Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Transformer-Based Cyclone Detection: Leveraging Spatiotemporal Features from Satellite Imagery

P. Karthikeyan

pkn@vcet.ac.in

Velammal college of Engineering & Technology

B.Sakthivel

sakthivelb.tce@gmail.com

Menakshi college of Engineering

S.S.Surya Varsiny

svarsiny.2502@gmail.com

Velammal college of Engineering & Technology

S.M.Tejesvini

tejasindhu1211@gmail.com

Velammal college of Engineering & Technology

Research Article

Keywords: Tropical cyclone detection, satellite remote sensing, Data-efficient Image Transformer (DeiT), spatiotemporal feature extraction, native temporal sequences, optical flow analytics, cloud morphology dynamics, transformer-based classification, North Indian Ocean

Posted Date: March 31st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9243670/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Transformer-Based Cyclone Detection: Leveraging Spatiotemporal Features from Satellite Imagery

P. Karthikeyan[1],
Department of ECE,
Velammal college of
Engineering & Technology,
Madurai, Tamil Nadu, India.
pkn@vcet.ac.in

B. Sakthivel[2],
Department of AI&DS,
Menakshi college of
Engineering
Chennai, Tamil Nadu, India.
sakthivelb.tce@gmail.com

S.S. Surya Varsiny[3],
Department of ECE,
Velammal College of
Engineering & Technology,
Madurai, Tamil Nadu, India.
svarsiny.2502@gmail.com

S.M. Tejesvini[4],
Department of ECE,
Velammal College of
Engineering & Technology,
Madurai, Tamil Nadu, India.
tejasindhu1211@gmail.com

***Abstract* — Tropical cyclone detection is critical for disaster preparedness and risk mitigation. This study introduces a transformer-based framework for identifying cyclones in meteorological satellite data, utilizing the Data-efficient Image Transformer (DeiT) architecture. Unlike traditional methods relying on temporally densified datasets, our approach operates on native sequential satellite observations, preserving cloud evolution dynamics. The primary observational method for tropical cyclone discovery and extensive atmospheric monitoring is satellite imaging. In contrast to many modern deep learning-based cyclone detection systems that depend on artificial temporal densification through optical-flow interpolation, frame synthesis, or constraint-based reconstruction, the suggested methodology only uses native sequential satellite data, maintaining the inherent temporal fidelity of cloud dynamics. For end-to-end cyclone classification, a transformer-based Data-efficient Image Transformer (DeiT) architecture is used, taking advantage of the long-range feature dependencies and global spatial context present in geostationary satellite data. In order to preserve historical cloud structure and motion continuity without adding artificial intermediate frames, the dataset, which consists of native North Indian Ocean satellite sequences, specifically incorporates previous temporal frames (t−1, t−2, etc.). Spatiotemporal dynamics are measured post-inference using physically interpretable descriptors that are only calculated between successive observations, as opposed to preprocessing-driven motion reconstruction. These include mean brightness variability, texture energy, and dense Farneback optical-flow magnitude, enabling objective characterization of cloud-top evolution and motion persistence. We integrate physically interpretable spatiotemporal descriptors, including mean brightness, texture energy, and dense optical flow magnitude, to characterize cloud-top evolution and motion persistence. Temporal differencing and lag-wise analysis support research-oriented visualization through line and scatter representations, facilitating examination of structural transitions and dynamical consistency across successive timesteps. Experiments on North Indian Ocean satellite sequences demonstrate competitive classification performance, with improved physical interpretability and reduced preprocessing complexity. Comparative evaluation against recent optical-flow-intensive cyclone detection studies indicates that the proposed framework achieves competitive classification performance while significantly reducing preprocessing complexity, eliminating interpolation-induced artifacts, and improving physical interpretability through explicit time-series characterization of cloud morphology and motion dynamics. Our framework offers a scalable solution for automated cyclone monitoring, enhancing early detection and continuous atmospheric surveillance.*
The implementation and code used in this study are publicly available at: https://doi.org/10.5281/zenodo.19249181

***Keywords* — Tropical cyclone detection; satellite remote sensing; Data-efficient Image Transformer (DeiT); spatiotemporal feature extraction; native temporal sequences; optical flow analytics; cloud morphology dynamics; transformer-based classification; North Indian Ocean.**

I. INTRODUCTION

Tropical cyclones (TCs) represent some of the most destructive and dynamically complex atmospheric phenomena, generating extreme winds, torrential precipitation, and storm surges that threaten coastal populations, maritime operations, and critical infrastructure. Early and reliable detection of cyclonic systems is therefore essential for disaster preparedness, risk mitigation, and the timely dissemination of warnings. Continuous in situ monitoring over vast oceanic expanses, however, is operationally constrained and economically prohibitive. Consequently, satellite-based remote sensing has become the primary observational modality for cyclone detection and large-scale atmospheric surveillance.

2015	01A (Ashobaa), 02B (Komen), 03A, 04A (Chapala), 05A (Megh)
2016	01B (Roanu), 02A, 03B (Kyant), 04B (Nada), 05B (Vardah)
2017	01B (Maarutha), 02B (Mora), 03B (Ockhi), 04B
2018	01A (Sagar), 02A (Mekunu), 04B (Daye), 05A (Luban), 06B (Titli), 07B (Gaja), 08B (Phethai)
2019	01B (Fani), 02A (Vayu), 03A (Hikaa), 04A (Kyaar), 05A (Maha), 06A (Pawan), 07A
2020	01B (Amphan), 02A (Nisarga), 03A (Gati), 04B (Nivar), 05B (Burevi)
2021	01A (Tauktae), 02B (Yaas), 03B (Gulab), 03A (Shaheen)
2022	01B, 02B (Asani)

Figure. 1. *Annual record of cyclonic systems over the North Indian Ocean basin (2015–2022) used for dataset construction and analysis.*

Satellite infrared (IR) imagery provides synoptic-scale coverage and captures critical indicators of cyclonic activity, including cloud-top temperature distributions, convective organization, and mesoscale structural evolution. Historically, cyclone identification has relied with expert-driven visual interpretation techniques such as the Dvorak method, which, despite their operational utility, are inherently subjective, labor-intensive, and difficult

to scale across data-intensive monitoring systems. These demerits underscore the need for automated, objective, and scalable detection frameworks capable of consistently identifying cyclonic signatures from satellite observations. Recent advances in context of deep learning and computer vision have accelerated progress in automated meteorological image analysis. Convolutional neural networks (CNNs) have been widely applied to cyclone detection tasks due to their effectiveness in extracting localized spatial features, including cloud textures, spiral rainbands, and convective clusters. However, CNN-based architectures often exhibit limited capability in modeling long-range spatial dependencies and global cloud organization, both of which are essential for robust cyclone identification in complex atmospheric environments. Transformer-based architectures have emerged as powerful alternatives by leveraging self-attention mechanisms to capture global contextual relationships across an image. In particular, the Data-efficient Image Transformer (DeiT) enables effective training on relatively limited labeled datasets while maintaining strong representational capacity. Despite their predictive strength, purely data-driven models frequently lack physical interpretability, a critical requirement for operational meteorological deployment and scientific analysis. our study presents a robust and interpretable cyclone detection framework for the North Indian Ocean using geostationary satellite imagery. Unlike approaches that rely on artificial temporal densification through optical-flow interpolation or frame reconstruction, the proposed framework operates exclusively on native sequential observations, preserving the physical integrity of cloud evolution. Historical temporal context (t−1, t−2, …) is retained to capture continuity in cloud morphology and motion without introducing synthetic intermediate frames. The framework integrates a transformer-based DeiT model for primary cyclone classification with a complementary spatiotemporal feature extraction pipeline designed to quantify physically meaningful atmospheric descriptors. Post-inference metrics include mean brightness variability, texture energy, and dense Farneback optical-flow magnitude computed between consecutive frames, enabling objective characterization of cloud-top evolution and motion persistence. Time-series alignment and lag-wise differencing facilitate consistent analysis of temporal cloud dynamics. To enhance transparency and operational usability, the system incorporates an interactive visualization module that supports satellite image ingestion, real-time prediction, annotated outputs, and metric-driven interpretation. By synergizing physical meteorological insight with transformer-based global pattern recognition, the proposed framework delivers an efficient, interpretable, and operationally viable solution for early cyclone detection and continuous atmospheric monitoring.

II. DATA AND DEIT DESIGN

A. Data

Satellite observations constitute the primary data source for tropical cyclone monitoring and analysis. In this study, geostationary satellite infrared imagery obtained from the Meteosat-8 Indian Ocean sector was utilized to analyze cyclonic systems over the North Indian Ocean (NIO) basin. The dataset was curated from the publicly available *Deep Learning Cyclone Intesity Estimation NIO Satellite* repository, which provides preprocessed cyclone-centered image samples along with associated metadata. Each image represents cloud-top infrared brightness temperature patterns essential for identifying convective organization and cyclone structure. The dataset primarily focuses on the North Indian Ocean basin, encompassing the Bay of Bengal (BoB) and Arabian Sea, with cropped spatial coverage spanning approximately 50°E–100°E longitude and 0°–30°N latitude. This domain of concept fully captures the principal cyclone genesis and evolution regions affecting the Indian subcontinent. Storm identifiers rely within the dataset (e.g., *01B*, *02A*) correspond to basin-specific classification conventions, where “B” denotes Bay of Bengal systems and “A” denotes Arabian Sea systems, ensuring accurate basin attribution. All images were pre-centered on storm precise locations and spatially cropped to cyclone-focused regions to reduce background noise and enhance structural feature learning.

Sl. No.	Cyclone Number	Other Name	Date Range	Satellite	Cropped Region	Temporal Resolution	Image Type	Basin
1	01B	—	20220305 – 20220306	Meteosat-8	50E to 80E, 0 to 30N	6 hr	Infrared	BoB
2	02B	Asani	20220507 – 20220512	Meteosat-8	70E to 100E, 0 to 30N	6 hr	Infrared	BoB

Table 1: *Summary of cyclone events and corresponding Meteosat-8 infrared imagery parameters.*

The satellite observations provide consistent radiometric measurements of cloud-top thermal structure, enabling detection of spiral banding, Cold cloud shields, convective bursts, and eye formation patterns associated with tropical cyclones. The imagery is provided in standardized raster format and was resized to a uniform spatial resolution of 224 × 224 pixels to ensure compatibility with deep learning model input requirements while preserving mesoscale cloud organization.

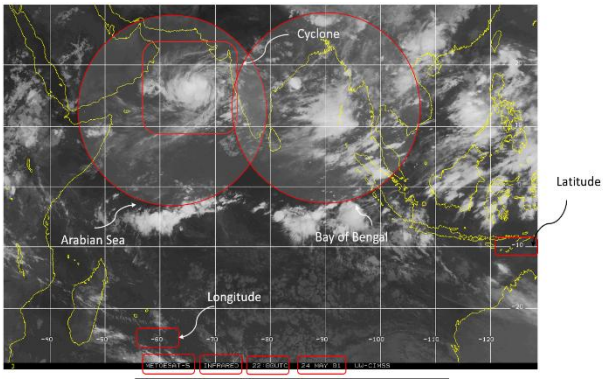


Figure 2. *Geographical context and satellite-based visualization of cyclonic activity over the North Indian Ocean (NIO) region.*

The dataset includes labeled samples categorized into cyclone and non-cyclone classes based on verified storm track information and associated meteorological records. These labels serve as a ground truth interference for supervised training and evaluation of the classification model. The imagery captures varying stages of cyclone evolution, including pre-genesis disturbances, intensification structures, mature storm bands, and dissipating systems, enabling robust model learning across diverse atmospheric conditions. To support temporal analysis, sequential satellite frames representing consecutive observation times were utilized without artificial interpolation or temporal densification. It preserves the native temporal continuity of atmospheric evolution while avoiding artifacts introduced by synthetic frame generation. Past time-series segments (e.g., t−8 to t−3 frames) were analyzed to characterize antecedent cloud organization and motion trends preceding the present frame. Spatiotemporal metrics derived from these sequences contributes interpretation of cyclone evolution dynamics within the NIO basin. The dataset provides a physically consistent and region-specific representation of tropical cyclone structure and evolution, making it suitable for transformer-based cyclone detection and interpretable spatiotemporal analysis.

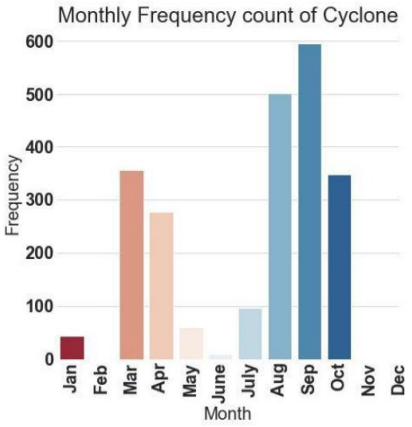


Figure 3: *Monthly frequency count showing two different cyclonic seasons in NIO.*

B. Brief Introduction to Vision Transformers and DeiT

Convolutional Neural Networks (CNNs) have long been the dominant architecture in computer vision due to their ability to learn hierarchical spatial features through convolutional filters and nonlinear activations. By utilizing local receptive fields and weight sharing, CNNs effectively gather textures, edges, and spatial patterns. However, their reliance on localized convolution kernels can regulate the modeling of long-range spatial dependencies and global structural relationships, which are essential for interpreting large-scale atmospheric patterns in satellite imagery. Vision Transformers (ViTs) introduce a fundamentally different paradigm by replacing convolutional operations with self-attention mechanisms that model global relationships across an image. Instead of scanning with convolution filters, ViTs divide an image into fixed-size patches and linearly align them into tokens. These tokens are processed through multi-head self-attention layers, assisting the model to learn interactions between distant regions of the image. This global receptive field allows transformers to better capture large-scale cloud organization, spiral band structures, and mesoscale circulation patterns critical for cyclone detection.

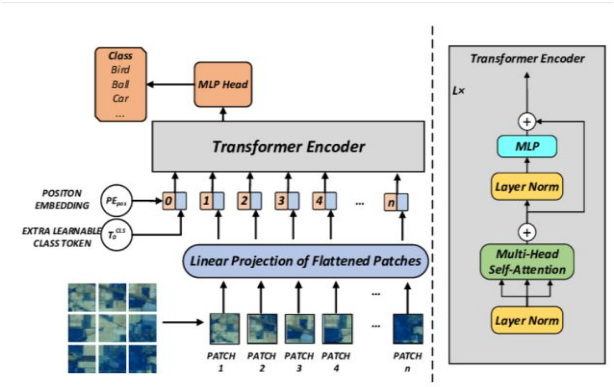


Figure 4. *Vision Transformer (ViT) architecture illustrating patch embedding, positional encoding, transformer encoder layers, and classification head.*

The Data-efficient Image Transformer (DeiT) is a effective Vision Transformer architecture designed to achieve high performance with comparatively limited training data. DeiT introduces a distillation token and a training strategy that transfers knowledge from a convolutional teacher network to the transformer model, improving convergence and data efficiency. This design enables robust learning without requiring extremely large datasets typically needed by standard transformers. Structurally, DeiT processes an input image by splitting it into non-overlapping patches, projecting them into an embedding space, and adding positional encodings to preserve spatial arrangement. The sequence is then passed through stacked transformer encoder blocks consisting of multi-head self-attention and feed-forward networks. Through iterative attention-based feature refinement, the model learns global contextual representations, which are mapped to classification outputs. Unlike CNNs that emphasize local feature extraction, DeiT captures global spatial dependencies and contextual coherence, making it particularly suitable for meteorological imagery where large-scale cloud systems and spatial organization are crucial indicators of cyclonic activity. To add on, transformer-based representations enhance interpretability when combined with physically meaningful post-inference metrics, supporting operational meteorological analysis. In this study, the DeiT architecture is employed for cyclone classification from satellite imagery, its global attention mechanism and data-efficient learning capability to improve detection robustness while preserving spatial context essential for atmospheric system interpretation.

C. Description of Deit model

The proposed cyclone detection framework is built upon the Data-efficient Image Transformer (DeiT) architecture, introduced by Hugo Touvron and colleagues at Facebook AI Research in 2021. DeiT was developed to make Vision Transformers practical for real-world applications by decreasing their dependence on extremely large training datasets and computational resources. The model achieves this efficiency through knowledge distillation, optimized training procedures, and data-efficient learning, allowing transformer-based vision systems to overcome convolutional neural networks (CNNs) while using substantially less data. Not similar with very deep convolutional networks which require extensive datasets and pretraining, DeiT leverages multi-head self-attention to model global spatial relationships and integrates a distillation token that transfers knowledge from a teacher network during training. This design improves convergence stability, enhances generalization, and preserves strong representational capability even when labeled datasets are limited. These characteristics make DeiT particularly suitable for meteorological satellite imagery, where precise cyclone datasets are relatively small compared to conventional computer vision.

Traditional convolutional architectures rely on localized receptive fields and hierarchical feature extraction, which can limit their ability to capture large-scale atmospheric organization. Cyclonic systems exhibit spatial structures spanning hundreds of kilometers, including spiral rainbands, symmetric eyewalls, and extensive cloud shields. DeiT addresses this limitation by modeling global spatial dependencies through self-attention, enabling the network to interpret mesoscale cloud organization and long-range structural coherence essential for reliable cyclone identification. By leveraging attention-based global feature modeling and data-efficient training, the DeiT architecture provides a powerful and scalable solution for automated cyclone detection from satellite imagery, bridging the gap between deep learning performance and meteorological interpretability.

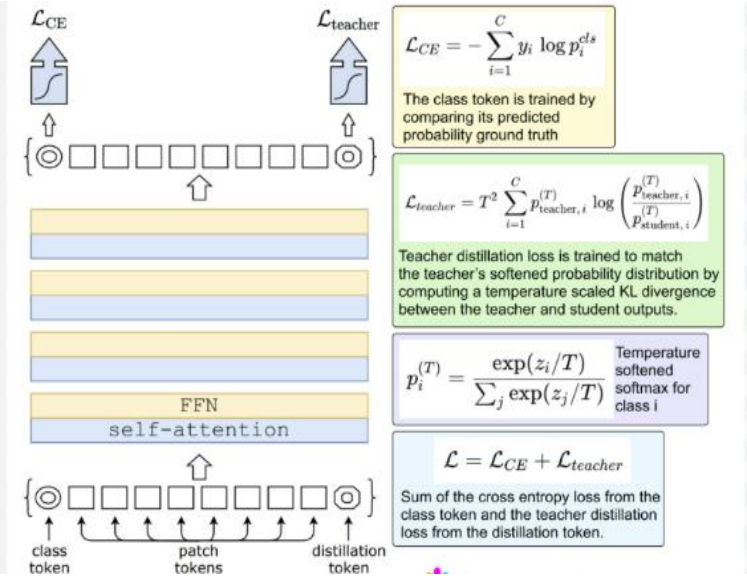


Figure 5: *DeiT-Based Transformer Architecture with Knowledge Distillation for Cyclone Classification*

In the context of cyclone detection, the adoption of DeiT is motivated by several operational and scientific advantages:

- **Global context awareness** — captures large-scale cloud organization and spiral band connectivity.
- **Data efficiency** — performs effectively with limited labeled cyclone imagery.
- **Robust feature representation** — learns spatial dependencies beyond local texture cues.
- **Reduced preprocessing requirements** — operates directly on native satellite frames.
- **Improved generalization** — handles diverse storm morphologies and atmospheric variability.

Stage	Technique	Parameters	Purpose
Preprocessing	Image Resizing	224 × 224 pixels	Match DeiT input requirements
Augmentation	Rotation	±20°	Improve rotational robustness
	Scaling	90% – 110%	Handle size variability
	Gaussian Noise	σ = 10	Increase noise tolerance
Sequence Handling	Frame Ordering	Temporal sequence preserved	Enable spatiotemporal analysis

Table 2: *Summary of preprocessing, augmentation, and sequence handling techniques applied to the cyclone dataset prior to model training.*

1) **Input Representation and Pre-processing:** Infrared (IR) satellite imagery constitutes the primary input to the proposed cyclone detection framework. Each image is resampled to **224 × 224 pixels**, matching the input dimensionality required by the DeiT architecture, while preserving the spatial structure necessary to capture mesoscale cloud organization. Pixel intensities are normalized to a [0, 1] floating-point range, improving numerical stability, accelerating convergence, and ensuring consistent gradient propagation during training and inference. To enhance model robustness while maintaining the physical integrity of the observed atmospheric structures, a structured preprocessing pipeline is applied:

A. Spatial Normalization

Raw satellite brightness values are converted to floating-point representations and scaled to a standardized range. This normalization:

- Stabilizes gradient updates during optimization,
- Prevents numerical overflow during self-attention computations, and
- Ensures consistent input distributions across training batches.
- Where required, channel-wise normalization can be applied to compensate for sensor calibration variability.

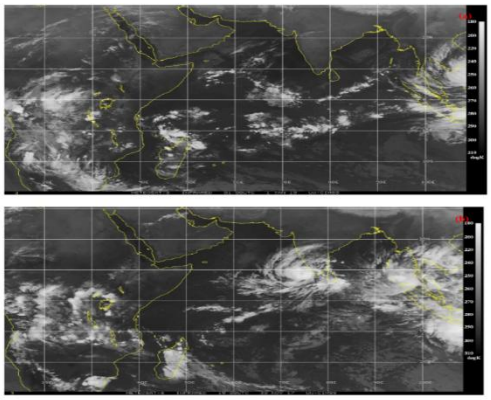


Figure 6. *Image preprocessing and normalization.*
(a) *Raw infrared satellite image over the North Indian Ocean.*
(b) *Preprocessed and normalized image after resizing and intensity standardization, enhancing cloud structures and ensuring consistent inputs for the transformer model.*

B. Data Augmentation for Generalization

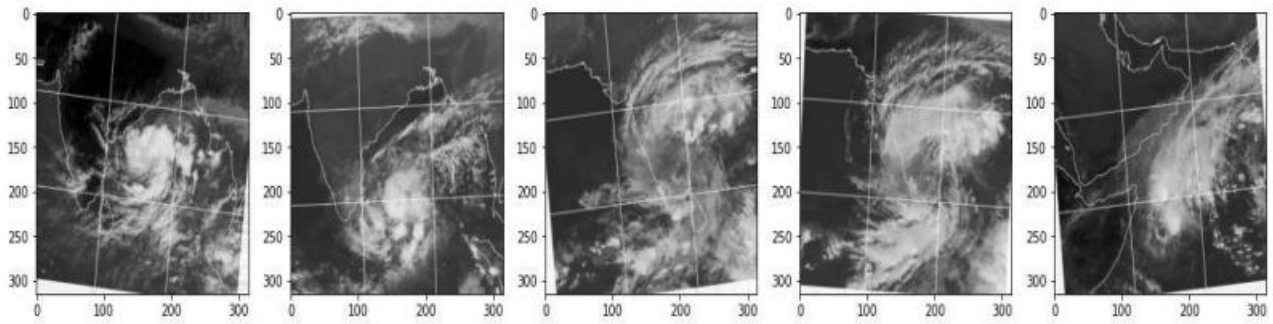


Figure 7*: Presents examples of augmented cyclone satellite imagery used for model training. The augmentation pipeline incorporates controlled rotational adjustments, scaling transformations, and Gaussian noise strengthening the model's generalization capability.*

To improve robustness under varying observational conditions and cyclone orientations, controlled augmentation is applied during training. These transformations simulate realistic variability without altering meteorological structure:

- **Random rotations** to account for arbitrary storm orientation,
- **Horizontal/vertical flips** to enhance rotational invariance,
- **Minor brightness and contrast adjustments** to reflect sensor and atmospheric variability, and
- **small affine perturbations** to simulate geolocation uncertainty.

These augmentations increase effective dataset diversity and reduce overfitting while preserving physically meaningful cloud morphology.

C. Centered Storm Alignment

When cyclone center coordinates are available from best-track metadata or detection algorithms, images are spatially aligned to center the storm core. This alignment:

- Emphasizes eyewall and inner-core structure,
- Reduces background noise and unrelated cloud systems, and
- Improves sensitivity to structural symmetry and convective organization. If center coordinates are unavailable, the model operates on uncropped frames, leveraging global attention to identify relevant structures.

D. Preservation of Native Temporal Structure

Unlike approaches that rely on optical-flow interpolation or synthetic temporal densification, the proposed framework processes native sequential satellite frames directly. This design choice:

- Preserves observational authenticity,
- Prevents interpolation-induced artifacts,

- Maintains physical consistency of cloud motion
- Enables reliable post-inference temporal analysis.

2) **Patch Embedding and Tokenization**: Instead of convolutional feature extraction, DeiT partitions each input image into fixed-size non-overlapping patches.

Stage	Layer Type	Output Shape	Parameters
Input	Input Layer	(224 × 224 × 3)	0
Patch Embedding	PatchEmbedding	(196, 192)	147,648
Transformer Block × 6	Multi-Head Attention	(196, 192)	444,288 per block
	Layer Normalization	(196, 192)	384 per block
	Feed Forward (Dense 192→384→192)	(196, 192)	148,032 per block
Token Aggregation	Global Average Pooling	(192)	0
Classification Head	Dense (192 → 256)	(256)	49,408
	Dropout	(256)	0
	Dense (256 → 1)	(1)	257

Table 3: *Network design and parameter allocation of the proposed cyclone detection framework.*

- Image size: **224 × 224**
- Patch size: **16 × 16**
- Number of patches: **196**
- Embedding dimension: **768**

Each patch is flattened and projected linearly into a vector representation. Positional embeddings are then added to preserve spatial ordering information lost during patch flattening. A special classification token (CLS token) is prepended to the token sequence to aggregate global contextual information for final classification.

3) **Transformer Encoder Architecture:** Following patch embedding and positional encoding, the input token sequence is processed through a stack of transformer encoder blocks, which form the core representational engine of the DeiT architecture. Unlike convolutional operations that aggregate information within fixed receptive fields, transformer encoders model global contextual relationships through attention mechanisms, enabling holistic interpretation of large-scale atmospheric structures. Each encoder block consists of the following components:

A. Multi-Head Self-Attention (MHSA)

The multi-head self-attention mechanism enables the model to compute pairwise relationships among all patch tokens, allowing the network to dynamically focus on spatially distant yet meteorologically related regions. For an input token sequence $X \in \mathbb{R}^{N \times D}$, attention is computed as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

where queries Q, keys K, and values V are learned projections of the token embeddings.

By employing multiple attention heads, the model simultaneously captures diverse structural dependencies, which is particularly critical for cyclone morphology analysis, including:

- **spiral rainband connectivity and curvature**,
- **eyewall symmetry and compactness**,
- **mesoscale convective clustering**, and
- **outer cloud shield expansion and anisotropy**.

This global attention capability allows the network to recognize cyclone organization patterns that may span the entire satellite frame.

B. Position-wise Feed-Forward Network (FFN)

Each encoder layer includes a fully connected feed-forward network applied independently to each token:

$$\text{FFN}(x) = \max(0, xW_1 + b_1)\,W_2 + b_2 \quad (2)$$

This two-layer multilayer perceptron introduces nonlinear transformations that enhance feature discrimination and refine latent representations learned via attention. The FFN expands the embedding dimension before projecting it back to the original size, enabling richer feature interactions.

Layer	Units / Output Shape	Patch / Head / FFN Size	Activation
Input	**(224, 224, 3)**	**–**	**–**
Patch Embedding	**(196, 192)**	**Patch size = 16 × 16**	**–**
Transformer Block 1	**(196, 192)**	**3 Heads, FFN: 192 → 384 → 192**	**GELU**
Transformer Block 2	**(196, 192)**	**3 Heads, FFN: 192 → 384 → 192**	**GELU**
Transformer Block 3	**(196, 192)**	**3 Heads, FFN: 192 → 384 → 192**	**GELU**
Transformer Block 4	**(196, 192)**	**3 Heads, FFN: 192 → 384 → 192**	**GELU**
Transformer Block 5	**(196, 192)**	**3 Heads, FFN: 192 → 384 → 192**	**GELU**
Transformer Block 6	**(196, 192)**	**3 Heads, FFN: 192 → 384 → 192**	**GELU**
Global Average Pooling	**(192)**	**–**	**–**
Dense	**256**	**–**	**ReLU**
Dropout	**0.5**	**–**	**–**
Output Dense	**1 (BC)**	**–**	**Sigmoid**

Table 4. *Layer-wise configuration and parameter allocation of the proposed DeiT-based cyclone detection model.*

C. Residual Connections and Layer Normalization

Residual skip connections surround both the MHSA and FFN sublayers:

$$x' = x + \mathrm{MHSA}\big(\mathrm{LN}(x)\big) \quad (3)$$
$$y = x' + \mathrm{FFN}\big(\mathrm{LN}(x')\big) \quad (4)$$

These pathways facilitate stable gradient propagation, mitigate vanishing gradient effects, and support deeper transformer stacks. Layer normalization standardizes intermediate activations, improving numerical stability and accelerating convergence during training. Through successive encoder layers, token representations are iteratively refined via attention-driven context aggregation and nonlinear transformation. This hierarchical refinement enables the network to learn global spatial semantics essential for cyclone detection, including structural coherence, symmetry, and cloud-field organization.

D. Knowledge Distillation for Data Efficiency and Classification Head

A defining innovation of DeiT is its distillation token, which enables supervised knowledge transfer from a convolutional teacher model during training. Introduced to improve data efficiency, this mechanism augments the standard class token by incorporating teacher-guided supervision. After the final encoder block, the class (CLS) token encodes the global representation of the input scene. This token is passed through a fully connected classification head to generate cyclone probability estimates.

Model outputs include:

- Probability of cyclone presence,
- Binary classification: Cyclone / Non-Cyclone.

Binary cross-entropy loss is employed during training:

$$\mathcal{L} = -[y \log(p) + (1 - y) \log(1 - p)] \quad (5)$$

Ensuring probabilistic calibration and robust decision boundaries.

E. Training Strategy and Optimization

The DeiT-based model is trained on sequential satellite imagery spanning multiple tropical cyclone events across the North Indian Ocean basin, capturing diverse structural and environmental conditions.

Optimization configuration:

- **Loss function:** Binary cross-entropy
- **Optimizer:** AdamW (decoupled weight decay for improved generalization)
- **Learning rate scheduling:** cosine decay or step scheduling to ensure stable convergence
- **Weight regularization:** implicit via normalization and attention smoothing
- **Early stopping:** prevents overfitting and preserves generalization performance

Transformer architectures benefit from normalization layers and attention-based feature aggregation, enabling efficient convergence and stable optimization even with moderate dataset sizes.

III. IMPLEMENTATION ARCHITECTURE AND KERAS-BASED DEPLOYMENT CONFIGURATION

A. Execution Environment and System Initialization

The cyclone detection framework is deployed within a Google Colab GPU-enabled environment, enabling scalable experimentation and near–real-time inference. Integration with cloud storage is achieved through Google Drive mounting, allowing persistent access to trained model artifacts and satellite imagery datasets.The implementation relies on a scientific computing stack including:

- **TensorFlow / Keras** for deep learning inference and model serialization
- **NumPy** for numerical operations
- **OpenCV** for motion analysis and texture computation
- **Pillow (PIL)** for image handling
- **Matplotlib** for visualization
- **Google Colab utilities** for file upload and interactive execution. This environment ensures reproducibility, portability, and compatibility with operational workflows.

B. Custom Patch Embedding Layer and Transformer Compatibility

Because the DeiT backbone utilizes a patch-based tokenization mechanism, a **custom Keras-serializable patch embedding layer** is defined to ensure compatibility during model deserialization.

Model Complexity

- **Total Parameters:** 3,755,841
- **Trainable Parameters:** 3,755,841
- **Model Size:** 14.33 MB

The Patch Embedding layer performs:

- Patch extraction via Conv2D projection
- kernel size = patch size
- stride = patch size
- Linear projection into embedding space
- Reshaping into token sequences suitable for transformer processing

Mathematically, for an input image $X \in \mathbb{R}^{H \times W \times C}$, the projection produces:

$$T \in \mathbb{R}^{N \times D} \qquad (6)$$

where:

- N= number of patches
- D= embedding dimension

Registering the layer with @register_keras_serializable () ensures the model can be safely loaded across environments without architecture loss.

C. DeiT Model Loading and Serialization Structure

The cyclone detection model is deployed using the native Keras Saved Model (keras) serialization format, which preserves the complete network topology, learned parameters, and runtime metadata in a modular, platform-agnostic structure. This format ensures reproducibility, version traceability, and compatibility across research, cloud, and operational environments. Unlike legacy HDF5-only storage, the keras container encapsulates architectural configuration and weights in a structured manner that supports long-term maintainability and seamless redeployment.

File Name	Last Modified	Size (bytes)
metadata.json	1980-01-01 00:00:00	64
config.json	1980-01-01 00:00:00	821,207
model.weights.h5	2026-01-07 05:55:54	45,359,888

Table 5. *File Structure and configuration of the trained DeiT cyclone detection framework.*

The serialized model package consists of three primary components. The config.json file encodes the architectural blueprint, which includes transformer encoder depth, embedding dimensionality, multi-head self-attention configuration, and classification head structure. The model. weights.h5 file stores optimized weight tensors representing the learned parameters across convolutional projections, attention layers, feed-forward networks, and normalization modules. The metadata.json file contains serialization versioning and environment descriptors, ensuring compatibility with the Keras

runtime and safeguarding against deserialization inconsistencies. Model restoration is performed using custom object registration to reconstruct the PatchEmbedding layer responsible for transformer tokenization. The model is loaded with compile is termed as "False", disabling training graph reconstruction and thereby reducing initialization overhead for inference-only deployment. TensorFlow employs poor graph construction, allocating computational strategies dynamically at runtime to optimize memory utilization. This modular loading strategy facilitates reproducible experimentation, efficient version control, and reliable deployment in real-time meteorological workflows.

D. Input Standardization and Tensor Preparation

Satellite infrared imagery undergoes standardized preprocessing prior to inference to ensure numerical stability, consistency with training conditions, and compatibility with transformer input constraints. Because transformer architectures are sensitive to input scaling and dimensional uniformity, strict preprocessing alignment is essential to preserve learned feature distributions and prevent inference drift. Initially, all satellite frames are spatially resampled to 224 × 224 pixels, matching the DeiT input resolution and ensuring consistent patch token dimensionality. Resizing is performed using interpolation methods that preserve spatial structure while maintaining the integrity of cloud morphology and mesoscale organization patterns. Images are then converted into RGB tensor format, ensuring compatibility with convolution-based patch projection layers. Although infrared imagery is intrinsically single-channel, RGB formatting maintains compatibility with pretrained architectural assumptions and facilitates tensor processing within the Keras pipeline. Pixel intensities are normalized to the floating-point range [0, 1], improving numerical conditioning, stabilizing gradient propagation, and ensuring consistency with training normalization regimes. Finally, a batch dimension is appended to produce a four-dimensional tensor of shape:

$$(1, 224, 224, 3) \qquad (7)$$

This tensor structure aligns with Keras inference pipelines and enables efficient GPU execution. Collectively, these preprocessing steps preserve radiometric fidelity while ensuring transformer compatibility and numerical robustness.

E. Frame-Level Cyclone Prediction Pipeline

Each satellite frame is independently evaluated using a forward pass through the trained DeiT transformer backbone. The model estimates cyclone presence probability as:

$$P_{\text{cyclone}} = f_\theta(I_t) \qquad (8)$$

where I_t denotes the satellite observation at time t, and f_θ represents the trained transformer model parameterized by θ. This formulation allows the network to extract global spatial dependencies and mesoscale organization patterns from each frame without requiring temporal sequence reconstruction. To support operational decision-making, probabilistic outputs are mapped into discrete decision categories using calibrated thresholds. Frames with probability values greater than or equal to 0.75 are classified as Cyclone, while those below or equal to 0.25 are classified as Not a Cyclone. Intermediate values are labeled Uncertain, reflecting convective organization or early-stage cyclogenesis. Confidence scores are derived from posterior probability magnitude, enabling uncertainty-aware classification. This probabilistic interpretation is particularly valuable during cyclone genesis, structural transitions, or weakening phases, where atmospheric organization may not exhibit fully developed cyclonic symmetry. The decision framework therefore enhances operational reliability while reducing false alarm rates and supporting risk-aware forecasting.

F. Image-Derived Physical Feature Extraction

To complement transformer-based predictions with physically interpretable diagnostics, additional image-derived metrics are computed from each satellite frame. These descriptors provide meteorologically meaningful insight into cloud-top structure, convective organization, and radiative variability, thereby improving interpretability for operational meteorology. Mean brightness is computed from grayscale intensity and serves as a proxy for cloud-top thermal structure. In infrared imagery, colder cloud tops associated with deep convection appear brighter, making mean brightness an effective indicator of sustained convective activity and vertical storm development. Brightness variability, measured as the standard deviation of grayscale intensity, captures radiative heterogeneity within the cloud field. Elevated variability often indicates mixed cloud-top heights and active convective clustering, whereas low variability is characteristic of stratiform cloud decks or dissipating systems. Texture energy is computed using Sobel gradient operators to quantify structural organization:

$$TE = \text{mean}\left(\sqrt{G_x^2 + G_y^2}\right) \qquad (9)$$

where $G_x$and $G_y$represent horizontal and vertical intensity gradients. High texture energy values indicate organized mesoscale convection, spiral rainbands, and sharp eyewall boundaries, all of which are characteristic of mature cyclonic systems. Together, these metrics provide physically meaningful descriptors that enhance the interpretability of automated predictions.

G. Dense Optical Flow for Cloud Motion Dynamics

Cloud motion dynamics between consecutive frames used the Farnebäck dense optical flow algorithm, which estimates pixel-wise displacement fields representing atmospheric motion. Unlike sparse motion estimation techniques, dense optical flow provides constant motion vectors across the entire cloud field, allowing comprehensive analysis of circulation patterns and mesoscale flow coherence and the algorithm is configured with a pyramid scale of 0.5, three pyramid levels, a window size of 15 pixels, three iterative refinements, polynomial expansion size of 5, and Gaussian smoothing parameter of 1.2. These parameters provide a balance between motion sensitivity and noise suppression, enabling robust estimation of large-scale cloud displacement. The resulting velocity field is converted into motion magnitude:

$$M = \frac{1}{N}\sum\sqrt{u^2 + v^2} \qquad (10)$$

where u and v represent horizontal and vertical motion components. The mean magnitude serves as an indicator of circulation strength, rotational coherence, and large-scale cloud transport. Increasing motion magnitude often corresponds to strengthening cyclonic circulation, while weak or inconsistent motion fields characterize non-cyclonic cloud regimes.

IV. METHODOLOGY

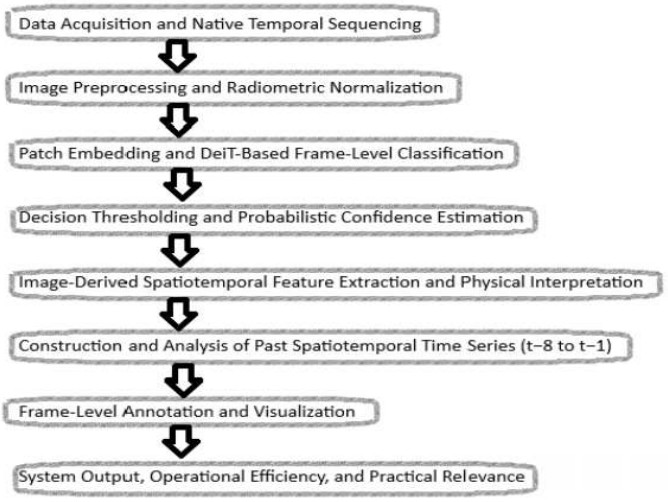


Figure 8. *End-to-End DeiT-Based Spatiotemporal Cyclone Detection and Analysis Pipeline*

A. Data Acquisition and Native Temporal Sequencing

The proposed cyclone detection framework is developed using geostationary satellite infrared (IR) imagery covering the North Indian Ocean (NIO) basin, a region characterized by pronounced spatiotemporal variability in tropical cyclone genesis and evolution. Satellite observations are obtained from the publicly available Deep Learning Cyclone Intensity Estimation NIO dataset, which provides cyclone-centered IR image sequences with verified storm metadata and ground-truth labels.

This is provisional Best Track. The Final will Be uploaded after review of the panel.

Serial Number of system during year	Basin of origin	Name	Date(DD/MM/YY YY)	Time (UTC)	Latitude (lat)	longitude (Long)	CI No [or "T. No"]	Estimated Central Pressure (hPa) [or "E.C.P"]	Maximum Sustained Surface Wind (kt)	Pressure Drop (hPa)[or "delta P"]	Grade (text)	Outermost closed Isobar (hPa)	Diameter/Size of outermost closed Isobar(in degree)
2	BOB		3/20/2022	0000	9.8	92.5	1.5	1002	25	3	D		
	BOB			0300	10.1	92.6	1.5	1002	25	3	D		
	BOB			0600	10.4	92.9	1.5	1002	25	3	D		
	BOB			1200	10.9	93.2	1.5	1002	25	3	D		
	BOB			1800	11.3	93.4	1.5	1001	25	4	D		
	BOB		3/21/2022	0000	11.9	93.7	2.0	1000	30	5	DD		
	BOB			0300	12.3	93.8	2.0	1000	30	5	DD		
	BOB			0600	12.6	93.9	2.0	1000	30	5	DD		
	BOB			1200	13.3	94.0	2.0	1000	30	5	DD		
	BOB			1800	14.0	94.0	2.0	1000	30	5	DD		
	BOB		3/22/2022	0000	14.9	94.2	2.0	1000	30	5	DD		
	BOB			0300	15.2	94.2	2.0	1000	30	5	DD		
	BOB			0600	15.8	94.2	2.0	1000	30	5	DD		
	BOB			Crossed Myanmar coast during 0800-0900 UTC near 16°N/94.2°E with maximum sustained wind speed of 30 kt gusting to 40 kt									
	BOB			0900	16.3	94.3	-	1000	30	5	DD		
	BOB			1200	16.6	94.4	-	1001	25	4	D		
	BOB			1800	17.6	94.5	1.5	1002	25	3	D		
	BOB			Crossed Myanmar coast during 1900-2000 UTC near 17.8°N/94.5°E with maximum sustained wind speed of 25 kt gusting to 30 kt									
	BOB		3/23/2022	0000	19.0	94.6	-	1003	20	3	D		
	BOB			0300	Weakened into a well marked low pressure area over Myanmar								
	BOB												

Figure 9: *Snapshot of Best Track Data taken from RSMC New Delhi*

The dataset includes both cyclonic and non-cyclonic atmospheric conditions, ensuring exposure to organized tropical systems as well as visually similar but dynamically incoherent cloud regimes such as stratiform cloud fields, fragmented convection, and weak mesoscale disturbances. In contrast to approaches that artificially regulate temporal resolution through frame interpolation or optical-flow-based reconstruction, the proposed framework performs explicitly on native sequential satellite observations acquired at regular temporal intervals. Every sample consists of a chronologically ordered image sequence preserving the true atmospheric evolution without introducing synthetic intermediate frames. The design ensures that temporal findings reflect legitimate physical processes rather than reconstruction tools. To incorporate historical atmospheric context, the framework explicitly utilizes past temporal frames spanning t−8 to t−1, representing antecedent cloud structure and motion preceding the present observation at time t. These past frames are not merged into a recurrent or attention-based temporal encoder; instead, they are retained as a physically grounded temporal window that supports post-inference spatiotemporal characterization. This strategy preserves model simplicity while enabling objective analysis of temporal continuity and organizational trends critical for early-stage cyclone detection. All image sequences are temporally sorted prior to preprocessing, and no temporal shuffling or interpolation is performed at any stage of the pipeline.

B. Image Preprocessing and Radiometric Normalization

Each satellite frame undergoes a standardized preprocessing pipeline designed to ensure spatial consistency, stabilize radiometric properties, and preserve physically meaningful cloud characteristics Raw infrared images are first resized to a uniform spatial resolution of 224 × 224 pixels to meet

the input dimensionality requirements of the Data-efficient Image Transformer (DeiT) backbone while retaining mesoscale cloud organization such as spiral banding, convective clustering, and large-scale cloud shields. Pixel intensities are normalized to the unit interval [0, 1] to mitigate variability arising from sensor calibration differences, illumination effects, and seasonal radiative fluctuations.

This normalization ensures that the transformer model learns relative spatial and structural contrasts rather than absolute brightness values.

	FileName	Size (W,H)	Width	Height	Aspect Ratio
0	20001127.12-45.jpg	(333, 339)	333	339	0.98
1	20001127.15-45.jpg	(332, 340)	332	340	0.98
2	20001127.18-45.jpg	(335, 339)	335	339	0.99
3	20001127.21-45.jpg	(333, 340)	333	340	0.98
4	20001128.00-55.jpg	(334, 339)	334	339	0.99
5	20001128.03-55.jpg	(334, 340)	334	340	0.98
6	20001129.12-55.jpg	(334, 340)	334	340	0.98
7	20001129.15-50.jpg	(332, 340)	332	340	0.98
8	20001129.18-45.jpg	(334, 340)	334	340	0.98
9	20001129.21-40.jpg	(334, 339)	334	339	0.99
10	20001225.06-45.jpg	(333, 339)	333	339	0.98
11	20001225.09-50.jpg	(334, 339)	334	339	0.99
12	20001225.12-55.jpg	(333, 340)	333	340	0.98
13	20190107.00_(1).jpg	(331, 334)	331	334	0.99
14	20190107.00_(2).jpg	(334, 332)	334	332	1.01
15	20190107.06_(1).jpg	(333, 331)	333	331	1.01
16	20190107.06_(2).jpg	(333, 331)	333	331	1.01
17	20190107.12_(1).jpg	(332, 333)	332	333	1.00
18	20190107.12_(2).jpg	(333, 333)	333	333	1.00
19	20190107.18_(1).jpg	(333, 333)	333	333	1.00
20	20190107.18_(2).jpg	(334, 333)	334	333	1.00
21	20190108.00_(1).jpg	(333, 332)	333	332	1.00
22	20190108.00_(2).jpg	(335, 332)	335	332	1.01
23	20190108.06_(1).jpg	(333, 333)	333	333	1.00
24	20190108.06_(2).jpg	(332, 333)	332	333	1.00
25	20190108.12_(1).jpg	(333, 332)	333	332	1.00
26	20190108.12_(2).jpg	(333, 332)	333	332	1.00
27	20190108.18_(1).jpg	(334, 333)	334	333	1.00
28	20190108.18_(2).jpg	(334, 333)	334	333	1.00

Figure 10. *Resolution consistency and aspect-ratio verification of Infrared satellite images during preprocessing and radiometric normalization.*

For physically interpretable analysis, grayscale representations are derived from the normalized RGB images. Infrared grayscale intensity directly corresponds to cloud-top radiative properties and convective depth, making it suitable for extracting image-derived spatiotemporal descriptors. Grayscale frames serve as the basis for brightness statistics, texture analysis, and optical-flow computation. Throughout preprocessing, strict temporal ordering of frames is preserved to ensure that inter-frame motion estimation reflects true atmospheric evolution rather than preprocessing-induced artifacts.

C. Patch Embedding and DeiT-Based Frame-Level Classification

The core classification component of the framework is a Data-efficient Image Transformer (DeiT), selected for its ability to capture global spatial dependencies while maintaining robust performance under limited labeled data conditions. In contrary to convolutional neural networks that rely on localized receptive fields, the transformer architecture processes each satellite frame as a sequence of patch tokens, enabling explicit modeling of long-range spatial relationships critical for cyclone identification. Each preprocessed image is divided into non-overlapping patches using a convolution-based patch embedding layer with kernel size and stride equal to the patch dimension. This operation works on spatial patches into a embedding space and reshapes them into a token sequence suitable for transformer processing. Positional encodings are added to maintain spatial arrangement. The embedded tokens are passed through stacked transformer encoder blocks composed of multi-head self-attention (MHSA) and feed-forward networks. Through self-attention, the model learns interactions between distant cloud regions, enabling detection of large-scale organizational cues such as spiral symmetry, eye-like features, and coherent rainband structure that extend beyond local texture patterns. Importantly, classification is performed at the frame level using the present-time image (t) only. Historical frames (t−8 to t−1) are not directly ingested by the transformer; instead, the model focuses on spatial organization at the current observation and ensuring architectural simplicity and inference efficiency. This design choice separates classification from temporal modeling while allowing temporal information to be analyzed explicitly through physically interpretable descriptors.
The transformer output corresponding to the classification token is passed to a lightweight classification head, producing a scalar probability indicating cyclone presence within the frame.

D. Decision Thresholding and Probabilistic Confidence Estimation

Following transformer-based inference, the proposed framework maps the continuous posterior probability output to discrete decision categories using a calibrated probabilistic thresholding strategy. Rather than enforcing a rigid binary classification, a tri-level decision scheme is adopted to explicitly manage epistemic uncertainty under ambiguous or transitional atmospheric conditions. This formulation performs the inherent continuum of cloud organization states observed in satellite imagery and particularly during cyclone genesis and decay phases. Specifically, frames with posterior probability exceeding an upper threshold are classified as *Cyclone*, while frames below a lower threshold are labeled as *No Cyclone*. Predictions within the intermediate range are categorized as *Uncertain*, indicating insufficient confidence for automated decision-making. This uncertainty-aware labeling mechanism is designed to reduce false alarms arising from transient convective clusters, stratiform cloud decks, or weakly organized mesoscale disturbances that may visually resemble early-stage cyclones yet lack sustained dynamical coherence. Prediction confidence is quantified as the normalized distance of the posterior probability from the decision boundary, providing a continuous measure of model certainty. In contrast to hard decision rules that discard probabilistic information, this formulation preserves confidence granularity and enables risk-sensitive deployment. From an operational perspective, uncertain predictions can be flagged for expert review, deferred classification, or integration with complementary observational sources, thereby enhancing reliability in early warning systems. This explicit encoding of uncertainty distinguishes the proposed framework from deterministic cyclone detection approaches and aligns with the requirements of operational meteorology, where conservative decision-making and transparency are critical.

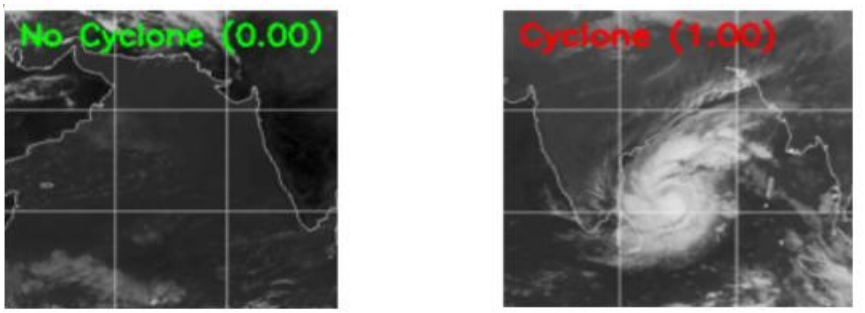


Figure. 11. *a) Non-Cyclonic Confidence Estimation Output b) Cyclonic Confidence Estimation Output*

E. Image-Derived Spatiotemporal Feature Extraction and Physical Interpretation

To complement transformer-based classification and enhance physical interpretability, the proposed framework extracts a set of image-derived spatiotemporal descriptors that explicitly quantify cloud radiative properties, structural organization, and dynamical motion. These features are computed independently of the learned transformer representations and serve as physically meaningful indicators of cyclone-related atmospheric processes.

Feature Type	Feature Name	Processing Method	Represents
Spatial	Mean Brightness	Average grayscale intensity	Cloud density
	Texture Energy	Sobel gradient magnitude	Structural cloud variation
Temporal	Optical Flow	Farnebäck method between consecutive frames	Cloud motion dynamics

Table 6: *Feature Descriptors Utilized for Cyclone Structure and Cloud- dynamic Characterization*

1) Mean Brightness

Mean brightness is computed from grayscale infrared imagery and represents the spatially averaged cloud-top radiative intensity. In infrared observations, lower brightness temperatures (higher radiative intensity in normalized form) are strongly associated with deep convection and vigorous vertical development. Persistent high brightness values across successive frames indicate sustained convective activity, a hallmark of organized cyclonic systems.

By summarizing large-scale radiometric behavior, mean brightness provides a robust, noise-resistant metric that captures bulk convective intensity while remaining insensitive to localized pixel-level fluctuations.

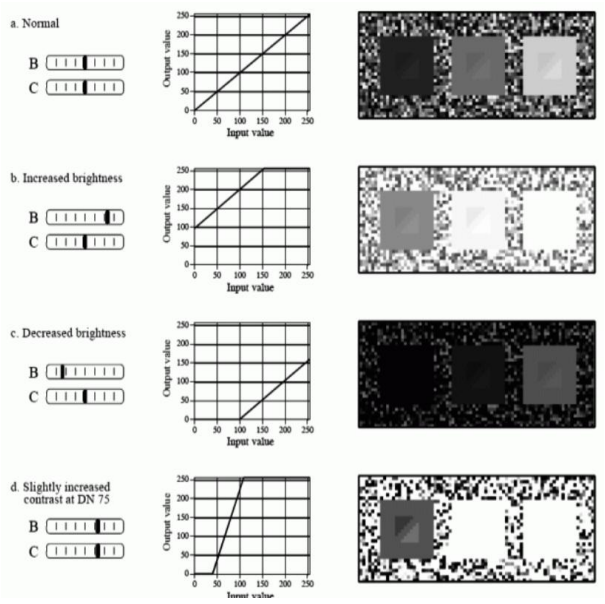


Figure 12. *Radiometric Normalization Examples:*
(a) Normal: Original intensity distribution without modification.
(b) Increased Brightness: Uniform upward shift of pixel intensities to enhance visibility.
(c) Decreased Brightness: Downward intensity adjustment to suppress overexposed regions.
(d) Increased Contrast: Local contrast stretching to emphasize cloud boundaries and texture

2) Texture Energy (Gradient-Based Structural Complexity)

Spatial structural organization is quantified using gradient-based texture analysis. Horizontal and vertical intensity gradients are estimated via Sobel operators, and texture energy is defined as the spatial average of gradient magnitude. This formulation captures organized spatial discontinuities associated with curved rainbands, eye-wall boundaries, and coherent cloud clustering. Cyclonic systems typically exhibit elevated texture energy due to organized mesoscale curvature and rotational symmetry, whereas non-cyclonic cloud regimes such as stratiform decks or fragmented convection display weaker and less coherent gradient structure. Importantly, texture energy is invariant to uniform brightness shifts, ensuring that structural complexity is emphasized over absolute radiometric values.

3) Dense Optical Flow Magnitude

Temporal dynamics are characterized using dense optical flow estimated between consecutive native satellite frames via the Farneback polynomial expansion method. The resulting displacement vectors represent local cloud motion, and their average magnitude provides a compact descriptor of motion persistence and circulation strength. Temporal dynamics are characterized using dense optical flow estimated between consecutive grayscale frames via Farneback's polynomial expansion method. The resulting displacement vectors $(w\ v)$represent local cloud motion, and the average flow magnitude,

$$F = \frac{1}{HW}\sum_{i,j}\sqrt{u(i,j)^2 + v(i,j)^2} \qquad (11)$$

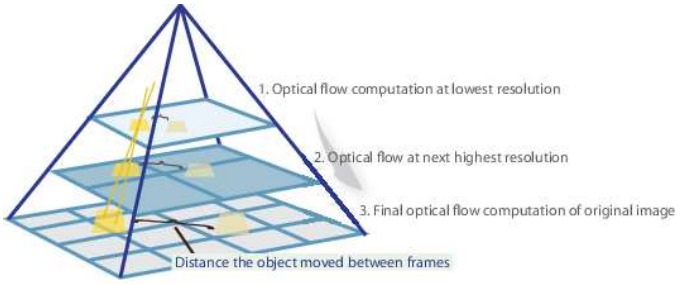


Figure 13. *Multi-scale optical flow estimation framework*

This distinction ensures that temporal motion characterization remains grounded in observed atmospheric displacement rather than inferred or interpolated dynamics.

F. Construction and Analysis of Past Spatiotemporal Time Series (t−8 to t−1)

	FileName	Size (W,H)	Width	Height	Aspect Ratio	File Size (bytes)
0	t-1.jpg	(333, 331)	333	331	1.01	18318
1	t-2.jpg	(331, 334)	331	334	0.99	17763
2	t-3.jpg	(334, 332)	334	332	1.01	20558
3	t-4.jpg	(335, 339)	335	339	0.99	20422
4	t-5.jpg	(333, 340)	333	340	0.98	20855
5	t-6.jpg	(334, 339)	334	339	0.99	20473
6	t-7.jpg	(332, 340)	332	340	0.98	20245
7	t-8.jpg	(333, 339)	333	339	0.98	20650

Figure 14. *Dimensional and aspect-ratio verification of historical frames (t−8 to t−1) used for past time-series construction.*

To capture temporal continuity beyond instantaneous frame-level appearance, the proposed framework constructs multivariate spatiotemporal time series from image-derived descriptors across the historical window spanning t−8 to t−1. For each frame in the sequence, mean brightness, texture energy, and optical-flow magnitude are aggregated in chronological order, preserving the native temporal evolution of cloud organization and motion. This explicit time-series construction enables analysis of persistence, trend consistency, and dynamical stability—key indicators of genuine cyclonic development. Sustained elevation in radiometric intensity and texture, coupled with increasing or stable optical-flow magnitude, reflects organized circulation and convective maintenance. In contrast, non-cyclonic regimes typically exhibit irregular, transient, or weakly correlated temporal profiles. Hence approaches that encode temporal dependency implicitly within recurrent or attention-based latent states, the proposed method retains temporal information in an interpretable form. This design choice avoids reliance on learned temporal abstractions that are difficult to validate physically and allows temporal reasoning to be directly inspected, visualized, and cross-validated against meteorological expectations. Furthermore, by separating temporal analysis from the classifier backbone, the framework remains robust to incomplete or variable-length sequences and avoids compounding classification errors across time.

G. Frame-Level Annotation and Visualization

To support qualitative interpretation and diagnostic assessment of model predictions, each processed satellite frame is annotated with the predicted class label, posterior probability, confidence score, and associated image-derived spatiotemporal feature values. Therefore, visual overlays are applied directly to the satellite imagery to contextualize classification outputs with respect to observed cloud structure and spatial organization. Bounding boxes are used to delineate the region of interest, whereas textual annotations display probabilistic outputs and confidence metrics corresponding to each frame. These annotated visualizations provide an intuitive means of examining differences between cyclonic and non-cyclonic atmospheric conditions, particularly in terms of cloud morphology, organization, and spatial coherence. High-confidence cyclone detections typically exhibit well-defined spiral structures, organized convective cores, or symmetric cloud shields, while non-cyclonic cases display fragmented or weakly organized cloud patterns. By presenting model outputs alongside the underlying imagery, the visualization framework enables direct visual verification of classification behavior without reliance on abstract performance metrics alone and in addition to frame-level annotations, the framework generates temporal visualizations of extracted spatiotemporal descriptors, including mean brightness, texture energy, and optical-flow magnitude, plotted across successive observation times. These plots simplify examination of temporal consistency and dynamical evolution in cloud structure and motion, allowing researchers to assess whether detected cyclonic signatures are supported by sustained physical trends rather than isolated visual artifacts. The visualization module is designed as an analytical tool to aid model evaluation, result interpretation, and post-event analysis. It does not alter model predictions or influence decision logic, but instead provides a transparent representation of how classification outputs relate to observable atmospheric features. This emphasis on diagnostic visualization enhances interpretability and reproducibility while remaining consistent with the scope and objectives of the proposed framework.

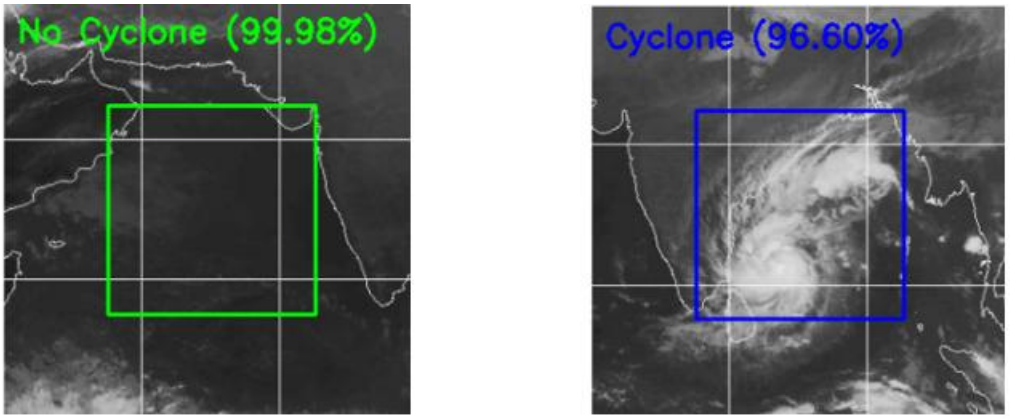


Figure. 15. *a) Frame-Level Annotation for Non-Cyclonic Scene(99.98% confidence) b) Frame-Level Annotation for Cyclonic Structure(96.60% confidence)*

H. System Output, Operational Efficiency, and Practical Relevance

The final system produces a unified output consisting of frame-level cyclone predictions, probabilistic confidence estimates, annotated satellite imagery, and spatiotemporal trend visualizations. By integrating transformer-based global spatial reasoning with explicit physical characterization of temporal dynamics, the framework delivers robust detection performance while maintaining interpretability and operational transparency. From a computational perspective, inference requires a single forward pass through the DeiT model per frame, followed by lightweight image-processing operations for feature extraction. This design ensures suitability for near-real-time deployment without reliance on computationally expensive temporal reconstruction or sequence modeling. The modular architecture allows straightforward extension to additional satellite channels or expanded temporal windows without fundamental redesign. Importantly, the framework avoids dependence on historical intensity records or bias-correction mechanisms, enabling applicability during early-stage or weakly organized systems where intensity information may be sparse or unavailable.
The explicit uncertainty labeling further supports conservative decision-making, reducing false alarms and enabling targeted expert intervention. Overall, the proposed methodology bridges the gap between high-performance deep learning and physically grounded atmospheric analysis, offering a transparent, scalable, and operationally viable solution for early cyclone detection and continuous satellite-based monitoring.

V. ALGORITHM AND MATHEMATICAL FORMULATION

Algorithm 1: Transformer-Based Cyclone Detection and Diagnostic Analysis

After describing the DeiT-based feature extraction, spatial attention modeling, and cyclone probability estimation, the operational inference workflow must be formalized to ensure reproducibility and diagnostic interpretability. The frame-level cyclone detection and descriptor extraction procedure is formalized in Algorithm 1. This algorithm transforms raw satellite imagery into structured atmospheric descriptors while preserving physical interpretability. Transformer-Based Frame Analysis and Decision Inferences

$$I_t \rightarrow \{P_c, L_t, D_t\} \qquad (12)$$

Step	Procedure
Initialization	Initialize trained DeiT model f_θ, cyclone thresholds $\tau_c$and τ_n, diagnostic storage T, and previous frame $I_{prev} = \emptyset$.
1	**For each satellite frame $I_t$do**
1.1	**Preprocessing Phase**
	Resize $I_t$to 224×224resolution.
	Normalize pixel intensities to ([0,1]).
	Convert image into tensor representation.
1.2	**Prediction Phase**
	Compute cyclone probability: $p_t = f_\theta(I_t)$.
	If $p_t \geq \tau_c \rightarrow L_t$ =Cyclone, confidence $C_t = p_t$.
	Else if $p_t \leq \tau_n \rightarrow L_t$ = non-cyclone, confidence $C_t = 1 - p_t$. |
	Else $\rightarrow L_t$ =Uncertain, confidence $C_t = \max(p_t, 1 - p_t)$.
1.3	**Structural Intensity Analysis**
	Convert frame to grayscale G_t.
	Mean brightness: $B_t = \text{mean}(G_t)$.
	Brightness deviation: $\sigma_t = \text{std}(G_t)$.
	Texture energy: $TE_t = \frac{1}{N}\sum\sqrt{G_x^2 + G_y^2}$.
1.4	**Motion Analysis Phase**
	If $I_{prev} \neq \emptyset$: |
	Compute dense optical flow between $I_{prev}$and I_t.
	Motion magnitude: $OF_t = \frac{1}{N}\sum\sqrt{u^2 + v^2}$.
	Else set $OF_t = 0$
1.5	**Store Diagnostics**
	Append $(B_t,\ TE_t,\ OF_t)$to diagnostic storage T.
1.6	**Visualization**
	Overlay label L_t, confidence C_t, and diagnostics on frame.
	Display annotated output.

1.7	**Update Frame Memory**
	Set $I_{prev} \leftarrow I_t$.
End	End loop.
Output	Cyclone detection labels, confidence scores, and temporal diagnostic series.

Table 7: *Algorithm 1: Frame Processing and Cyclone Detection Workflow*

Where at equation (12),

- I_t= input satellite frame at time t
- P_c= cyclone probability predicted by the transformer
- L_t= decision label (Cyclone / Non-Cyclone / Uncertain)
- D_t= descriptor vector

$$D_t = \{B_t, T_t, F_t\} \qquad (13)$$

with:

- B_t= mean brightness
- T_t= texture energy
- F_t= optical flow magnitude

Descriptor Formulations Included in Algorithm Explanation

Mean Brightness

$$B_t = \frac{1}{N}\sum_{i=1}^{N} I_t\,(i) \qquad (14)$$

where $N$is the number of pixels.

Texture Energy (Gradient Magnitude)

$$T_t = \frac{1}{N}\sum_{i=1}^{N}\sqrt{G_x^2 + G_y^2} \qquad (15)$$

where $G_x, G_y$are Sobel gradients representing horizontal and vertical intensity variations.

Optical Flow Magnitude

$$F_t = \frac{1}{N}\sum_{i=1}^{N}\sqrt{u_i^2 + v_i^2} \qquad (16)$$

where u, vrepresent motion vectors between consecutive frames.

Decision Rule Embedded in Algorithm

$$L_t = \begin{cases} \text{Cyclone}, & P_c \geq \tau_c \\ \text{Non-Cyclone}, & P_c \leq \tau_n \\ \text{Uncertain}, & \text{otherwise} \end{cases} \qquad (17)$$

where $\tau_c = 0.75$and $\tau_n = 0.25$.

This algorithm bridges transformer inference with physically interpretable atmospheric descriptors, enabling structured decision-making and diagnostic transparency.

To formalize the learning objective, consider a dataset

$$X = \{x_1, x_2, \ldots, x_n\}, x_i \in \mathbb{R}^d \qquad (18)$$

The transformation function is:

$$f: \mathbb{R}^d \to \mathbb{R}^k \qquad (19)$$

optimized with:

$$\min_{\theta} L(\theta) = \frac{1}{n}\sum_{i=1}^{n} l(f(x_i;\theta), y_i) + \lambda\Omega(\theta) \qquad (20)$$

where:

- $l(\cdot)$= loss function
- $\Omega(\theta)$= regularization term
- λ= regularization weight
- θ= model parameters

This formulation ensures balance between predictive accuracy and generalization.

Algorithm 2: Temporal Dynamics and Cyclone Evolution Assessment

After establishing frame-level detection, sequential analysis is required to determine cyclone evolution, structural consolidation, and vortex intensification.

The temporal evolution analysis and cyclone development assessment procedure is described in Algorithm 2.

Step	Procedure
Initialization	Initialize: Google Drive path P trained DeiT model fθ image size S = 224 × 224 uploaded frame sequence I = {I1, I2, … , IN} previous frame Iprev ← ∅ diagnostics storage D ← ∅
1	Mount Google Drive at path P
2	Load trained DeiT cyclone model fθ
3	**For each uploaded frame It in sequence I do**
3.1	**Pre-processing phase** Resize It to S × S Convert It to RGB tensor Normalize pixel values to [0,1]
3.2	**Cyclone prediction phase** pt ← fθ(It) if pt ≥ 0.75 then label ← Cyclone else label ← No Cyclone end
3.3	**Structural feature extraction** Convert frame to grayscale Gt Bt ← mean(Gt) Gx, Gy ← Sobel(Gt) TEt ← mean(√(Gx² + Gy²))
3.4	**Motion estimation phase** if Iprev ≠ ∅ then Compute dense optical flow using Farnebäck method OFt ← mean motion magnitude else OFt ← 0 end
3.5	**Store diagnostics** Append (Bt, TEt, OFt) to D
3.6	**Visualization** Overlay label and probability on frame Store annotated frame for display
3.7	**Update previous frame** Iprev ← It
End Loop	End
4	Display annotated frames in a single horizontal sequence
5	**Past Evolution Analysis (if N ≥ 8 frames)**
5.1	Select past sequence: P ← {I(N-8), … , I(N-3)}
5.2	Compute brightness, texture, and optical flow for P
5.3	Compute: AvgFlow ← mean(flow values) FlowTrend ← flow last − flow first
5.4	if AvgFlow > 1.2 AND FlowTrend > 0 then PastLabel ← Cyclone Evolution Detected else PastLabel ← No Cyclone Evolution end
6	**Temporal Visualization**
6.1	Extract series: BrightnessSeries TextureSeries FlowSeries
6.2	Plot temporal evolution curves for analysis
Return	cyclone detection labels, confidence scores, structural metrics (Average Flow, Flow Trend, Past Sequence Result), temporal evolution assessment

Table 8: *Algorithm 2: Deit-Based Cyclone Detection and Past Series Sequence Analysis Pipeline*

Technical Role

Algorithm 2 evaluates structural evolution across a sequence:

$$S = \{I_{t-k}, \dots, I_t\} \qquad (21)$$

and derives temporal indicators:

$$\bar{F} = \frac{1}{m}\sum_{j=1}^{m} F_j \qquad (22)$$

$$\Delta F = F_{end} - F_{start} \qquad (23)$$

where:

- $\bar{F}$= average flow magnitude
- ΔF= flow trend (circulation strengthening indicator)

Cyclone evolution is inferred when:

$$\bar{F} > \gamma \wedge \Delta F > 0 \qquad (24)$$

where $\gamma$is an empirically determined threshold.

Brightness persistence and texture amplification are also evaluated:

$$\frac{dT}{dt} > 0 \text{and} B_t \approx \text{stable} \qquad (25)$$

indicating strengthening convective organization.

Physical Interpretation

- Increasing $T_t \rightarrow$ consolidation of spiral cloud bands
- Stable high $B_t \rightarrow$ persistent deep convection
- Rising $F_t \rightarrow$ vortex intensification and circulation strengthening

This algorithm transforms sequential satellite observations into measurable indicators of cyclone development and atmospheric organization.

Model parameters are updated iteratively using gradient-based optimization:

$$\theta_{t+1} = \theta_t - \eta \nabla_\theta L(\theta_t) \qquad (26)$$

Adaptive optimization (Adam) improves convergence:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1)\nabla L(\theta_t) \qquad (27)$$
$$v_t = \beta_2 v_{t-1} + (1 - \beta_2)(\nabla L(\theta_t))^2 \qquad (28)$$

These moment estimates stabilize updates and accelerate convergence in high-dimensional parameter spaces.

If each iteration processes all samples, time complexity is:

$$O(nd) \qquad (29)$$

Mini-batch training reduces cost to:

$$O(bd), b \ll n \qquad (30)$$

Memory complexity:

$$O(dk) \qquad (31)$$

Gradient clipping improves stability:

$$g \leftarrow \frac{g}{\max(1, \| g \| / c)} \qquad (32)$$

Normalization improves robustness:

$$x' = \frac{x - \mu}{\sigma} \qquad (33)$$

Robustness and Convergence Considerations: Sparse or noisy satellite inputs may degrade performance. Preprocessing steps including normalization, noise filtering, and contrast enhancement improve descriptor reliability. Handling missing frames, atmospheric artifacts, and illumination inconsistencies ensures temporal continuity. Regularization strategies, dropout mechanisms, and data augmentation prevent overfitting and improve generalization across diverse meteorological conditions.

Step	Procedure
Initialization	Initialize model parameters θ
1	Repeat until convergence
1.1	Sample mini-batch B
1.2	Compute gradients: g = ∇θ L(θ; B)
1.3	Apply stabilization (clipping / normalization)
1.4	Update parameters: θ ← θ − η g
1.5	Adjust learning rate if needed
End Loop	End repeat
Return	Final optimized parameters θ

Table 9: *Iterative Optimization Procedure*

Summary

The proposed transformer-based cyclone detection framework integrates deep vision transformers with physically interpretable atmospheric descriptors to enable accurate, explainable, and temporally aware cyclone analysis. Algorithm 1 establishes a robust frame-level inference pipeline by combining DeiT-based probability estimation with brightness, texture energy, and optical-flow-derived motion descriptors. This integration ensures that predictions are not treated as black-box outputs but are supported by measurable structural and dynamical indicators. Algorithm 2 extends this capability by modeling temporal evolution, enabling the detection of cyclone development through motion intensification, structural consolidation, and convective persistence. By incorporating flow trend analysis and temporal descriptor dynamics, the system captures precursor signatures of vortex strengthening and atmospheric organization. Mathematically grounded optimization, adaptive gradient methods, and stability enhancements ensure reliable convergence and computational efficiency, while robustness strategies improve performance under noisy and variable satellite imaging conditions. The fusion of transformer attention mechanisms with physically meaningful descriptors provides both predictive strength and meteorological interpretability. Overall, this integrated spatiotemporal framework transforms sequential satellite imagery into actionable cyclone diagnostics, supporting early detection, evolution monitoring, and decision-support applications in meteorological forecasting and disaster preparedness.

VI. MODEL OPTIMIZATION AND PRELIMINARY RESULTS

A. Model Optimization

Model optimization in the proposed framework is achieved through architectural efficiency, probabilistic calibration, and computationally lightweight spatiotemporal augmentation within a TensorFlow–Keras deployment environment. The DeiT-based transformer backbone, integrated with a custom convolutional patch embedding layer, performs spatial tokenization using a stride-equal-to-patch projection. This reduces redundant spatial sampling while preserving high-frequency structural information critical for cyclone morphology recognition. Input normalization to the [0,1] intensity range ensures numerical stability and consistent feature scaling across satellite scenes with heterogeneous radiometric characteristics. The fixed spatial resolution of 224 × 224 represents a deliberate trade-off between structural fidelity and inference throughput, enabling efficient GPU execution without compromising mesoscale cloud organization patterns. Optimization at the decision level is achieved via a calibrated dual-threshold probabilistic framework. Rather than relying on a single hard classification boundary, the system employs asymmetric confidence thresholds ($\geq$0.75 for Cyclone, $\leq$0.25 for No Cyclone). This expands the margin around the decision boundary, mitigating high-variance predictions under ambiguous convective structures and improving robustness to transitional storm phases. The intermediate uncertainty band functions as a reliability buffer, effectively reducing false alarm rates while preserving sensitivity to organized systems. Computational optimization is further reinforced by isolating the neural network to single-frame forward inference, thereby eliminating the overhead associated with recurrent or sequence-based architectures. Temporal dynamics are instead quantified through deterministic post-processing metrics, including gradient-based texture energy and dense optical flow magnitude using the Farneback algorithm. By decoupling motion estimation from network training, the framework maintains low parameter complexity while enabling physically interpretable spatiotemporal characterization. The modular separation between transformer inference and physics-informed post-processing enhances scalability and deployment efficiency. Since optical flow and texture computations are non-trainable and computationally bounded, the overall pipeline maintains near real-time performance with controlled memory footprint. Collectively, the optimization strategy emphasizes computational parsimony, probabilistic reliability, and structural interpretability, ensuring stable cyclone detection performance under operational meteorological constraints without introducing additional trainable complexity.

B. Prediction Visualization and Interpretability Assessment

To evaluate operational reliability and qualitative performance, a representative set of satellite infrared test frames from the North Indian Ocean basin was examined alongside their predicted cyclone classifications. Special emphasis was placed on frames exhibiting ambiguous cloud morphology or probability values near decision thresholds, as these conditions present the highest risk of misclassification. Beyond numerical accuracy, the evaluation prioritized interpretability by examining how model predictions aligned with observable atmospheric structures and motion patterns.

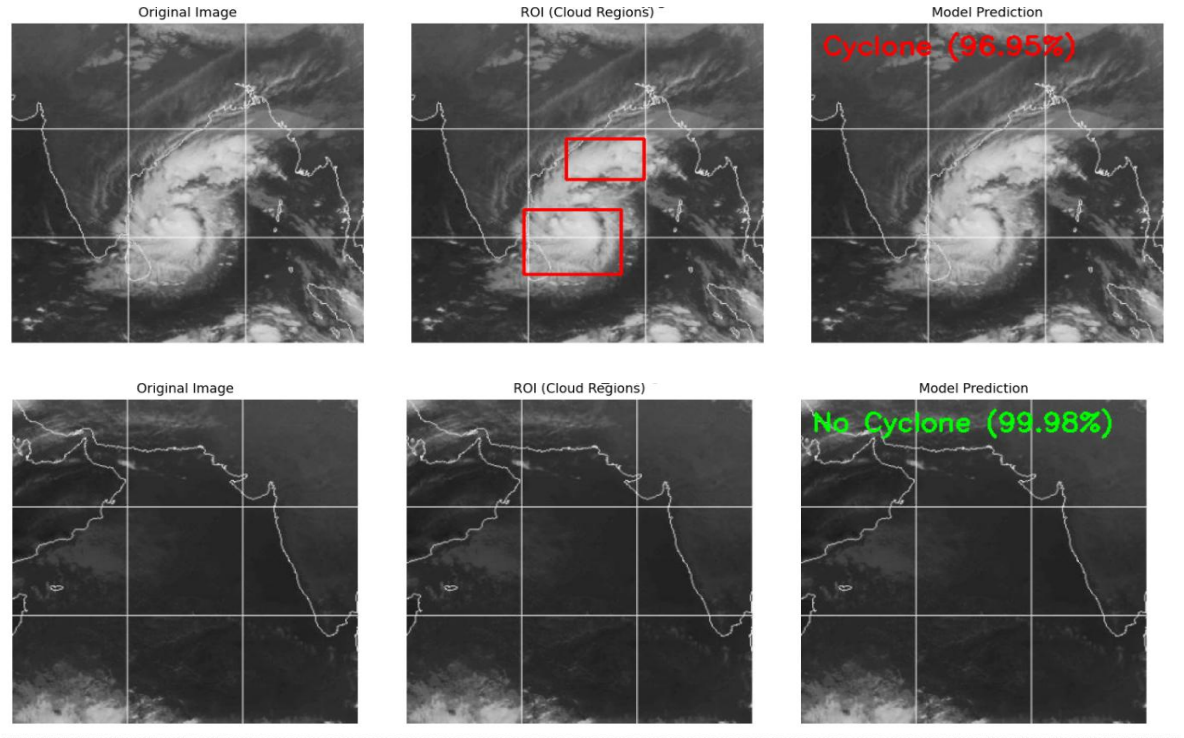


Figure. 16. *a) Prediction Visualization for Cyclonic System (with ROI localization and model prediction ~96.95% confidence of the original input image) b) Prediction Visualization for Non-Cyclonic Scene (with ROI localization and model prediction ~99.98% confidence of the original input image)*

Following training, the transformer-based DeiT classifier was applied to previously unseen satellite imagery to assess generalization performance under diverse meteorological conditions. For every frame, the model produced a cyclone probability score that was mapped into two decision categories: Cyclone and No Cyclone. Confidence levels were derived directly from posterior probabilities, enabling differentiation between high-confidence detections and borderline classifications. This probabilistic framework is particularly valuable during cyclone symmetry, structural transitions, and weakening phases when cloud organization may be incomplete and to enhance interpretability, predicted outputs were overlaid directly on satellite imagery. Region-of-interest bounding boxes were rendered to spatially delineate areas associated with cyclone detection, typically including the central dense overcast region and surrounding spiral rainbands. Along with the bounding box, physically interpretable descriptors—including mean brightness, texture energy, and optical flow magnitude—were displayed. This integration provides a productive linkage between model decisions and observable cloud characteristics without displaying internal transformer activations, thereby maintaining transparency while preserving model abstraction. Frame-level visualization was complemented by spatiotemporal analysis using sequential satellite frames and temporal validation of brightness, texture energy, and optical flow magnitude were generated to characterize cloud evolution and motion dynamics. These time-series plots support qualitative validation by correlating structural organization and circulation coherence with cyclone predictions, offering observation into storm formation and progression. To facilitate operational usability, a lightweight interactive interface was developed to allow users to upload satellite imagery and receive real-time predictions, annotated visual outputs, and spatiotemporal plots. This interface prioritizes clarity, responsiveness, and accessibility, making it suitable for both meteorological analysts and non-expert decision-makers. By combining automated classification with interpretable visual cues, the system supports early cyclone monitoring and informed decision-making in both research and operational environments.

C. Comparative Technical Analysis: Transformer-Based Detection vs. AI with Historical Bias Correction

Using geostationary infrared satellite imagery over the North Indian Ocean, the proposed transformer-based framework demonstrates robust detection of organized cyclonic systems across multiple developmental stages. Visual inspection indicates that the DeiT model consistently emphasizes large-scale spatial coherence, capturing defining cyclone features such as curved convective rainbands, azimuthal symmetry, eyewall consolidation, and the central dense overcast structure during mature phases. The global self-attention mechanism enables the model to integrate contextual information across the entire cloud field, allowing reliable discrimination between organized cyclonic systems and non-cyclonic convective clusters. For well-defined cyclones, detection probabilities frequently exceed 0.99, with bounding boxes accurately encapsulating both the storm core and surrounding spiral bands, thereby providing spatially meaningful localization and enhanced situational awareness.

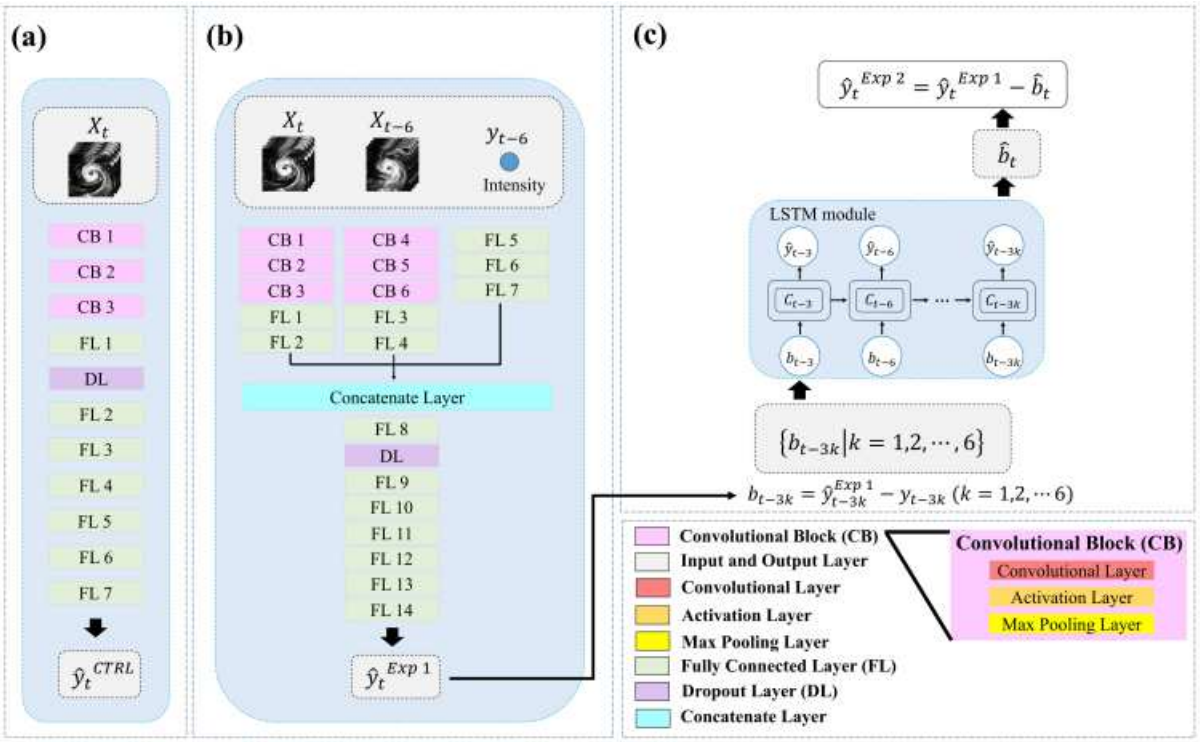


***Figure 17**: Architectures of (a) CTRL model, which uses only current TC images (), (b) Experiment 1 (Exp 1) model, which uses TC images () and intensity information () from 6 h earlier as additional inputs, and (c) Exp 2 model, which incorporates a self-diagnosis algorithm that corrects intensity biases () from the past 18 h (). The meaning of the colors used for each layer are shown in the bottom right corner of the figure.*
[DOI: 10.1109/TGRS.2025.3632317]

In contrast, the methodology presented by Jung et al. focuses primarily on tropical cyclone intensity estimation using convolutional feature extraction integrated with historical intensity trajectories and track analysis. Their evaluation emphasizes statistical agreement between predicted and reference wind speeds through density scatter plots, correlation coefficients, root mean square error (RMSE), mean absolute error (MAE), and bias metrics. While this approach demonstrates high accuracy in intensity regression, it does not explicitly model spatial cloud organization or provide frame-level interpretability of cyclone morphology. A key distinction lies in spatial localization capability. The transformer-based framework produces bounding-box-level localization and explicit cyclone presence classification, enabling real-time situational awareness. The Jung et al. framework, by design, assumes the presence of a cyclone and focuses on intensity estimation, thereby not addressing false-positive detection or early-stage cyclone screening. Beyond classification outputs, the proposed framework computes physically interpretable descriptors, including mean brightness, texture energy, and inter-frame optical flow magnitude. Cyclonic frames exhibit moderate brightness values associated with cold, optically thick cloud tops, elevated texture energy reflecting tightly organized spiral bands, and coherent optical flow magnitudes indicative of rotational dynamics. These descriptors provide direct insight into cloud organization and atmospheric motion. Conversely, Jung et al. incorporate temporal evolution implicitly through historical intensity inputs but do not compute or visualize physical image-derived descriptors. As a result, temporal interpretability is limited to intensity-based regression trends rather than observable structural evolution. The transformer-based framework performs temporal interpretation in measurable atmospheric motion and cloud structure, enhancing physical interpretability. The DeiT model also demonstrates strong suppression of false positives. Even when localized cloud elements exhibit high brightness or texture, the model rejects non-organized convection due to its global

contextual reasoning capability. In contrary, intensity estimation frameworks assume cyclone presence and therefore do not explicitly address false detection. Time-series analysis of brightness, texture energy, and optical flow across cyclone sequences reveals persistent elevation and structured variability, consistent with sustained organization and rotational coherence throughout the storm lifecycle. In combined sequences, these metrics clearly separate cyclonic and non-cyclonic regimes, providing physically interpretable temporal signatures.

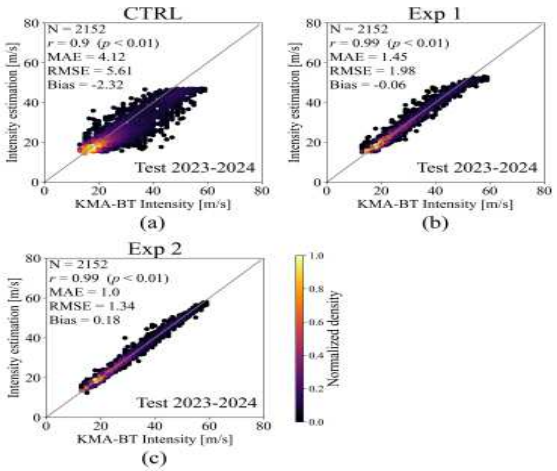


Figure 18 : *Density scatter plots of TC intensity estimation for GK2A test data using (a) CTRL model, (b) Exp 1 model, and (c) Exp 2 model. In each panel, the -axis is the KMA-BT data (KMA-BT) TC intensity, and the -axis is the intensity estimated by the models.*
[DOI: 10.1109/TGRS.2025.3632317]

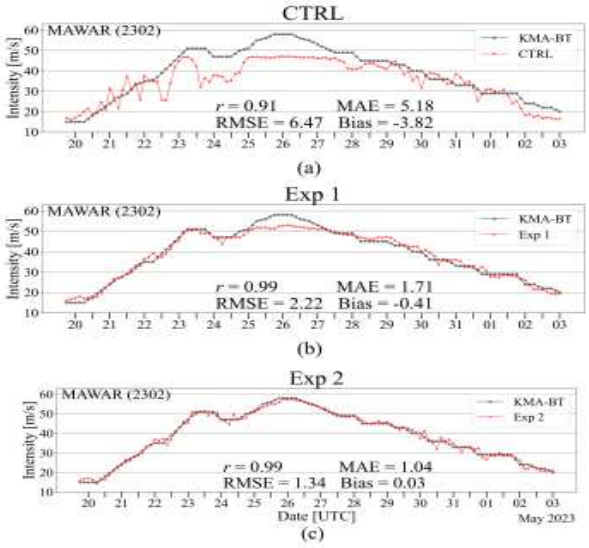


Figure 19 : *Time series of TC intensity estimated by (a) CTRL model, (b) Exp 1 model, and (c) Exp 2 model for the 2023 TC Mawar. The TC name, followed by the last two digits of the year of occurrence and the TC number in parentheses, is shown in the upper left corner.*
[DOI: 10.1109/TGRS.2025.3632317]

While Jung et al. report high correlation coefficients (approximately 0.97–0.98) and low RMSE values for intensity estimation, these metrics do not quantify cloud morphology or motion dynamics. The proposed transformer-based framework extends interpretability beyond regression accuracy by linking probabilistic detection with measurable physical descriptors and spatial organization patterns. Overall, the transformer-based approach complements intensity-focused methodologies by delivering robust cyclone detection with explicit spatial localization and interpretable spatiotemporal diagnostics. By leveraging global self-attention and image-derived physical metrics, the system provides actionable, frame-level insight into cyclone presence and evolution. This capability supports early warning applications, operational screening, and situational awareness, whereas intensity-focused approaches excel in refining wind speed estimates for already identified cyclones.

VII. RESULTS AND DISCUSSIONS

A. Qualitative Results: Cyclone Stage Representation

The reference study presents qualitative visualization of cyclone stages using infrared satellite imagery from the CIMSS dataset, emphasizing representative spatial patterns across intensity classes. The dataset used in this study was obtained from the publicly available repository [manishmawatwal, *Deep_Learning_Cylcone_Intensity_Estimation_NIO_Satellite*, GitHub] which contains labeled satellite imagery for cyclonic and non-cyclonic conditions over the North Indian Ocean. Early-stage depressions are characterized by loosely organized convective clusters with fragmented cloud tops and weak curvature signatures. As system intensity increases, mature cyclones exhibit tightly wound spiral rainbands, enhanced azimuthal symmetry, and expansive central dense overcast (CDO) regions with persistently cold cloud-top temperatures. In contrast, non-cyclonic scenes display irregular cloud distributions, lacking radial symmetry and mesoscale structural coherence.

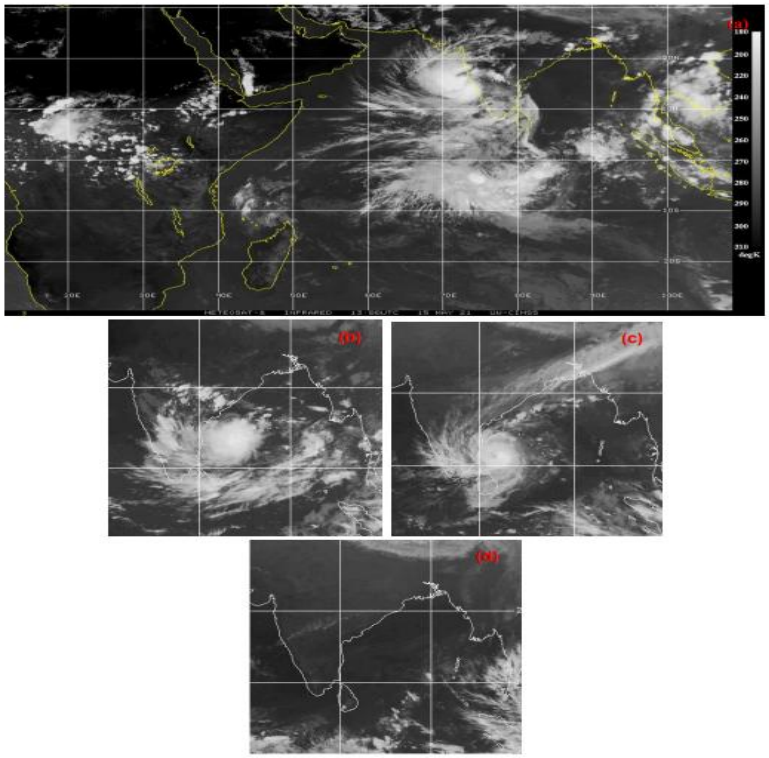


***Figure 20**: (a) Raw CIMSS infrared image, (b) Depression stage, (c) Very Severe Cyclonic Storm, and (d) non-cyclonic cloud field.*

These representative examples illustrate the intrinsic visual complexity associated with cyclone detection, where structural organization evolves continuously across intensity stages. Increasing storm strength is accompanied by enhanced spatial coherence, convective consolidation, and expansion of high-altitude cloud shields, whereas non-cyclonic environments exhibit disorganized convection and weak structural continuity. In contrast to purely visual interpretation, the proposed framework augments qualitative visualization with transformer-based inference outputs and physically interpretable diagnostics. Each frame is enriched with predicted cyclone probability, confidence level, spatial localization, and derived descriptors including mean brightness, gradient-based texture energy, and optical flow magnitude. This integrated representation enables direct correspondence between observed cloud morphology and algorithmic decision-making. Furthermore, while conventional qualitative analyses rely solely on visual inspection, the proposed system associates visual patterns with quantified spatiotemporal signatures. This linkage enables measurable characterization of cyclone organization, convective persistence, and circulation development, thereby transforming qualitative visualization into an interpretable diagnostic tool. The enriched qualitative representation improves interpretability and operational relevance by bridging visual meteorological assessment with model-driven inference.

B. Cyclone Prediction Visualization and Structural Diagnostics

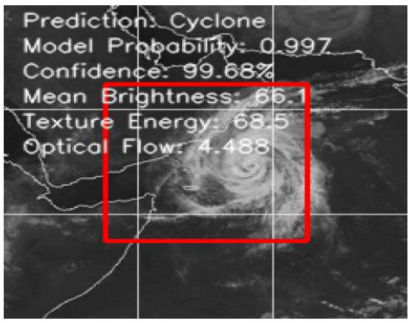


***Figure 21:** High-confidence cyclone detection with coherent spiral structure, elevated texture energy, and strong rotational motion.*

Unlike intensity-regression frameworks such as Jung et al., where cyclone characterization is inferred indirectly through wind-speed estimation and bias correction, the proposed transformer-based system performs direct frame-level cyclone detection supported by spatial localization and physically interpretable diagnostics. For the mature cyclonic system illustrated in Fig. 12, the model produces a cyclone probability of 0.997 (99.68% confidence), accompanied by bounding-box localization that encloses the cyclone core and surrounding spiral rainbands. The DeiT architecture's global self-attention mechanism enables holistic evaluation of cloud-field organization, allowing discrimination between organized rotational systems and disordered convection. Quantitative descriptors extracted from the detected frame reveal:

- Moderate mean brightness consistent with cold, optically thick cloud tops
- Elevated texture energy reflecting tightly organized mesoscale gradients
- Significant optical flow magnitude (~4.49) indicating coherent rotational motion

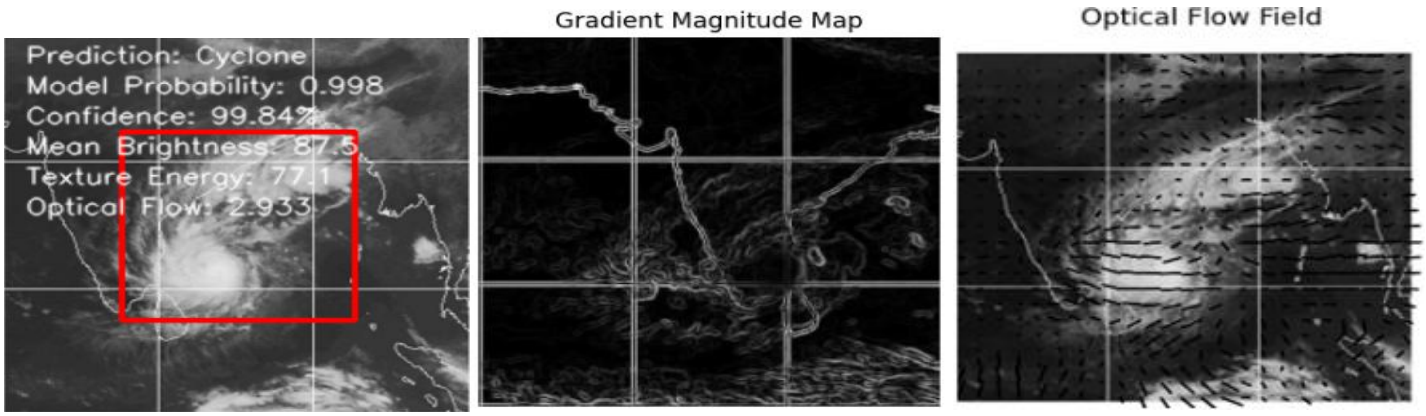


Figure 22: *Cyclone Detection Using Integrated Gradient and Optical Flow Analysis*

(a) Original satellite image with ROI showing cyclone prediction (Probability = 0.998, Confidence = 99.84%) and extracted features: Mean Brightness = 87.5, Texture Energy = 77.1, Optical Flow Magnitude = 2.933.
(b) Gradient magnitude map highlighting strong edge intensities and organized spiral cloud structures.
(c) Optical flow field illustrating coherent rotational vector patterns, included to capture vortex-based atmospheric circulation.

These indicators align with established meteorological signatures of mature tropical cyclones, including deep convection persistence and organized circulation. The approaches dependent on historical intensity trajectories, the proposed framework validates cyclone presence directly through measurable structural and motion characteristics. This capability is critical for early detection scenarios where historical intensity data may be sparse or unavailable.

C. Non-Cyclone Prediction Visualization and False-Positive Suppression

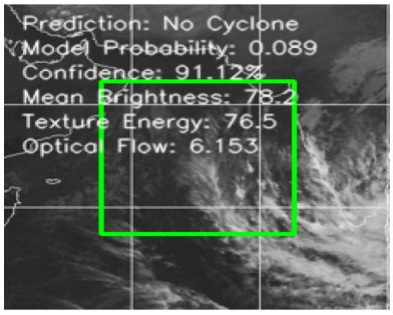


Figure 23: *Non-cyclone classification with high brightness and texture but absence of large-scale rotational coherence.*

The analyzed satellite frame is classified as No Cyclone with a predicted cyclone probability of 0.089 and confidence of 91.12%. Although the scene exhibits moderate atmospheric activity, the spatial organization does not satisfy cyclonic structural criteria.

Measured descriptors include:

- **Mean brightness: 78.2** (dense cloud validation)
- **Texture energy: 76.5** (localized structural gradients)
- **Optical flow magnitude: 6.15** (cloud motion dynamics)

Despite these indicators, the absence of radial symmetry, closed circulation, and spiral band organization preserves positive classification. Although certain non-cyclonic frames may exhibit elevated instantaneous values of brightness, texture energy, or motion magnitude due to deep convective clusters, mesoscale convective systems, or strong environmental winds, these rises are spatially fragmented and temporally inconsistent. In contrast to cyclonic systems, they lack radial symmetry, organized spiral banding, and persistent rotational circulation. The distinction highlights that cyclone identification depends not on absolute parameter magnitude but on structural coherence and spatiotemporal persistence, enabling the proposed framework to suppress false positives while maintaining sensitivity to organized cyclonic development and from a transformer perspective, global self-attention evaluates contextual relationships across all image patches. While localized regions exhibit cyclone-like features, holistic context suppresses cyclone probability due to lack of mesoscale rotational coherence. Consequently, the final classification logit remains negative. In contrast to bias-corrected regression frameworks that rely on temporal smoothing of historical intensity values, the proposed model resolves ambiguity directly at the image level using spatial structure and motion cues. This approach reduces false positives while preserving sensitivity to transitional atmospheric states. The probabilistic output reflects a near-boundary decision state, demonstrating the model's capacity to detect partial structural similarity while maintaining classification reliability.

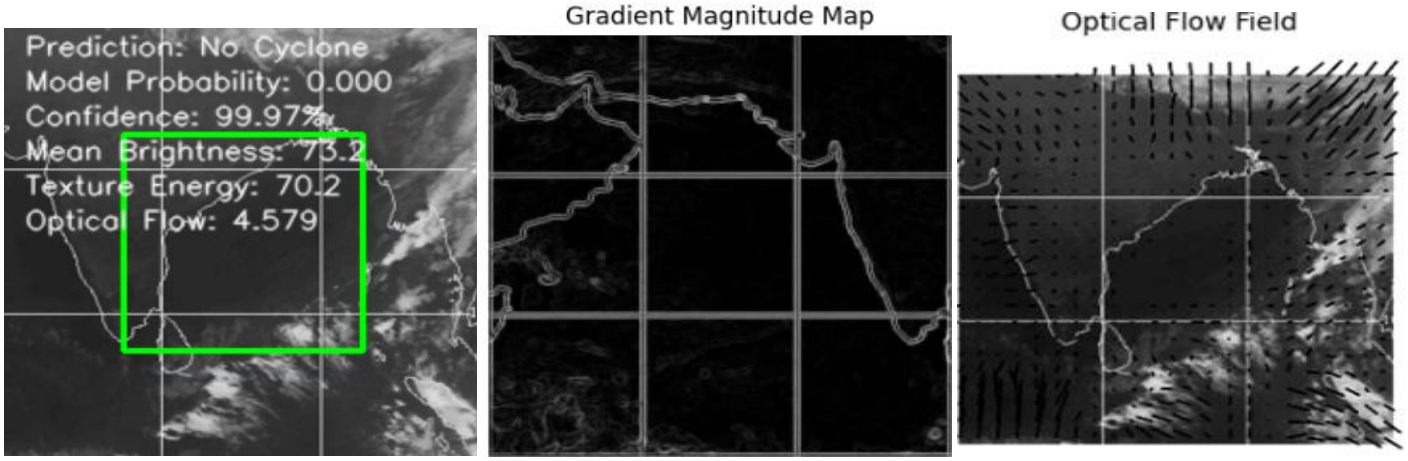


Figure 24: *Non- Cyclonic Detection Using Integrated Gradient and Optical Flow Analysis*

(a) Original satellite image with ROI indicating non-cyclone prediction (Probability = 0.000, Confidence = 99.97%) and extracted features: Mean Brightness = 73.2, Texture Energy = 70.2, Optical Flow Magnitude = 4.579.
(b) Gradient magnitude map showing weak, non-structured edge distributions without spiral organization.
(c) Optical flow field displaying irregular, non-vortical motion vectors, confirming absence of cyclonic circulation.

D. Spatiotemporal Evolution of Cyclone Sequences

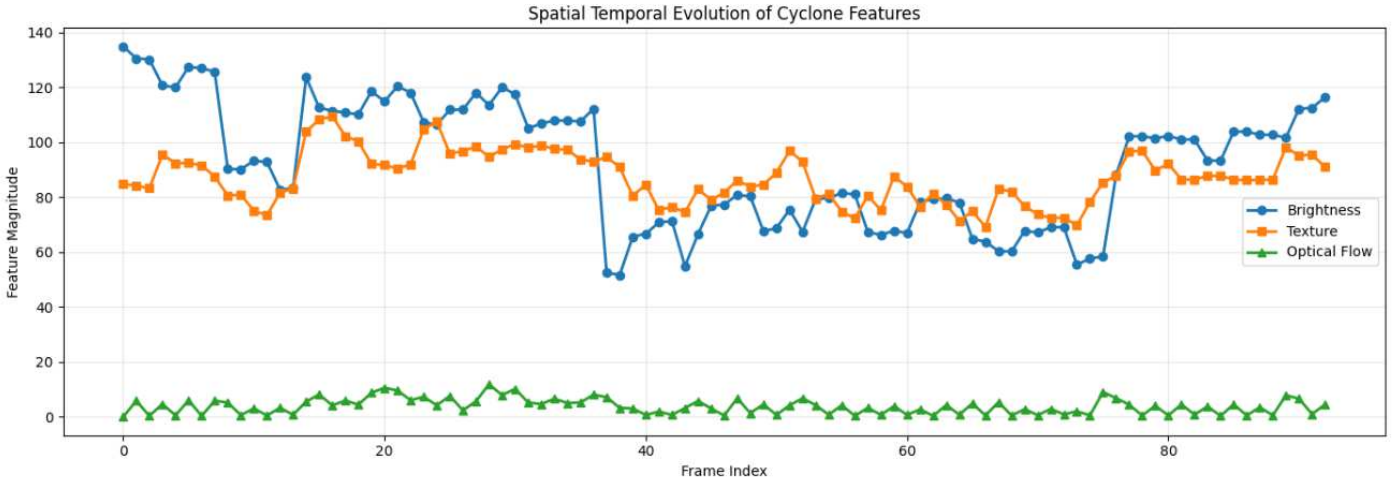


Figure 25: *Cyclone sequence exhibiting persistent convection, strong structural coherence, and enhanced rotational dynamics.*

This experiment analyzes spatiotemporal descriptors derived from sequential satellite frames representing a cyclonic event over the North Indian Ocean.

Radiometric Behavior

Mean brightness remains persistently elevated, indicating sustained cold cloud-top temperatures associated with deep convection. Mid-sequence fluctuations correspond to transient eyewall reorganization and convective bursts.

Structural Organization

Texture energy remains consistently high, confirming stable mesoscale organization and strong spatial gradients characteristic of spiral rainbands and the central dense overcast.

Motion Dynamics

Optical flow magnitude exhibits intermittent peaks corresponding to intensification of rotational circulation and cloud advection. These peaks reflect evolving vortex dynamics and structural consolidation.

Past-Sequence Temporal Parameter Behavior

Analysis of the past temporal sequence (t−8 → t−1) reveals clear distinctions between non-cyclonic cloud fields and the onset of organized cyclonic structure. During the initial non-cyclone frames, mean brightness values remain comparatively lower whereas indicating warmer cloud-top temperatures and the absence of deep convective towers typically associated with organized tropical systems. Concurrently, texture energy values appear relatively elevated and spatially irregular, reflecting fragmented cloud distributions and mesoscale convection rather than a coherent vortex structure. Optical flow magnitude during these early frames remains weak and spatially inconsistent, suggesting limited rotational coherence and predominantly advective cloud motion driven by ambient atmospheric flow. As the sequence progresses toward cyclogenesis, a marked increase in mean brightness is observed, corresponding to colder cloud-top temperatures and the development of deep convective cores. Texture energy stabilizes, indicating enhanced structural organization and banding features. Simultaneously, optical flow magnitude increases and exhibits improved directional coherence, consistent with intensifying rotational dynamics and organized circulation. This transition pattern confirms that elevated texture variability and weak flow coherence characterize non-cyclonic conditions, while cyclone evolution is associated with increasing brightness, structural consolidation, and coherent motion fields. These physically interpretable metrics provide an objective basis for distinguishing disorganized convection from organized cyclonic development.

- Progressive increase in texture energy preceding cyclone intensification,
- Stabilization of brightness values indicating convective persistence,
- Gradual amplification of motion magnitude signaling circulation strengthening.

These temporal precursors provide measurable indicators of structural organization prior to peak cyclone maturity. Intensity-based temporal modeling approaches, the proposed method directly measures cloud-field evolution, enabling physical interpretation of cyclone development without reliance on historical intensity labels.

E. Spatiotemporal Characteristics of Non-Cyclonic Sequences and Regime Separation

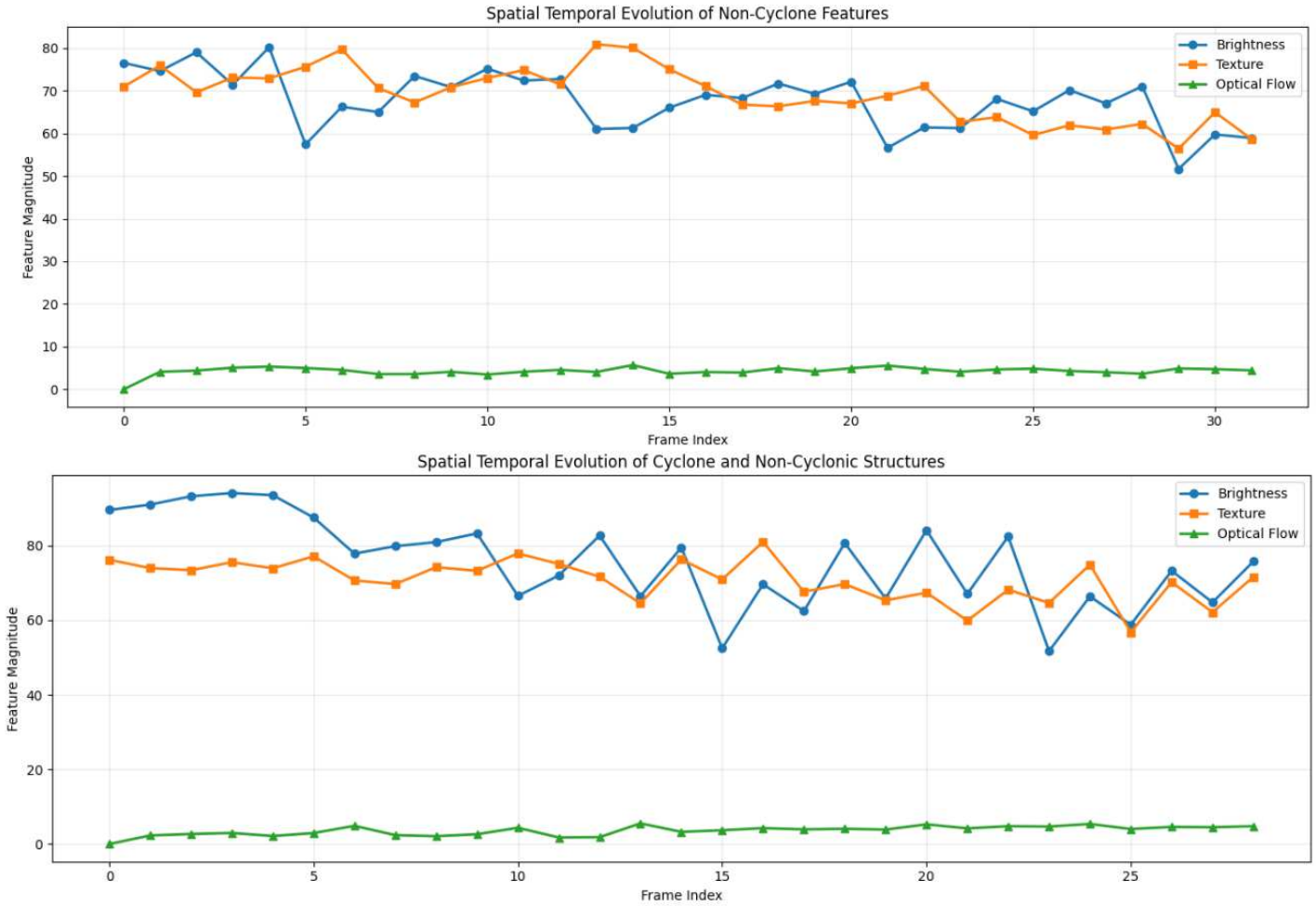


***Figure 26**: a) non-cyclonic sequence showing weak structural organization and limited motion dynamics. b) Spatial Temporal separation between cyclonic and non-cyclonic regimes.*

Temporal analysis of non-cyclonic cloud sequences demonstrates distinct dynamical behavior compared to cyclone-associated frames.

Radiometric Variability

Brightness values exhibit short-term fluctuations without sustained elevation, reflecting the absence of persistent deep convection.

Structural Coherence

Texture energy remains low and stable, indicating weak spatial gradients and lack of mesoscale organization typical of stratiform or fragmented convective clouds.

Motion Patterns

Optical flow magnitude remains minimal, confirming absence of coherent circulation or sustained cloud displacement.

Past-Sequence Temporal Trends

Preceding-frame analysis shows:

- Inconsistent brightness evolution,
- Absence of progressive texture amplification,
- Weak and non-persistent motion patterns.

These behaviors contrast sharply with cyclone precursors, reinforcing regime separation.

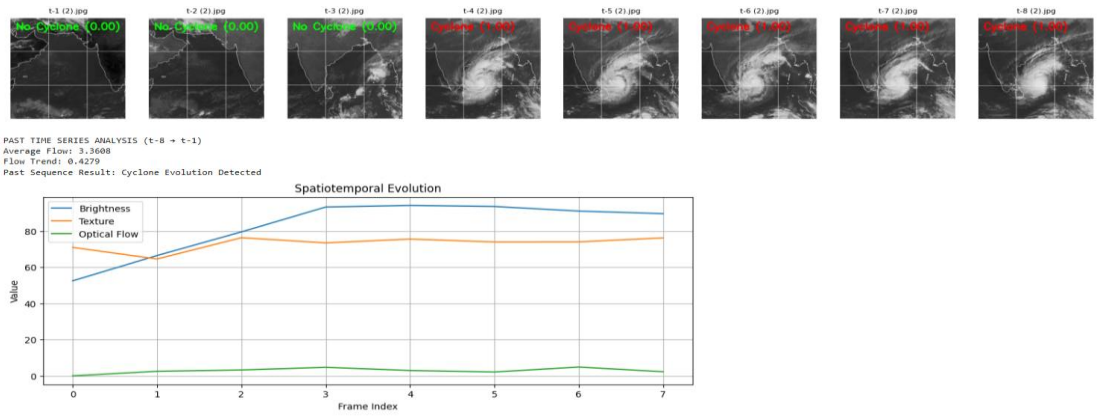


Figure 27: *Eight-Frame Past-Sequence Temporal Analysis Showing Average Flow (3.3608), Positive Flow Trend (0.4279), and Cyclone Evolution Detection*

The eight-frame sequence (t–1 to t–8) illustrates the transition from non-cyclonic to cyclonic conditions. In the initial frames (t–1 to t–3), the model predicts *No Cyclone* with probability = 0.00, indicating absence of organized vortex structure. From t–4 onward, the prediction shifts to *Cyclone* with probability = 1.00, confirming mature system detection.

Average Optical Flow (t–8 to t–3): 3.3608
This moderate flow magnitude reflects sustained cloud motion preceding full cyclone classification.

Flow Trend: 0.4279
The positive trend value indicates a consistent increase in motion intensity across preceding frames, suggesting progressive organization of atmospheric circulation.

Cyclone Evolution Result: Cyclone Evolution Detected
The increasing brightness, stabilized texture energy, and increasing optical flow collectively indicate structured vortex formation rather than random cloud drift.

The spatiotemporal plot further demonstrates:

- Steady rise in brightness (52 → 90), signifying intensifying convective cloud mass.
- Moderate variation in texture energy (65–78), reflecting structural consolidation.
- Gradual increase in optical flow prior to classification shift.

Hence, the eight-frame temporal progression confirms that cyclone detection is preceded by measurable increases in organized motion and cloud intensity and validating the effectiveness of temporal trend analysis for early-stage cyclone identification.

Regime Discrimination

Joint temporal evaluation across cyclonic and non-cyclonic frames demonstrates clear separation:

Descriptor	Cyclonic Regime	Non-Cyclonic Regime
Brightness	sustained, convective, persistent	variable, transient, unstable
Texture Energy	organized, curved, coherent gradients	irregular, fragmented gradients
Optical Flow	coherent circulation & rotational consistency	advective, inconsistent motion
Temporal Trend	persistent & structured evolution	transient & random variation
Spatial Organization	spiral symmetry & mesoscale coherence	disorganized cloud fields

Table 10: *Qualitative Comparison of Extracted Features Across Atmospheric Regimes*

Because these descriptors operate independently of the transformer latent space, they serve as physically grounded corroborative evidence supporting model predictions. This explicit quantification of radiative intensity, structural organization, and motion dynamics enhances both robustness and interpretability, particularly in visually ambiguous or transitional atmospheric scenarios.

VIII. FUTURE GLIMPSE

1. Physics-Constrained Vision Transformers

A major future advancement in developing physics-aware spatiotemporal Vision Transformer architectures efficient of modeling cyclone morpho dynamics under atmospheric constraints. Instead of treating frames independently, a unified transformer encoder with temporal self-attention and cross-frame token propagation can be designed to learn continuous cloud evolution fields. Configuring differentiable physical priors — including vorticity conservation, hydrostatic balance approximations, and thermodynamic stability constraints — into the attention weighting or loss regularization may guide the transformer toward meteorologically consistent representations. Such a system could develop flow-aware positional encodings derived from optical flow vectors or estimated circulation fields, enabling tokens to carry motion-consistent spatial context across time. Implementing this framework requires solving challenges related to memory scaling for long sequences, stabilizing physics-guided optimization, and ensuring numerical consistency between learned representations and atmospheric dynamics.

2. Hierarchical Multi-Scale Attention Modeling

Future transformer designs may employ hierarchical multi-scale tokenization strategies to capture cyclone structures spanning synoptic-scale cloud shields to mesoscale convective cores. Adaptive token merging, deformable attention windows, and dynamic resolution routing can enable selective refinement of regions exhibiting rotational symmetry or convective intensification. By efficiently integrating scale-aware positional embeddings and anisotropic attention kernels, the transformer can prioritize spiral rainbands, eyewall formation zones, and central dense overcast regions without excessive computational overhead. Advanced implementations may combine Swin-style shifted windows with global context tokens and learned scale-transition operators to preserve long-range dependencies while maintaining local structural fidelity. Such architectures demand complex memory scheduling, multi-resolution feature alignment, and stability control during training due to token densities and dynamic receptive field adaptation.

3. Self-Supervised Spatiotemporal Transformers

A challenging and promising research direction involves large-scale self-supervised pretraining of vision transformers using masked spatiotemporal token reconstruction and temporal consistency objectives tailored for geostationary satellite imagery. Masking spatial patches and temporal segments simultaneously, the model can learn invariant representations of cloud organization, rotational symmetry emergence, and mesoscale convective aggregation without explicit cyclone labels. Contrastive predictive coding across temporal windows and motion-consistent latent alignment losses may further enforce evolution-aware representations. Incorporating spectral channel reconstruction and cross-modal token prediction could allow the transformer to gather latent atmospheric dynamics from incomplete observations. Such pretraining paradigms require massive unlabeled datasets, careful masking policy design to preserve physical coherence, and mitigation of representation collapse, but they offer the potential to produce cyclone-aware embeddings transferable across sensors, and atmospheric regimes.

IX. CONCLUSION

This study introduced a transformer-driven cyclone detection framework that integrates Data-Efficient Image Transformers (DeiT) with physically interpretable spatiotemporal diagnostics derived from sequential satellite imagery. Usually, traditional cyclone detection methodologies that rely on temporally densified datasets, handcrafted interpolation, or tightly coupled spatiotemporal deep networks, the proposed framework operates directly on naturally ordered satellite observations, preserving authentic cloud evolution dynamics while avoiding interpolation-induced artifacts. The DeiT architecture captures long-range spatial dependencies and global cloud organization, enabling robust discrimination of cyclonic systems under diverse structural configurations. Complementary post-inference extraction of physically interpretable descriptors including mean radiometric brightness, Sobel-based texture energy, and dense Farnebäck optical flow magnitude provides an independent characterization of cloud intensity, structural complexity, and motion dynamics. The resulting time-series diagnostics supply quantitative evidence of cyclone presence and evolution, enhancing the interpretability and meteorological credibility of transformer-based predictions. By separating spatial learning from temporal motion analysis, the framework improves generalization, reduces computational overhead, and enhances analytical transparency properties essential for operational meteorological deployment. Experimental results depict competitive detection performance while performing computational efficiency relative to optical-flow-intensive and temporally reconstructed learning pipelines. Therefore, the proposed framework establishes that transformer-based satellite image analysis, augmented with lightweight and physically grounded spatiotemporal diagnostics, constitutes a reliable and scalable approach for automated cyclone monitoring. This work provides a strong foundation for future extensions toward multi-modal satellite fusion, long-horizon temporal modeling, cyclone intensity estimation, and physics-guided learning, supporting the development of explainable and operational AI systems for tropical cyclone early warning, risk assessment, and climate resilience planning.

X. REFERENCE

1) CAFormer: a connectivity-aware vision transformer for road extraction from remote sensing images. The Visual Computer, 2025: 1-17

2) Benchmarking computer vision architectures for cloud detection from lidar ceilometer backscatter data. The Visual Computer, 2025: 1-18

3) V. F. Dvorak, “Tropical cyclone intensity analysis and forecasting from satellite imagery,” Monthly Weather Rev., vol. 103, no. 5, pp. 420–430, May 1975.

4) V. F. Dvorak, “A technique for the analysis and forecasting of tropical cyclone intensities from satellite pictures,” NOAA Tech. Memorandum NESS, vol. 45, p. 19, May 1973.

5) V. Dvorak and S. Wright, “Tropical cyclone intensity analysis using enhanced infrared satellite data,” in Proc. 11th Tech. Conf. Hurricanes Trop. Meteorol., May 1977, pp. 268–273.

6) Zhang, Mengting, and Xiuxia Tian. "Transformer architecture based on mutual attention for image-anomaly detection." Virtual Reality & Intelligent Hardware 5.1 (2023): 57-67.

7) Lin, Chuchao, Changjun Zou, and Hangbin Xu. "SCNet: A Dual-Branch Network for Strong Noisy Image Denoising Based on Swin Transformer and ConvNeXt." Computer Animation and Virtual Worlds 36.3 (2025): e70030.

8) Pan, Fengli, et al. "Simulation of Ocean Waves With Spectrum-Based Wave Particle." Computer Animation and Virtual Worlds 36.2 (2025): e70014.

9) SAT-Net: Structure-Aware Transformer-Based Attention Fusion Network for Low-Quality Retinal Fundus Images Enhancement, IEEE Transactions on Multimedia, 2025

10) EAPT: efficient attention pyramid transformer for image processing, IEEE Transactions on Multimedia 25 (2021): 50-61

11) FFFN: Frame-by-frame feedback fusion network for video super-resolution, IEEE Transactions on Multimedia 25 (2022): 6821-6835

12) Gers, F. A., N. N. Schraudolph, and J. Schmidhuber (2003). Learning precise timing with lstm recurrent networks. J. Mach. Learn. Res., 3(null), 115–143. ISSN 1532-4435. URL https://doi.org/10.1162/153244303768966139. 10, 25, 40

13) Hall, T. M. and S. Jewson (2007). Statistical modeling of north atlantic tropical cyclone tracks. Tellus, 59A, 486–498. 34, 50

14) Z. Ma, Y. Yan, J. Lin, and D. Ma, "A multiscale and multilayer feature extraction network with dual attention for tropical cyclone intensity estimation," IEEE Trans. Geosci. Remote Sens., vol. 62, 2024, Art. no. 4101215, doi: 10.1109/TGRS.2024.3349416.

15) M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," Science, vol. 349, no. 6245, pp. 255–260, Jul. 2015, doi: 10.1126/science.aaa8415.

16) Hersbach, H., B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, A. Simmons, C. Soci, S. Abdalla, X. Abellan, G. Balsamo, P. Bechtold, G. Biavati, J. Bidlot, M. Bonavita, G. De Chiara, P. Dahlgren, D. Dee, M. Diamantakis, R. Dragani, J. Flemming, R. Forbes, M. Fuentes, A. Geer, L. Haimberger, S. Healy, R. J. Hogan, E. Hólm, M. Janisková, S. Keeley, P. Laloyaux, P. Lopez, C. Lupu, G. Radnoti, P. de Rosnay, I. Rozum, F. Vamborg, S. Villaume, and J.-N. Thépaut (2020). The era5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146(730), 1999– 2049. URL https://rmets.onlinelibrary.wiley.com/doi/abs/10. 1002/qj.3803. 6, 21, 35, 51, 63

17) Z. Chen et al., "A novel general semisupervised deep learning framework for classification and regression with remote sensing images," in Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS), Sep. 2020, pp. 1323–1326, doi: 10.1109/IGARSS39084.2020.9323932.

18) Ioffe, S. and C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift. In International conference on machine learning. PMLR, 2015. 12, 41

19) M. Maskey et al., "Deepti: Deep-learning-based tropical cyclone intensity estimation system," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 13, pp. 4271–4281, 2020, doi: 10.1109/ JSTARS.2020.3011907.

20) R. Pradhan, R. S. Aygun, M. Maskey, R. Ramachandran, and D. J. Cecil, "Tropical cyclone intensity estimation using a deep convolutional neural network," IEEE Trans. Image Process., vol. 27, no. 2, pp. 692–702, Feb. 2018, doi: 10.1109/TIP.2017.2766358.

21) B.-F. Chen, B. Chen, H.-T. Lin, and R. L. Elsberry, "Estimating tropical cyclone intensity by satellite imagery utilizing convolutional neural networks," Weather Forecasting, vol. 34, no. 2, pp. 447–465, Apr. 2019, doi: 10.1175/waf-d-18-0136.1.

22) J. Lee, J. Im, D.-H. Cha, H. Park, and S. Sim, "Tropical cyclone intensity estimation using multi-dimensional convolutional neural networks from geostationary satellite data," Remote Sens., vol. 12, no. 1, p. 108, Dec. 2019, doi: 10.3390/rs12010108.

23) C. Wang, G. Zheng, X. Li, Q. Xu, B. Liu, and J. Zhang, "Tropical cyclone intensity estimation from geostationary satellite imagery using deep convolutional neural networks," IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 1–16, 2022, doi: 10.1109/TGRS.2021.3066299.

24) Knapp, K. R., M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neumann (2010). The international best track archive for climate stewardship (ibtracs): Unifying tropical cyclone data. Bulletin of the American Meteorological Society, 91(3), 363 – 376. URL https://journals.ametsoc.org/view/journals/bams/ 91/3/2009bams2755_1.xml. 6, 21, 35, 61

25) R. Zhang, Y. Liu, L. Yue, Q. Liu, and R. Hang, "Estimating tropical cyclone intensity using an STIA model from Himawari-8 satellite images in the western north Pacific basin," IEEE Trans. Geosci. Remote Sens., vol. 62, 2024, Art. no. 4101813, doi: 10.1109/TGRS.2024.3352704.

26) Mohapatra, M., D. Nayak, M. Sharma, R. Sharma, and B. Bandyopadhyay (2015). Evaluation of official tropical cyclone landfall forecast issued by india meteorological department. Journal of Earth System Science, 124. 35

27) D. Ma, L. Wang, S. Fang, and J. Lin, "Tropical cyclone intensity prediction by inter- and intra-pattern fusion based on multi-source data," Environ. Res. Lett., vol. 18, no. 1, Jan. 2023, Art. no. 014020, doi: 10.1088/1748-9326/aca9e2.

28) Krizhevsky, A., I. Sutskever, and G. Hinton (2012). Imagenet classification with deep convolutional neural networks. Neural Information Processing Systems, 25. viii, 12, 25, 29, 30, 40, 45, 68

29) S. M. Griffin, A. Wimmers, and C. S. Velden, "Predicting shortterm intensity change in tropical cyclones using a convolutional neural network," Weather Forecasting, vol. 39, no. 1, pp. 177–202, Jan. 2024, doi: 10.1175/waf-d-23-0085.1.

30) T. Li, M. Lai, S. Nie, H. Liu, Z. Liang, and W. Lv, "Tropical cyclone trajectory based on satellite remote sensing prediction and time attention mechanism ConvLSTM model," Big Data Res., vol. 36, May 2024, Art. no. 100439, doi: 10.1016/j.bdr.2024.100439.

31) B. Tong, X. Wang, J. Y. Fu, P. W. Chan, and Y. C. He, "Short-term prediction of the intensity and track of tropical cyclone via ConvLSTM model," J. Wind Eng. Ind. Aerodynamics, vol. 226, Jul. 2022, Art. no. 105026, doi: 10.1016/j.jweia.2022.105026.

32) C. Zhang and C. Zhang, "Rapid intensification forecast for tropical cyclone by combining multi-ConvLSTM model with infrared and microwave satellite images," in Proc. Conf. Infr., Millim., Terahertz Waves Appl. (IMT), Apr. 2023, p. 45, doi: 10.1117/12.2661748.

33) R. Chen, X. Wang, W. Zhang, X. Zhu, A. Li, and C. Yang, "A hybrid CNN-LSTM model for typhoon formation forecasting," GeoInformatica, vol. 23, no. 3, pp. 375–396, May 2019, doi: 10.1007/s10707-019- 00355-0.

34) M. Choo, Y. Kim, J. Lee, J. Im, and I.-J. Moon, "Bridging satellite missions: Deep transfer learning for enhanced tropical cyclone intensity estimation," GISci. Remote Sens., vol. 61, no. 1, pp. 1–18, Mar. 2024, doi: 10.1080/ 15481603.2024.2325720.

35) J. Kim, S. O. Hwang, S. S. Kim, I. Oh, and D. J. Ham, "Algorithms for determining Korea meteorological administration (KMA)'s official typhoon best tracks in the national typhoon center," Atmosphere, vol. 32, no. 4, pp. 381–394, Dec. 2022, doi: 10.14191/Atmos.2022.32.4.381. 4111912 IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. 63, 2025

36) Korea Meteorological Admin. (2018). Annual Report 2017. Accessed: Jun. 19, 2025. [Online]. Available: https://www.kma.go.kr/ download 01/Annual Report 2017.pdf

37) K. Ito, "Errors in tropical cyclone intensity forecast by RSMC Tokyo and statistical correction using environmental parameters," SOLA, vol. 12, pp. 247–252, Dec. 2016, doi: 10.2151/sola.2016-049.

38) Milletari, F., N. Navab, and S. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation. In 2016 Fourth International Conference on 3D Vision (3DV). 2016. 12, 25, 40

39) T. L. Olander and C. S. Velden, "The advanced Dvorak technique (ADT) for estimating tropical cyclone intensity: Update and new capabilities," Weather Forecasting, vol. 34, no. 4, pp. 905–922, Aug. 2019, doi: 10.1175/waf-d-19-0007.1.

40) T. L. Tsui and R. J. Miller, "Evaluation of western north Pacific tropical cyclone objective forecast aids," Weather Forecasting, vol. 3, no. 1, pp. 76–85, Mar. 1988.

41) D. H. Kim and M. H. Ahn, "Introduction of the in-orbit test and its performance for the first meteorological imager of the communication, ocean, and Meteorological Satellite," Atmos. Meas. Techn., vol. 7, no. 8, pp. 2471–2485, Aug. 2014, doi: 10.5194/amt-7-2471-2014.

42) S.-R. Chung, M.-H. Ahn, K.-S. Han, K.-T. Lee, and D.-B. Shin, "Meteorological products of geo-KOMPSAT 2A (GK2A) satellite," Asia Pacific J. Atmos. Sci., vol. 56, no. 2, p. 185, Apr. 2020, doi: 10.1007/ s13143-020-00199-x.

43) D. Kim, M. Gu, T.-H. Oh, E.-K. Kim, and H.-J. Yang, "Introduction of the advanced meteorological imager of Geo-Kompsat-2a: In-orbit tests and performance validation," Remote Sens., vol. 13, no. 7, p. 1303, Mar. 2021, doi: 10.3390/rs13071303.

44) S. Kaufman, S. Rosset, C. Perlich, and O. Stitelman, "Leakage in data mining: Formulation, detection, and avoidance," ACM Trans. Knowl. Discovery Data, vol. 6, no. 4, pp. 1–21, Dec. 2012, doi: 10.1145/ 2382577.2382579.

45) Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830. 25, 42, 54, 66

46) C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," in Proc. Int. Conf. Artif. Neural Netw., vol. 11141, 2018, pp. 270–279, doi: 10.1007/978-3-030-01424-7.

47) T. Song, K. Yang, X. Li, S. Peng, and F. Meng, "Probabilistic estimation of tropical cyclone intensity based on multi-source satellite remote sensing images," Remote Sens., vol. 16, no. 4, p. 606, Feb. 2024, doi: 10.3390/rs16040606.

48) R. Longadge and S. Dongre, "Class imbalance problem in data mining review," 2013, arXiv:1305.1707.

49) Z. Chen and X. Yu, "A novel tensor network for tropical cyclone intensity estimation," IEEE Trans. Geosci. Remote Sens., vol. 59, no. 4, pp. 3226–3243, Apr. 2021, doi: 10.1109/TGRS.2020.3017709.

50) Z. Zhao, Z. Zhang, Q. Wang, L. Cui, and P. Tang, "TFA-Net: A temporal feature aggregation framework for tropical cyclone intensity estimation from satellite images," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 18, pp. 12008–12023, 2025, doi: 10.1109/ JSTARS.2025.3563472.

51) M. B. Ba and S. E. Nicholson, "Analysis of convective activity and its relationship to the rainfall over the rift valley lakes of east Africa during 1983–90 using the Meteosat infrared channel," J. Appl. Meteorol., vol. 37, no. 10, pp. 1250–1264, Oct. 1998, doi: 10.1175/ 1520-0450(1998)0372.0.co;2.

52) H. Norouzi et al., "Inferring land surface parameters from the diurnal variability of microwave and infrared temperatures," Phys. Chem. Earth, A/B/C, vols. 83–84, pp. 28–35, Feb. 2015, doi: 10.1016/ j.pce.2015.01.007.

53) Y.-H. Baek, I.-J. Moon, J. Im, and J. Lee, "A novel tropical cyclone size estimation model based on a convolutional neural network using geostationary satellite imagery," Remote Sens., vol. 14, no. 2, p. 426, Jan. 2022, doi: 10.3390/rs14020426.

54) P. Purwono, A. Ma'arif, W. Rahmaniar, H. I. K. Fathurrahman, A. Z. K. Frisky, and Q. M. U. Haq, "Understanding of convolutional neural network (CNN): A review," Int. J. Robot. Control Syst., vol. 2, no. 4, pp. 739–748, Jan. 2023, doi: 10.31763/ijrcs.v2i4.888.

55) Pradhan, R., R. S. Aygun, M. Maskey, R. Ramachandran, and D. J. Cecil (2018). Tropical cyclone intensity estimation using a deep convolutional neural network. IEEE Transactions on Image Processing, 27(2), 692–702. 34.

56) Ramachandran, P., B. Zoph, and Q. V. Le (2017). Searching for activation functions. arXiv preprint arXiv:1710.05941. URL https://arxiv.org/abs/1710. 05941. 55.

57) Schumacher, A. B., M. DeMaria, and J. A. Knaff (2009). Objective estimation of the 24-h probability of tropical cyclone formation. Weather and Forecasting, 24(2), 456 – 471. URL https://journals.ametsoc.org/view/journals/wefo/ 24/2/2008waf2007109_1.xml. 18.

58) Schuster, M. and K. K. Paliwal (1997). Bidirectional recurrent neural networks. IEEE Transactions on Signal Processing, 45(11), 2673–2681. 11.

59) Shakya, S., S. Kumar, and M. Goswami (2020). Deep learning algorithm for satellite imaging-based cyclone detection. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13, 827–839. 20, 31, 34.

60) Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov (2014). Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15, 1929–1958. 13.

61) Sugi, M., A. Noda, and N. Sato (2002). Influence of the global warming on tropical cyclone climatology: An experiment with the jma global model. Journal of the Meteorological Society of Japan. Ser. II, 80(2), 249–272. 20.

62) Leroux, M.-D., K. Wood, R. L. Elsberry, E. O. Cayanan, E. Hendricks, M. Kucas, P. Otto, R. Rogers, B. Sampson, and Z. Yu (2018). Recent advances in research and forecasting of tropical cyclone track, intensity, and structure at landfall. Tropical Cyclone Research and Review, 7(2), 85 – 105. ISSN 2225-6032. URL http://www. sciencedirect.com/science/article/pii/S2225603219300189. 14, 15, 4

63) Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.