

RESEARCH ARTICLE

ConvLSTM-based tropical cyclone intensity estimation and classification using satellite imagery over the North Indian ocean

Manju M. S. , Harsh Pateriya², Rajeev Kumar Gupta³, Deepak Singh Tomar¹, Punit Gupta^{3,4*}, Asmir Butkovic                        <img alt="ORCID icon" data-bbox="798

Competing interests: The authors have declared that no competing interests exist.

surges, and widespread flooding. Since climate change has increased the frequency and strength of tropical cyclones in recent years [1,2], early warning systems, disaster planning, and mitigation activities depend heavily on the precise detection and estimation of cyclone intensity. Forecasting the intensity of cyclones is inherently complex and difficult because it demands a deep understanding of the dynamic interactions between oceanic and atmospheric systems. All these include sea surface temperature, humidity, vertical wind shear, and atmospheric pressure gradients in developing, moving, and intensifying a cyclone.

The most conventional and used traditional intensity estimation method is the Dvorak technique, used since the 1970s as a guideline in TCs assessment [3]. The Dvorak method is based on a satellite image interpretation done by means of visible and infrared determination. Meteorologists estimate intensity based on certain cloud patterns, storm eye formations, and temperature gradients. Despite its proven record for decades, the Dvorak technique relies on human skill and judgment in interpretation, allowing significant room for differences in severity classification, especially for weakly structured or rapidly intensifying cyclones [4]. It is generally less effective for cloud-filled or irregularly shaped cyclones, where the structural characteristics necessary for classification are not easily discernible. Besides the Dvorak, since then Numerical Weather Prediction [5] has also been extensively used in the cyclone intensity estimation. These models use complex mathematical equations and atmospheric physics to simulate the behaviour of cyclones according to a series of meteorological parameters, including wind speed, pressure variation, levels of humidity, and ocean temperatures. These models are very computationally intensive and thus require a capacity for high-performance computing and an extensive dataset of real-time meteorological data [5].

The traditional numerical models often fail to exploit the spatial and temporal patterns in the satellite imagery to the fullest extent possible, thereby limiting their predictability. Deep learning and data-driven approaches have transformed many aspects of meteorology, such as cyclone detection and intensity estimation [6]. CNNs are particularly well-suited for analysing high-resolution satellite images of tropical storms since they have demonstrated remarkable success in automated image analysis [7]. Fig 1 shows an example of a satellite image from the CIMSS tropical cyclone data archive, highlighting complex cloud structures associated with a tropical cyclone. Such features serve as critical inputs for training spatial models like CNNs. Unlike traditional methods that rely on handcrafted features and manual interpretation, CNNs automatically extract hierarchical spatial features from cyclone images, allowing them to differentiate cyclone structures from non-cyclone atmospheric formations with minimal human intervention [8]. This automation reduces subjectivity, increases consistency, and enhances efficiency. Therefore, CNN-based models are highly valuable for operational cyclone forecasting. However, several challenges persist in deep learning-based approaches. One of the major issues is that generalizability models trained on cyclone data from one region (e.g., the North Indian Ocean) may not perform well on cyclones from other basins (e.g., Atlantic or Pacific) due to regional differences in cyclone morphology and environmental conditions [9].

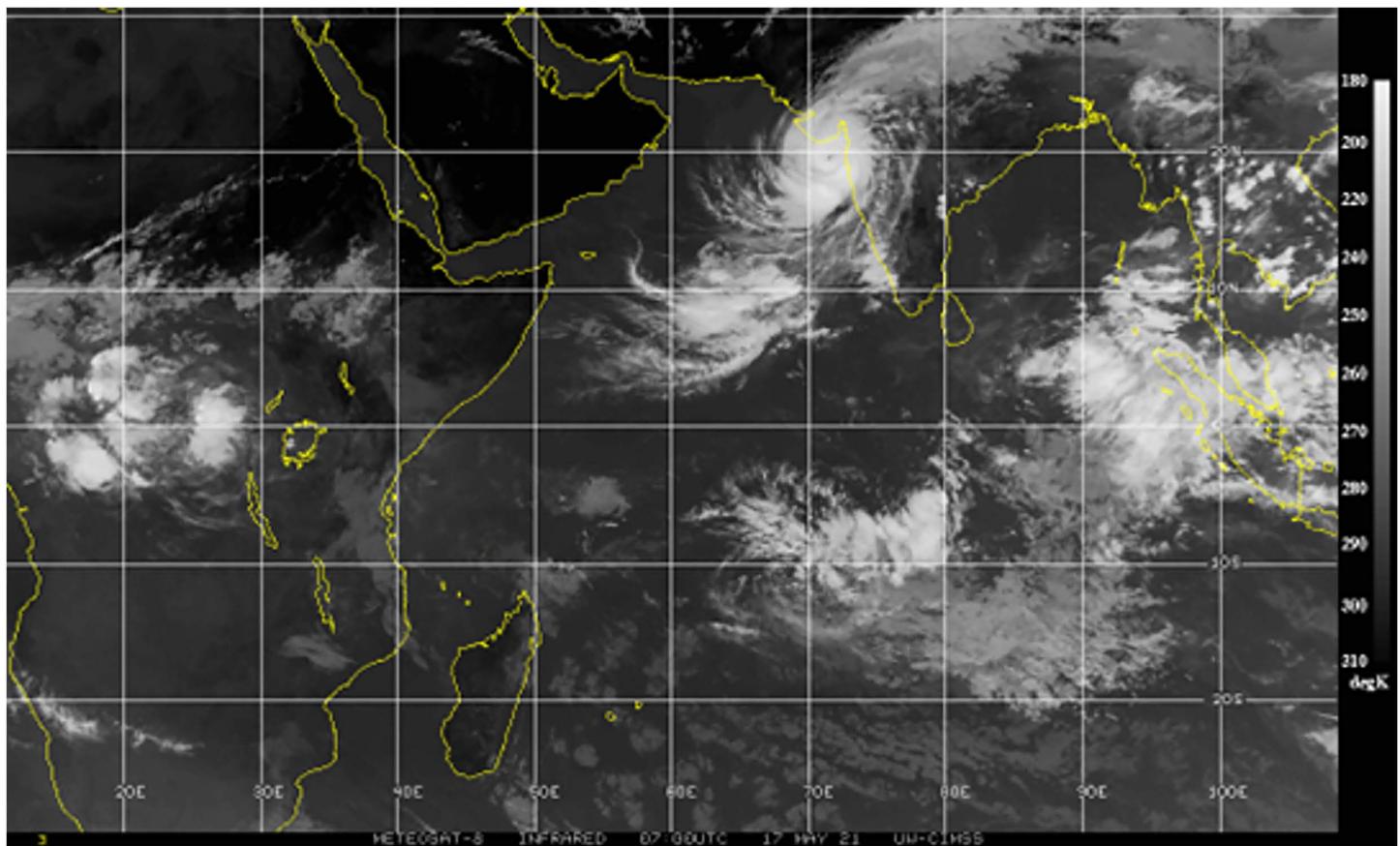


Fig 1. Original satellite image of a tropical cyclone sourced from the cooperative institute for meteorological satellite studies (CIMSS).

<https://doi.org/10.1371/journal.pone.0330705.g001>

Additionally, interpretability remains a challenge, prompting the development of explainable AI (XAI) techniques to make these models more transparent and trustworthy to meteorologists [10]. While this study does not directly incorporate XAI techniques, its development underscores the importance of model interpretability in operational forecasting. Several recent studies have attempted to incorporate deep learning for TC intensity estimation. Many existing models either ignore temporal dependencies or treat sequential image frames independently, limiting their ability to capture evolving storm dynamics. Furthermore, existing approaches often neglect challenges such as image resolution variability, missing frames, and dataset imbalance, which are prevalent in real-world cyclone data.

This study aims to estimate and classify the intensity of tropical cyclones using deep learning models, with a focus on the north Indian Ocean region, a cyclone-prone area with densely populated coastlines, where timely forecasts are crucial for minimizing human and economic losses. The primary objectives include: Binary classification of cyclone and non-cyclone satellite images. Multi-class classification of cyclones into predefined intensity levels and intensity estimation to predict wind speeds. To achieve these objectives, the research work proposes a novel deep learning architecture that combines convolutional layers for spatial feature extraction with ConvLSTM layers for modelling temporal evolution. The proposed methodology addresses several challenges, including class imbalance, noisy backgrounds, and data scarcity in sequential satellite imagery.

The unique contributions and novelty of this research work are:

1. Spatiotemporal Feature Learning: Jointly capture spatial and temporal patterns, improving intensity estimation over models using either alone.
2. Cyclone Region Isolation: A clustering-based method isolates the cyclone area by reducing background noise to enhance feature relevance.
3. Sequence-Level Augmentation: Temporally-consistent data augmentation to better preserve sequence integrity during training.
4. Class Imbalance Handling: The impact of imbalanced cyclone grade classes is addressed using the SMOTE technique, improving classification robustness.
5. Hybrid Architecture: The model architecture combines CNNs for spatial analysis and ConvLSTM for temporal dynamics, offering better representational efficiency than standalone models.

The findings of this study have significant implications for early warning systems, disaster response strategies, and operational meteorology, contributing to the development of more accurate and automated cyclone classification models. The rest of this paper is structured as follows: The related works section reviews related works and existing approaches. The methodology section presents the dataset and methodology used in this work. The result and discussion section includes details experimental results and discussions. The future scope and challenges sections contain the limitations, along with directions for future research. Finally, the conclusions section concludes the study by summarizing the main findings, contributions, limitations, and potential future directions.

Related works

There are many traditional approaches to forecasting tropical cyclones (to include the Dvorak method [3] and numerical weather prediction (NWP) [5] models) which have been very successful in many respects for operational meteorology, but they also come with several limitations associated with their core assumptions. One major issue is that many of these methods, like the Dvorak method, are predicated upon subjective interpretation/analysis, and this can and does introduce intra- and inter-analyst biases [11,12]. Moreover, these methods have challenges that make it difficult to understand qualitatively what 'real' tropical cyclone behavior is where the highly complex, nonlinear, and time-dependent nature of tropical cyclones is concerned. For instance, with the proof-of-concept work using mixed methods, forecasting of rapid intensification and structural evolution requires analysis of short time scales not adequately addressed by Dvorak or NWP methods, an analytical gap which is substantive by previous research [13]. Since these methods have limited generalization that spans multiple basins because of some limitations in data collection and meteorological dynamics, they simply don't work as well outside of their training areas [14]. Additionally, traditional methods fail to provide the spatial-temporal resolution that a reliable real-time intensity estimate would require [15,16]. These limitations make a transition to machine learning (ML) and deep learning (DL) approaches for addressing intensity forecast issues especially relevant in tropical cyclone analysis, as these techniques are designed to handle high-dimensional datasets, temporal dependence, and various environmental conditions of storms.

Recent advancements in deep learning techniques, in particular CNNs and recurrent neural networks (RNNs), have facilitated the estimation of tropical cyclone (TC) intensity: C. Zhang et al. (2021) [12] proposed a TC intensity estimation model called TCICENet which seamlessly combines classification (of TC's) and CNN-based regression on infrared (IR) satellite images; the results achieved RMSE 8.60 kt and MAE 6.67 kt, which outperformed traditional studies and previous DL models. Sattar et al. (2025) [17] made a GRU (Gated Recurrent Unit) model that used high-resolution meteorological data that spans 37 isobaric (pressure) levels, achieved MAE 4.16, 34.35% better than the mean absolute error of the Saf-Net. These models illustrate the advantages brought by spatial and temporal aspects of the data, which help to provide a better tropical cyclone intensity classification estimation. To fill in gaps of either purely data-driven or physics-based models, hybrid models have been explored.

Varalakshmi et al. (2023) [18] combined CNN models (AlexNet, VGG16) with regressors (e.g., SVR – Support Vector Regressor, and Linear Regression) and achieved RMSE values of 4.88 kt. Z. Ma et al. (2024) [19] presented a dual-attention model based on an Xception backbone, which fused multi-scale IR and water vapor images with a Laplacian pyramid fusion and attention modules and achieved 11.4% improvement in RMSE and 8% improvement in MAE. These multi-branch, multimodal models provide a balance of accuracy and complexity, which shows potential for operational forecasting. Transformer-based models and attention mechanisms were explored and improved TC intensity estimation due to their ability to capture long-range dependencies and complex spatial-temporal patterns. Zhao et al. (2025) [15] introduced TFA-Net, a dual-branch transformer net with gated feature fusion on sequential satellite images, resulting in an RMSE of 7.21 kt, and R^2 of 0.93. Tian et al. (2024) [20] developed ViT-TC, a vision transformer with multitask learning, outperforming CNN baselines with MAE 6.49 kt. Zhang et al. (2025) [21] proposed STIA, integrating spatial-to-temporal and temporal-to-spatial attention modules on Himawari-8 image sequences, achieving RMSE 3.61 m/s and MAE 2.83 m/s, surpassing state-of-the-art models by over 10% in RMSE reduction. These works demonstrate transformers' potential for precise and responsive TC intensity prediction.

Multi-task learning (MTL) and cross-basin generalization remain critical challenges. Ding et al. (2024) [22] proposed CBIL-TCIE, a cross-basin incremental learning model predicting maximum sustained wind (MSW) and minimum sea-level pressure (MSLP), mitigating catastrophic forgetting and improving performance by 19.2% over fine-tuning. Zhao et al. (2024) [23] developed MT-GN, integrating shared CNNs with graph-based task embedding to learn task relationships and spatial correlations, achieving 90.37% classification accuracy and RMSE 9.50 kt. These highlight MTL and graph learning as key to robust, transferable TC forecasting. Multi-source data fusion also advances TC estimation accuracy. Xu et al. (2024) [24] proposed FHDTIE, combining satellite imagery and reanalysis data via clustering, U-Net, and graph convolutional networks, achieving lowest RMSE and MAE across 10 datasets. W. Tian et al. (2024) [25] introduced TC-Rolling, fusing multi-source satellite imagery with deviation-angle variance (DAV) via 3D CNN and Convolutional Block Attention Module, reporting RMSE 4.48–13.94 kt and outperforming baselines. W. Tian et al. (2024) [25] introduced TC-Rolling, fusing multi-source satellite imagery with deviation-angle variance (DAV) via 3D CNN and Convolutional Block Attention Module, reporting RMSE 4.48–13.94 kt and outperforming baselines.

Generative models and augmentation are used to address data scarcity. Pang et al. (2021) [26] incorporated DCGAN and YOLOv3 to synthetically augment data. The study achieved 97.78% classification accuracy and 81.39% average precision. Ibrar et al. (2025) [13] applied an LSTM autoencoder and Gaussian filters to detect disturbances in estuarine environments. The tests produced MSE 0.0359, illustrating the ability of unsupervised methods like autoencoders for anomaly detection. Regional models enhance local applicability. Mawatwal and Das (2024) [14] incorporated CNN and YOLO and trained a hybrid architecture based on images from North Indian Ocean satellite data. The resulting model achieved 98.4% accuracy for cyclone detection and 63.83% accuracy for intense classifying of five categories with RMSE 16.2 kt. Pal et al. (2024) [27] proposed Small Skip Net (SSN) to use as a lightweight model where skip connections and residual blocks mitigated vanishing gradients and reduced data load. In testing, SSN achieved 92.35% classification accuracy on 80 bytes of small imagery data from INSAT-3D satellite images, outperforming deeper networks, also offered for real-time operations. Previous ML model types also exhibited effectiveness while being more interpretable than deep learning due to features available to the study models. For example, Kar and Banerjee (2021) [28] demonstrated Random Forest classifiers were trained from a combination of geometric IR satellite features achieved 86.66% accuracy; model correctly classified more than Dvorak. Kar et al. (2019) [11], incorporated multilayer perceptrons and used statistics extracted from spatiotemporal aspects of the satellite to classify intensities with 84% accuracy.

The incorporation of physical models with ML builds upon previous approaches to improve performance. Niu et al. (2025) [29] created a hybrid model by linking the ML-based Pangu with the physics-based WRF model through spectral nudging along with a data assimilation approach that used observed FY-4B satellite data, leading to an overall improvement of 20.3% for typhoon track forecasts, and a reduction of 12.5% typographical intensity error. Cheng et al. (2025) [16]

implemented a convolutional neural network (CNN) model with layers that were rotation-invariant and multi-frame satellite motion, achieving improvements of RMSE and Mean Absolute Error > 22%. Forecasting of rapid intensification (RI) was still problematic. Sharma et al. (2025) [30] used SMOTE to develop an interpretable support vector machine (SVM) model on aspects of the SHIPS dataset achieving a Probabilities of Detection of 0.88 and Heidke Skill Score of 0.492 for the Indian Ocean, equaling the SHIPS-RII-C model, which illustrated the interest in ML and the potential for RIs to be detected with ML. Regardless we propose several limitations to the body of research. Many studies fail to compare their models using standardized datasets for the community to compare the performance. Furthermore, while older methods are often criticized, the majority of works do not appropriately create a barrier of entry to compare old vs new deep learning models. The supply of models that embody lightweight and operationally feasible outcomes that are between low-to-high fidelity – like the event-based super-resolved model (SSN) proposed by Pal et al. (2024) [27] – is sparse. Further, there is little discussion regarding whether interpretability will always be feasible or not, (e.g., most transformer and hybrid CNN-ML frameworks lack interpretability).

In conclusion, the literature from 2019 to 2025 reflects a clear shift from manual and shallow-learning methods to complex, data-driven deep learning approaches. These include multi-task learning, attention mechanisms, transformer architectures, and physically informed modeling strategies. The summary of this methodology is given in [Table 1](#) categorized by approach. Each method is evaluated based on its technique, key advantages, and limitations, providing insight into trends and research gaps in cyclone forecasting. While these methods have led to substantial gains in performance, challenges related to benchmarking, explainability, computational efficiency, and generalization across basins remain open avenues for future research. Bridging these gaps is essential for transitioning from high-accuracy research models to reliable, real-time operational forecasting systems.

Methodology

The methodology begins by inputting and preprocessing satellite images to enhance quality and prepare data for analysis. A binary classification model is then used to detect the presence of a cyclone. If no cyclone is detected, the process ends;

Table 1. Summary of methodologies in cyclone intensity estimation (2019–2025).

Methodology Type	Representative Studies	Technique used	Advantage	Limitation
Traditional Methods	Dvorak Technique, NWP [3,5]	Manual interpretation, rule-based analysis	Operationally established, interpretable	Subjective, low spatial-temporal resolution, basin-specific limitations
CNN & RNN-based Deep Learning	TCICENet [12], GRU-based model [17]	CNNs, residual blocks, 3D CNNs, GRU-based sequence modeling	Learns spatiotemporal patterns; outperforms traditional models	High computational cost; limited explainability
Hybrid CNN+ML Models	Varalakshmi et al. (2023) [18], Z. Ma et al. (2024) [19]	CNN+SVR/LR, attention fusion, multi-branch networks	Enhances accuracy via multimodal fusion and classic ML interpretability	Limited generalization; fusion complexity
Transformer & Attention-based Models	TFA-Net [15], ViT-TC [20], STIA [21]	Vision Transformer, attention fusion, spatial-temporal interactions	Captures long-range dependencies, saliency-aware learning	Resource-intensive; interpretability and real-time viability concerns
Multi-Task Learning (MTL)	CBIL-TCIE [22], MT-GN [23]	Shared CNNs, cross-basin incremental learning, graph-based task modeling	Simultaneous MSW & MSLP prediction; avoids forgetting when adapting	Requires more data and training complexity
Multi-source Fusion Models	FHDTIE [24], TC-Rolling [25]	Satellite+reanalysis fusion, U-Net, CBAM, GCN	Integrates spatial/temporal/environmental features	Fusion design can be ad hoc; difficult to optimize weights

<https://doi.org/10.1371/journal.pone.0330705.t001>

otherwise, detected cyclone images undergo clustering and filtering to refine the data. For detected cyclones, the system performs two parallel tasks: intensity estimation using a regression model and intensity classification using a multiclass classifier, where SMOTE is applied to handle class imbalance. The process outputs both estimation and classification results for further meteorological analysis. [Fig 2](#) illustrates this systematic approach for cyclone intensity estimation and classification using satellite imagery.

Data collection and source

The satellite (long wave Infrared) images are collected from the University of Wisconsin – (CIMSS, <https://tropic.ssec.wisc.edu/>) [31]. Satellite images of 15 tropical cyclones of the North Indian Ocean focusing on the last 10 years of data (2013–2022) are selected for the study. The dataset provides a spatial resolution of 10 km, with a temporal resolution of 3 or 6 hours for continuous monitoring of tropical cyclone development and progression. The IMD Cyclone Best Track Data provided by the India Meteorological Department (IMD) through the (<https://mausam.imd.gov.in/>) [32], maintained by the Regional Specialized Meteorological Centre (RSMC) in Delhi is used for labelling the intensity. The best track data contains information on tropical cyclone parameters, including latitude, longitude, central pressure, and maximum sustained wind speed, recorded at regular intervals. The cyclone intensity measurements are provided with a resolution of 5 knots to ensure accurate classification and analysis of different storm categories.

Data preprocessing

The satellite images undergo preprocessing to enhance feature extraction and classification accuracy. The steps include:

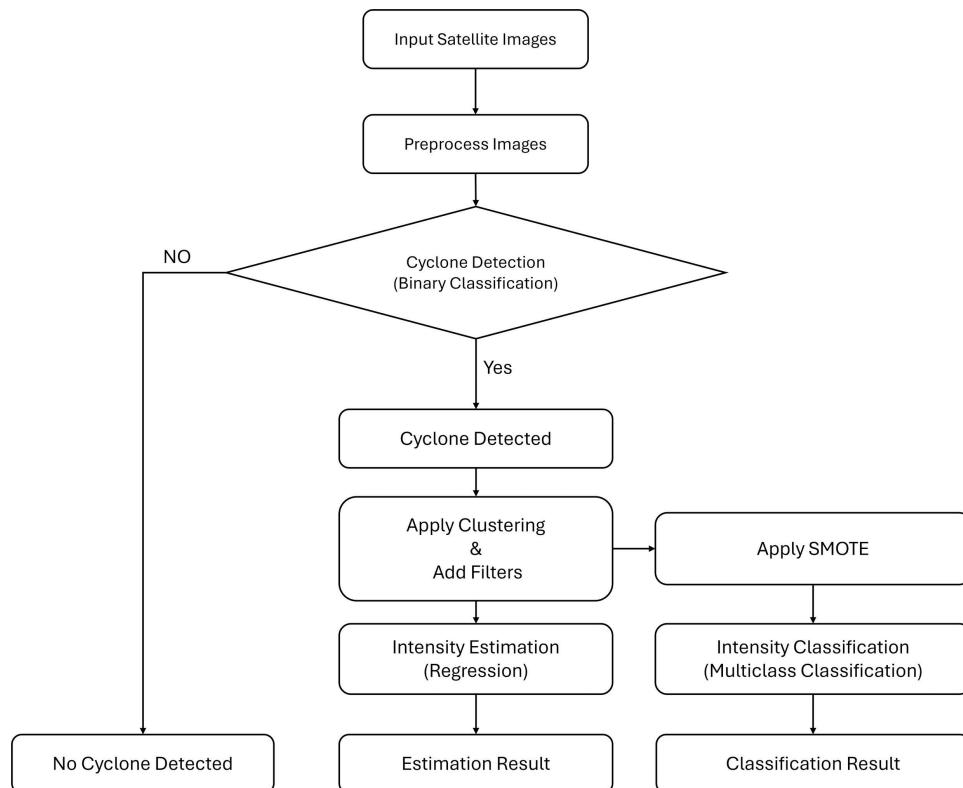


Fig 2. Workflow diagram illustrating the complete cyclone intensity estimation and classification framework proposed in this study.

<https://doi.org/10.1371/journal.pone.0330705.g002>

1. Cropping & Resizing: Each image is cropped to focus on the cyclone region and resized to 310×310 pixels.
2. Gray Scaling: Converts images to a single-channel format, reducing computational complexity.
3. Gaussian Blur: Applied to smooth pixel variations and remove noise.
4. Erosion: Enhances cyclone boundary visibility by reducing small cloud artifacts.

These preprocessing steps aim to reduce background noise and enhance model robustness. [Fig 3](#) illustrates the workflow: (a) the cyclone region is initially cropped to focus on the Indian Ocean only, (b) the satellite images were manually cropped and resized to 310×310 pixels while preserving a 1:1 aspect ratio. This specific resolution helps exclude the cyclone eye from the centre, enhancing model robustness by incorporating more varied, non-cyclonic regions. It strikes a practical balance between spatial detail and computational efficiency, as larger sizes increase memory usage and training time without meaningful performance improvement. Therefore, 310×310 was selected as an effective input size for capturing spatiotemporal cyclone patterns. (c) Morphological erosion is applied to remove grid lines, artificial borders, and other high-contrast artifacts commonly present in satellite images. These elements do not carry meaningful meteorological information and can mislead the model during training. (d) Gaussian blur is used to smooth the image, reducing high-frequency noise and helping the model focus on broader spatial patterns rather than isolated pixel-level variations. Together, these steps enhance the quality of input data and support more stable and generalizable learning. The dataset contains a total of 1,200 images, which are split into training and testing sets using an 80–20 ratio. The split is stratified based on cyclone intensity levels to ensure that each class is proportionally represented in both subsets. To enhance the training dataset, data augmentation techniques such as rotation, horizontal flipping, and zooming are applied, following the approach described in [\[33\]](#).

Binary classification of cyclones

The binary classification (cyclone vs. non-cyclone), utilizes transfer learning with VGG16, ResNet50, and InceptionV3 [\[34–36\]](#). These models are chosen because:

- VGG16: Captures fine-grained spatial details.
- ResNet50: Mitigates vanishing gradient issues through residual connections.
- InceptionV3: Utilizes multi-scale feature extraction for better pattern recognition.

Algorithm 1 describes the training procedure of the transfer learning-based cyclone classification model, making use of pretrained CNNs like VGG16, ResNet50, or InceptionV3. The model takes advantage of extracting spatial features from

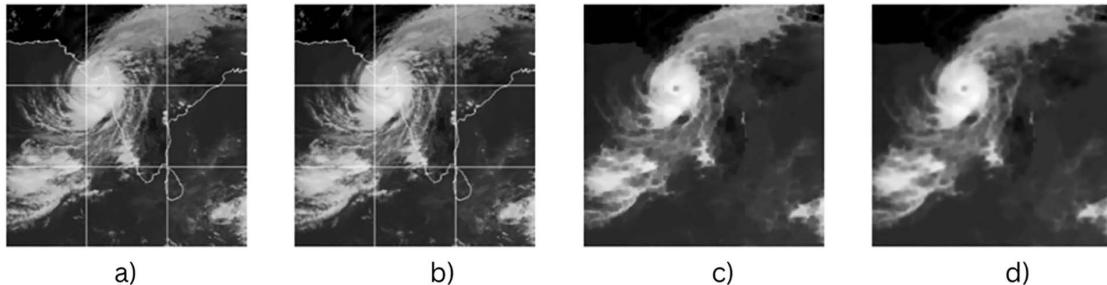


Fig 3. (a) Original satellite image of a cyclone, (b) resized version (310×310 pixels), (c) eroded image highlighting key features, and (d) Gaussian-blurred image used for noise reduction.

<https://doi.org/10.1371/journal.pone.0330705.g003>

the pretrained models, while only tuning the last few for the cyclone task. The models retain their convolutional layers while replacing the output layer with a fully connected layer (256 neurons), dropout layer, and sigmoid output. Training is performed using Adam optimizer ($LR=0.0001$) for 40 epochs with batch size 32. Performance is evaluated using accuracy, precision, recall, F1 score, and a confusion matrix [37]. Fig 4 shows the feature maps extracted from the first convolutional layer (Conv1) of the VGG16 model. (a) shows a non-cyclone image, while (b) depicts a cyclone image. These maps reveal how the model captures low-level spatial features like edges and cloud formations, essential for cyclone detection. Different filters highlight distinct patterns, such as outer cloud bands or the cyclone's eye [34]. The varying activation intensities indicate that VGG16's initial layers focus on edge detection and texture representation, forming the basis for deeper layers to learn cyclone-specific features.

Algorithm 1. Training of transfer learning-based cyclone classification.

Initialize pretrained CNN M (VGG16, ResNet50, InceptionV3) with frozen convolutional base, retaining spatial feature extraction.

1	Load & preprocess dataset $\mathcal{D} = \{(X_i, Y_i)\}_{i=1}^N$, where X represents cyclone/non-cyclone images, and Y represents labels. Resize X to 224×224 , normalize $X \leftarrow X/255.0$, augment using random rotation, flipping, zooming, and split into $\mathcal{D}_{train}, \mathcal{D}_{val}$
2	Extract hierarchical features via $F_M(X)$, flatten convolutional outputs from layer L , and pass through dense layers with ReLU activation, dropout $p = 0.5$, and batch normalization
3	Attach classifier $C(x)$ with fully connected layers, ReLU activations, and final sigmoid activation for binary classification.
4	Define loss function $\mathcal{L} = -\frac{1}{N} \sum_{i=1}^N y_i \log \hat{y}_i + (1 - y_i) \log (1 - \hat{y}_i)$, optimize using Adam ($\alpha = 10^{-4}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$).
5	Train model for T epochs using mini-batch SGD (batch size $m = 32$), fine-tuning last k convolutional layers, employing early stopping and checkpointing the best weights based on validation loss.

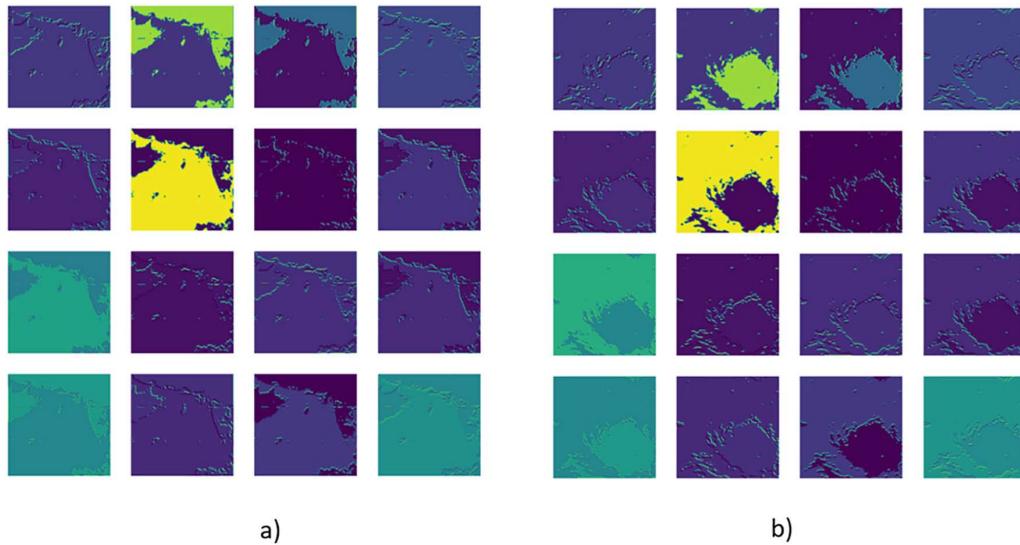


Fig 4. Feature maps from the first convolutional layer (Conv1) showing activations for (a) a non-cyclone image and (b) a cyclone image, illustrating early layer feature extraction differences.

<https://doi.org/10.1371/journal.pone.0330705.g004>

Cyclone segmentation techniques

To isolate cyclone regions from satellite images, various clustering techniques are applied:

- K-Means Clustering
- DBSCAN (Density-Based Spatial Clustering of Applications with Noise) [38]
- Agglomerative Clustering
- Gaussian Mixture Models (GMM)
- Mean Shift

A comparative analysis is conducted to determine the most effective segmentation technique, using the silhouette score to measure performance based on cohesion and separation. Texture-based feature extraction methods further enhance segmentation accuracy and reduce background noise. Gabor filtering detects swirling cyclone structures by convolving the grayscale image with orientation-specific kernels [39], while Local Binary Pattern (LBP) captures fine-grained texture variations to distinguish the cyclone from surrounding cloud formations. The Canny edge detector enhances intensity gradients, refining the cyclone's boundary [40].

To achieve precise segmentation, the extracted feature maps are combined with the K-Means mask using pixel-wise logical operations. Morphological filtering is used to remove small cloud artifacts and noise. Techniques such as opening (erosion followed by dilation) eliminate minor unwanted regions while closing (dilation followed by erosion) ensures cyclone structures remain intact, minimizing false positives and improving segmentation reliability. [Fig 5](#) illustrates this process: (a) the original satellite image, (b) K-Means segmentation output, (c) the feature-based mask integrating texture and edge features. The feature mask, refined using Gabor filtering, LBP, and Canny edge detection, ensures better segmentation and is used for training classification models. After segmentation, contour detection identifies the cyclone's ROI, with the largest contour representing the cyclone. Furthermore, classification models trained on labelled cyclone data categorize the cyclone as a tropical depression, storm, or severe cyclone [41].

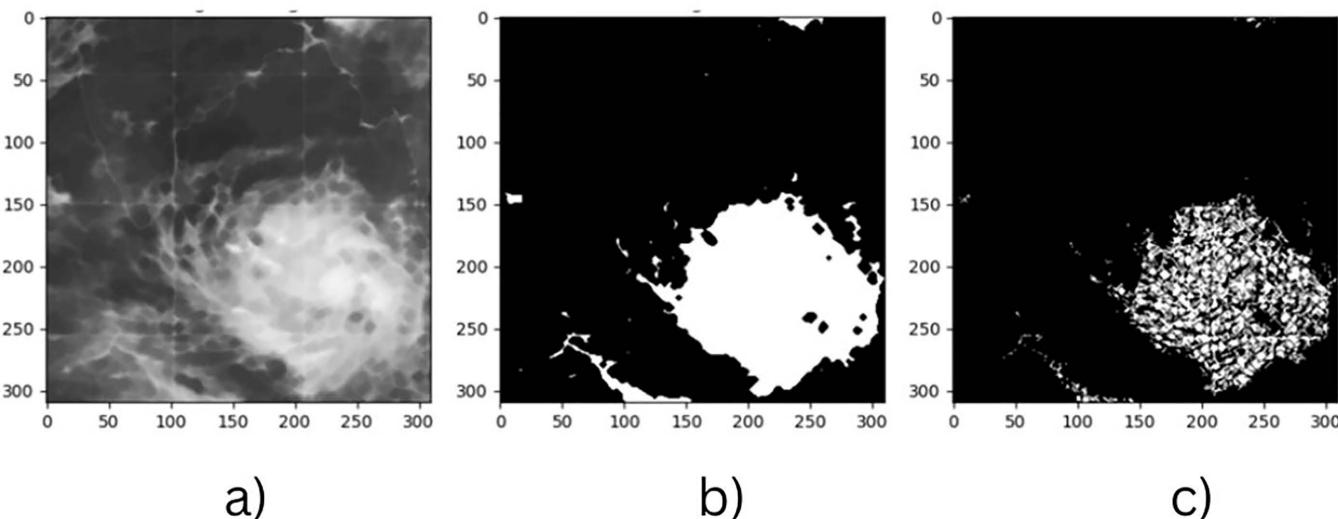


Fig 5. (a) Pre-processed satellite image, (b) K-Means clustered segmentation output, and (c) feature-based mask highlighting cyclone-relevant regions extracted.

<https://doi.org/10.1371/journal.pone.0330705.g005>

Cyclone intensity estimation and classification

The proposed cyclone intensity and classification framework is designed to estimate and categorize multiple cyclone intensity levels using sequences of satellite images and a deep spatiotemporal learning model. The architecture integrates TimeDistributed convolutional layers with ConvLSTM, forming a hybrid model tailored to the spatiotemporal nature of cyclone evolution. ConvLSTM networks are particularly suited for spatiotemporal data, where both the spatial structure and their temporal evolution are critical. Traditional CNNs or LSTMs, when used in isolation, fail to simultaneously capture these joint dependencies. The ConvLSTM layer, by applying convolutional operations within the recurrence, allows the model to learn motion-aware spatial features across sequential frames, an essential aspect of tropical cyclone dynamics. The process begins by providing a cyclone intensity label to categorical text in the dataset. It is numerically encoded using label encoding to make it compatible with the softmax-based classification output. To model temporal patterns of cyclone evolution, images are grouped into sequential frames, where each sample consists of four consecutive satellite images. Because we are taking satellite images with a six-hourly gap, only four satellite images will be there for a day. The target label for a sequence corresponds to the cyclone intensity of the last frame in that sequence. To enhance the model's ability to generalize and learn robust spatial features, frame-wise data augmentation is performed [33].

Specifically, each frame in a sequence undergoes random transformations independently. These include horizontal flipping, rotations up to 30 degrees, zooming, and width/height shifts of up to 30%. This technique allows the model to learn features invariant to orientation, scale, and position, while preserving the temporal structure of the cyclone evolution across sequences. Unlike standard augmentation approaches applied only once per image, this frame-wise augmentation is embedded directly into the preprocessing pipeline, increasing the diversity of samples while maintaining sequence integrity. To ensure generalization and minimize bias, the model is evaluated using 5-fold cross-validation and is averaged across folds for both classification and estimation. Early stopping is applied to terminate training once convergence is detected, reducing the risk of overfitting.

As shown in [Fig 6](#), the estimation model begins by extracting spatial features using TimeDistributed Conv2D layers, which apply convolutional operations independently to each frame in a four-image sequence. [Table 2](#) describes the Layer-wise comparison of the ConvLSTM-based architectures used for cyclone wind speed estimation and intensity classification. Differences in convolutional depth, filter size, and dropout rates highlight trade-offs in model complexity and target task. The estimation network includes three convolutional layers with 32, 64, and 128 filters (3×3 kernels, ReLU activation), each followed by 2×2 max pooling and dropout layers (rates of 0.2, 0.3, and 0.4, respectively) to downsample feature maps and reduce overfitting. These frame-level features are passed to a ConvLSTM2D layer with 128 filters and a 3×3 kernel, which captures both spatial patterns and their temporal evolution. Unlike traditional LSTMs, ConvLSTM retains spatial structure, enabling it to detect motion cues like spiral rainbands or eye development. The output from the ConvLSTM layer is flattened and passed through a Dense layer with 256 units and a 0.5 dropout rate, followed by a final output neuron that predicts the normalized wind speed. The model is trained using the Huber loss function, chosen for its robustness to outliers, and optimized with Adam (learning rate = 0.001).

Algorithm II. Training of ConvLSTM for Cyclone Intensity estimation.

Initialize model parameters θ , ω , learning rate α , batch size m , penalty coefficient λ , and Adam optimizer hyperparameters β_1 , β_2 .

1	Load dataset $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N$, where x_i represents input images and y_i represents wind speeds.
2	Normalize wind speeds $y = \text{StandardScaler}(y)$
3	Construct time-series sequences $S_t = \{x_{t-k}, \dots, x_t\}$
4	For $k = 1, 2, \dots, N$ do
5	For $t = 1, 2, \dots, n_{\text{critic}}$ do
6	For $i = 1, 2, \dots, m$ do

Initialize model parameters θ , ω , learning rate α , batch size m , penalty coefficient λ , and Adam optimizer hyperparameters β_1 , β_2 .

7	Sample a sequence $S_t \sim p(S)$ and corresponding wind speed label y_t .
8	Pass S_t through TimeDistributed CNN layers with ReLU activations and dropout regularization.
9	Extract temporal features using ConvLSTM with 3D kernels.
10	Flatten features and pass through fully connected dense layers.
11	Compute loss $L^{(i)} = \text{HuberLoss}(y_{\text{pred}}, y_t)$
12	End For
13	Update critic parameters $\omega \leftarrow \text{Adam}(\nabla_{\omega} \frac{1}{m} \sum_{i=1}^m L^{(i)}, \omega, \alpha, \beta_1, \beta_2)$.
14	End For
15	Sample new mini-batch $\{S_t^{(i)}\}_{i=1}^m$
16	Update generator parameters $\theta \leftarrow \text{Adam}(-\nabla_{\theta} \frac{1}{m} \sum_{i=1}^m L^{(i)}, \theta, \alpha, \beta_1, \beta_2)$
17	End For
18	Store the best model weights based on validation loss.
19	Train until early stopping criteria is met.

Algorithm II describes the training process in detail, including sequence sampling, feature extraction, and parameter updates. A known challenge in cyclone datasets is class imbalance, where certain intensity classes may be underrepresented. To mitigate this, the Synthetic Minority Over-sampling Technique (SMOTE) is applied to the training data [42].

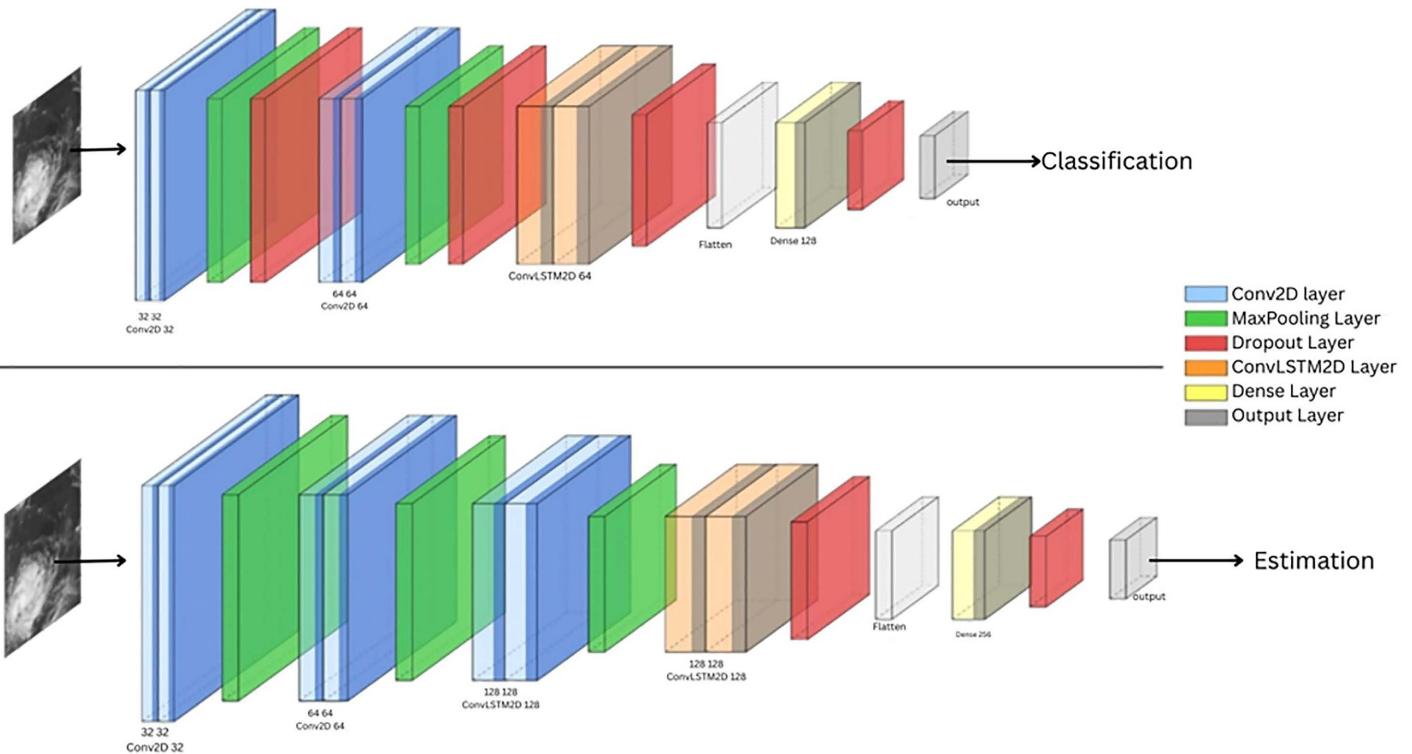


Fig 6. Schematic block diagram of the ConvLSTM-based deep learning architecture employed for cyclone intensity estimation and classification.

<https://doi.org/10.1371/journal.pone.0330705.g006>

Table 2. Model architecture comparison of estimation vs classification.

Layer Index	Layer Type	Estimation Model	Classification Model
1	TimeDistributed Conv2D	Filters: 32, Kernel: (3,3), Activation: ReLU	Filters: 32, Kernel: (3,3), Activation: ReLU, L2(0.01)
2	TimeDistributed MaxPooling2D	Pool Size: (2,2)	Pool Size: (2,2)
3	TimeDistributed Dropout	Rate: 0.2	Dropout (not TimeDistributed), Rate: 0.3
4	TimeDistributed Conv2D	Filters: 64, Kernel: (3,3), Activation: ReLU	Filters: 64, Kernel: (3,3), Activation: ReLU, L2(0.01)
5	TimeDistributed MaxPooling2D	Pool Size: (2,2)	Pool Size: (2,2)
6	TimeDistributed Dropout	Rate: 0.3	Dropout (not TimeDistributed), Rate: 0.4
7	TimeDistributed Conv2D	Filters: 128, Kernel: (3,3), Activation: ReLU	<i>Not included</i>
8	TimeDistributed MaxPooling2D	Pool Size: (2,2)	<i>Not included</i>
9	TimeDistributed Dropout	Rate: 0.4	<i>Not included</i>
10	ConvLSTM2D	Filters: 128, Kernel: (3,3), Activation: ReLU	Filters: 64, Kernel: (3,3), Activation: ReLU, L2(0.01)
11	Dropout	Rate: 0.5	<i>Not included</i>
12	Flatten	Yes	Yes
13	Dense	Units: 256, Activation: ReLU, L2 Regularization	Units: 128, Activation: ReLU, L2(0.01)
14	Dropout	Rate: 0.5	Rate: 0.5
15	Output Dense	Units: 1 (Regression)	Units: len(label_encoder.classes_), Activation: Softmax (Classification)

<https://doi.org/10.1371/journal.pone.0330705.t002>

Before resampling, image sequences are reshaped into flat feature vectors. SMOTE then generates synthetic examples for the minority classes to balance the class distribution. After resampling, the data is reshaped back into the original temporal sequence format suitable for deep learning.

Fig 7 shows the class distribution before and after applying SMOTE in this study, showing how class imbalance among cyclone intensity categories is corrected through synthetic oversampling. After data augmentation and before applying SMOTE, the class distributions were 451, 202, 450, 306, and 347. After applying SMOTE, each

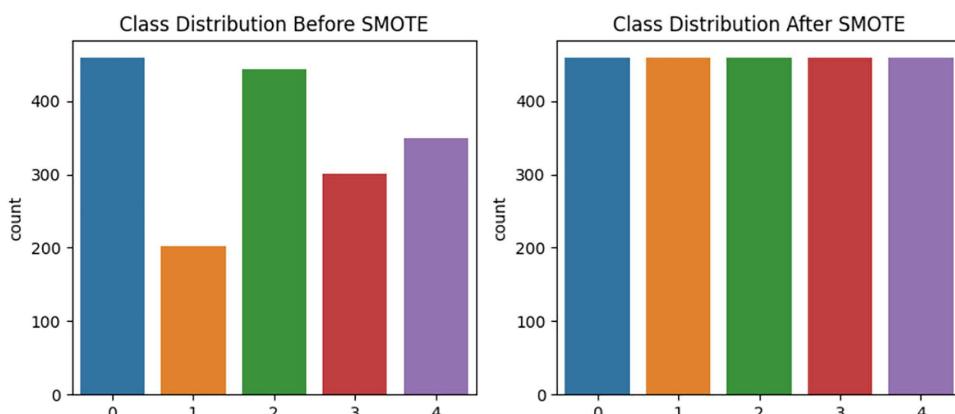


Fig 7. Class distribution of cyclone intensity categories before and after applying SMOTE to the training dataset.

<https://doi.org/10.1371/journal.pone.0330705.g007>

class was balanced to 451 samples. The classification model shares a similar architecture with the estimation model, with key differences in depth and complexity. It includes only two TimeDistributed Conv2D layers (with 32 and 64 filters, respectively), compared to the deeper convolutional stack in the estimation model. Additionally, the ConvLSTM2D layer in the classification network uses 64 filters, whereas the estimation model employs 128 filters. These modifications reduce the model's complexity while preserving its ability to capture essential spatiotemporal features required for classifying cyclone intensity. The output of the ConvLSTM is flattened and passed to a dense layer with 128 neurons and ReLU activation, followed by a dropout layer with a rate of 0.5. Finally, a softmax output layer maps the output to multiple cyclone intensity classes.

Algorithm III. Training of ConvLSTM for Cyclone Intensity Classification

Input: $\{X_n, y_n, g_n\}$, where $n = 1, 2, \dots, N$; hyperparameters α, λ, β maximum epochs E

Output: Trained model $f(X; \theta)$

1	Initialize θ randomly
2	Load dataset $\{X_n, y_n\}$ from source and apply preprocessing $\mathcal{P}(X_n)$, including resizing, normalization, and augmentation
3	Encode labels y_n using label mapping $\mathcal{L}(y_n)$
4	Construct temporal sequences $\mathcal{S}_t = \{X_t^{(i)}, y_t^{(i)}\}_{i=1}^M$ with time steps T
5	Split \mathcal{S}_t into training set \mathcal{D}_{train} and testing set \mathcal{D}_{test} using stratified sampling
6	Reshape X_n into (M, T, H, W, C) format for ConvLSTM processing
7	while θ not converged & maximum epochs E not reached do
8	for each mini-batch $\mathcal{B} \subset \mathcal{D}_{train}$ do
9	Pass \mathcal{B} through ConvLSTM model: $c_t = \sigma(W_c * X_t + b_c)$ for convolutional layers $h_t = \sigma(W_h * c_t + U_h * h_{t-1} + b_h)$ for LSTM layers
10	Compute loss $\mathcal{L}(\theta)$ as sparse categorical cross-entropy with regularization: $\mathcal{L}(\theta) = -\sum_i y_i \log(\hat{y}_i) + \lambda \ \theta\ ^2$
11	Compute gradient $\nabla_{\theta} \mathcal{L}(\theta)$ via backpropagation
12	Update model parameters using Adam optimizer: $\theta \leftarrow \theta - \alpha \nabla_{\theta} \mathcal{L}(\theta)$
13	end for
14	Compute validation loss $\mathcal{L}_{val}(\theta)$ on \mathcal{D}_{test}
15	Apply learning rate scheduler and early stopping conditions if necessary
16	end while
17	Evaluate final model $f(X; \theta)$ on \mathcal{D}_{test} using accuracy, precision, recall, and F1-score
18	Return trained model θ

Algorithm III outlines the procedure for training the ConvLSTM model for cyclone intensity classification. It includes input preprocessing, sequence construction, data augmentation, loss computation, and performance evaluation. Training is performed using the Adam optimizer with a learning rate of 0.0005. The loss function is sparse categorical cross-entropy, which is regularized with L2 penalties to prevent overfitting. Early stopping (patience = 3) and a learning rate reduction mechanism (factor = 0.5, patience = 2) are used to optimize training convergence [43]. Training occurs over a maximum of 50 epochs with a batch size of 32. All experiments were executed on a high-performance computing (HPC) cluster using the max_dgx queue with a job configuration of 1 node, 10 CPUs, and 2 NVIDIA GPUs on a DGX. The allocated GPUs ensured efficient training and evaluation of the deep learning model, with hardware availability and utilization verified dynamically during runtime.

Results and discussion

Binary classification of cyclone

The binary classification of cyclone intensity was performed using three pre-trained deep learning models, VGG16, ResNet50, and InceptionV3, alongside a classical baseline model, Logistic Regression. [Fig 8](#) illustrate a visual comparison across different models used for binary classification of cyclone vs. non-cyclone images. VGG16 outperforms others, highlighting its strong feature extraction capability for cyclone detection. [Table 3](#) provides Binary classification performance of different models including VGG16, ResNet50, InceptionV3, and Logistic Regression. VGG16 achieves the highest accuracy and F1 score, demonstrating its effectiveness in distinguishing cyclone from non-cyclone imagery. VGG16 achieved the highest accuracy (99.16%), followed by InceptionV3 (98.33%) and ResNet50 (96.67%), whereas Logistic Regression obtained a lower accuracy of 93.52%. The deep learning models significantly outperformed the baseline, demonstrating their superior ability to capture complex spatial patterns in cyclone imagery. While ResNet50 introduces residual learning, its performance slightly lags due to its complex architecture, which may require more data for optimal generalization [\[44\]](#).

Clustering algorithm performance in cyclone segmentation

A comparative analysis of different clustering algorithms K-Means, Mean Shift, Agglomerative Clustering, DBSCAN, and Gaussian Mixture, based on their average Silhouette Scores is done and shown in [Fig 9](#). Average Silhouette Scores for different clustering algorithms used in cyclone region isolation. Mean Shift achieved the highest score, indicating superior cluster separation and cohesion for segmenting cyclone-relevant regions in satellite imagery. K-Means followed with a score of 0.7398, showing strong clustering performance. Agglomerative Clustering and Gaussian Mixture had moderate scores of 0.7182 and 0.6899, respectively. DBSCAN performed the worst with a score of 0.6472, suggesting that it may

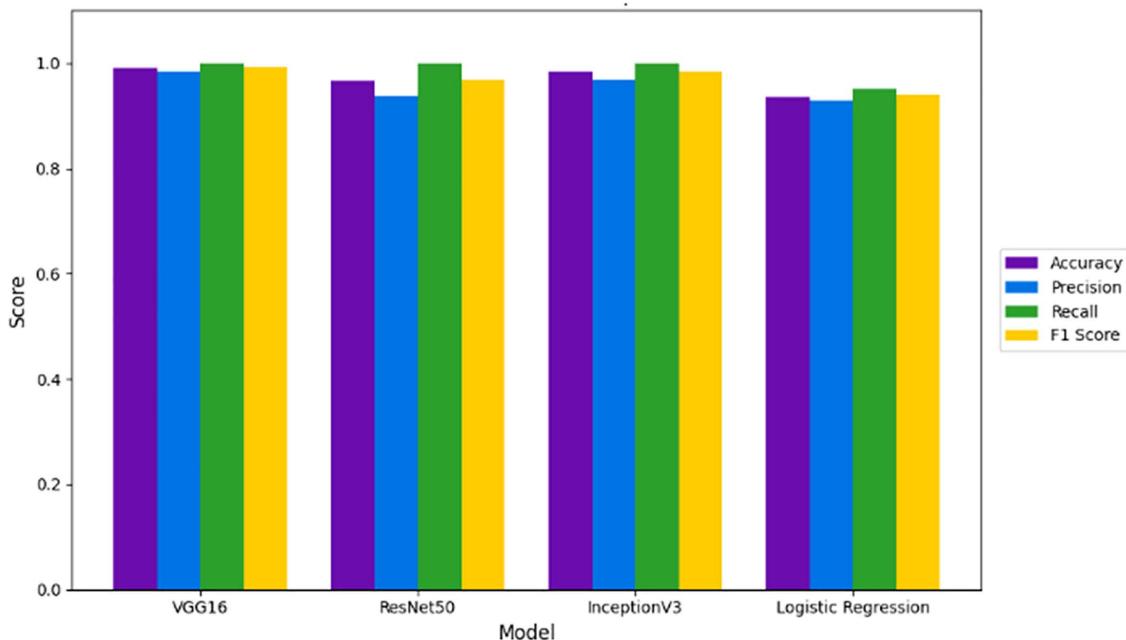


Fig 8. Comparative performance of different models for binary cyclone classification.

<https://doi.org/10.1371/journal.pone.0330705.g008>

Table 3. Performance metrics of various models for binary cyclone intensity classification.

Model	Accuracy	Precision	Recall	F1 Score
VGG16	0.991667	0.983607	1.0	0.991736
ResNet50	0.966667	0.937500	1.0	0.967742
InceptionV3	0.983333	0.967742	1.0	0.983607
Logistic Regression	0.9352	0.93	0.95	0.94

<https://doi.org/10.1371/journal.pone.0330705.t003>

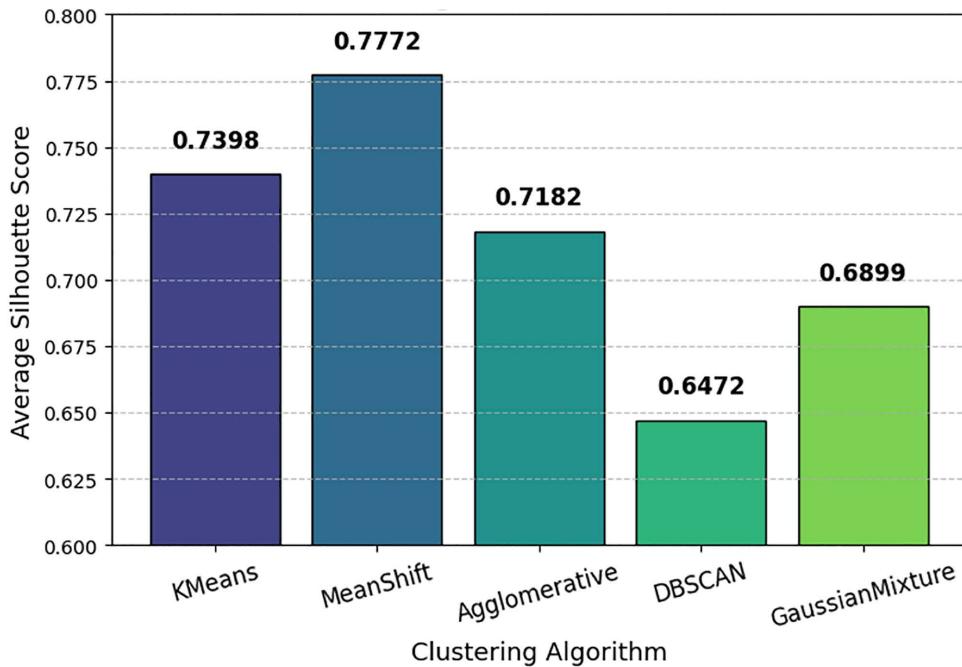


Fig 9. Comparison of clustering algorithms (K-Means, MeanShift, etc.) based on the average Silhouette Score.

<https://doi.org/10.1371/journal.pone.0330705.g009>

struggle with clear boundary definition in cyclone segmentation. This comparison helps in selecting the most effective clustering technique for cyclone feature extraction.

Cyclone intensity estimation and classification model evaluation

Both K-Means and MeanShift clustering algorithms were employed to preprocess the dataset before estimation and classification, and their impact on performance was evaluated.

[Fig 10](#) and [Fig 11](#) present the training and validation error curves - Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), respectively across 5-fold cross-validation on the Meanshift dataset. In both Figs, the error consistently decreases across epochs for each fold, indicating effective convergence of the model during training. The validation curves closely follow the training trends in most folds, suggesting good generalization with minimal overfitting. While minor fluctuations in validation metrics are observed, especially in MAE plots for folds 2 and 5, the overall trend confirms stable model learning and robustness across all folds. The model achieved an average RMSE of 8.31 ± 1.10 over five folds, with a total training time of 9742.57 seconds, reflecting both the computational cost and predictive accuracy of the proposed spatiotemporal deep learning framework.

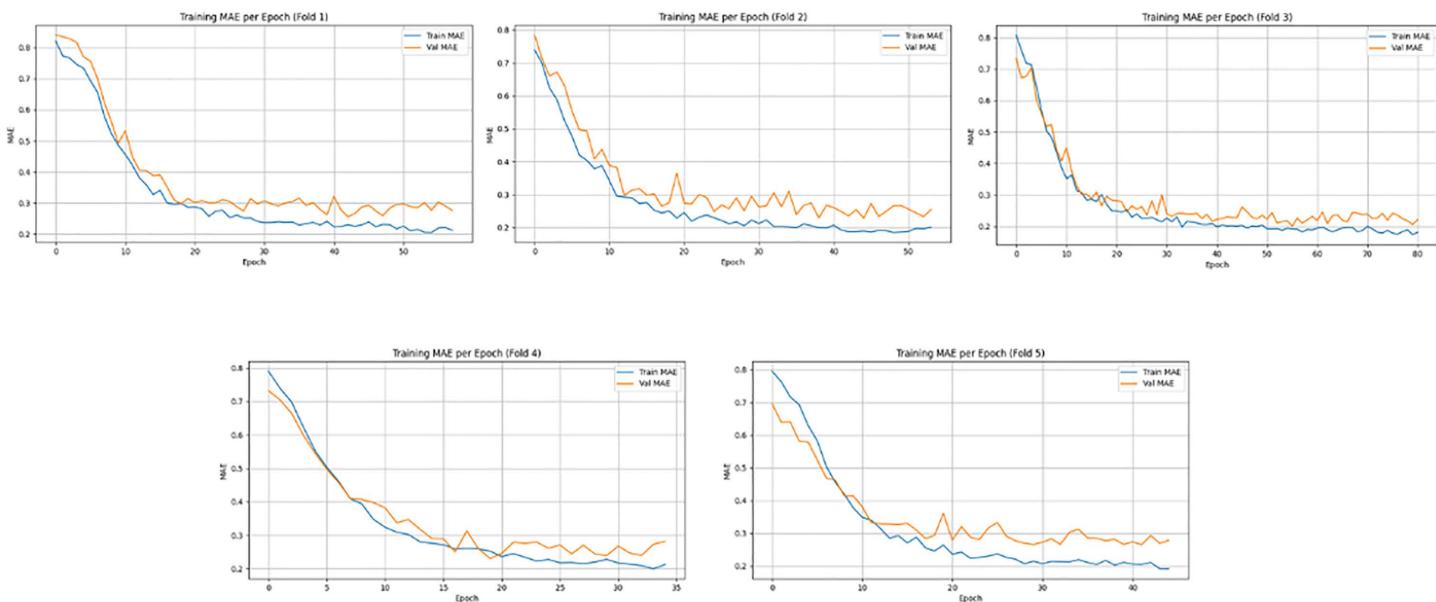


Fig 10. Training and validation Mean Absolute Error (MAE) per epoch across five-fold cross-validation using the MeanShift clustered dataset.

<https://doi.org/10.1371/journal.pone.0330705.g010>

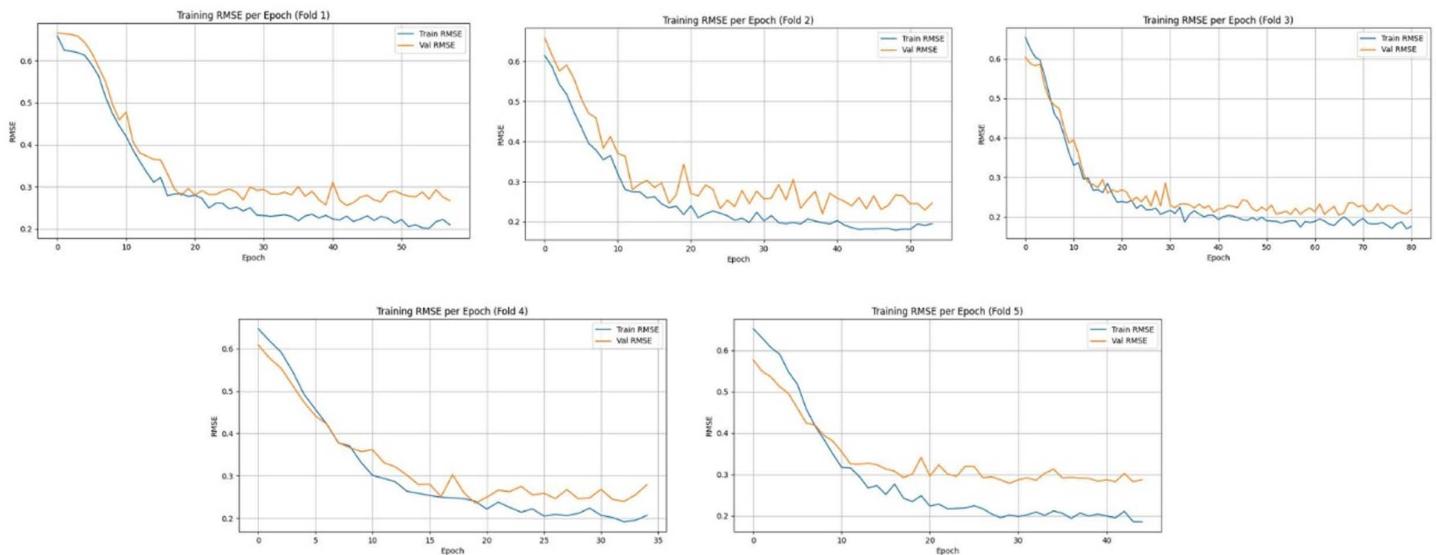


Fig 11. Training and validation Root Mean Square Error (RMSE) per epoch across five-fold cross-validation using the MeanShift clustered dataset.

<https://doi.org/10.1371/journal.pone.0330705.g011>

[Fig 12](#) and [Fig 13](#) illustrate the training and validation MAE and RMSE per epoch, respectively, across 5-fold cross-validation for the enhanced experimental setup. In all five folds, both MAE and RMSE show a consistent decreasing trend, indicating effective model convergence and generalization. The close alignment between training and validation curves suggests that the model is not overfitting and is robust across different validation splits. Minor fluctuations in the validation RMSE, especially in folds 2 and 5, are within acceptable bounds and highlight

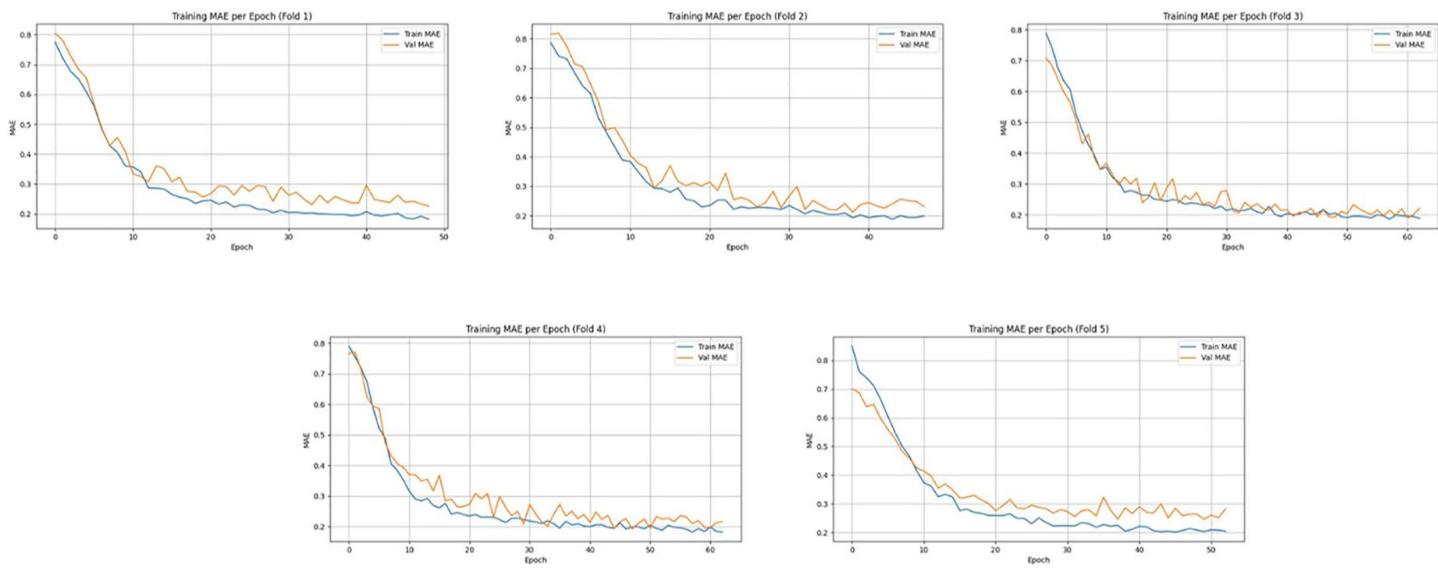


Fig 12. Training and validation MAE per epoch across five-fold cross-validation using the K-Means clustered dataset.

<https://doi.org/10.1371/journal.pone.0330705.g012>

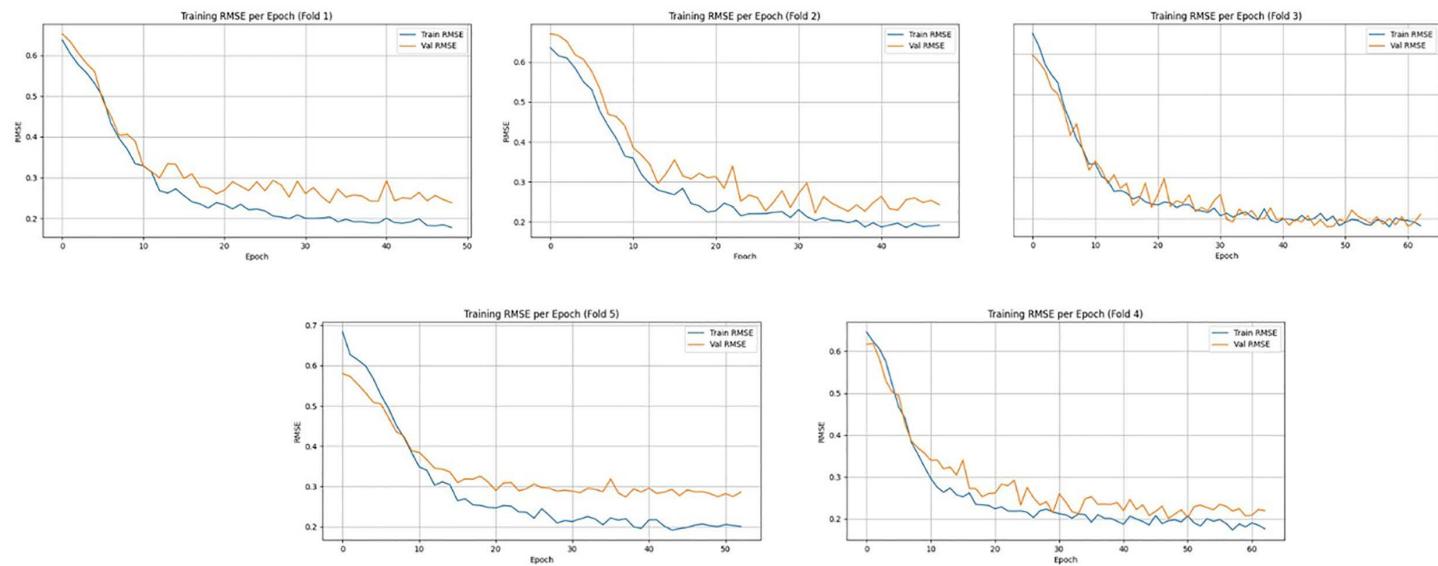


Fig 13. Training and validation RMSE per epoch across five-fold cross-validation using the K-Means clustered dataset.

<https://doi.org/10.1371/journal.pone.0330705.g013>

the natural variability in temporal satellite-based intensity estimation. The model achieved an average RMSE of 7.79 ± 1.27 with a total training time of 7743.23 seconds, demonstrating improved performance and computational efficiency over prior configurations.

Based on the experimental results, the K-Means-based framework outperformed the MeanShift-based setup, achieving a lower average RMSE of 7.79 ± 1.27 compared to 8.31 ± 1.10 . Additionally, K-Means required less training time (7743.23 sec) than MeanShift (9742.57 sec), highlighting its computational efficiency. While MeanShift is more adaptive

to complex cluster shapes, K-Means proved more effective and stable for cyclone region isolation in this context. Hence, K-Means was selected as the optimal clustering approach for the proposed model.

The MeanShift-clustered dataset achieved slightly higher average accuracy (0.8110) than the KMeans-based approach (0.8015), suggesting it forms more effective clusters for classification tasks. However, this came at the cost of a longer average training time per fold (1776.18s vs. 1504.80s). Despite having more misclassified samples (106 vs. 94), MeanShift demonstrated better class separation, especially in the confusion matrix for D and SCS classes. On the other hand, K-Means had fewer misclassifications and trained faster, making it more computationally efficient but slightly less accurate. [Fig 14](#) shows a misclassified sample from the K-Mean cluster dataset where the true class is 'CS' (Cyclonic Storm), but the model predicted 'VSCS' (Very Severe Cyclonic Storm). The evaluation results are summarized in [Table 4](#), and corresponding confusion matrices are presented in [Fig 15 a](#) (KMeans) and [Fig 15 b](#) (MeanShift). The confusion matrix for K-Means ([Fig 15 a](#)) shows better precision for the CS and DD classes, while MeanShift ([Fig 15 b](#)) improves classification for D and SCS, suggesting it handles more ambiguous classes slightly better. Yet, MeanShift also demonstrates a higher tendency toward class confusion, particularly between neighboring intensity levels like D and DD. MeanShift is preferable when accuracy and deeper class separation are the priorities, ideal for applications requiring high reliability in cyclone prediction. K-Means is a better choice when efficiency and lower training overhead are critical, such as in time-sensitive or resource-constrained forecasting systems.

The proposed spatiotemporal deep learning framework demonstrated superior performance in both cyclone intensity estimation and classification tasks compared to baseline models. While MeanShift achieved a slightly higher classification accuracy (81.10%) than K-Means (80.15%), it incurred greater training time and more misclassifications, particularly in regions with ambiguous class boundaries. K-Means proved to be more computationally efficient and more balanced in precision across multiple classes, making it ideal for time-sensitive applications. [Table 5](#) describes the performance comparison of the proposed framework against baseline models (LSTM, GRU, CNN) for both cyclone intensity classification and wind speed estimation. The proposed model outperforms others in classification accuracy and estimation error (RMSE), achieving the highest classification accuracy (81.1 ± 4.33) and competitive estimation RMSE (8.31 \pm 1.10 RMSE).

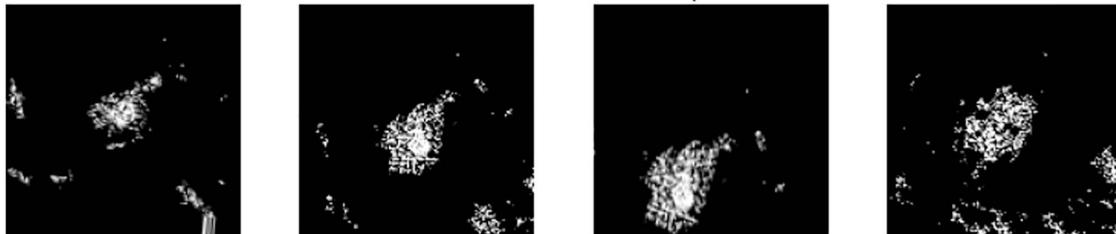


Fig 14. Example of misclassification, showing four consecutive cyclone images where the true class (Cyclonic Storm, CS) was incorrectly predicted as Very Severe Cyclonic Storm (VSCS).

<https://doi.org/10.1371/journal.pone.0330705.g014>

Table 4. K-Means vs. Meanshift clustered datasets for cyclone intensity classification.

Metric	KMeans	MeanShift
Misclassified Samples	94	106
Accuracy (Avg over 5 folds)	0.8015 ± 0.0154	0.8110 ± 0.0433
Training Time (Fold 5)	1254.05 seconds	1319.94 seconds
Avg Training Time per Fold	1504.80 seconds	1776.18 seconds

<https://doi.org/10.1371/journal.pone.0330705.t004>



Fig 15. Confusion matrices illustrating cyclone intensity classification performance using (a) the K-Means clustered dataset and (b) the Mean-Shift clustered dataset.

<https://doi.org/10.1371/journal.pone.0330705.g015>

Table 5. Performance comparison of the proposed framework with baseline architectures.

Sl.no	Architecture	Intensity classification	Intensity estimation
1	Proposed	81.1 ± 4.33	7.79 ± 1.27
2	LSTM	68.18 ± 4.65	9.35 ± 1.9
3	GRU	69.09 ± 4.6	8.18 ± 0.7
4	CNN [14]	63.83 ± 1.3	16.2 ± 0.9

<https://doi.org/10.1371/journal.pone.0330705.t005>

These results confirm the effectiveness of the proposed hybrid framework in capturing cyclone dynamics with higher precision and robustness, making it a strong candidate for operational tropical cyclone analysis systems.

Future scope and challenges

In this study, one of the principal issues is the relatively small dataset (1,200 images) which restricts a model's capability to learn robust and generalizable spatiotemporal features. This problem is significant for rare or extreme cyclone intensity classifications, as greater representation is required due to their relatively low frequency to avoid biased learning and subsequently poor classification performance. Moreover, increased risk of overfitting is another consequence of limited representation, particularly for deep learning models which rely heavily on large datasets. To partially mitigate this limitation, image augmentation techniques such as rotation, zoom, and flipping were applied to artificially expand the training dataset and introduce variability. However, further dataset expansion through data synthesis remains necessary.

While deep learning models have drastically improved cyclone intensity estimation and classification, data availability and quality are among the key challenges left to address, where future research could be directed [20]. To enhance cyclone intensity classification, model architecture is one of the most insightful-providing future research directions. ConvLSTM has been confirmed to be effective for capturing spatio-temporal dependencies [45], while developing hybrid deep

learning models could further improve performance. For instance, merging ConvLSTM with Transformer-based architectures such as Vision Transformers (ViTs) could enhance the ability to extract long-range features. The dataset would be augmented by providing features to the model that can learn more generalizable patterns. Additionally, future work could incorporate Monte Carlo dropout during inference to estimate predictive uncertainty through multiple stochastic forward passes.

Alternatively, Bayesian deep learning techniques, such as variational inference, can be used to model parameter uncertainty, which offers more reliable forecasts for operational meteorology.

Although SMOTE was applied to reduce class imbalance, there is still minimal representation of cyclone intensity categories, particularly Very Severe Cyclonic Storms (VSCS) or above, contributing to a classification bias and thereby decreasing model sensitivity to these potential high-impact events. Model underrepresentation in severe cyclone categories leads to the misclassification of severe cyclones, which is detrimental to early warning systems. Perhaps the next step can help alleviate this feature underrepresentation through more advanced handling of imbalance such as using adaptive synthetic oversampling, cost sensitive learning, or category specific weighted loss functions. A future consideration could also be to develop either class-conditional synthetic data generation using GANs or variational autoencoders to produce realistic storm examples from the rare categories, which could better support model robustness and recall on the severe cyclone categories. A detailed consideration of spatial, temporal, and other environmental factors would assist in developing a thorough presentation of cyclone behaviour and provide for better accuracy in forecasting. Another problem any studied cyclone will face is the temporal resolution of images taken by satellites. Due to time intervals set when taking photographs, satellite images may fail to capture rapid fluctuations in cyclone intensity due to their limited temporal resolution. Future studies can contrast proposed higher-frequency satellite images, probable over-spatial analyses, or interpolation methods that would help assess intensity variation between images. This usually helps bring out better representation for cyclone evolution in a continuous sense, hence increasing predictive modelling.

To broaden the applicability of the proposed model, future research should explore its adaptability to other ocean basins such as the Atlantic and Pacific. Variations in satellite sensors, image resolutions, and distinct atmospheric dynamics across basins can introduce distributional differences between source and target datasets. Transfer learning and domain adaptation techniques can be employed to fine-tune models trained on North Indian Ocean data for use in other basins, thereby increasing their generalizability and operational value. Another major concern is modelling generalisation across different ocean basins. Cyclone characteristics vary in various regions due to differences in oceanic and atmospheric conditions. Models trained on North Indian Ocean cyclone data may not generalise well to cyclones in the Atlantic or Pacific basins due to regional climatic differences. Domain adaptation and transfer learning techniques can be explored to overcome this [45]. These approaches let you fine-tune the model from one region's data into another so it's more equipped for different climatic conditions.

Despite advancements in deep learning, computational complexity remains a challenge. Training ConvLSTM-based models is computationally intensive, and their high inference cost poses challenges for real-time deployment [20]. Such inference currently relies on heavy computation and cannot be operational because swift and efficient prediction is always desired. To address the feasibility of real-time deployment, future work can explore model compression techniques, such as pruning or quantization, and adopt lighter architectures like MobileNet-based ConvLSTM variants. Additionally, real-time forecasting can benefit from deploying the trained model on edge devices or cloud-integrated platforms, which enable low-latency predictions critical for disaster management.

In terms of integration into operational pipelines, future studies must evaluate not just model accuracy but also latency, scalability, and ease of deployment under varying environmental conditions. Solutions like edge deployment, modular architectures, and scalable cloud computing environments can facilitate this integration. Facing these challenges and gazing ahead into future advancements, cyclone classification and intensity estimation models would eventually become more certain and reliable. Integration of advanced deep learning techniques, multi-modal data fusion, domain adaptation,

and real-time deployment strategies will add great value to the accuracy and usability of these models. These developments will further contribute to efficient early warning systems, incentives to work towards proactive measures, and a consequent reduction in the impact of cyclones on vulnerable communities. May AI-driven cyclone forecasting lead to great insights into new horizons, greatly transforming the circulation and tracking of storms from the eyes of meteorologists. Bridging current gaps and improving existing models will enable researchers to develop cyclone monitoring systems capable of real-time application.

Conclusion

This research provides a cutting-edge deep learning model and procedure for detecting, segmenting, estimating intensity, and classifying cyclones using satellite images. Transfer learning models (VGG16, ResNet50, and InceptionV3) were analyzed for a comparative performance whereby VGG16 had the highest accuracy of 99.16%, which was attributed to better feature extraction. Mean Shift achieved a higher Silhouette Score of 0.7772 during cyclone segmentation in comparison to K-Means, thereby demonstrating its suitability of being applied for cyclone isolation. ConvLSTM was used for cyclone intensity estimation and classification, where the clustering-based dataset obtained improved generalization. The model trained on the Mean Shift-clustered dataset produced a RMSE of 8.31 ± 1.10 whereas K-Means had 7.79 ± 1.27 , implying a better RMSE in intensity estimation. While Meanshift based clustering provided better distinction between cyclone intensity levels at $81.1 \pm 4.33\%$, K Mean performed better in intensity estimation. Other challenges such as data imbalance, computational complexity, and model generalization across different ocean basins still derive further advancement in the study. This study proposes future research directions, including hybrid deep learning models combining ConvLSTM and Vision Transformers coupled with Bayesian deep learning for uncertainty estimation and domain adaptation techniques for better generalization across different climatic conditions. Moreover, the applications of lightweight architectures combined with cloud implementations can improve the real-time forecasting of AI-driven cyclone prediction to practically feasible operational levels. With the advent of these primary challenges and the application of ascendant AI-based techniques, cyclone monitoring and forecasting should be prioritized in future meteorological research. Such advancements will contribute in no small way to increased accuracy in early warning systems, hence achieving better disaster preparedness and mitigation, significantly reducing the adverse effects of tropical cyclones on vulnerable communities.

Author contributions

Conceptualization: Manju M. S.

Investigation: Punit Gupta.

Methodology: Harsh Pateriya, Rajeev Kumar Gupta, Asmir Butkovic.

Validation: Deepak Singh Tomar.

References

1. Knutson TR, McBride JL, Chan J, Emanuel K, Holland G, Landsea C, et al. Tropical cyclones and climate change. *Nature Geosci*. 2010;3(3):157–63. <https://doi.org/10.1038/ngeo779>
2. Hurricanes of the North Atlantic: climate and society. *Choice Reviews Online*. 2000;37(05):37–2832–37–2832. <https://doi.org/10.5860/choice.37-2832>
3. Dvorak VF. A technique for the analysis and forecasting of tropical cyclone intensities from satellite pictures. 1973.
4. Velden C, Harper B, Wells F, Beven JL II, Zehr R, Olander T, et al. The Dvorak Tropical Cyclone Intensity Estimation Technique: A Satellite-Based Method that Has Endured for over 30 Years. *Bull Amer Meteor Soc*. 2006;87(9):1195–210. <https://doi.org/10.1175/bams-87-9-1195>
5. Lu X, Wong WK, Au-Yeung KC, Choy CW, Yu H. Verification of tropical cyclones (TC) wind structure forecasts from global NWP models and ensemble prediction systems (EPSs). *Tropical Cyclone Research and Review*. 2022;11(2):88–102. <https://doi.org/10.1016/j.tcrr.2022.07.002>
6. 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 Sensing*. 2023;15(17):4188. <https://doi.org/10.3390/rs15174188>

7. Jung H, Baek Y-H, Moon I-J, Lee J, Sohn E-H. Tropical cyclone intensity estimation through convolutional neural network transfer learning using two geostationary satellite datasets. *Front Earth Sci.* 2024;11. <https://doi.org/10.3389/feart.2023.1285138>
8. Pradhan R, Aygun RS, Maskey M, Ramachandran R, Cecil DJ. Tropical Cyclone Intensity Estimation Using a Deep Convolutional Neural Network. *IEEE Trans Image Process.* 2018;27(2):692–702. <https://doi.org/10.1109/TIP.2017.2766358> PMID: 29185987
9. Wenwei X, Karthik B, Andrew A, Nicholas L, Nathan H, Mark D, et al. Deep Learning Experiments for Tropical Cyclone Intensity Forecasts. *Weather and Forecasting.* 2021. <https://doi.org/10.1175/waf-d-20-0104.1>
10. Zhuo J-Y, Tan Z-M. Physics-augmented Deep Learning to Improve Tropical Cyclone Intensity and Size Estimation from Satellite Imagery. *Monthly Weather Review.* 2021. <https://doi.org/10.1175/mwr-d-20-0333.1>
11. Kar C, Kumar A, Banerjee S. Tropical cyclone intensity detection by geometric features of cyclone images and multilayer perceptron. *SN Appl Sci.* 2019;1(9). <https://doi.org/10.1007/s42452-019-1134-8>
12. Zhang C-J, Wang X-J, Ma L-M, Lu X-Q. Tropical Cyclone Intensity Classification and Estimation Using Infrared Satellite Images With Deep Learning. *IEEE J Sel Top Appl Earth Observations Remote Sensing.* 2021;14:2070–86. <https://doi.org/10.1109/jstars.2021.3050767>
13. Ibrar MA, Usama M, Salman AM. A machine learning model for detecting and quantifying tropical cyclone related disturbance and recovery in estuaries. *Sci Rep.* 2025;15(1):5230. <https://doi.org/10.1038/s41598-025-89196-6> PMID: 39939689
14. Mawatwal MK, Das S. An End-to-End Deep Learning Framework for Cyclone Intensity Estimation in North Indian Ocean Region Using Satellite Imagery. *J Indian Soc Remote Sens.* 2024;52(10):2165–75. <https://doi.org/10.1007/s12524-024-01929-8>
15. Zhao Z, Zhang Z, Wang Q, Cui L, Tang P. TFA-Net: A Temporal Feature Aggregation Framework for Tropical Cyclone Intensity Estimation From Satellite Images. *IEEE J Sel Top Appl Earth Observations Remote Sensing.* 2025;18:12008–23. <https://doi.org/10.1109/jstars.2025.3563472>
16. Cheng X, Kong X, Fei J, Huang X. Boost Cyclone Intensity Estimation via Processing Motion Information and Embedding Rotation Invariance. *IEEE J Sel Top Appl Earth Observations Remote Sensing.* 2025;18:9893–903. <https://doi.org/10.1109/jstars.2025.3556475>
17. Sattar K, Muhammad Saad Missen M, Saher N, Nawaz Bashir R, Zoupar Zahra S, Faheem M, et al. Tropical Cyclone Intensity Prediction Using Spatio-Temporal Data Fusion. *IEEE Access.* 2025;13:70095–104. <https://doi.org/10.1109/access.2025.3561355>
18. Varalakshmi P, Vasumathi N, Venkatesan R. Tropical Cyclone intensity prediction based on hybrid learning techniques. *J Earth Syst Sci.* 2023;132(1). <https://doi.org/10.1007/s12040-022-02042-5>
19. Ma Z, Yan Y, Lin J, Ma D. A Multiscale and Multilayer Feature Extraction Network With Dual Attention for Tropical Cyclone Intensity Estimation. *IEEE Trans Geosci Remote Sensing.* 2024;62:1–15. <https://doi.org/10.1109/tgrs.2024.3349416>
20. Tian Y, Zhou W, Cheung PKY, Liu Z. Vision Transformer for Extracting Tropical Cyclone Intensity from Satellite Images. *Adv Atmos Sci.* 2024;42(1):79–93. <https://doi.org/10.1007/s00376-024-3191-1>
21. Zhang R, Liu Y, Yue L, Liu Q, Hang R. Estimating Tropical Cyclone Intensity Using an STIA Model From Himawari-8 Satellite Images in the Western North Pacific Basin. *IEEE Trans Geosci Remote Sensing.* 2024;62:1–13. <https://doi.org/10.1109/tgrs.2024.3352704>
22. Ding J, Hang R, Zhang R, Yue L, Liu Q. Cross-basin incremental learning for tropical cyclone intensity estimation. *Atmospheric Research.* 2025;315:107887. <https://doi.org/10.1016/j.atmosres.2024.107887>
23. Zhao Z, Zhang Z, Tang P, Wang X, Cui L. MT-GN: Multi-Task-Learning-Based Graph Residual Network for Tropical Cyclone Intensity Estimation. *Remote Sensing.* 2024;16(2):215. <https://doi.org/10.3390/rs16020215>
24. Xu G, Ng MK, Ye Y, Zhang B. FHDTIE: Fine-Grained Heterogeneous Data Fusion for Tropical Cyclone Intensity Estimation. *IEEE Trans Geosci Remote Sensing.* 2024;62:1–15. <https://doi.org/10.1109/tgrs.2024.3489674>
25. Tian W, Song P, Chen Y, Zhang Y, Wu L, Zhao H, et al. Short-Term Rolling Prediction of Tropical Cyclone Intensity Based on Multi-Task Learning with Fusion of Deviation-Angle Variance and Satellite Imagery. *Adv Atmos Sci.* 2024;42(1):111–28. <https://doi.org/10.1007/s00376-024-3301-0>
26. Pang S, Xie P, Xu D, Meng F, Tao X, Li B, et al. NDFTC: A New Detection Framework of Tropical Cyclones from Meteorological Satellite Images with Deep Transfer Learning. *Remote Sensing.* 2021;13(9):1860. <https://doi.org/10.3390/rs13091860>
27. Pal S, Das U, Bandyopadhyay O. SSN: A Novel CNN-Based Architecture for Classification of Tropical Cyclone Images From INSAT-3D. *IEEE Trans Geosci Remote Sensing.* 2024;62:1–8. <https://doi.org/10.1109/tgrs.2024.3441729>
28. Kar C, Banerjee S. Tropical cyclone intensity classification from infrared images of clouds over Bay of Bengal and Arabian Sea using machine learning classifiers. *Arab J Geosci.* 2021;14(8). <https://doi.org/10.1007/s12517-021-06997-5>
29. Niu Z, Huang W, Zhang L, Deng L, Wang H, Yang Y, et al. Improving Typhoon Predictions by Integrating Data-Driven Machine Learning Model With Physics Model Based on the Spectral Nudging and Data Assimilation. *Earth and Space Science.* 2025;12(2). <https://doi.org/10.1029/2024ea003952>
30. Sharma T, Mawatwal M, Das S. A Machine Learning Approach for Detecting Rapid Intensification in Tropical Cyclones. *J Indian Soc Remote Sens.* 2025;53(6):1965–78. <https://doi.org/10.1007/s12524-024-02101-y>
31. Kossin JP, Knapp KR, Olander TL, Velden CS. Global increase in major tropical cyclone exceedance probability over the past four decades. *Proc Natl Acad Sci U S A.* 2020;117(22):11975–80. <https://doi.org/10.1073/pnas.1920849117> PMID: 32424081
32. India Meteorological Department. Cyclone Best Track Data. 2023. <https://mausam.imd.gov.in>

33. Shorten C, Khoshgoftaar TM. A survey on Image Data Augmentation for Deep Learning. *J Big Data*. 2019;6(1). <https://doi.org/10.1186/s40537-019-0197-0>
34. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *Computer Vision and Pattern Recognition*. 2014.
35. Shafiq M, Gu Z. Deep Residual Learning for Image Recognition: A Survey. *Applied Sciences*. 2022;12(18):8972. <https://doi.org/10.3390/app12188972>
36. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the Inception Architecture for Computer Vision. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016. 2818–26. <https://doi.org/10.1109/cvpr.2016.308>
37. Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*. 2020;21(1). <https://doi.org/10.1186/s12864-019-6413-7>
38. Ester M, Kriegel H, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. *Knowledge Discovery and Data Mining*. 1996;226–31.
39. Manjunath BS, Ma WY. Texture features for browsing and retrieval of image data. *IEEE Trans Pattern Anal Machine Intell*. 1996;18(8):837–42. <https://doi.org/10.1109/34.531803>
40. Canny J. A Computational Approach to Edge Detection. *IEEE Trans Pattern Anal Mach Intell*. 1986;PAMI-8(6):679–98. <https://doi.org/10.1109/tpami.1986.4767851>
41. Ma C, Li T. An Empirical Model of Tropical Cyclone Intensity Forecast in the Western North Pacific. *J Meteorol Res*. 2022;36(5):691–702. <https://doi.org/10.1007/s13351-022-2016-3>
42. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. *jair*. 2002;16:321–57. <https://doi.org/10.1613/jair.953>
43. Moishin M, Deo RC, Prasad R, Raj N, Abdulla S. Designing Deep-Based Learning Flood Forecast Model With ConvLSTM Hybrid Algorithm. *IEEE Access*. 2021;9:50982–93. <https://doi.org/10.1109/access.2021.3065939>
44. Mukesh Ch, Likhita A, Yamini A. Performance Analysis of InceptionV3, VGG16, and Resnet50 Models for Crevices Recognition on Surfaces. *Lecture Notes in Networks and Systems*. Springer Nature Singapore. 2024: 161–72. https://doi.org/10.1007/978-981-99-7817-5_13
45. Ma T-Y, Li H-C, Zheng Y-B, Du Q, Plaza A. Fully Tensorized Lightweight ConvLSTM Neural Networks for Hyperspectral Image Classification. *IEEE Trans Neural Netw Learn Syst*. 2025;36(8):14213–27. <https://doi.org/10.1109/TNNLS.2024.3511575> PMID: 40030676