35.8	39.8	40.77	-0.37
ResNet50	128×64×1	30.22	43.8	26.7	36.69	50.33	32.84	0.12
ResNet50	128×64×32×1	27.07	29.272	27.96	32.04	39.54	32.73	0.12

3.2.3 Neural Architecture Search on Transfer Learning without Data Augmentation

Based on the produced results, VGG16 consistently showed dependable performance across all methods and Fully Connected Layer (FCL) configurations. Therefore, Neural Architecture Search is implemented for VGG16 as the foundational model.

Table 5: The Top 5 architectures from NAS.

NAS Suggested Top Architectures							
				MAE		**RMSE**	
CNN Model	**FCL**	**Dropout Output Layer**	**Val Loss**	**Train**	**Val**	**Train**	**Val**
VGG-16	32×32×1	3%	0.0042	3.86	6.36	4.93	8.49
VGG-16	32×160×32×1	3%	0.0056	5.1	7.28	6.58	9.7
VGG-16	160×32×1	3%	0.0058	4.86	7.54	6.18	9.83
VGG-16	96×1	3%	0.0064	5.15	7.81	6.62	10.41
VGG-16	128×128×1	3%	0.0064	5.85	7.99	7.58	10.47

Employing Neural Architecture Search, as discussed earlier, with VGG16 as the foundational model in a transfer learning setting, we aim to identify the best Fully Connected Layer (FCL) format for the typical task of cyclone intensity estimation. Following NAS, where each architecture experienced a maximum of 20 trials and a maximum of 10 epochs, we selected the top 5 performing FCL architectures based on their validation loss, especially the mean-squared error.

Table 6: Results of the Transfer Learning on Top NAS Architectures without Data-Augmentation.

Transfer Learning on Top NAS Architectures without Data-Augmentation									
			MAE			**RMSE**			
CNN Model	**FCL**	**Dropout**	**Train**	**Val**	**Test**	**Train**	**Val**	**Test**	**R2-Score**
VGG-16	32×32×1	3%	5.44	7.24	7.38	7.64	9.54	9.44	0.93
VGG-16	32×160×32×1	3%	5.01	5.83	7.81	9.14	7.85	9.83	0.92
VGG-16	160×32×1	3%	3.17	7.12	6.93	9.49	9.20	9.10	0.93
VGG-16	96×1	3%	5.8	6.60	6.80	9.80	8.80	8.68	0.94
VGG-16	**128×128×1**	**3%**	**5.67**	**6.77**	**6.525**	**7.25**	**8.88**	**8.26**	**0.945**

Upon vast training of the architectures suggested by Neural Architecture Search (NAS) with an extended number of epochs, we noted an improvement in performance of few models. Few architectures outperformed the earlier results from the transfer learning-based VGG16-128×64×1 without any data augmentation. The selection of the best-performing model is still based on the earlier mentioned conditions of ability towards handling overfitting and performance of metrics over training, validation, and test data sets. The performance of NAS suggested architectures is summarized in Table-6.The below graph shows how near the metrics of the model VGG16-128×128×1 is distributed in comparison to other models. This also shows the model's ability to handle overfitting. The final generalized (on the test set) MAE and RMSE values achieved through NAS on VGG16-based transfer learning are 6.25 kts and 8.26 kts respectively.

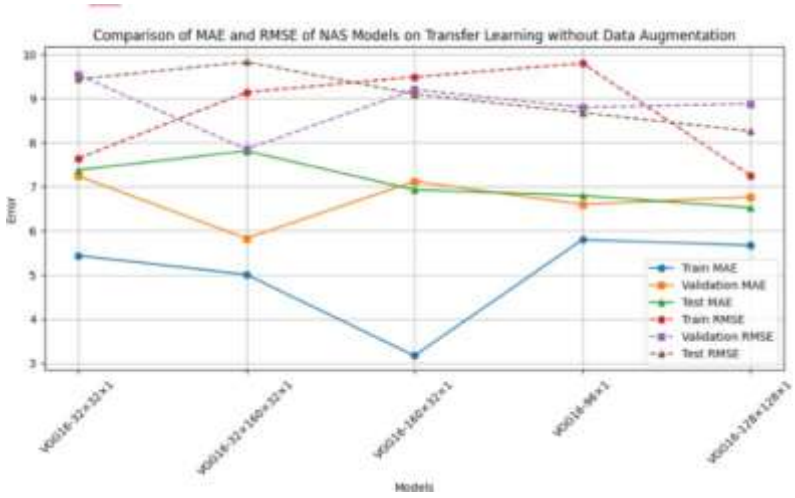

Figure 10. Comparison of MAE and RMSE of NAS models with Transfer Learning without data augmentation.

3.3 Limitations and Future Scope

3.3.1 Limitations

The proposed methodology utilizes complex architectures such as VGG16, VGG19, and ResNet50, which come with pre-trained knowledge from large datasets like ImageNet. However, the findings and comparisons above indicate that simpler architectures outperformed their more complex networks. Overfitting is also evident in some of the architectural designs. Therefore, a less complex architecture trained from scratch on cyclone data has the potential to be more advantageous. Due to the skewed data still, there is a considerable difference between MAE and RMSE. Working on making balanced data through the augmentation of samples selectively might be beneficial.

Future Scope

Expanding Neural Architecture Search could involve integrating the convolution layers into the search space to identify the optimal architecture for this problem. Given the signs from the results preferring a less complex architecture, this study could contribute to defining the search space for convolution layers. Most of the research primarily focuses on Infrared (IR) images. However, Microwave/Brightness Temperature (BT) images have also proved certain unique cyclone features. Therefore, models based on fusion, which combine multiple images or integrate features from various images, can be advantageous by providing additional data. An approach involving fusion models, improving feature sets through the combination of features from IR images and Microwave/BT images can be beneficial.

CONCLUSION

The neural architecture search with transfer learning-suggested VGG16-128×128×1 architecture has proven to be the most effective among the techniques and architectures that have been deployed. Current study has shown characteristics of the architecture design, such as the use of FCL and Convolution layers that are simpler. The need for a better data augmentation method for cyclone data is further emphasized by this study. In addition, by combining the Fusion models with infrared and microwave/brightness temperature images, we can enhance the work that is already being done in this field. It is also more appropriate to use a NAS with full search (Conv + FCL) to identify a high-performing model.

REFERENCES

1. Velden, C., Harper, B., Wells, F., Beven, J.L., Zehr, R., Olander, T., Mayfield, M., Guard, C.C., Lander, M., Edson, R., Avila, L., Burton, A., Turk, M., Kikuchi, A., Christian, A., Caroff, P., McCrone, P.: The Dvorak tropical cyclone intensity estimation technique: A satellite-based method that has endured for over 30 years. Bulletin of the American Meteorological Society 87(9), 1195–1210 (2006) [https://doi.org/10.1175/BAMS-87-9-1195]
2. Knaff, J.A., Brown, D.P., Courtney, J., Gallina, G.M., Beven, J.L.: An evaluation of Dvorak technique–based tropical cyclone intensity estimates. Weather and Forecasting 25(5), 1362–1379 (2010) [https://doi.org/10.1175/2010waf2222375.1]

3. Ritchie, E.A., Wood, K.M., Rodriguez-Herrera, O.G., Pi˜neros, M.F., Tyo, J.S.: Satellite-derived tropical cyclone intensity in the north Pacific Ocean using the deviation-angle variance technique. Weather and Forecasting 29(3), 505–516 (2014) [https://doi.org/10.1175/waf-d-13-00133.1]
4. Olander, T.L., Velden, C.S.: The advanced Dvorak technique (adt) for estimating tropical cyclone intensity: Update and new capabilities. Weather and Forecasting 34(4), 905–922 (2019) [https://doi.org/10.1175/WAF-D-19-0007.1]
5. Ahmed, R., Mohapatra, M., Giri, R.K., Dwivedi, S.: An evaluation of the advanced Dvorak technique (9.0) for the tropical cyclones over the north Indian ocean. Tropical Cyclone Research and Review 10(4), 201–208 (2021) [https://doi.org/10.1016/j.tcrr.2021.11.003]
6. Kotal, S.D., Bhowmik, S.K.R., Kundu, P.K., Kumar, A.D.: A statistical cyclone intensity prediction (scrip) model for the Bay of Bengal 117(2), 157–168 (2008) [https://doi.org/10.1007/s12040-008-0006-1]
7. DeMaria, M., Kaplan, J.: A statistical hurricane intensity prediction scheme (ships) for the Atlantic basin. Wea. Forecasting 9, 209–220 (1994) [https://doi.org/10.1175/1520-0434(1994)009⟨0209: ASHIPS⟩2.0.CO;2]
8. Lee, Y.-J., Hall, D., Liu, Q., Liao, W.-W., Huang, M.-C.: Interpretable tropical cyclone intensity estimation using Dvorak-inspired machine learning techniques. Engineering Applications of Artificial Intelligence 101, 104233 (2021) [https://doi.org/10.1016/j.engappai.2021.104233]
9. Asif, A., Dawood, M., Jan, B., Khurshid, J., DeMaria, M., Minhas, F.u.A.A.: Phurie: hurricane intensity estimation from infrared satellite imagery using machine learning. Neural Computing and Applications 32, 4821–4834 (2020) [https://doi.org/10.1007/s00521-018-3874-6]
10. Jin, Q., Fan, X., Liu, J., Xue, Z., Jian, H.: Using extreme gradient boosting to predict changes in tropical cyclone intensity over the western north pacific. Atmosphere 10, 341 (2019) [https://doi.org/10.3390/atmos10060341]
11. Wang, Z., Zhao, J., Huang, H., Wang, X.: A review on the application of machine learning methods in tropical cyclone forecasting. Frontiers in Earth Science 10, 902596 (2022) [https://doi.org/10.3389/feart.2022.902596]
12. Lee, C.-Y., Tippett, M.K., Camargo, S.J., Sobel, A.H.: Probabilistic multiple linear regression modeling for tropical cyclone intensity. Monthly Weather Review 143(3), 933–954 (2015) [https://doi.org/10.1175/mwr-d-14-00171.1]
13. Meng, F., Yao, Y., Wang, Z., Peng, S., Xu, D., Song, T.: Probabilistic forecasting of tropical cyclones intensity using machine learning model. Environmental Research Letters 18(4), 044042 (2023) [https://doi.org/10.1088/1748-9326/acc8eb]
14. Pradhan, R., Aygun, R.S., Maskey, M., Ramachandran, R., Cecil, D.J.: Tropical cyclone intensity estimation using a deep convolutional neural network. IEEE Transactions on Image Processing 27(2), 692–702 (2018) [https://doi.org/10.1109/TIP.2017.2766358]
15. Maskey, M., Ramachandran, R., Ramasubramanian, M., Gurung, I., Freitag, B., Kaulfus, A., Bollinger, D., Cecil, D., Miller, J.J.: Deepti: Deep learning-based tropical cyclone intensity estimation system. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 1–1 (2020) [https://doi.org/10.1109/jstars.2020.3011907]
16. Tian, W., Lai, L., Niu, X., Zhou, X., Zhang, Y., Kenny, L.K.S.T.C.: Estimation of tropical cyclone intensity using multi-platform remote sensing and deep learning with environmental field information. Remote Sens. 15, 2085 (2023) [https://doi.org/10.3390/rs15082085]
17. Tian, W., Huang, W., Yi, L., Wu, L., Wang, C.: A CNN-based hybrid model for tropical cyclone intensity estimation in meteorological industry. IEEE Access 8, 59158–59168 (2020) [https://doi.org/10.1109/ACCESS.2020.2982772]
18. Tong, B., Fu, J., Deng, Y., Huang, Y., Chan, P., He, Y.: Estimation of tropical cyclone intensity via deep learning techniques from satellite cloud images. Remote Sens. 15, 4188 (2023) [https://doi.org/10.3390/rs15174188]
19. Punjabi, V.D., Borane, K.V., Sarnaik, S.A., Patil, K.N., Patil, R.G.: Deep learning-based cyclone intensity estimation using insat-3d ir imagery. International Journal of Research Publication and Reviews 5(1), 4891–4902 (2024) [DOI not provided]
20. Shakya, S., Kumar, S., Goswami, M.: Deep learning algorithm for satellite imaging-based cyclone detection. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 13, 827–839 (2020) [https://doi.org/10.1109/JSTARS.2020.2970253]
21. Patil, H.: Cyclone prediction from remote sensing images using hybrid deep learning approach based on AlexNet. International Journal of Image, Graphics and Signal Processing 13(1), 96–107 (2024) [https://doi.org/10.5815/ijigsp.2024.01.07]

22. Sahoo, U.K., Ghosh, T.: A statistical model for predicting tropical cyclone intensity in north Indian ocean (nio) (2023). PREPRINT (Version 1) available at Research Square [https://doi.org/10.21203/rs.3.rs-2400950/v1]
23. Xue, L., Li, Y., Yao, S.: A statistical analysis of tropical cyclone-induced low-level winds near Taiwan island. Atmosphere 14, 715 (2023) [https://doi.org/10.3390/atmos14040715]
24. Wu, T., Duan, Z.: A new and efficient method for tropical cyclone detection and tracking in gridded datasets. Weather and Climate Extremes 42, 100626 (2023) [https://doi.org/10.1016/j.wace.2023.100626]
25. Park, M.-S., Kim, M., Lee, M.-I., Im, J., Park, S.: Detection of tropical cyclone genesis via quantitative satellite ocean surface wind pattern and intensity analyses using decision trees. Remote Sensing of Environment 183, 205–214 (2016) [https://doi.org/10.1016/j.rse.2016.06.006]
26. Biswas, K., Kumar, S., Pandey, A.K.: Tropical cyclone intensity estimations over the Indian ocean using machine learning (2021) [https://doi.org/10.48550/arXiv.2107.05573]
27. Zhang, C.-J., Luo, Q., Dai, L.-J., Ma, L.-M., Lu, X.-Q.: Intensity estimation of tropical cyclones using the relevance vector machine from infrared satellite image data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 12(3), 763–773 (2019) [https://doi.org/10.1109/JSTARS.2019.2894654]
28. Wang, P., Wang, P., Wang, D., Xue, B.: A conformal regressor with random forests for tropical cyclone intensity estimation. IEEE Transactions on Geoscience and Remote Sensing 60, 1–14 (2022) [https://doi.org/10.1109/TGRS.2021.3139930]
29. Pradhan, R., Aygun, R.S., Maskey, M., Ramachandran, R., Cecil, D.J.: Tropical cyclone intensity estimation using a deep convolutional neural network. IEEE Transactions on Image Processing 27(2), 692–702 (2018) [https://doi.org/10.1109/TIP.2017.2766358]
30. Lee, J., Im, J., Cha, D.-H., Park, H., Sim, S.: Tropical cyclone intensity estimation using multi-dimensional convolutional neural networks from geostationary satellite data. Remote Sens. 12, 108 (2020)
31. Punjabi, V.D., Borane, K.V., Sarnaik, S.A., Patil, K.N., Patil, R.G.: Deep learning-based cyclone intensity estimation using insat-3d ir imagery. International Journal of Research Publication and Reviews 5(1), 4891–4902 (2024)
32. Dharpure, H.N., Mohod, T.S., Malani, R.V., Chandak, J., Belge, A.S., Ambedkar, P.R., Pande, A.: Deep learning-based cyclone intensity estimation using insat-3d ir imagery: A comparative study. International Journal of Research Publication and Reviews 4(4), 4359–4365 (2023)