Information Sciences 741 (2026) 123258

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Multi-branch local global fusion network for cyclone classification using domain aware pooling

Soumi Sarkar [a,*], Subhadip Boral [b], Sarbani Palit [c], Ashish Ghosh [d]

[a] *Machine Intelligence Unit, Indian Statistical Institute, 203 Barrackpore Trunk Road, Dunlop, Bonhooghly Government Colony, Baranagar, West Bengal 700108, India*

[b] *Department of Computer Science and Technology, Siksha 'O' Anusandhan, ITER College Rd, Jagmohan Nagar, Bhubaneswar, Odisha 751030, India*

[c] *Computer Vision and Pattern Recognition Unit, Indian Statistical Institute, 203 Barrackpore Trunk Road, Dunlop, Bonhooghly Government Colony, Baranagar, West Bengal 700108, India*

[d] *International Institute of Information Technology, Gothapatana, Malipada, Bhubaneswar, Odisha 751003, India*

HIGHLIGHTS

- SSIM-based grouping, reduces model complexity and enabling domain-aware pooling.
- Domain-aware pooling extracts better physical features for each group in the model.
- Triple attention mechanisms learn both local and global cyclone structures.
- Multi-branch design preserves dynamics and enhances 8-class cyclone classification.

ARTICLE INFO

Keywords:
Tropical cyclone
Machine learning
Meteorological data
Classification
Attention mechanisms
Pooling

ABSTRACT

Precise cyclone classification is essential for disaster preparedness, early warning, and mitigation. However, existing methods face three key challenges: coarse-grained categorization, difficulty in capturing both local and global meteorological patterns, and limited incorporation of physical structures in deep learning (DL) models. To address these issues, this work proposes the Structural Similarity Index Measure (SSIM)-based Grouped Attention Network (SGA-Net), a multi-branch deep neural architecture that integrates localized and global contextual information to classify cyclones into eight categories. SGA-Net comprises SSIM-based feature grouping for domain-aware pooling, spatial attention for localized feature refinement, global feature extraction blocks with multi-head self-attention to learn long-range spatial dependencies, and channel-attention modules to model intra-feature relationships. The model is evaluated across nine metrics, emphasizing the Matthews Correlation Coefficient (MCC) for its suitability under class imbalance. On the eight-class task, SGA-Net achieves an MCC of 0.60, a 13% improvement over state-of-the-art (SOTA) methods on ERA5 and IBTrACS data from the Bay of Bengal (BoB). Additional analyses demonstrate robustness to spatial noise, cross-regional generalization to the Arabian Sea, and high computational efficiency. Grad-CAM visualizations further confirm the interpretability and structure-aware learning capability of SGA-Net. The code is available at https://github.com/soumisarkar02/SGA-Net.git.

* Corresponding author.
Email addresses: soumi.s_r@isical.ac.in (S. Sarkar), subhadipboral@soa.ac.in (S. Boral), sarbanip@isical.ac.in (S. Palit), ashish@iiit-bh.ac.in, ash@isical.ac.in (A. Ghosh).

https://doi.org/10.1016/j.ins.2026.123258
Received 21 October 2025; Received in revised form 15 February 2026; Accepted 16 February 2026
Available online 19 February 2026

1. Introduction

Tropical cyclones (TCs) [1] are intense low-pressure systems that develop over warm tropical oceans and are characterized by closed low-level circulation, strong winds, spiral thunderstorm bands, and heavy precipitation [2]. These systems cause significant socio-economic and ecological damage, particularly in densely populated coastal regions. Different countries categorize tropical cyclones according to their characteristics in order to take precautionary measures to mitigate their impacts. Similarly, the Indian Meteorological Department (IMD), acting as the Regional Specialized Meteorological Center (RSMC), categorizes TCs in the North Indian Ocean (NIO) using an eight-grade scale based on three-minute sustained wind speed.

The NIO, comprising the Arabian Sea (AS) and Bay of Bengal (BoB), is one of the most cyclone-prone regions globally [3]. While cyclone frequency has remained relatively stable, recent decades have seen a notable rise in the intensity and destructiveness of these storms [3]. Coastal India, with over 340 million people living within 100 km of the shoreline [4], is particularly vulnerable. Economic losses from TCs are estimated at nearly $6 billion annually, with major storms between 2005 and 2020 causing thousands of deaths and damages exceeding $40 billion [5]. Most importantly, post-monsoon cyclones over the BoB show rapid intensification and have shifting landfall patterns toward northeastern India and Bangladesh [3]. This trend has amplified the need for precise and timely classification systems to support early warning, disaster response, and risk mitigation.

An effective cyclone classifier should accurately distinguish classes within highly imbalanced datasets, maintain computational efficiency, and generalize to evolving meteorological patterns. It must be capable of capturing both fine-grained local features and global spatial patterns. Local features capture fine-scale patterns such as cyclone eye structure that are critical for distinguishing closely related classes. Global features, in contrast, encode large-scale spatial dependencies that provide contextual cues for accurate classification. The classifier must adapt to the heterogeneous characteristics of meteorological features.

The TC classification methods can be grouped into two categories. The first traditional approach relies on expert knowledge, such as the Dvorak [6] and DAV-T methods [7], which use satellite images to estimate cyclone intensity based on cloud structure patterns. Other expert-based methods like the AMSU algorithm [8], which utilizes brightness temperature data. These approaches, however, depend heavily on expert judgment, making them prone to subjective errors.

The advent of deep learning (DL) techniques [9] as powerful data-driven alternatives capable of learning complex, nonlinear patterns from high-dimensional meteorological data has inspired the application of DL to the context of cyclone analysis. Here, it is treated as a regression or classification problem. Regression models estimate wind speeds from satellite or meteorological features [10], and classification models predict cyclone types. They automatically extract cyclone patterns from satellite images or meteorological features and then feed them to regressors or classifiers and achieve better performance compared to expert-based approaches. The authors of CNN_V0 [11], CNN_V1 [12] have performed binary classification, whereas CNN + DT [13], and residual CNNs [3] on INSAT-3D data have performed multi-class classification. The author in [14] has proposed a two-stage DL pipeline. In the first stage it uses Co-DETR [15] to identify potential TC regions , and then in the next stage it is followed by a U-Net-based model that performs multi-task learning to predict the cyclone center and classify its intensity with state-of-the-art (SOTA) performance. The authors in [16] have proposed a 2D-3D CNN-based intensity estimation framework, which is also based on a density map generation. TFG-Net, proposed in [17], is a DL graph convolutional network that estimates TC intensity from satellite images by combining fine-grained feature extraction and general wind speed transition rules. Transformer-based approaches [18,19] have also been explored for intensity estimation and classification.

However, both traditional domain-based expert methods and DL methods offer distinct advantages—expert methods incorporate physical insights, whereas DL models excel at feature extraction—a combination of these two explicitly integrates domain knowledge into DL architecture, as demonstrated in [17] for improved classification. The authors in [20] proposed a Dvorak-inspired machine learning technique for intensity estimation using the satellite cloud pattern data.

Despite recent advances, SOTA architectures exhibit notable limitations primarily in localized feature extraction and do not consider the global spatial patterns hence miss out on global contextual cues. Some models [11,12] address the inherent class imbalance problem by merging higher-category cyclones, but this approach inevitably sacrifices fine-grained class information. Employment of traditional data augmentation techniques, including contrast enhancement, brightness adjustment, translation, and rotations of 90°, 180° and 270° [3] are potentially beneficial; such augmentations must be meticulously applied to avoid introducing distortions that could impair model generalization. The authors in [18] propose another transformer-based deep learning architecture for tropical cyclone intensity classification and estimation. The main focus of this paper is on using a deep learning-based model using a ConvNet architecture as a means of learning the spatial hierarchy of tropical cyclones at a given time step; then the extracted feature is given as input to ViT for intensity estimate and classification. RTCC-Net [19] is another transformer-based deep learning framework designed specifically for the classification and prediction of tropical cyclone genesis within 24 hours or less. It follows the architecture of using a strategy of multi-source information fusion; the features are learned using convolutional subnetworks operating in parallel that focus on learning local spatial patterns; then these learned feature maps are then integrated using the ViT with the Attention mechanism. Transformer-based architectures [18,19] often require large volumes of balanced training data for generalizing well. These architectures have been observed to overfit to higher categories and cannot detect the lower categories. The domain-aware intensity estimation methods [17,20] are based on the formation of clouds, which is an aftereffect of cyclones, but the genesis of cyclones is due to meteorological factors, so the incorporation of meteorological patterns is crucial.

The use of the Co-DETR model in [14] to initially extract cyclone regions requires training with ground-truth masks that accurately delineate cyclone-affected areas. The training masks are generated by centering a fixed-radius [21] circular region around the cyclone eye coordinates provided by the International Best Track Archive for Climate Stewardship (IBTrACS) dataset [22]. The fixed radius approach ignores the natural variability in cyclone sizes and imposes a uniform spatial extent on all cyclone instances. The authors in [16] proposed an intensity estimation framework with a dynamic radius density map. These maps, unlike the previous method, are prepared based on the size of the cyclones occurring in the western North Pacific. However, the IBTrACS data corresponding to India does not have the radius information; thus, using this approach is not feasible.

Furthermore, most of the SOTA architectures explore the cyclone intensity prediction framework. However, cyclone intensity becomes available only after cyclone formation. Consequently, intensity prediction models depend exclusively on post-formation data and are incapable of detecting earlier stages such as normal, depression, and deep-depression conditions. In addition, the statistical and structural characteristics of the underlying meteorological features differ substantially; however, existing SOTA architectures typically overlook these differences. They often employ uniform pooling strategies that disregard variable-specific dynamics and apply identical feature extractors across all features, despite their distinct statistical and structural properties. These limitations of SOTA architectures motivate the proposed work to address the majority of them.

Meteorological features exhibit different spatial behaviors and physical characteristics. Treating all variables identically [3,12] assumes they share similar spatial structure and benefit from the same feature extraction pathway—an assumption that is physically incorrect and often degrades model performance. For example, precipitation is sparse and discontinuous, whereas SST is smooth and slowly varying. Applying a single uniform operation feature extraction strategy across such heterogeneous fields leads to oversmoothing, loss of fine-scale gradients, and attenuation of extremes.

To address this, variables must be grouped according to their spatial similarity so that each group can receive an appropriate, physically consistent processing strategy. The structural similarity index measure (SSIM) provides a principled way to form these groups because it evaluates similarity in three key dimensions: luminance (captures background mean similarity), contrast (captures spatial variability similarity), and structure (captures pattern and gradient alignment). Unlike point-wise metrics such as mean absolute error or root mean square error, SSIM captures how well the underlying spatial organization of two fields matches, which is crucial for this purpose. By grouping variables based on SSIM, meteorological features with comparable spatial patterns can be processed together, reducing conflicting gradients during learning and improving representational efficiency. This grouping enables the next phase of the proposed work, i.e., application of domain-appropriate pooling and feature extraction tailored to each group of variables, ultimately preserving physically meaningful structures and enhancing predictive skill.

Thus, the proposed SGA-Net, a multi-branch local-global CNN architecture for cyclone classification across all eight categories, operates in two phases. It first groups meteorological features based on SSIM analysis. Each group is then processed in a dedicated CNN branch with a domain-specific pooling strategy—max pooling for wind speed, min pooling for precipitation, average pooling for moisture, and no pooling for the remaining slowly varying features—to preserve physically relevant spatial patterns. Spatial attention modules emphasize the important local regions of extracted features. Channel attention enables the architecture to assess the relative importance of individual features within the multi-feature groups. A Global Feature Extraction (GFE) block with multi-head attention captures global feature dependencies from intermediate convolution layers. Finally, local and global representations from all branches are fused and fed into fully connected layers for cyclone classification. The following core innovations of the proposed framework are dedicated to different challenges:

1. SSIM-based feature grouping: Climate variables are grouped based on SSIM. This grouping strategy systematically organizes heterogeneous climate variables according to their structural similarity rather than treating all variables uniformly and independently, unlike naive concatenation approaches in SOTA. This approach both reduces architectural complexity and also enables the application of a domain-aware pooling strategy.
2. Domain-aware pooling: Domain-specific pooling operations for each group in the multi-branch architecture are better at extracting better discriminating groups' physical characteristics. Building upon the above grouping strategy, the domain-aware pooling mechanism is proposed, which, instead of employing a uniform pooling operation across all branches, uses an adaptive pooling strategy based on the physical characteristics of each variable group. This design choice allows the network to preserve critical spatial cues and extract more discriminative physical features relevant to cyclone structure and balance the model complexity.
3. Triple Attention Mechanism: The combination of Channel Attention (CA), Spatial Attention (SA), and GFE block enables SGA-Net to learn the local and global structure of cyclones. While CA and SA modules enable the network to focus on the most informative variables and spatial regions, respectively, the GFE block complements these local attentional cues by modeling long-range spatial dependencies and global cyclone morphology. The integration of these three components allows the model to effectively capture both localized convective patterns and large-scale storm structures that are crucial for accurate cyclone intensity classification.
4. Multi-branch architecture: The multi-branch representation helps in preserving both the intra-group dynamics and the cyclone structures, while enabling more effective downstream fusion that enhances class discrimination and introduces 8-class classification. Such a balanced fusion strategy is particularly important for fine-grained multi-class cyclone classification, where subtle intensity transitions must be accurately distinguished. Moreover, tight integration of these components within a unified and computationally efficient framework clearly differentiates the proposed method from existing approaches, which typically rely on monolithic architectures or simple feature concatenation.

The SGA-Net is trained using ERA5 [23] reanalysis data and IBTrACS best-track [22] labels. Data augmentation is applied to generate and reduce training samples depending on the class cardinality to address class imbalance. Due to the presence of class imbalance in the test data, the Matthews Correlation Coefficient (MCC) is used to have a balanced and interpretable evaluation. Accuracy, RMSE, and MAE are also used for evaluation.

Experimental performance shows that SGA-Net outperforms SOTA methods, achieving $\sim$ 13% higher MCC and $\sim$ 6% higher Accuracy with minimal variation in RMSE and MAE. Extensive experiments further validate the robustness of the proposed model to spatial noise, lower class dispersion, and its strong cross-regional generalization ability, as evidenced by consistent performance when evaluated on cyclones from the Arabian Sea. Moreover, detailed seasonal analysis demonstrates that SGA-Net maintains stable and consistent performance across different cyclone seasons, highlighting its robustness to seasonal variability in cyclone characteristics.

In addition to performance gains, the proposed framework is designed with practical deployment in mind and has a very low inference time of around 0.0.06997 secs. It achieves an effective balance between model complexity, inference efficiency, and predictive Accuracy, with an overall computational complexity of $O(H^2W^2C)$ and low inference latency (0.07 secs), making it suitable for real-world operational settings.

The paper is organized as follows. Section 2 reviews cyclone classification literature. Section 3 describes the datasets along with preprocessing and augmentation mechanisms for class imbalance. Section 4 details the proposed SGA-Net framework and its complexity, and the experimental setup and performances are presented in Section 5. Section 6 summarizes the findings and outlines potential directions for future research.

2. Literature survey

CNN_V0 [11] classifies nine meteorological feature image frames in cyclone and non-cyclone classes using CNN architecture. CNN_V1 [12] enhances the classification performance using hyperparameter tuning. A hybrid CNN + Decision Tree architecture (CNN_DT) [13] extracts spatial characteristics using CNN, and the decision tree classifies them into five cyclone grades. CNN_DT is dependent on post-formation intensity data, which is addressed by the residual CNN architecture [3]. It experiments on multispectral INSAT-3D imagery and merges the classes with limited samples and performs six-class classification instead of eight classes (as shown in Table 1). The authors in [18] propose a transformer-based framework that combines convolutional feature extraction with a Vision Transformer for tropical cyclone intensity estimation and classification. Similarly, RTCC-Net [19] employs parallel convolutional subnetworks to learn local spatial patterns, which are subsequently integrated using a ViT-based attention mechanism for cyclone classification. However, these transformer-based architectures typically require large, well-balanced datasets to generalize effectively and often exhibit overfitting toward higher-intensity categories, leading to poor detection of lower categories.

Pooling operations in CNN-based architectures are designed to reduce dimensionality and computational load by retaining salient patterns and discarding irrelevant details. However, traditional pooling methods like max pooling [24] (which retains the strongest activation) may suppress subtle but informative structures. Average pooling [25] smooths features, potentially diluting strong signals. Min pooling [26] is rarely used due to its high information loss. Mixed pooling [27] alternates between max and average pooling, introducing randomness to reduce overfitting. Spatial Pyramid Pooling (SPP) [28] combines information at multiple scales; its variants, like SP Average and SP Max pool, use average or max across each level. LP Pooling [29] uses the L^P norm to interpolate between average and max pooling. Weighted pooling [30] assigns trainable weights to each spatial location. LEAP pooling [31] learns linear filters for pooling windows, optimizing directly for downstream task performance. LBP Pooling [32] encodes local texture information through binary patterns in 3×3 neighborhoods, preserving fine spatial structure.

Besides pooling, the application of a domain-aware mechanism in weather-related domains, like the authors in [33] proposes a soft sensor framework called DKST-DLSTM that integrates domain knowledge with a spatio-temporal deep LSTM model to better predict key variables in industrial processes by aligning time delays and decoupling subprocesses. The approach effectively captures complex, nonlinear interactions and improves model Accuracy for mineral processing applications. Moreover, attention mechanisms in DL have also been shown to be very promising in computer vision [34]. Among them, the most popular ones are channel attention, spatial attention, self-attention, and multi-head attention, each guiding the model on what and where to focus. Channel attention refines feature strength across channels, spatial attention localizes important regions in the input feature map, self-attention captures global interdependencies, and multi-head attention models multiple feature interactions to enhance representational diversity.

For channel attention, methods like Squeeze-and-Excitation (SE) blocks introduced in SENet [35] use global average pooling (GAP) followed by two fully connected layers and a sigmoid activation to adaptively recalibrate feature responses across channels. A major advantage of SE blocks over other channel attention mechanisms lies in their simplicity, modularity, and computational efficiency, while still being able to capture global context. Variants such as GSoP-Net [36] extend SE blocks by incorporating second-order statistics via covariance pooling; ECANet [37] replaces fully connected layers with lightweight 1D convolutions for more efficient

Table 1
Class distribution of cyclone instances in BoB according to the Indian grading scale.

Wind Speed (ws) (in Knots)	Cyclone Category (grade)	Class Distribution (%)
ws < 17	Normal / Low Pressure (0)	2496 (46.4%)
17 ≤ ws < 28	Depression (1)	1130 (21%)
28 ≤ ws < 34	Deep Depression (2)	694 (12.9%)
34 ≤ ws < 48	Cyclonic Storm (3)	603 (11.2%)
48 ≤ ws < 64	Severe Cyclonic Storm (4)	234 (4.3%)
64 ≤ ws < 91	Very Severe Cyclonic Storm (5)	181 (3.4%)
91 ≤ ws < 120	Extremely Severe Cyclonic Storm (6)	35 (0.7%)
ws ≥ 120	Super Cyclonic Storm (7)	8 (0.1%)

channel modeling; and FcaNet [38] uses Discrete Cosine Transform (DCT) [39] to improve global representation learning through frequency-domain analysis.

In the spatial attention domain, the Convolutional Block Attention Module (CBAM) [40] applies attention in two sequential stages: first using SE-based channel attention, followed by spatial attention computed from global max and average pooling, concatenated and passed through a convolutional layer and a sigmoid function to produce a spatial attention map. This dual attention framework enables CBAM to focus simultaneously on both the most informative channels and their corresponding spatial locations. Other spatial attention approaches include STN [41], which predicts affine transformation parameters to align significant spatial regions.

For capturing global dependencies, self-attention, which is implemented in Vision Transformers (ViT) [42] for modeling long-range feature interactions via dot-product attention across all spatial positions, is best suited. Multi-Head Attention (MHA), a key component in transformers, computes attention across multiple learned subspaces in parallel to represent diverse semantic relations among features. These attention strategies significantly enhance model expressiveness and are particularly useful for learning global spatial contextual relations in structurally complex data such as meteorological imagery.

3. Dataset

This study experiments on meteorological features relevant to cyclone analysis, which are obtained from the ERA5 reanalysis dataset [23] provided by the European Center for Medium-Range Weather Forecasts (ECMWF). The features include zonal ($u10$) and meridional ($v10$) wind components, total precipitation (tp), sea surface temperature (sst), surface pressure (sp), 2-meter dew point temperature ($d2m$), 2-meter air temperature ($t2m$), and total column water vapor ($tcwv$). These features are extracted at a spatial resolution of $0.25° \times 0.25°$, covering the BoB region. The cyclones in this region that occurred between the years 2002 and 2021 have been used in this manuscript in order to have at least one of the rare higher-category cyclones. Each meteorological feature was annotated with the corresponding cyclone grade from the best track dataset by IBTrACS [22] which includes storm track coordinates, wind speed, central pressure, and class labels at 3-hour intervals. As cyclones are rare events, there is a huge class imbalance between normal and cyclone instances. So, for each cyclone, meteorological features from 3 days before formation to dissipation at 3-hour intervals are considered; this reduced the imbalance and the data volume. The final class distribution of the data is given in Table 1.

The cyclones occurring between 2002 and 2017 are used for training, 2018–19 for the validation, and 2020–21 for the testing. The other problem is different cyclones with different intensities can occur simultaneously at different locations, which might confuse the model regarding the grade. So, for each cyclone instance, approximately $10.5° \times 10.5°$ an area (as the cyclone's diameter is observed to not extend beyond 1000km normally [21], which approximates to $10°$; an extra $0.5°$ is added as padding) centered on the cyclone eye (obtained from IBTrACS) was extracted, which creates a spatial patch of 42×42.

3.1. Data augmentation

In the context of TC classification, the distribution of samples across different cyclone intensity grades is highly skewed (as observed in Table 1). This imbalance causes models to become biased toward the more frequently occurring lower-grade or non-cyclonic events, reducing their ability to accurately predict rare but critical high-intensity cyclones.

A targeted data augmentation strategy was employed to synthetically increase the number of samples in underrepresented classes to mitigate this issue. The augmentation process (adopted from [3]) involved applying transformations like small random rotations (limited to $\pm$ 5 degrees), adaptive histogram equalization (CLAHE) [43] to enhance contrast, and controlled adjustments to brightness and contrast to the input data. Such small, physically realistic augmentation operations introduce variability in the dataset while preserving the underlying structure and spatial integrity of the cyclonic patterns and reducing bias toward high-frequency categories. Extreme radiometric changes, large translations, or mirror flips (which invert swirl direction) are avoided to preserve physical realism. Additionally, the rotation limit was kept deliberately small because each pixel in the input data corresponds to a specific geographic location. Applying large rotations could distort the spatial structure of cyclones and misalign important geophysical patterns, ultimately reducing the reliability of the data.

The augmentation is only applied to the training data. Each channel was independently processed through the augmentation pipeline, and the modified channels were recombined to form the augmented sample. This approach ensures that feature-wise relationships are maintained without introducing unrealistic distortions. To balance the dataset, a minimum sample threshold (500 samples) was set for each class. Classes with fewer samples than this threshold were identified, and their existing samples were repeatedly augmented until they met or exceeded the target count. This iterative process continued until all classes had a sufficient number of samples for effective training.

Finally, to maintain uniformity across classes and prevent the dominance of overrepresented categories, the number of samples in the more frequent classes was randomly reduced to match the desired sample count. This ensured a balanced training dataset across all cyclone grades.

4. Proposed methodology

4.1. Problem formulation

The objective is to determine, given a set of meteorological features at a particular time step, whether a TC event is present and, if so, then classify it into the appropriate class. Let the input feature vector be defined as, $\mathcal{X}_f \in \mathbb{R}^{N \times H \times W \times C}$, where $f \in$

$\{u10, v10, tp, tcwv, sst, sp, t2m, d2m\}$, H and W denote spatial dimensions and C is the number of input channels.

$$\mathcal{X}^f = [x_f(1), x_f(2), \dots, x_f(N)] \tag{1}$$

and

$$\mathcal{X} = [\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_M] \tag{2}$$

where M denotes the number of features and N denotes the number of samples. The classification model $h(\cdot)$ produces a prediction given by,

$$\hat{\mathcal{Y}} = h(\mathcal{X}; \theta), \quad \hat{y} \in \{0, 1, 2, \dots, NC\}, \tag{3}$$

where $\hat{\mathcal{Y}}$ is the predicted class label, $\mathcal{Y}$ is the ground truth label, NC represents the total number of classes, and θ is the weight parameter learned during training.

The proposed work addresses the task of cyclone classification using image-based meteorological data discussed in Section 3. Such data inherently contain rich spatial structures, making CNNs a natural choice, as they are well-suited for extracting discriminative spatial features critical for accurate categorization. Learning these distinct characteristics of each feature typically necessitates training a dedicated feature extractor; the resulting representations are then fused and forwarded to an 8-class classifier. Although effective, this design leads to considerable computational overhead. Conversely, combining all features into one group is imprudent, given their distinct physical and structural characteristics. In the proposed framework, features are partitioned into groups, each serving as input to different branches of the proposed multi-branch CNN for improved representation learning. The features are grouped based on their structural patterns (discussed in Section 4.2), and this approach balances representational fidelity with computational efficiency.

CNN architectures used as feature extractors typically consist of a sequence of convolution and pooling operations. When the data exhibit heterogeneous characteristics, convolutions can extract salient features, but a uniform pooling function does not serve its purpose. Therefore, the proposed architecture employs a domain-appropriate pooling function, which enables models to more effectively capture such structures. Section 4.3 discusses these proposed domain-aware pooling techniques applicable to cyclones.

4.2. Feature grouping using SSIM

First, the similarity between each feature pair, $\binom{M}{2}$, (M is the number of features), was evaluated for every instance using the SSIM [44]. SSIM across N samples and $\binom{M}{2}$ feature pairs produces a similarity matrix of shape $(N, \binom{M}{2})$, where N denotes the number of training samples. The SSIM value theoretically is bounded by $[-1, 1]$ but in practice (with standard image data) it lies within $[0, 1]$. The same was observed in this case, i.e., the SSIM values were always non-negative, and hence, for this study, it is assumed that the SSIM values $\in [0, 1]$. Now, an unbiased estimator is required to obtain a representative SSIM value for each feature pair. To identify this, first, the distribution of SSIM values for each individual feature pair is identified.

The beta distribution [45] is a family of probability distributions defined on the interval $[0, 1]$. With appropriate choices of shape parameters $\alpha(> 0)$ and $\beta(> 0)$, it can approximate a wide range of continuous probability density functions (PDF), while those not expressible using its form can be modeled with beta mixture distributions.

Since the SSIM values of each pair of features do not follow any particular probability distribution but they are in the range $[0, 1]$, so the beta distribution is fitted. The shape parameters α and β are computed in accordance with the SSIM values between each pair of features. Hence, the beta distribution's mean, $\hat{\mu}$ is used as an unbiased estimator of SSIM value for each feature pair. Higher mean SSIM values indicate greater structural similarity between features, whereas lower values reflect weak or negligible similarity.

The fitted beta distribution of pairwise SSIM values is shown in Fig. 1. Further, the log-likelihood values of eleven distributions and the Q–Q plots illustrating the fit of the Beta distribution for each pair were also checked for completeness and are detailed in Section A of the Appendix.

Thus, these $\binom{M}{2}$ estimators help in the grouping of features with similar spatial integrity. It has been observed that only sp & $t2m$, $t2m$ & $d2m$ and $d2m$ & sp have SSIM > 0.8, others have low levels of similarity in structure, i.e., SSIM< 0.4. This clear separation in the SSIM value motivates the choice of 0.8 as a threshold to identify structurally similar features. Thus, the ones with SSIM > 0.8 are considered in the same group, giving rise to 6 groups as follows:

- Group $C_1 = \{\mathrm{u10}\}$
- Group $C_2 = \{\mathrm{v10}\}$
- Group $C_3 = \{\mathrm{tp}\}$
- Group $C_4 = \{\mathrm{tcwv}\}$
- Group $C_5 = \{\mathrm{sst}\}$
- Group $C_6 = \{\mathrm{sp}, \mathrm{t2m}, \mathrm{d2m}\}$

As stated, $u10$ and $v10$ are the zonal and meridional wind vector components, and wind speed ws is the magnitude of these vectors. Thus, the following are the final five groups, which are fed to the individual branches of SGA-Net.

- Group $G_1 = \{\mathrm{ws}\}$, where $ws = \sqrt{(u10)^2 + (v10)^2}$
- Group $G_2 = \{\mathrm{tp}\}$

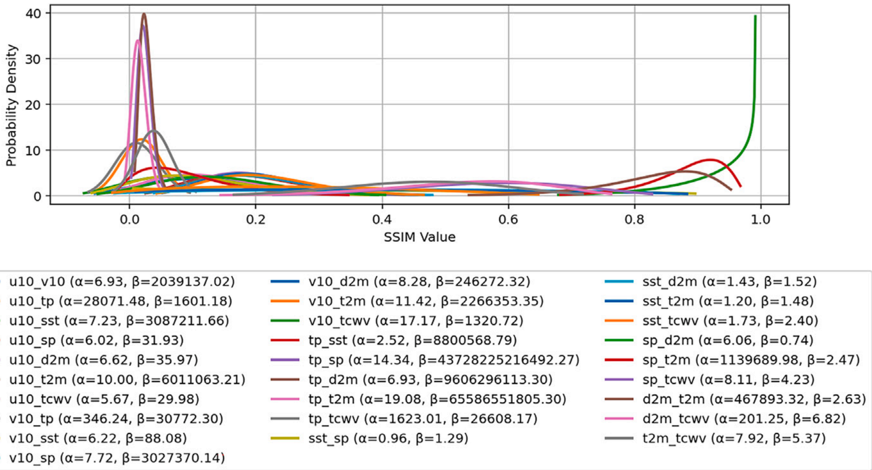

Fig. 1. Probability density function of fitted beta distribution of SSIM values for every feature pair and corresponding shape parameter values.

- Group $G_3 = \{\text{tcwv}\}$
- Group $G_4 = \{\text{sst}\}$
- Group $G_5 = \{\text{sp}, \text{t2m}, \text{d2m}\}$

4.3. Domain-specific pooling strategy

Let us consider a feature map $F \in \mathbb{R}^{H' \times W' \times C'}$, where H', W', and C' denote the height, width, and number of channels, respectively. Pooling operations apply a function $\mathcal{P}$ over neighborhoods of F to produce a downsampled representation F' with translation invariance and retain the most informative summary statistics. It reduces the dimension to approximately $\frac{H'}{k} \times \frac{W'}{k}$ with pooling window k and the output at location (i, j) is

$$F'_{i,j} = \mathcal{P}\left(\{F_{m,n} \mid m \in [i, i+k),\ n \in [j, j+k)\}\right). \tag{4}$$

The spatial characteristics of different meteorological features vary significantly depending on the physical processes involved. Application of the same pooling operation in this scenario will not yield optimal performance. Therefore, this design decision is guided by the physical understanding of the data, which also mitigates the inherent lossy compression.

4.3.1. Group G_1

The most intense winds often occur near the eyewall or spiral rainbands and are critical indicators of cyclone strength. Max pooling is employed to retain these peak wind values within each local region, which allows the model to focus on extreme wind intensities, that correlate directly with cyclone category and potential damage.

4.3.2. Group G_2

The eyewall region experiences extremely heavy rainfall, but the eye often exhibits near-zero precipitation. However, the eye typically covers only a small fraction of the entire cyclone domain. If repeated max pooling or average pooling is applied, the extremely high precipitation values from the eyewall may dominate the representation, causing the relatively dry eye region to be underrepresented or even lost. By applying min pooling, the network explicitly retains information about the minimum precipitation within each pooling region, thus preserving the existence of subsidence or dry zones such as the eye. Nevertheless, regions outside the eye exhibit significant spatial variability and sharp changes in precipitation distribution that also need to be preserved. Moreover, if min pooling is used with a large pooling window, min pooling would consistently select the smallest value in each kernel, potentially suppressing these variations and eliminating critical edge information outside the eye. To prevent this loss, the pooling window size is kept small (2×2).

4.3.3. Group G_3

Cyclone development and intensification are heavily dependent on the ambient moisture content of the atmosphere. Average pooling is applied to extract moisture levels, allowing the model to capture the broader moist environment that sustains cyclone intensification.

4.3.4. Group G_4 and Group G_5

The characteristic spatial correlation length for these features is large compared to the kernel size of pooling operations. Applying pooling to such fields would unnecessarily average out subtle but physically meaningful variations in the spatial structure, leading to loss of important discriminating information. Moreover, these features have low spatial variability, which is studied using the Moran's Index. The Moran's Index ranges from -1 to $+1$, where a more negative value toward -1 indicates low spatial correlation

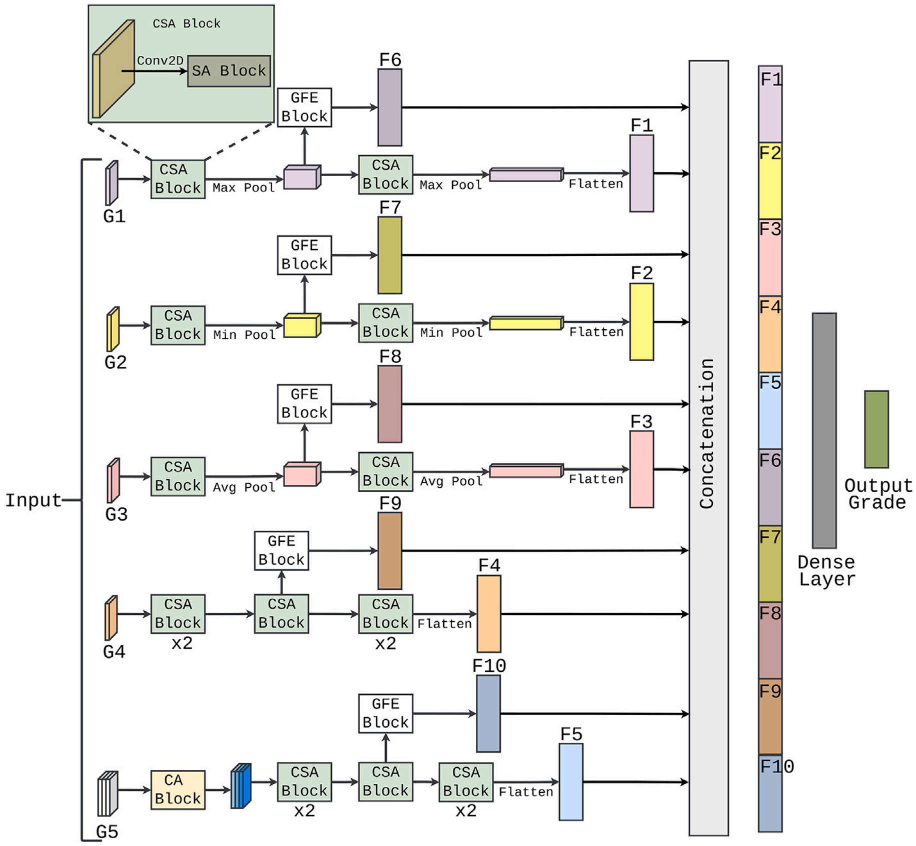

Fig. 2. Proposed SGA-Net Model: The CSA Block, which comprises of 2D convolution and the SA Block (Fig. 3), the CA Block (Fig. 4) and the novel GFE Block (Fig. 7) are its feature extractor components.

and a high positive value toward +1 indicates high spatial correlation. It was observed that for group G_4, the index value is, 0.944 and for group G_5, the corresponding value is 0.909. These high values closer to +1 indicate that these features in these regions have high spatial correlation, i.e., the features vary slowly over time. Moreover, if spatial variability is low, then pooling produces diminishing returns: the pooled value often approximates the original value itself due to low local variance, making pooling redundant.

This targeted pooling approach improves both feature efficiency and the physical interpretability of learned representations.

4.4. Proposed SGA-net architecture

The proposed model consists of a multi-branch CNN architecture which inputs these groups into five distinct branches. The overall architecture (Fig. 2) is capable of jointly learning both localized and global structural patterns in cyclone systems.

4.4.1. Convolution, pooling, and attention for G_1 to G_3

Groups 1–3 are processed through two successive convolutional layers, each followed by feature-specific pooling operations:

$$F_i^g = \sigma(\mathrm{Conv}_i(F_{i-1}^g)), \quad g \in \{G_1, G_2, G_3\} \;\; \& \;\; i \in \{1, 2\} \tag{5}$$

$$F_i^g = \mathrm{Pool}_g(F_i^g), \quad g \in \{G_1, G_2, G_3\} \;\; \& \;\; i \in \{1, 2\} \tag{6}$$

where, $\mathrm{Pool}_{G_1} = MaxPool$, $\mathrm{Pool}_{G_2} = MinPool$, $\mathrm{Pool}_{G_3} = AvgPool$, and for $i = 1$, $F_{i-1}^g = F_0^g = \mathcal{X}_g$. Here, $\sigma(\cdot)$ denotes the activation function. Scaled Exponential Linear Unit (SELU) is used as activation, which has self-normalizing properties, reducing internal covariate shift and stabilizing training across heterogeneous meteorological features.

To lay emphasis on the significant regions of the feature maps extracted by the convolution operations, SA (Fig. 3), as described in Eqs. (7) and (8), is applied before each group-specific pooling. The spatial attention module computes:

$$M_i^{SA} = \sigma\big(Conv([\,\mathrm{AvgPool}(F_i^g) \odot \mathrm{MaxPool}(F_i^g)\,])\big), \quad \begin{array}{l} g \in \{G_1, G_2, G_3\}, \\ i \in \{1, 2\}. \end{array} \tag{7}$$

$$F_i^g = F_i^g \otimes M_i^{SA}, \quad g \in \{G_1, G_2, G_3\} \;\; \& \;\; i \in \{1, 2\} \tag{8}$$

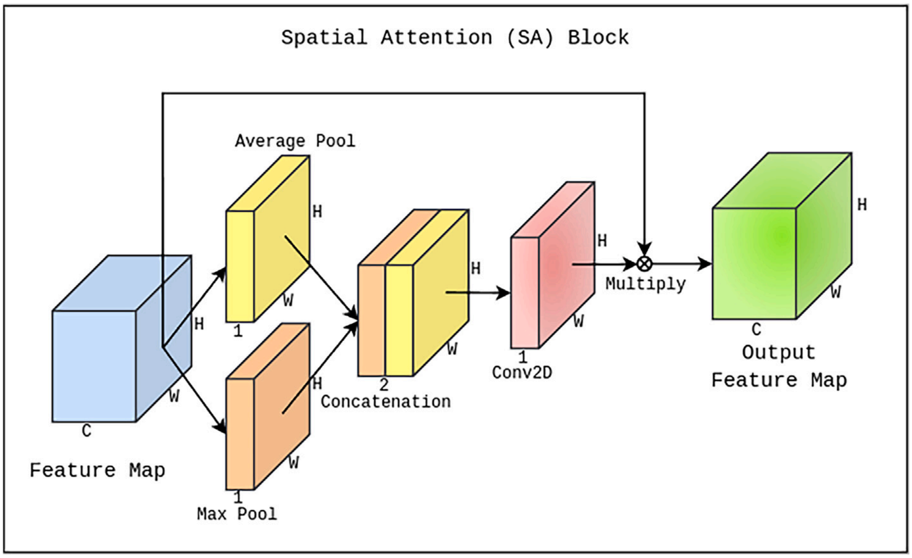

Fig. 3. Graphical representation of the spatial attention block architecture.

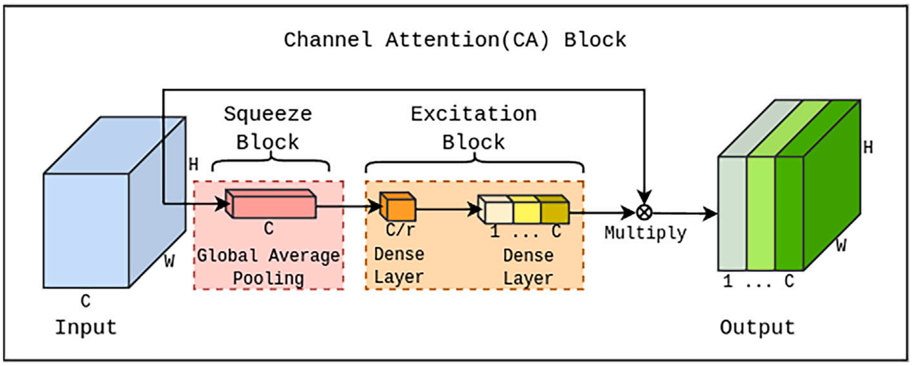

Fig. 4. Graphical representation of the channel attention block architecture.

where, M_i^{SA} is the weighted spatial attention matrix, i indicates the input convolution layer, $Conv$ operation has a kernel of size 7×7, $\odot$ indicates channel-wise concatenation, and $\otimes$ indicates element-wise multiplication.

4.4.2. Deep convolution and attention for G_4 & G_5

Groups 4 and 5 are processed through deeper convolution pipelines without pooling to preserve fine-scale gradients:

$$F_i^g = \sigma(\mathrm{Conv}_i(F_{i-1}^g)), \quad i \in \{1,2,3,4,5\} \ \ \& \ \ g \in \{G_4, G_5\} \tag{9}$$

Here, for $i = 1$, $F_{i-1}^{G_4} = F_0^{G_4} = \mathcal{X}_{G_4}$. As earlier, spatial attention (discussed above in Eqs. (7) and (8)) is applied after each convolution layer in Eq. (9).

Now only G_5 has more than one feature, and to understand the impact of each feature on classification efficiency, a squeeze-and-excitation CA block (Fig. 4) is used at the input. These adaptively reweight feature channels during training according to their importance. The CA block operates as follows:

$$z_c = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}\mathcal{X}_c(i,j), \quad c \in G_5 \quad \& \quad i \in \{1,2,3,4,5\} \tag{10}$$

$$e = \sigma(Dense(Dense(z))) \tag{11}$$

$$\mathcal{X}_c{}' = e_c \cdot \mathcal{X}_c, \quad c \in G_5 \tag{12}$$

Here, σ is the sigmoid activation function, H is the height of the input feature map, W is the width of the input feature map, and c represents the 3 features of G_5 the group. z is the array of all z_c. This step of computing z from the input is the squeeze step and e is the excitation step. Thus, in Eq. (9) when $i = 1$, $F_{i-1}^{G_5} = F_0^{G_5} = \mathcal{X}'_{G_5}$.

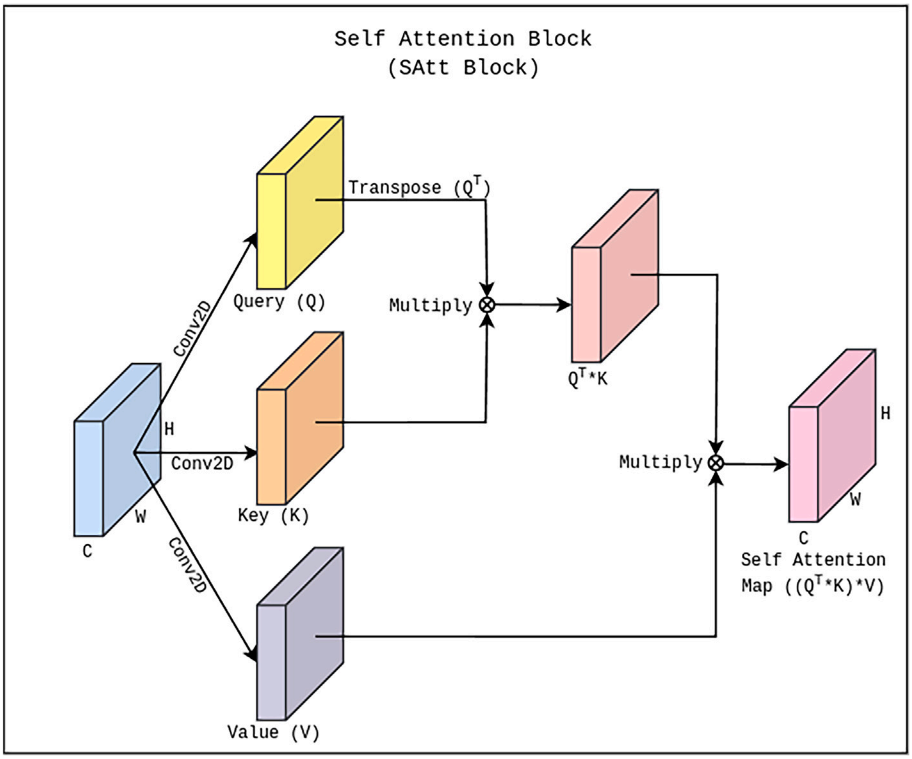

Fig. 5. Graphical representation of the self-attention network.

4.4.3. Global feature extraction (GFE) block

CNN theoretically extracts some portion of the cyclone images and computes feature maps. Spatial attention considers only a single feature map and tries to lay more emphasis on the important neighborhoods. For efficient classification, learning the relationship between each of the feature maps is important. For example, one feature map may capture the core of the cyclone, and another feature map may capture some portion of the outer spiral band of the cyclone, so the relationship between them is important to find the grade of the cyclone. Thus, the global relationship, i.e., the relation between different feature maps, is learned in the GFE block (Fig. 7) using an MHA mechanism.

The MHA block (Fig. 6) is composed of four heads, each of which is a single self-attention (SAtt) block (Fig. 5). The SAtt block tries to find the relation between two feature maps by computing attention weights based on dot-product similarity. These weights indicate how much each position in one feature map (called the query) should relate to positions in another (called the key/value maps). This mechanism helps produce context-aware representations that capture the required spatial relationships between the two inputs. Repeating this across multiple heads helps in obtaining a richer representation.

The GFE blocks for each group take the convolutional feature maps as input. Specifically, it considers feature maps after the first convolutional layer for groups 1–3 and after the third convolutional layer for groups 4–5. As most of the groups contain a single feature, there is only one query, key, and value. Thus, the purpose of capturing the spatial relationship between two features will fail. Now, as the resolution of the images are too high, the initial convolutional feature maps are not considered as inputs for groups 4–5 to reduce the processing time. On the other hand, it is also not taken from the last convolutional layer, as the abstraction becomes so high and it becomes difficult for the MHA block to find the relation between them. To balance performance, abstraction, and resolution, the input to the GFE block is approximately the middle convolution feature maps. For groups 1–3, a pooling operation follows the first convolutional layer, reducing the spatial resolution by half. For the same reasons as discussed for groups 4–5, feature maps from both the initial convolutional layer (due to excessively high resolution) and the final convolutional layer (due to overly abstract representations) are avoided. Consequently, the feature maps after the first convolution–pooling stage effectively correspond to the middle-most level of abstraction within these branches, making them a suitable and consistent choice for input to the GFE block.

The MHA in GFE is given by:

$$\text{MHA} = Conv([SAtt_0 \odot \ldots \odot SAtt_4]) \tag{13}$$

where each head is a self-attention network computed by:

$$\text{SAtt}_i = \text{softmax}\left(\frac{Q_i K_i^T}{\sqrt{d_k}}\right) V_i, \quad i \in \{1, 2, 3, 4, 5\} \tag{14}$$

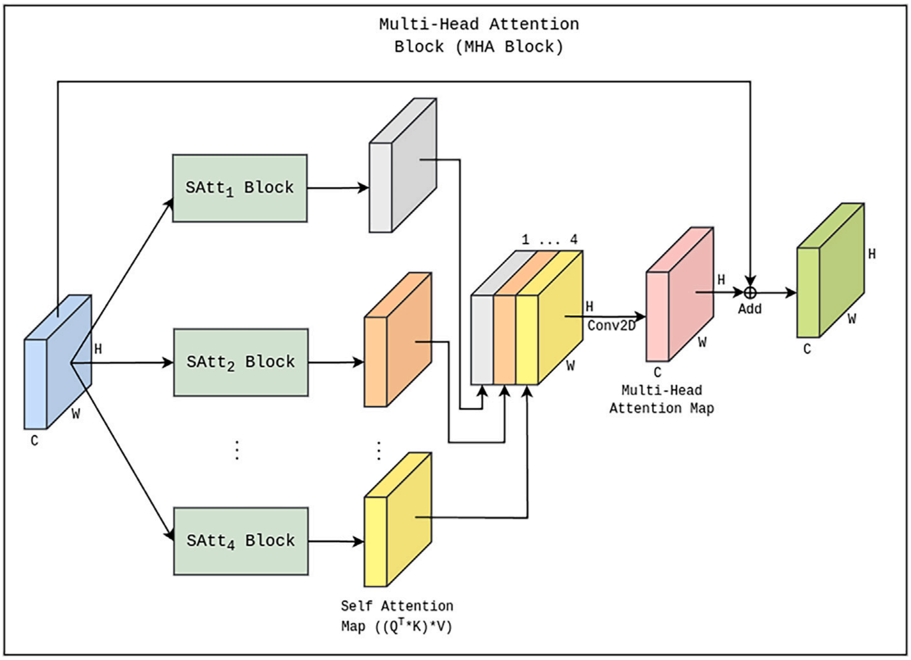

Fig. 6. Graphical representation of the multi-head attention network.

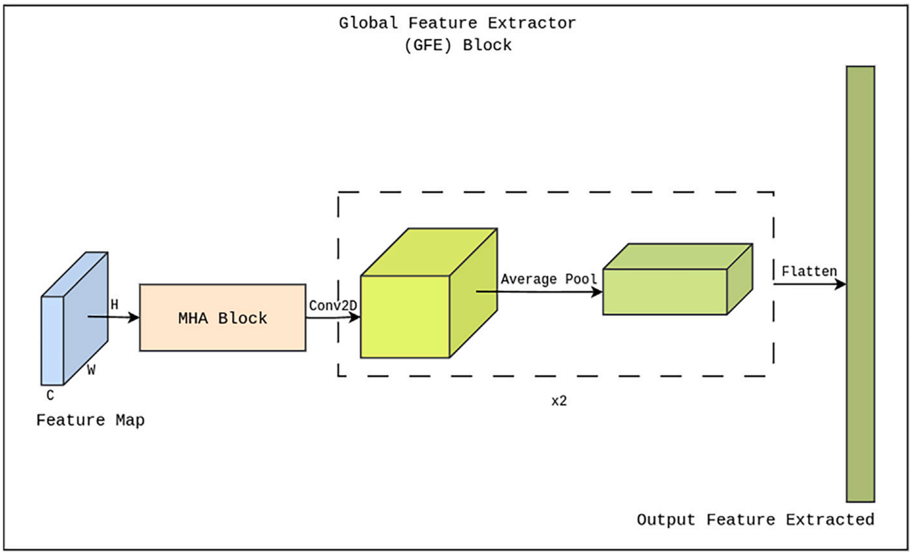

Fig. 7. Graphical representation of the proposed global feature extractor, which utilises the multi-head attention block (Fig. 6).

Here Q, K, V carry the usual definition of query, key, and value vector, respectively, and d is the dimension of the vectors, which is the same as the number of feature maps input to the attention head.

In the GFE block, the MHA output passes through two additional convolution layers and an average pooling layer for decreasing the resolution of the obtained representations.

$$
\begin{aligned}
&GFE_g = \mathrm{AvgPool}\Big(\sigma\Big(\mathrm{Conv}\Big(\mathrm{AvgPool}\Big(\sigma\Big(\mathrm{Conv}(\overline{F^g})\Big)\Big)\Big)\Big)\Big),\\
&\text{where } \overline{F^g} = MHA(F_1^g), \quad g \in \{G_1, G_2, G_3\},\\
&\qquad\quad \overline{F^g} = MHA(F_3^g), \quad g \in \{G_4, G_5\}.
\end{aligned} \tag{15}
$$

4.4.4. Feature fusion and classification

The feature maps from each of the groups are now combined and connected to a classifier to obtain cyclone labels among 8 classes. First, the feature maps from the final convolutional layers (i.e., $F_2^g \quad g \in \{G_1, G_2, G_3\}$ and $F_5^g \quad g \in \{G_4, G_5\}$) and the GFE blocks of all branches are flattened and concatenated into a unified feature vector.

$$F_{\text{local}} = [\text{Flatten}(F_2^{G_1}) \odot \text{Flatten}(F_2^{G_2}) \odot \text{Flatten}(F_2^{G_3}) \odot \text{Flatten}(F_5^{G_4}) \odot \text{Flatten}(F_5^{G_5})] \tag{16}$$

$$F_{\text{global}} = [\text{Flatten}(GFE_{G_1}) \odot \text{Flatten}(GFE_{G_2}) \odot \text{Flatten}(GFE_{G_3}) \odot \text{Flatten}(GFE_{G_4}) \odot \text{Flatten}(GFE_{G_5})] \tag{17}$$

$$F_{\text{final}} = [F_{\text{local}} \odot F_{\text{global}}] \tag{18}$$

Then this vector is passed through a fully connected layer and output layer to get the final grade.

$$y_{\text{pred}} = \text{Softmax}(Dense(F_{\text{final}})) \tag{19}$$

4.5. Training procedure

This section discusses the training details of the proposed SGA-Net. For each epoch, the training set of size $N(= 4000)$ is divided into mini-batches of size $B(= 448)$, resulting in $\frac{N}{B}$ iterations per epoch. For every mini-batch, a forward pass through the network produces the class logits, and a weighted cross-entropy loss is applied to address class imbalance. Gradients are computed via backpropagation, and the parameters are updated using the Adam optimizer with learning-rate scheduling. Training is performed for $E(= 80)$ epochs, and early stopping based on the validation MCC is used to avoid overfitting. The final model is selected as the checkpoint achieving the highest validation performance.

4.6. Computational complexity

The computational complexity of the proposed SGA-Net is analyzed at a high level in this Section, while detailed derivations for each architectural component are provided in Section B of the Appendix. SGA-Net consists of multiple convolution-dominated branches with domain-aware pooling, SA, CA, and a Global Feature Extractor (GFE) block incorporating Multi-Head Self-Attention (MHSA).

Most network components, including convolutional layers, pooling operations, SA, CA, and dense layers, exhibit linear complexity with respect to the spatial resolution, i.e., $\mathcal{O}(HW)$, where H and W are the current layer input feature height and width, respectively. In contrast, the MHSA module reshapes the feature map into a sequence of length HW and computes pairwise attention, resulting in a quadratic complexity of $\mathcal{O}(H^2W^2C)$, where C is the current layer input channels.

Since kernel sizes, channel dimensions, and dense layer widths are much smaller compared to image size and the reduction ratios of channel attention blocks are treated as constants, the overall computational complexity of SGA-Net is dominated by the quadratic powers of the input image dimensions in the MHSA operation in the GFE block. Consequently, the asymptotic complexity of the proposed model scales as

$$\mathcal{O}\left(H^2W^2C\right). \tag{20}$$

5. Experimental performance analysis & comparative study

Experimental performance analyses are conducted in a Red Hat Enterprise Linux Server 7.9 system with a 64-bit operating system, equipped with an Intel®Xeon®Gold 5317 CPU @ 3.00 GHz (48 cores), 500 GB of RAM, and 13 TB of storage. In this experiment, the categorical cross-entropy loss function is employed with the Adam optimizer (learning rate $\eta = 0.0001$) and SeLU activation. To prevent overfitting, an L2 regularization term is added to the loss function.

5.1. Metrics

The proposed model's performance is assessed using metrics Accuracy (Accuracy), mean absolute error (MAE), root mean square error (RMSE), Precision, Recall, G-mean, F1-score, Cohen's Kappa (Kappa), and MCC. MAE and RMSE measure the deviation of the predicted class label from the actual class label by computing their differences and then taking the sum of the absolute differences and the square root of the sum of the squared differences, respectively. Accuracy measures the proportion of correctly classified samples relative to the total predictions. Precision and Recall capture discrimination ability; Precision measures the proportion of correctly predicted samples within each predicted class, while Recall measures the proportion of correctly identified samples within each actual class. The F1-score, defined as the harmonic mean of Precision and Recall, provides a balanced measure that is robust to class imbalance. The G-mean evaluates the geometric mean of class-wise Recall values and is particularly suited for imbalanced classification tasks, ensuring that poor performance on minority classes is reflected in the final score. Kappa and MCC evaluate the quality of predictions using all four elements of the confusion matrix, yielding a balanced performance metric. To ensure reliability, most experiments are repeated ten times, and the averaged results are reported. Data augmentation during training is also performed anew in each run to maintain stability. The test data exhibits a similar class imbalance like the training data, but for these data, augmentation is not performed to evaluate their performance on real data only.

5.2. Group validation

The group formed in Section 4.2 is validated using 2D Wasserstein distance (2D WD) [46]. 2D WD measures the distributional dissimilarity between different features over the spatial grid. It is based on pointwise comparison and is less sensitive to noise or extreme values. Pairwise WD is computed across each feature pair, and the distance validates the grouping if the following are true.

- Intra-group 2D-WD values are low (indicating high similarity within groups).
- Inter-group 2D-WD values are high (indicating clear separation between groups).

Table 2
Intra-Cluster 2D WD score of Group 5 (C_5).

Feature_Pair	*sp_d2m*	*sp_t2m*	*d2m_t2m*
2D WD	0.2920	0.5311	0.6027

Table 3
Inter-Cluster 2D WD Score.

Cluster Pair	C_1_C_2	C_1_C_3	C_1_C_4	C_1_C_5	C_1_C_6	C_2_C_3	C_2_C_4	C_2_C_5	C_2_C_6	C_3_C_4	C_3_C_5	C_3_C_6	C_4_C_5	C_4_C_6	C_5_C_6
2D WD	1.6931	3.2507	2.3564	2.8208	2.2284	3.3769	2.3404	2.7715	2.1262	3.3001	5.6153	4.8221	2.9303	2.3412	1.1886

It is observed in Table 3 that the intergroup 2D-WD values are high, indicating clear separation between groups, and in Table 2 the intragroup 2D-WD values are low, indicating high similarity within groups. This justifies the grouping obtained in Section 4.2.

5.3. Comparison with SOTA

Table 4 reports a detailed comparative performance analysis which highlights the better efficiency of the proposed SGA-Net model over several SOTA approaches across nine key evaluation metrics, which include Kappa, G-mean, Precision, Recall, and F1-score. SGA-Net achieves the highest Accuracy of 0.7471 (↑ 7.5% over [13]), indicating improved classification capability over other methods. The model also records the lowest MAE (0.0902), reflecting better classification Precision with reduced average deviation from ground truth. Though [12] slightly surpasses SGA-Net in RMSE (0.2149 vs. 0.2159), the performance difference is marginal.

The Precision and Recall of [12] (0.8932) and [19] (0.6837) respectively are better than proposed, but with respect to F1-score (0.7189) and MCC (0.5968) the proposed model is best. High in Precision but low in Recall means it fails to detect many true positives, and high in Recall but low in Precision means it detects many positives but triggers many false alarms. The proposed model best balances Precision and Recall, i.e., it becomes high only when both Precision and Recall are reasonably strong at the same time.

The G-mean is a performance metric commonly used in imbalanced classification, as it evaluates the balance between sensitivity and specificity. The G-mean values indicate that three models, Liu et al. [11], Gardoll et al. [12] and Pal et al. [3], completely fail to balance sensitivity and specificity and achieve a G-mean of 0, implying that at least one cyclone intensity class (typically higher-grade categories) is entirely misclassified. This behavior is explicitly evidenced in Figs. 8–10, where higher-intensity cyclones are consistently assigned to lower categories. Furthermore, Table 5 in Section 5.4 clearly shows that these models fail to detect higher cyclone grades, directly resulting in a G-mean of 0. Pal et al. [3] and Tian et al. [18] performs slightly better (0.1420 and 0.0.191) but still lacks class-balance capability.

The recent RTCC model, a CNN-transformer hybrid architecture, by Tian et al. [19] improves the G-mean to 0.5950, indicating a better sensitivity-specificity trade-off; however, its overall performance, represented by Accuracy, F1-score, Kappa, and MCC, remains is poor compared to the proposed method. This higher G-mean arises from better performance for grades 4–6 compared to the proposed approach, as reported in Table 5. However, this advantage comes at the expense of performance in lower-intensity categories (grades 1–3), suggesting that the model prioritizes discrimination of higher grades while exhibiting limited capability in accurately distinguishing weaker cyclone grades, i.e., it misclassifies lower grades often, as can be observed in the confusion matrix in Fig. 13.

In contrast, the proposed SGA-Net achieves a G-mean of 0.4994 alongside better Accuracy, F1-score, Kappa, and MCC values, demonstrating a more stable and balanced classification across all cyclone grades (shown in Table 5 and Fig. 14). This indicates that SGA-Net effectively mitigates class collapse in higher-intensity categories while maintaining strong performance on majority classes, making it substantially more reliable under highly imbalanced operational settings.

Table 4
Performance comparison of the proposed SGA-Net with existing SOTA methods for cyclone classification.

Name	Accuracy ↑	MAE ↓	RMSE ↓	Precision↑	Recall↑	G-mean ↑	F1-score ↑	Kappa ↑	MCC ↑
Varalakshmi et al. [13]	0.671474	0.105128	0.478191	0.628466	0.635542	0	0.631984	0.455908	0.478191
Liu et al.[11]	0.699359	0.100200	0.226017	0.512145	0.566566	0	0.688411	0.492583	0.512145
Gardoll and Boucher [12]	0.712179	0.095557	**0.214867**	**0.893213**	0.532530	0	0.667249	0.473396	0.529705
Pal et al. [3]	0.656731	0.103800	0.231515	0.231200	0.554518	0.142090	0.663181	0.478400	0.453260
ViT [18]	0.430723	0.662651	0.928141	0.678615	0.430723	0.019144	0.442639	0.274061	0.323735
RTCC [19]	0.683735	0.466867	0.919992	0.678699	**0.683735**	**0.595011**	0.680010	0.533296	0.533642
Proposed SGA-Net	**0.747115**	**0.090184**	0.215904	0.831255	0.633434	0.499353	**0.718985**	**0.610994**	**0.596806**

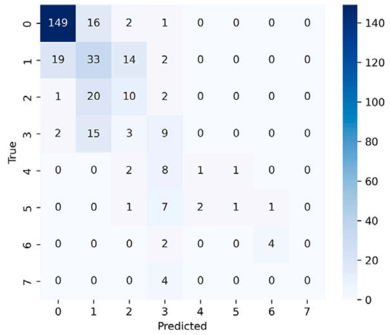

Fig. 8. Confusion matrix of Varalakshmi et al. [13].

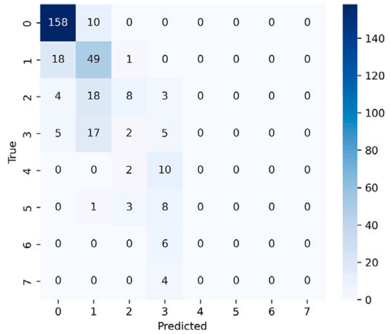

Fig. 9. Confusion matrix of Liu et al. [11].

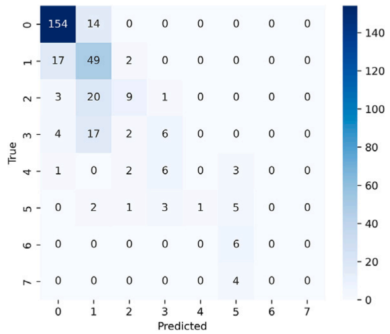

Fig. 10. Confusion matrix of Gardoll and Boucher [12].

Further, Kappa and MCC account for class imbalance and provide a balanced evaluation of multi-class classification, and therefore Kappa and MCC are relatively better performance evaluators of SGA-Net. SGA-Net outperforms other methods with a Kappa value of 0.6110 and a MCC value of 0.5968. Overall, the proposed SGA-Net exhibits strong and consistent performance, confirming its efficiency over existing methods.

5.4. Class-wise comparison with SOTA

Table 5 presents a detailed class-wise performance comparison of the proposed SGA-Net against SOTA methods. The results highlight that SGA-Net demonstrates strong and competitive performance across all eight cyclone intensity classes and nine evaluation metrics, particularly those suited for highly imbalanced datasets.

For lower and moderate intensity categories (Classes 0, 1, 2, and 3), the proposed SGA-Net consistently outperforms all competing approaches across almost all metrics. It achieves better Precision, Recall, F1-score, G-mean, Kappa, and MCC values while simultaneously yielding lower MAE and RMSE. These results indicate that SGA-Net effectively captures subtle spatial characteristics associated with weaker cyclone stages.

For Class 4-6, [19] achieves the best overall performance across most metrics, particularly Recall, F1-score, G-mean, Kappa, and MCC. Although SGA-Net attains competitive Accuracy and error measures for this class, its comparatively lower Recall suggests that a portion of true positive samples could not be correctly identified. Nevertheless, the proposed SGA-Net remains competitive and

Table 5
Class-wise performance comparison of the proposed SGA-Net with existing SOTA methods for cyclone classification.

Class	Name	Accuracy↑	MAE↓	RMSE↓	Precision↑	Recall↑	G-mean↑	F1-score↑	Kappa↑	MCC↑
0	Varalakshmi et al. [13]	0.876506	0.123494	0.351417	0.871345	0.886905	0.876316	0.879056	0.752922	0.753045
	Liu et al. [11]	0.888554	0.111446	0.333835	0.854054	**0.940476**	0.886364	0.895184	0.776801	0.780918
	Gardoll and Boucher [12]	0.882530	0.117470	0.342739	0.860335	0.916667	0.881437	0.887608	0.764838	0.766526
	Pal et al. [3]	0.885542	0.114458	0.338316	0.849462	**0.940476**	0.883123	0.892655	0.770752	0.775337
	ViT [18]	0.653614	0.346386	0.588545	1	0.315476	0.394763	0.479638	0.312864	0.430629
	RTCC [19]	0.882530	0.117470	0.342739	0.872832	0.898810	0.666326	0.885630	0.764941	0.765288
	Proposed SGA-Net	**0.903614**	**0.096386**	**0.310460**	**0.877778**	**0.940476**	**0.902394**	**0.908046**	**0.807033**	**0.809154**
1	Varalakshmi et al. [13]	0.740964	0.259036	0.508956	0.392857	0.485294	0.625735	0.434211	0.268648	0.271169
	Liu et al. [11]	0.804217	0.195783	0.442474	0.515789	**0.720588**	0.771383	0.601227	0.476163	0.487852
	Gardoll and Boucher [12]	0.783133	0.216867	0.465690	0.480392	**0.720588**	0.758897	0.576471	0.438451	0.454731
	Pal et al. [3]	0.780120	0.219880	0.468913	0.462687	0.455882	0.627468	0.459259	0.321273	0.321286
	ViT [18]	0.524096	0.475904	0.689858	0.252747	0.676471	0.733428	0.368000	0.099437	0.130816
	RTCC [19]	0.792169	0.207831	0.455885	0.491803	0.441176	0.592296	0.465116	0.336615	0.337380
	Proposed SGA-Net	**0.852410**	**0.147590**	**0.384175**	**0.620253**	**0.720588**	**0.799189**	**0.666667**	**0.572570**	**0.575224**
2	Varalakshmi et al. [13]	0.864458	0.135542	0.368161	0.312500	0.303030	0.529843	0.307692	0.232587	0.232621
	Liu et al. [11]	0.900602	0.099398	0.315274	0.500000	0.242424	0.485734	0.326531	0.279779	0.301284
	Gardoll and Boucher [12]	0.906627	0.093373	0.305571	0.562500	0.272727	0.516084	0.367347	0.323429	0.348288
	Pal et al. [3]	0.861446	0.138554	0.372229	0.303030	0.303030	0.528886	0.303030	0.226107	0.226107
	ViT [18]	0.885542	0.114458	0.338316	0.432432	0.484848	0.660799	0.457143	0.393403	0.394208
	RTCC [19]	0.870482	0.129518	0.359886	0.361111	0.393939	0.595636	0.376812	0.304695	0.305054
	Proposed SGA-Net	**0.918675**	**0.081325**	**0.285176**	**0.636364**	**0.424242**	**0.642566**	**0.509091**	**0.466683**	**0.478105**
3	Varalakshmi et al. [13]	0.861446	0.138554	0.372229	0.257143	0.310345	0.532649	0.281250	0.205328	0.206440
	Liu et al. [11]	0.834337	0.165663	0.407017	0.138889	0.172414	0.393413	0.153846	0.063205	0.063660
	Gardoll and Boucher [12]	0.900602	0.099398	0.315274	0.375000	0.206897	0.447290	0.266667	0.218099	0.229245
	Pal et al. [3]	0.912651	0.087349	0.295549	0.500000	0.344828	0.577449	0.408163	0.362722	0.370031
	ViT [18]	0.876506	0.123494	0.351417	0.342105	0.448276	0.639624	0.388060	0.320758	0.324385
	RTCC [19]	0.864458	0.135542	0.368161	0.214286	0.206897	0.434539	0.210526	0.136416	0.136441
	Proposed SGA-Net	**0.945783**	**0.054217**	**0.232845**	**0.789474**	**0.517241**	**0.714432**	**0.625000**	**0.597142**	**0.612684**
4	Varalakshmi et al. [13]	0.960843	0.039157	0.197880	0.333333	0.083333	0.287772	0.133333	0.120619	0.152043
	Liu et al. [11]	0.963855	0.036145	0.190117	0	0	0	0	0	0
	Gardoll and Boucher [12]	0.960843	0.039157	0.197880	0	0	0	0	−0.005592	−0.010644
	Pal et al. [3]	0.960843	0.039157	0.197880	0.428571	0.250000	0.496865	0.315789	0.297068	0.308559
	ViT [18]	0.972892	**0.027108**	0.164646	0.636364	**0.583333**	**0.749833**	0.608696	0.594683	0.595287
	RTCC [19]	**0.975904**	0.024096	**0.155230**	**0.700000**	**0.583333**	**0.749833**	**0.636364**	**0.624009**	**0.626784**
	Proposed SGA-Net	0.963855	0.036145	0.190117	0.500000	0.083333	0.288224	0.142857	0.133913	0.193470
5	Varalakshmi et al. [13]	0.963855	0.036145	0.190117	0.500000	0.083333	0.288224	0.142857	0.133913	0.193470
	Liu et al. [11]	0.963855	0.036145	0.190117	0	0	0	0	0	0
	Gardoll and Boucher [12]	0.939759	0.060241	0.245440	0.277778	0.416667	0.632250	0.333333	0.303107	0.309956
	Pal et al. [3]	0.969880	0.030120	0.173553	**1**	0.166667	0.408248	0.285714	0.278261	0.402015
	ViT [18]	0.963855	0.036145	0.190117	0	0	0	0	0	0
	RTCC [19]	**0.990964**	**0.009036**	**0.095059**	0.909091	**0.833333**	**0.896221**	**0.869565**	**0.864894**	**0.865773**
	Proposed SGA-Net	0.960843	0.039157	0.197880	0.454545	0.416667	0.639417	0.434783	0.414542	0.414963
6	Varalakshmi et al. [13]	0.990964	0.009036	0.095059	0.800000	0.666667	0.815243	0.727273	0.722717	0.725824
	Liu et al. [11]	0.981928	0.018072	0.134433	0	0	0	0	0	0
	Gardoll and Boucher [12]	0.981928	0.018072	0.134433	0	0	0	0	0	0
	Pal et al. [3]	0.978916	0.021084	0.145204	0.454545	0.833333	0.904431	0.588235	0.578374	0.606534
	ViT [18]	0.993976	**0.006024**	0.077615	**1**	0.666667	0.809085	0.800000	0.797066	0.814003
	RTCC [19]	**0.996988**	0.003012	**0.054882**	0.857143	**1**	**0.990923**	**0.923077**	**0.921550**	**0.924399**
	Proposed SGA-Net	0.969880	0.030120	0.173553	0.357143	0.833333	0.900182	0.500000	0.487021	0.534065
7	Varalakshmi et al. [13]	0.987952	0.012048	0.109764	0	0	0	0	0	0
	Liu et al. [11]	0.987952	0.012048	0.109764	0	0	0	0	0	0
	Gardoll and Boucher [12]	0.987952	0.012048	0.109764	0	0	0	0	0	0
	Pal et al. [3]	0.993976	0.006024	0.077615	0.666667	**1**	0.996947	0.800000	0.797066	0.814003
	ViT [18]	0.990964	0.009036	0.095059	0.571429	**1**	0.993958	0.727273	0.723026	0.752464
	RTCC [19]	0.993976	0.006024	0.077615	0.666667	**1**	0.993958	0.800000	0.797066	0.814003
	Proposed SGA-Net	**0.996988**	**0.003012**	**0.054882**	**0.800000**	**1**	**0.998474**	**0.888889**	**0.887381**	**0.893063**

outperforms several earlier SOTA methods, demonstrating improved robustness and stability compared to conventional CNN-based approaches.

For the highest intensity category (Class 7), SGA-Net achieves the best overall performance across all reported metrics, including Accuracy, error measures, and imbalance-aware indicators. This demonstrates its strong capability to identify the most severe cyclone events, even under extreme class imbalance. Furthermore, for higher-intensity classes, several traditional SOTA methods [11–13] report zero G-mean, Kappa, and MCC values, indicating complete failure to correctly identify these rare but operationally critical cyclone categories, as also reflected in Figs. 8–10.

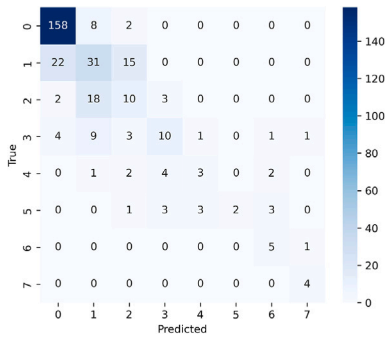

Fig. 11. Confusion matrix of Pal et al. [3].

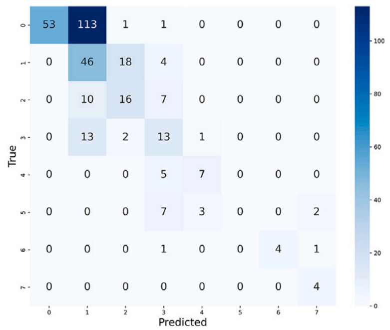

Fig. 12. Confusion matrix of ViT [18].

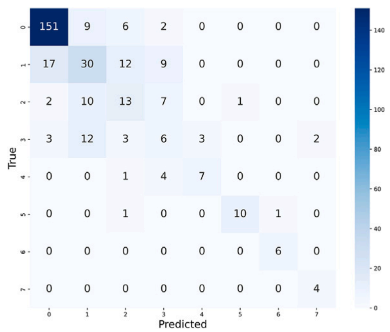

Fig. 13. Confusion matrix of RTCC [19].

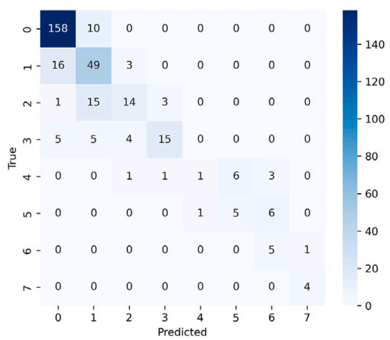

Fig. 14. Confusion matrix of the proposed SGA-Net.

Overall, SGA-Net performs best for five intensity grades (Classes 0, 1, 2, 3, and 7) and achieves competitive results for the remaining grades (Classes 4–6). While RTCC excels in Classes 4–6, its performance degrades noticeably in lower and moderate intensity categories, indicating reduced generalization across the full intensity spectrum. Thus, the proposed SGA-Net demonstrates a more balanced and consistent classification behavior across all cyclone intensity levels, with lower class dispersion and a practically tolerable error distribution. This balanced performance highlights the suitability of SGA-Net as a reliable and operationally viable architecture for real-world cyclone intensity classification.

5.5. Analysis of class-wise confusion matrices

Figs. 8–14 present the confusion matrices corresponding to four SOTA and the proposed SGA-Net for the multi-class cyclone intensity classification task. A comparative analysis of the distribution of errors over different classes reveals several consistent and contrasting patterns across these models.

Across all configurations, a strong diagonal dominance is observed for the lower-intensity classes (Class 0 and Class 1), indicating robust and stable classification performance for the majority classes. In particular, Liu et al. [11] and Pal et al. [3] and SGA-Net achieve the highest Accuracy for Class 0, reflecting a conservative prediction tendency toward lower cyclone categories. However, this conservatism of SOTA models leads to higher misclassification of lower-intensity cyclones into adjacent classes.

Intensity classes (Class 2 and Class 3) exhibit noticeable confusion across all models, primarily with neighboring categories. SGA-Net demonstrates comparatively improved balance in these classes, reducing excessive backward confusion into lower categories. In contrast, [11,18], and [19] show increased dispersion for these classes.

Intensity classes (Class 4, Class 5, and Class 6) also exhibit noticeable confusion across all models, primarily with neighboring categories except for [19]. The proposed SGA-Net has lower dispersion for class 5 and 6 and shows comparable results , but for Class 4 [19] is the best.

The most significant contrast is observed in the handling of higher-intensity class (Class 7). Liu et al. [11], Gardoll et al. [12] and Varalakshmi et al. [13] performs poorly in this regime, with most higher-class samples collapsing into lower categories, indicating a failure to capture intensity escalation. Pal et al. [3], Tian et al. [18] and W. Tian et al. [19], and SGA-Net shows more structured confusion patterns, while SGA-Net achieves the best relative separation among higher classes.

Overall, SGA-Net provides the most balanced performance across all cyclone intensity levels, effectively mitigating both over-conservatism and excessive class diffusion. These results highlight the reliability of the proposed SGA-Net on multi-class cyclone intensity classification, particularly under severe class imbalance conditions.

5.6. Class-wise density map visualization using Grad-CAM

To qualitatively assess model interpretability, Grad-CAM visualizations [47] are used to analyze the class-wise attention behavior of SGA-Net in comparison with existing SOTA models. Representative Grad-CAM maps for grades 0, 2, 5, and 7 are shown in Figs. 15–18, while comprehensive visualizations for all grades and discussion corresponding to each meteorological feature are provided in Section C of the Appendix. Each of the density maps has been annotated with (x) and the feature maps are annotated with (.) to indicate cyclone eye.

The SGA-Net consistently demonstrates more coherent and physically meaningful attention localization, particularly around the cyclone core and associated structural patterns. In contrast, baseline methods frequently exhibit diffuse or spatially misaligned attention, especially at lower and intermediate intensity grades. The improved interpretability of SGA-Net can be attributed to its local–global feature fusion strategy, which enables effective integration of fine-grained intensity cues with broader contextual information. Overall, the Grad-CAM analysis confirms that SGA-Net yields more structured and reliable attention distributions across cyclone intensities compared to existing approaches.

5.7. Ablation study of SGA-Net components

This Section presents an ablation study evaluating the individual and combined contributions of three key components in the proposed SGA-Net architecture: CA, SA, and GFE. The goal is to determine how each module impacts model performance. The ablation study in Table 6 demonstrates the importance of each architectural component in the proposed SGA-Net. The full model, incorporating CA, SA, and GFE, achieves the best performance across all evaluation metrics, indicating the complementary strengths of local attention and global context modeling. Removing GFE causes a slight drop in Accuracy and MCC, emphasizing its role in capturing long-range dependencies. Among the attention modules, SA contributes significantly better than CA, due to the spatially localized nature of cyclone patterns. The baseline model, with none of these enhancements, performs the worst, underscoring the critical importance of these design choices in achieving high-quality cyclone classification.

5.8. Ablation study using homogeneous pooling strategy

This Section presents a comparative study of the proposed SGA-Net with domain-aware pooling against some well-known and recent pooling techniques.

It is evident from Table 7 that the proposed domain-aware pooling method is far better than other methods for all four parameters. This confirms its efficacy in aggregating useful features for the purpose of cyclone classification. It gives the best Accuracy of 0.7471, the best MAE of 0.0902, the best RMSE of 0.2159, and the best MCC of 0.5968, making it the best method for classifying cyclones. Well-known methods of pooling, including all, max, average, min, and mixed, although moderately good, lack consistency compared to the proposed method. LEAP pooling and Normalized LBP, although recent, are less effective, indicating the advantage of incorporating prior knowledge compared to data-driven models for handling atmospheric data.

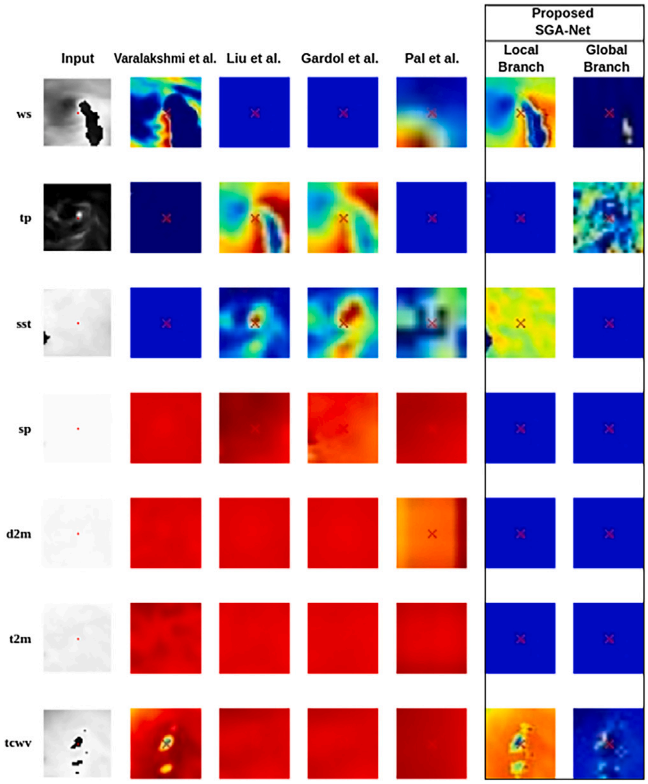

Fig. 15. Visualization of grade 0 attention patterns of SGA-Net and SOTA using GradCAM.

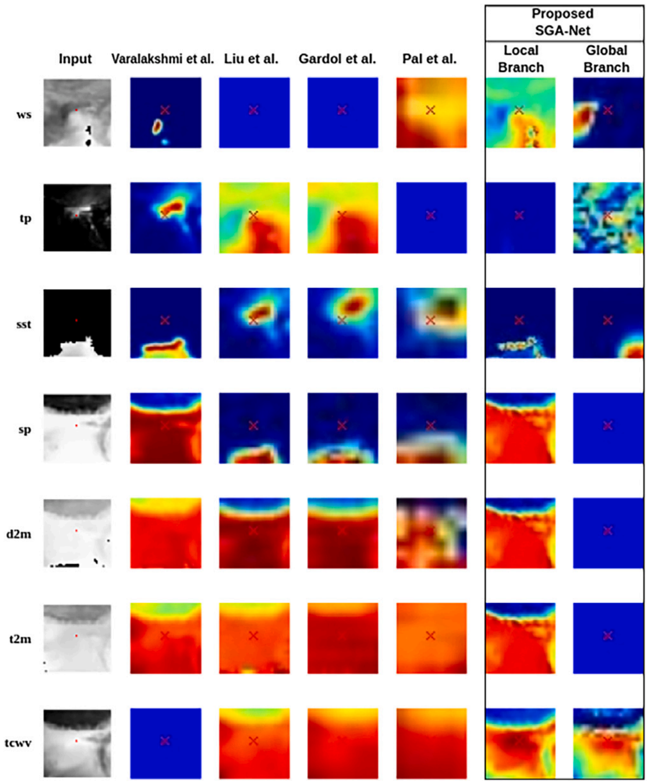

Fig. 16. Visualization of grade 2 attention patterns of SGA-Net and SOTA using GradCAM.

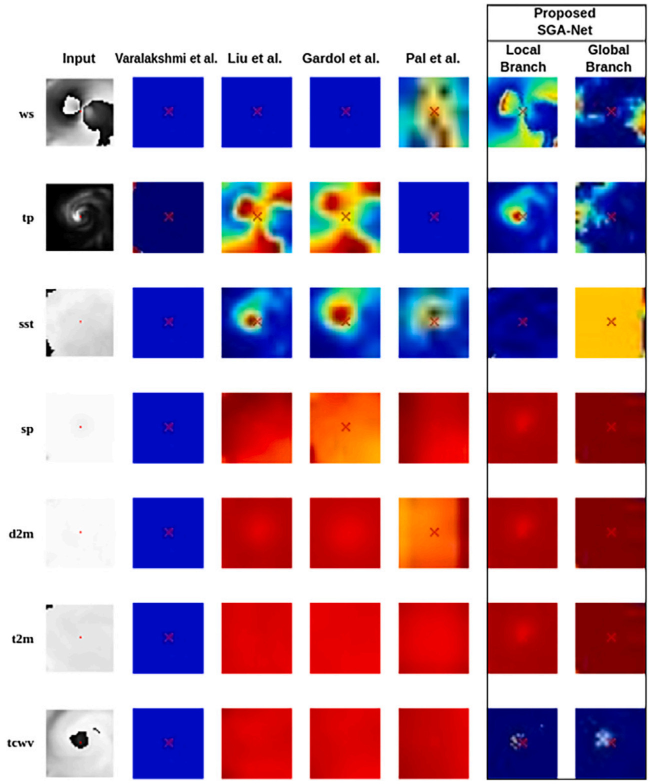

Fig. 17. Visualization of grade 5 attention patterns of SGA-Net and SOTA using GradCAM.

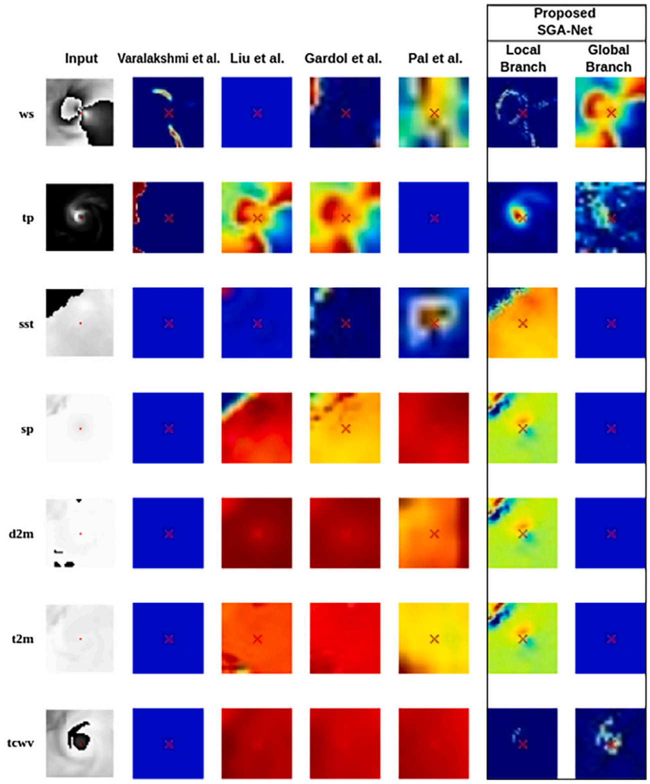

Fig. 18. Visualization of grade 7 attention patterns of SGA-Net and SOTA using GradCAM.

Table 6
An ablation study showing the impact of CA, SA, and GFE blocks.

CA	SA	GFE	Accuracy ↑	MAE ↓	RMSE ↓	MCC ↑
✓	✓	✓	**0.747115**	**0.090184**	**0.215904**	**0.596806**
✓	✓	✗	0.733654	0.092430	0.217342	0.576957
✗	✓	✗	0.717949	0.090503	0.216367	0.550484
✓	✗	✗	0.708333	0.095412	0.223502	0.537303
✗	✗	✗	0.682692	0.093029	0.220826	0.496389

Table 7
Performance comparison of SGA-Net pooling strategy with different pooling methods.

Name	Accuracy ↑	MAE ↓	RMSE ↓	MCC ↑
Proposed domain-aware Pooling	**0.747115**	**0.090184**	**0.215904**	**0.596806**
Max Pooling [24]	0.717308	0.092012	0.216702	0.532402
Average Pooling [25]	0.707051	0.091727	0.220433	0.512730
Min Pooling [26]	0.705769	0.091441	0.217789	0.512476
Mixed Pooling [27]	0.704808	0.091744	0.222280	0.510852
SP Average Pooling [28]	0.692628	0.087896	0.224401	0.496575
LP Pooling [29]	0.691026	0.094535	0.225937	0.487109
SP Max Pooling [28]	0.688782	0.086016	0.223629	0.500870
Weighted Pooling [30]	0.675641	0.093352	0.226246	0.476878
LEAP Pooling [31]	0.668269	0.092088	0.229166	0.467553
Normalized LBPPooling [32]	0.660256	0.103738	0.234984	0.431649

5.9. Ablation study using heterogeneous pooling strategy in the groups

The ablation study is presented in Table 8 quantitatively evaluates the effectiveness of the proposed domain-aware pooling strategy by systematically altering the pooling operation applied to each feature group while keeping all other architectural components unchanged. The (*) in the table indicates that the pooling operation in that group is kept as proposed.

The experimental results indicate that the full domain-aware configuration attains the highest performance across all metrics, with the exception of Precision and G-mean, for which it ranks second, exhibiting only a negligible difference from the best-performing configuration. Importantly, modifying the pooling operation for any single group consistently degrades performance, indicating that physically motivated combination of pooling for the groups contributes uniquely.

Table 8
Comprehensive ablation study on the pooling strategy of SGA-Net, where each group is assigned a distinct pooling method.

Groups					Metrics								
G1	G2	G3	G4	G5	Accuracy ↑	MAE ↓	RMSE ↓	Precision↑	Recall↑	G-mean ↑	F1-score ↑	Kappa ↑	MCC ↑
Avg	*	*	*	*	0.677711	0.096615	0.223219	0.813765	0.605422	0.450060	0.694300	0.516503	0.519451
Min	*	*	*	*	0.710843	0.092700	0.217760	0.802326	0.623494	0.489011	0.701695	0.567779	0.569951
Mixed	*	*	*	*	0.704819	0.094142	0.221462	0.824490	0.608434	0.480562	0.700173	0.551176	0.556577
NLBP	*	*	*	*	0.683735	0.098677	0.224488	0.807229	0.605422	0.411394	0.691910	0.518422	0.527249
No	*	*	*	*	0.671687	0.092045	0.228380	0.741573	0.596386	0.474300	0.661102	0.506915	0.509886
*	**Avg**	*	*	*	0.722892	0.093099	0.216905	0.831276	0.608434	0.460420	0.702609	0.583068	0.585546
*	**Max**	*	*	*	0.692771	0.095496	0.222406	0.801587	0.608434	0.459006	0.691781	0.541256	0.547951
*	**Mixed**	*	*	*	0.716867	0.093861	0.223520	0.808765	0.611446	0.481526	0.696398	0.567894	0.573018
*	**NLBP**	*	*	*	0.731928	0.095457	0.221973	0.828326	0.581325	0.485474	0.683186	0.594355	0.598591
*	**No**	*	*	*	0.686747	0.095956	0.223399	0.785992	0.608434	0.468626	0.685908	0.533140	0.536762
*	*	**Max**	*	*	0.710843	0.092402	0.218616	0.809339	0.626506	0.429074	0.706282	0.564685	0.570786
*	*	**Min**	*	*	0.725904	0.091902	0.219607	0.810277	0.617470	0.478409	0.700855	0.585529	0.589143
*	*	**Mixed**	*	*	0.695783	0.095352	0.221236	0.823770	0.605422	0.451089	0.697917	0.547403	0.549232
*	*	**NLBP**	*	*	0.710843	0.093493	0.216569	**0.850622**	0.617470	0.497638	0.715532	0.570678	0.572735
*	*	**No**	*	*	0.650602	0.095676	0.225471	0.745318	0.599398	0.421540	0.664441	0.479596	0.483094
*	*	*	**Avg**	*	0.740964	0.092176	0.217943	0.826613	0.617470	0.462780	0.706897	0.605869	0.609802
*	*	*	**Max**	*	0.722892	0.093555	0.217997	0.818548	0.611446	0.446848	0.700000	0.585248	0.591308
*	*	*	**Min**	*	0.710843	0.093475	0.218812	0.810277	0.617470	0.453258	0.700855	0.566178	0.571458
*	*	*	**Mixed**	*	0.713855	0.091266	0.217781	0.814961	0.623494	0.467307	0.706485	0.576019	0.580780
*	*	*	**NLBP**	*	0.659639	0.104893	0.229766	0.769547	0.563253	0.442559	0.650435	0.502216	0.511684
*	*	*	*	**Avg**	0.689759	0.093591	0.217622	0.824701	0.623494	0.419323	0.710120	0.536539	0.540647
*	*	*	*	**Max**	0.713855	0.091456	0.219786	0.820717	0.620482	0.493971	0.706690	0.564936	0.568599
*	*	*	*	**Min**	0.707831	0.093817	0.223042	0.800000	0.614458	0.435637	0.695060	0.559519	0.563480
*	*	*	*	**Mixed**	0.716867	0.091015	0.216739	0.816733	0.617470	0	0.703259	0.573615	0.576421
*	*	*	*	**NLBP**	0.722892	0.094171	0.221159	0.814815	0.596386	**0.527457**	0.688696	0.583438	0.585652
*	*	*	*	*	**0.742170**	**0.090750**	**0.216300**	0.832150	**0.633430**	0.499350	**0.719318**	**0.610900**	**0.614600**

Table 9
Performance with Seasonal Variation.

Season	Name	Accuracy↑	MAE↓	RMSE↓	Precision↑	Recall↑	G-mean↑	F1-score↑	Kappa↑	MCC↑
WinterMonsoon (December–January)	Varalakshmi et al. [13]	0.654010	**0.086497**	0.680262	0.675029	**0.654010**	0.058425	0.664354	0.445219	0.465439
	Liu et al. [11]	0.703610	0.101989	0.214151	**0.905615**	0.558263	0.163222	0.690729	0.533419	0.560074
	Gardoll and Boucher [12]	0.713804	0.096967	**0.211023**	0.883528	0.598605	0.129318	0.713680	0.547083	0.575801
	Pal et al. [3]	0.659469	0.104323	0.226748	0.823140	0.557723	0.098069	0.664924	0.469011	0.500643
	Proposed SGA-Net	**0.747889**	0.090472	0.211051	0.836993	0.634460	0	**0.721788**	**0.604202**	**0.627070**
Spring (February–March)	Varalakshmi et al. [13]	0.687494	**0.078126**	0.637066	0.715240	0.687494	0.111290	0.701093	0.437095	0.455983
	Liu et al. [11]	0.735745	0.097296	0.205748	**0.914944**	0.586610	0.193314	0.714880	0.496054	0.522963
	Gardoll and Boucher [12]	0.743186	0.091425	0.202080	0.893126	**0.628170**	0.127299	0.737575	0.471507	0.498159
	Pal et al. [3]	0.711146	0.093438	0.206661	0.853835	0.613322	0.217733	0.713864	0.492206	0.520411
	Proposed SGA-Net	**0.769025**	0.084682	**0.200761**	0.851886	0.659711	0.020000	**0.743582**	**0.505921**	**0.528963**
Summer (April–May)	Varalakshmi et al. [13]	0.695511	**0.076122**	0.510255	0.719557	0.695511	0	0.707329	0.452106	0.468638
	Liu et al. [11]	0.765983	0.088802	0.197451	**0.898584**	0.630820	0.326444	0.741263	0.598495	0.607164
	Gardoll and Boucher [12]	0.779977	0.083221	**0.192431**	0.883472	0.666540	0.258636	**0.759826**	0.620260	0.628913
	Pal et al. [3]	0.719195	0.090760	0.208658	0.854813	0.620611	0.139644	0.719124	0.525656	0.539210
	Proposed SGA-Net	**0.783050**	0.079673	0.200840	0.827659	**0.681053**	0	0.747233	**0.628252**	**0.639123**
SummerMonsoon (June–September)	Varalakshmi et al. [13]	0.673296	**0.081676**	0.666631	0.689756	0.673296	0.077901	0.681427	0.446151	0.458267
	Liu et al. [11]	0.738147	0.092071	0.204522	**0.897517**	0.612300	0.217629	0.727968	0.561773	0.575331
	Gardoll and Boucher [12]	0.753661	0.087331	**0.200717**	0.887046	0.648140	0.172424	0.749004	0.584279	0.597560
	Pal et al. [3]	0.695958	0.096510	0.218438	0.837234	0.600683	0.093096	0.699501	0.496711	0.516122
	Proposed SGA-Net	**0.776672**	0.083181	0.203554	0.848562	**0.676715**	0	**0.752958**	**0.628708**	**0.640947**
Autumn (PostMonsoon) (October–November)	Varalakshmi et al. [13]	0.764706	**0.058824**	0.257353	0.866516	0.764706	0	0.812434	0.484835	0.508134
	Liu et al. [11]	0.916176	0.075017	0.159117	**0.996036**	0.720588	0.652888	0.836213	0.807483	0.811325
	Gardoll and Boucher [12]	**0.922059**	0.067116	**0.149679**	0.990668	0.786765	0.461586	**0.877021**	**0.816114**	**0.817772**
	Pal et al. [3]	0.864706	0.073374	0.167479	0.938565	0.719118	0.279288	0.814316	0.708052	0.722313
	Proposed SGA-Net	0.902941	0.061981	0.154567	0.946245	**0.800000**	0	0.866999	0.777324	0.780312

5.10. Performance on seasonal-variation

The cyclones from the year 2021 are considered and, based on their time of occurrence, are categorized into five standard seasons commonly used in climatological studies over the Indian subcontinent: spring, summer or pre-monsoon, summer monsoon, autumn or post-monsoon, and winter monsoon. Moreover, in the past two decades, cyclones or depressions have rarely occurred in the BoB during spring, with only one such event recorded in 2011. This shows the tendency of cyclones to not occur during spring in the BoB basin due to insufficient initial conditions in this period. Therefore, the 2011 cyclone is included alongside the 2021 cyclones to ensure complete seasonal representation. The seasonal-wise performance of all models is summarized in Table 9. The 2021 dataset contains no grade-7 cyclones; thus, the corresponding confusion-matrix row is expected to be zero, implying a G-mean of zero. This expected behavior is consistently observed for the proposed SGA-Net across all seasons dominated by 2021 samples, whereas several SOTA methods report non-zero G-mean values, indicating erroneous misclassification into an absent class.

Across the summer, summer monsoon, and winter monsoon seasons, the proposed model achieves the best or near-best performance in terms of Accuracy, F1-score, Kappa, and Matthews correlation coefficient, demonstrating better robustness under seasonal variability. For the autumn (post-monsoon) season, although some SOTA methods achieve marginally higher values for select metrics, their non-zero G-mean values remain inconsistent with the dataset composition. For the spring season, the available sample corresponds to depression, which has a maximum grade of 1. The proposed model yields a slightly positive G-mean, suggesting occasional misclassification into neighboring grades. The SOTA models also produce positive lower G-mean values, but they are more than proposed, indicating more misclassification. Overall, the results in Table 9 indicate that SGA-Net provides more reliable and seasonally consistent classification performance compared to existing approaches.

5.11. Performance under spatially correlated noise

To establish a baseline assessment of robustness in Section 5.11 of the main manuscript, the classifier is evaluated with some perturbations. In this context, white Gaussian noise provides one of the most basic forms of additive variability. However, white Gaussian noise consists of independent perturbations at each grid point, i.e., $Z(s_1) \perp Z(s_2)$ for all $s_1 \neq s_2$, where s_1 and s_2 are two randomly chosen spatial points. While such noise is mathematically simple, it is physically unrealistic for meteorological fields, as they rarely exhibit pixel-wise independent variability. Applying white noise directly to gridded weather data introduces abrupt, high-frequency fluctuations that do not resemble true atmospheric structures and may artificially degrade the signal by injecting unphysical discontinuities.

To obtain more realistic perturbations, the white noise field is smoothed with a Gaussian kernel, producing a spatially correlated noise field. This imposes a correlation structure that decays smoothly with distance, such that nearby grid points receive similar perturbations, which closely resembles physical errors in meteorological datasets. Such perturbations allow us to assess the stability of the model under plausible noise scenarios and to examine its sensitivity to small but spatially structured deviations. The following paragraphs discuss how these noises are added to the data to study its robustness.

Let the input feature vector be defined as, $\mathcal{X}_f \in \mathbb{R}^{N\times H\times W\times C}$, where f is the climate features, H and W denote spatial dimensions, and C is the number of input channels. For each sample i, feature f, and channel c, the perturbed field is defined as

$$\tilde{X}^i_f(:,:,c) = X^i_f(:,:,c) + \epsilon_{i,f,c} \tag{21}$$

where the noise field $\varepsilon_{i,f,c}$ is constructed by smoothing white Gaussian noise. Specifically,

$$\varepsilon_{i,f,c} = G_{\sigma_{f,c}} * Z, \tag{22}$$

with $Z \sim \mathcal{N}(0, \sigma^2 I)$ representing zero-mean, independent Gaussian noise of variance σ^2, and $G_{\sigma_{f,c}}$ denoting a two-dimensional Gaussian kernel with standard deviation $\sigma_{f,c}$. The operator $*$ denotes spatial convolution.

In the experimental study, the σ^2 varied from 0.1 to 0.8, and $G_{\sigma_{f,c}}$ varied from 1 to 3 and each of the performance metrics is observed. Fig. 19 presents a comparative performance analysis between the proposed method and the SOTA approaches with respect to MCC reported across varying levels of spatially correlated Gaussian noise. It shows that the performance slightly fluctuates as noise is added to models, but the relative performance among them remains almost constant. It is observed for all models that when σ^2 is 0.8 and $G_{\sigma_{f,c}}$ is 1 or 1.5 the degradation is maximum. For the proposed model, in spite of noise, the MCC remains more than 0.50 (except when spatial noise is 0.8 and spatial correlation (cor) is 1.5, then the MCC is degraded by 18% which is still better than the others), and even under noisy conditions the proposed model performs the best. Fig. 19 shows the performance with respect to MCC.

Similarly, Figs. 20–26 shows that the performance slightly fluctuates as noise is added to models, but the relative performance among them remains almost constant. It is observed for all models that when σ^2 is 0.8 and $G_{\sigma_{f,c}}$ is 1 or 1.5 the degradation is maximum. The G-mean of [11–13] is 0 irrespective of the noise added, meaning that one or more classes go completely undetected. Pal et al. [3] and the proposed SGA-Net has a positive G-mean. But when σ^2 is 0.8 and $G_{\sigma_{f,c}}$ is 1.5 the performance of [3] decreases drastically, and the Accuracy is 0.1572. For the proposed model, in spite of noise, the Accuracy remains more than 0.70 except when spatial noise is 0.8 and spatial cor is 1.5. So, even under noisy conditions, the proposed model performs the best.

5.12. Generalizability analysis of SGA-Net

To establish the generalizability of SGA-Net, a cross-regional validation experiment is carried out in the AS region. For this region, the same experimental setup as described for the BoB in Section 3 is adopted to ensure methodological consistency. Specifically, the same set of ERA5 meteorological features, identical spatial resolution (0.25° × 0.25°), identical temporal resolution (3-hours) and the same temporal span (2002–2021) are used. Cyclone events in the AS region during this period are annotated using the IBTrACS dataset, following the same cyclone grading criteria, spatial patch extraction strategy, and train–validation–test split as employed for the BoB region in Section 3. Furthermore, the same preprocessing pipeline and data augmentation strategy described in Section 3.1

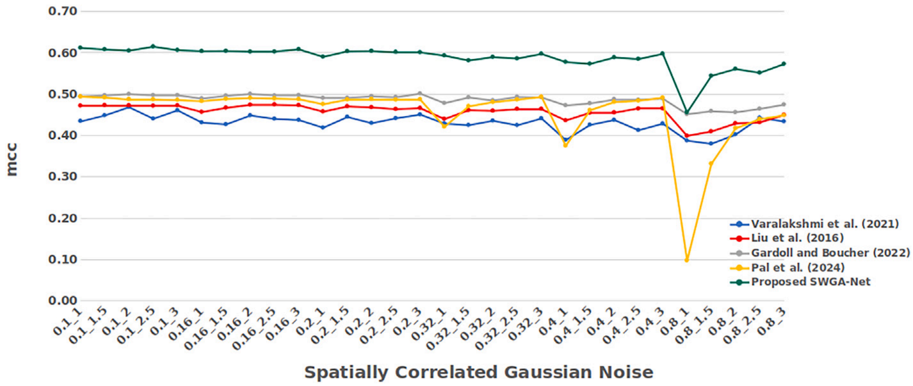

Fig. 19. MCC of all the methods against the addition of spatially correlated Gaussian noise.

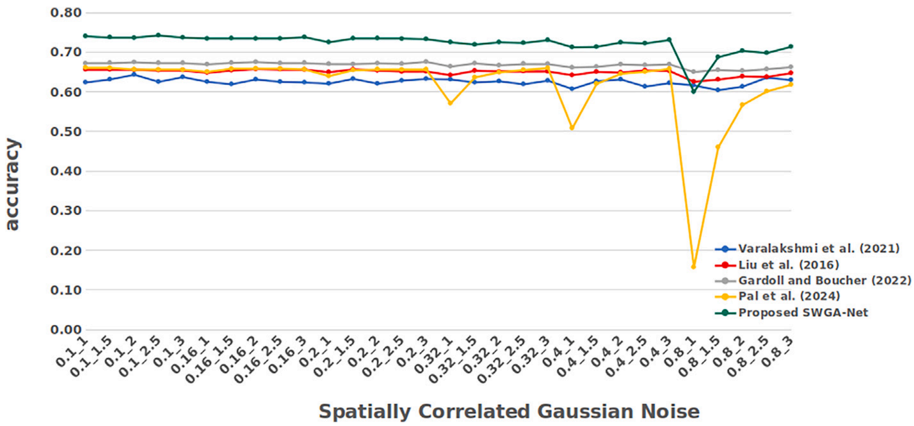

Fig. 20. Accuracy of all the methods against the addition of spatially correlated Gaussian noise.

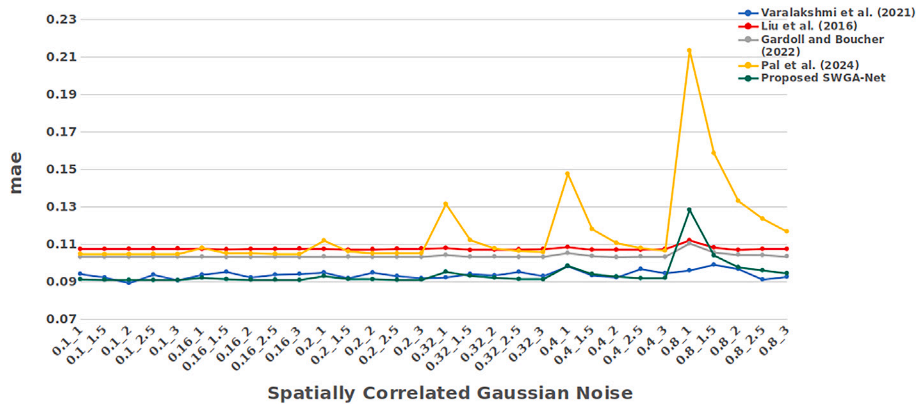

Fig. 21. MAE of all the methods against the addition of spatially correlated Gaussian noise.

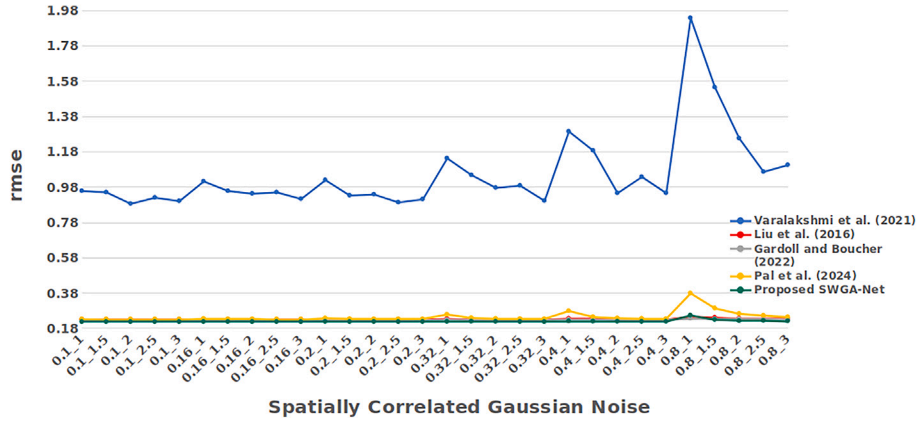

Fig. 22. RMSE of all the methods against the addition of spatially correlated Gaussian noise.

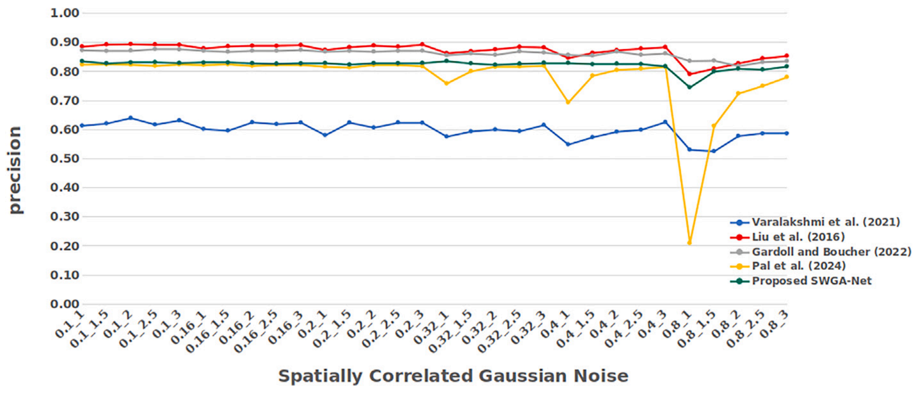

Fig. 23. Precision of all the methods against the addition of spatially correlated Gaussian noise.

are applied to the AS dataset to address class imbalance and ensure a balanced training set across all cyclone intensity grades. The final class distribution of the data obtained after curation is given in Table 10.

The generalizability analysis is conducted in two phases: first, it is analyzed without regional meteorological information, i.e., zero-shot learning, and in the second phase, the performance is analyzed after training with regional meteorological information.

5.12.1. Comparative performance analysis on AS region with zero-shot learning

Table 11 presents the performance of SOTA models and the proposed SGA-Net under zero-shot learning. It is evident from Table 11 that every model exhibits a significant comparative performance drop.

This is because although both the AS and the BoB exhibit a bimodal seasonal cycle of tropical cyclone genesis, the governing large-scale environmental conditions differ substantially between the two basins. In particular, more favorable thermodynamic and moisture conditions prevail over the Bay of Bengal, while subsidence and less favorable vertical circulation dominate over the Arabian

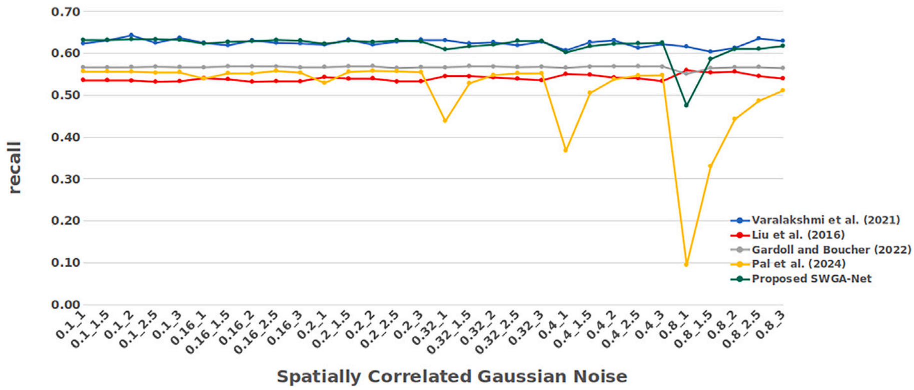

Fig. 24. Recall of all the methods against the addition of spatially correlated Gaussian noise.

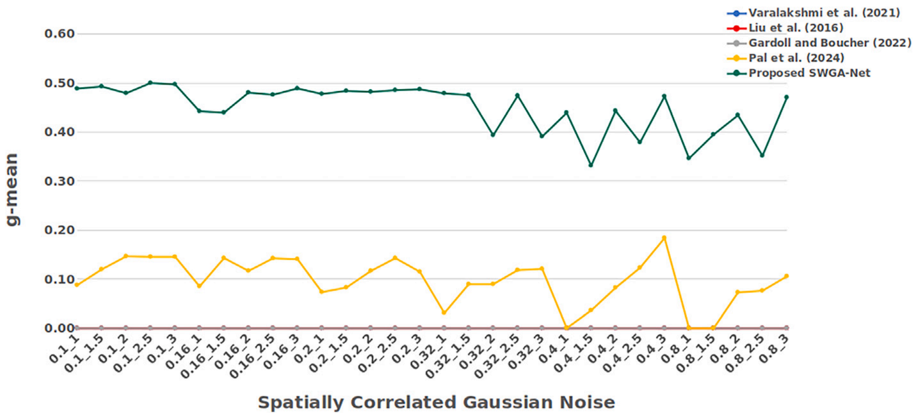

Fig. 25. G-mean of all the methods against the addition of spatially correlated Gaussian noise.

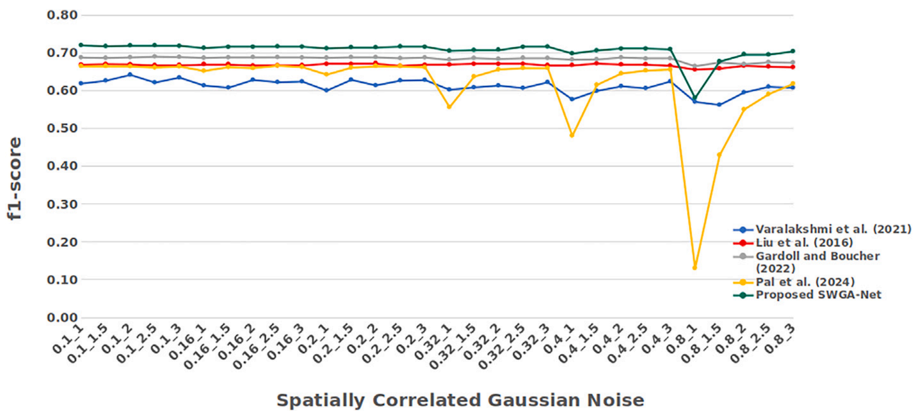

Fig. 26. F1-score of all the methods against the addition of spatially correlated Gaussian noise.

Sea, especially during the post-monsoon season [48]. These semantic feature differences of the AS and the BoB region are also shown by Fan et al. [49]. As a result, the models fail to perform in the unknown application area. The proposed SGA-Net, while performing best in the trained setting, does not perform best under the zero-shot scheme. It indicates that the deeper multi-branch architecture, although powerful, relies on learned cross-scale attention patterns that do not transfer optimally without information about the application region.

It is also important to note that all models have a G-mean = 0, which means that they completely fail to identify at least one class of cyclone grade. This confirms that zero-shot classification is insufficient for cyclone classification due to strong class imbalance and significant domain shifts.

Table 10
Class distribution of cyclone instances in the AS region according to the Indian grading scale.

Wind Speed (ws) (in Knots)	Cyclone Category (grade)	Class Distribution (%)
ws < 17	Normal / Low Pressure (0)	1715 (48.54%)
17 ≤ ws < 28	Depression (1)	706 (19.98)
28 ≤ ws < 34	Deep Depression (2)	335 (9.48%)
34 ≤ ws < 48	Cyclonic Storm (3)	331 (9.37%)
48 ≤ ws < 64	Severe Cyclonic Storm (4)	189 (4.7%)
64 ≤ ws < 91	Very Severe Cyclonic Storm (5)	166 (5.35%)
91 ≤ ws < 120	Extremely Severe Cyclonic Storm (6)	72 (2.04%)
ws ≥ 120	Super Cyclonic Storm (7)	19 (0.54%)

Table 11
Performance comparison of the proposed SGA-Net with existing SOTA methods for cyclone classification in the AS region with a zero-shot setting.

Name	Accuracy ↑	MAE ↓	RMSE ↓	Precision↑	Recall↑	G-mean ↑	F1-score ↑	Kappa ↑	MCC ↑
Varalakshmi et al. [13]	**0.422000**	**0.144500**	4.088000	0.312457	**0.422000**	0	**0.359059**	**0.278798**	**0.293104**
Liu et al. [11]	0.286400	0.177480	0.334024	0.462958	0.241200	0	0.317160	0.059397	0.074035
Gardoll and Boucher [12]	0.289600	0.175743	0.324410	**0.533416**	0.235200	0	0.326455	0.076369	0.090387
Pal et al. [3]	0.296800	0.179051	**0.328357**	0.505689	0.220800	0	0.307386	0.112149	0.126217
Proposed SGA-Net	0.289600	0.179756	0.345197	0.418818	0.235600	0	0.301561	0.088000	0.098573

Table 12
Performance comparison of the proposed SGA-Net with existing SOTA methods for cyclone classification in the AS region with transfer learning.

Name	Accuracy ↑	MAE ↓	RMSE ↓	Precision↑	Recall↑	G-mean ↑	F1-score ↑	Kappa ↑	MCC ↑
Varalakshmi et al. [13]	0.390000	0.152500	1.742000	0.393928	0.390000	0	0.391954	0.266411	0.272421
Liu et al. [11]	0.416000	0.159630	0.285590	**0.802326**	0.276000	0	0.410714	0.300994	0.340071
Gardoll and Boucher [12]	0.476000	0.153748	0.281973	0.788462	0.328000	0	0.463277	0.358460	0.375970
Pal et al. [3]	0.582400	0.146180	0.269489	0.785723	0.320800	0.226465	0.455589	0.491432	0.497325
Proposed SGA-Net	**0.592000**	**0.134887**	**0.259697**	0.727273	**0.448000**	**0.455596**	**0.554455**	**0.507636**	**0.513308**

5.12.2. Comparative performance analysis on AS region with training on regional meteorological features

Table 12 presents the performance of all models (trained on the BoB region) after training on the meteorological features of the AS region. Consistent with its behavior on the BoB, the proposed SGA-Net achieves the best overall performance for the AS region as well. It achieves the highest Accuracy (0.5920), G-mean (0.4556), Kappa (0.5076), and MCC (0.5133); and the lowest MAE (0.1349) and RMSE (0.2597), indicating better balanced classification across classes—including minority intensities that other models fail to detect (G-mean = 0). Although [11] achieves higher Precision, its low Recall (0.2760) and G-mean value 0 reveal weaker sensitivity to multiple classes. This two-phase generalizability analysis highlights the cross-regional applicability of the proposed SGA-Net under transfer learning and shows that it learns the regional semantic features with efficacy.

5.13. Efficiency measures

Table 13 summarizes the computational benchmarks, including parameter count, gflops, inference time, memory usage, throughput, and Accuracy-related metrics, measured under the experimental setup described in Section 5. A cost–benefit analysis is performed using Accuracy, MCC, their percentage gains achieved by SGA-Net, and comparative computational complexity.

SGA-Net attains the highest Accuracy (0.7471) and MCC (0.5968), outperforming all baselines by at least 5% and 12%, respectively. Although lighter models [11,12] exhibit lower parameter counts, gflops, and faster inference, these gains come at a substantial cost to predictive performance. The proposed model has a slightly higher inference time (0.06 s per sample), which remains well within practical limits for operational deployment. Recent research has also shown that there are systematic differences between the outputs of climate models and actual satellite measurements [50]. Thus, throughout testing, no synthetic data generation or augmentation was used to ensure that the assessment settings were more realistic and focused on deployment.

The SOTA baselines primarily rely on standard convolutional operations. The theoretical complexity of these SOTA methods is dominated by $\mathcal{O}(HWK^2CO_F)$. SGA-Net, on the other hand, exhibits a higher theoretical complexity –$\mathcal{O}(H^2W^2C)$ – due to the multi-head attention module, enabling richer feature interactions and better predictive performance.

In summary, the performance improvement achieved by SGA-Net makes its computational complexity justified, as it offers an efficient trade-off between performance and computational complexity without the need for further optimization methods such as pruning and quantization.

Table 13
Efficiency measures of all the models.

Name	Accuracy ↑	%Accuracy gain by SGA-Net	MCC ↑	%MCC gains by SGA-Net	#params↓	gflops↓	Avg forward pass time per batch (s)↓	Avg infer time (s)↓	Avg mem usage (mb)↓	Throughput↑	Avg forward pass time per sample (s)↓	Complexity↓
Varalakshmi et al. [13]	0.671474	11.26492	0.478191	24.80494	9,740,872	560.0121	0.327558	0.072904	0.081843	13	0.010236	$\mathcal{O}\left(HWK^2CO_F\right)+\mathcal{O}\left(NF_{flatten}D_{depth}\right)$
Liu et al. [11]	0.699359	4.905508	0.512145	16.53067	462,794	3.809735	**0.130471**	0.058795	0.000413	17	**0.00051**	$\mathcal{O}\left(HWK^2CO_F\right)$
Gardoll and Boucher [12]	0.712179	6.828539	0.529705	12.66762	**298,826**	**2.047525**	0.142666	**0.053884**	**3.80E-05**	**18**	0.000557	$\mathcal{O}\left(HWK^2CO_F\right)$
Pal et al. [3]	0.656731	13.76271	0.45326	31.66968	5,565,561	290.3732	0.57502	0.05979	0.000876	16	0.001284	$\mathcal{O}\left(HWK^2CO_F\right)$
Proposed SGA-Net	**0.747115**	**0**	**0.596806**	**0**	1,217,876	436.0237	1.603806	0.06997	0.080191	14	0.00358	$\mathcal{O}\left(H^2W^2C\right)$

6. Conclusion & future works

In this manuscript, SGA-Net, a novel multi-branch local–global architecture tailored for cyclone classification, is introduced. Through SSIM-based feature grouping, domain-aware pooling, and a triple attention mechanism, the proposed model effectively addresses the major challenges outlined in the introduction. The method mitigates class imbalance and enables efficient 8-class cyclone classification. The triple attention mechanism facilitates the learning of both local and global cyclone structures, rather than relying solely on local features. Moreover, Grad-CAM visualizations further validate that the model captures discriminative local and global patterns, enhancing interpretability.

With significantly improved MCC, F1-score, Kappa, and Accuracy compared to SOTA methods, SGA-Net represents an important advancement for cyclone classification and establishes a strong foundation for integrating domain knowledge into DL architectures for geophysical applications. The framework can be extended to other geographic regions, provided region-specific training is performed, since models trained on one domain do not generalize well to regions with differing climatic conditions. Additionally, the model offers an effective balance between computational complexity, inference time, and performance gains—achieving a complexity of $\mathcal{O}(H^2W^2C)$—and demonstrates robustness to spatial noise, where performance degradation remains minimal. These advantages make SGA-Net highly suitable for real-world operational deployment.

Although SGA-Net shows better efficiency and probable real-life applicability, various future directions remain open for enhancement. Firstly, the prediction of cyclone evolution over time should be explored through spatiotemporal forecasting frameworks, which can better capture dynamical variations in cyclone structure and motion. Within such frameworks, the proposed model can be integrated as a spatial feature extractor or cyclone state classifier, operating on future climate instances generated by temporal models to support the analysis of cyclone intensification and weakening patterns. Although the current work addresses class imbalance using a traditional method, future studies should investigate more advanced imbalance-handling strategies. In particular, the use of synthetic oversampling techniques—such as generating realistic minority-class samples via conditional generative models or mixup-based augmentation—can be explored. Additionally, factors such as cyclone translational speed, size, and asymmetry should be integrated into the generative modeling process to improve robustness and predictive reliability. Also, the incorporation of specialized loss functions pertaining to class imbalance may further enhance model generalization. This study can also be further extended to learn region-specific meteorological semantics of two important basins—the western North Pacific and North Atlantic basins—and evaluate its cross-regional generalization capability.

CRediT authorship contribution statement

Soumi Sarkar: Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Conceptualization. **Subhadip Boral:** Writing – review & editing, Validation, Methodology, Formal analysis. **Sarbani Palit:** Writing – review & editing, Supervision. **Ashish Ghosh:** Writing – review & editing, Supervision, Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Validation of the beta distribution assumption for SSIM

In Section 4.2 of the main manuscript, the rationale for adopting the Beta distribution was introduced. To further validate this choice, an analysis of the statistical distribution of SSIM values by fitting them to eleven commonly used distributions: Cauchy, Chi-squared, Exponential, Exponentiated Power, Gamma, Log-Normal, Normal, Power-law, Rayleigh, Uniform, and Beta was performed.

For each feature pair, the maximum log-likelihood value was computed across these distributions. A higher log-likelihood indicates a better fit, suggesting that the distribution more accurately represents the observed SSIM values. As summarized in Table A.1, the Beta distribution consistently achieves the highest log-likelihood values for all feature pairs, providing strong empirical evidence that SSIM values follow a Beta distribution.

This conclusion is further supported by the Q-Q plots in Fig. A.1, which visually confirm the alignment of SSIM values with the theoretical Beta distribution.

Table A.1
Log-likelihood values for 11 common distributions.

Feature	cauchy	chi2	expon	exponpow	gamma	lognorm	norm	powerlaw	rayleigh	uniform	beta
u10_v10	3026.14	1370.22	1529.33	3563.89	4047.92	4051.05	3768.63	1340.82	3724.92	1290.08	4047.89
u10_tp	9706.78	10177.02	5193.53	9509.47	10196.92	10202.52	10204.56	5908.6	8223.3	5882.35	10206
u10_sst	3535.16	1197.25	1674.88	4177.86	4745.39	4765.52	4427.74	2085.72	4124.54	2085.61	4744.67
u10_sp	4536.16	5279.66	2413.46	4490.44	5279.66	5301.22	4982.21	1954.55	4768.7	1868.09	5242.98
u10_d2m	4830.12	5562.86	3037.87	4910.96	5660.59	5675.85	5297.86	2235.22	5265.6	2055.43	5618.16
u10_t2m	4248.82	5083.36	1627.88	4232.78	-787.99	5100.94	4821.61	1637.93	4237.9	1626.6	5083.27
u10_tcwv	4475.09	2001.59	2552.31	4534.76	5305.62	5323.06	4999.58	1753.38	4858.22	1582.37	5273.62
v10_tp	9526.1	4498.95	5015.21	8753.57	9872.01	9831.82	9861.71	4653.28	8016.88	4617.91	9872.91
v10_sst	3825.69	4961.43	2413.88	4546.51	4961.43	4955.35	4726.77	2330.9	4640.11	2295.56	4961.93
v10_sp	4132.23	596.25	1932.06	4318.39	5006.5	5032.5	4676.04	1823.49	4376.44	1796.53	5005.58
v10_d2m	4569.28	5392.74	2460.29	4775.59	5379.85	5405.46	5117.7	2149.66	4858.27	2092.85	5392.69
v10_t2m	4355.01	5176.65	1101.52	4328.01	5176.65	5201.54	4928.76	2011.2	3909.6	1969.35	5174.73
v10_tcwv	3731.06	484.12	975.54	4165.41	4673.76	4674.9	4568.73	1671.43	3695.06	1655.66	4673.65
tp_sst	5517.07	4351.13	5456.23	5535.03	6430.39	6767.73	4908.91	2663.69	5767.06	711.14	6430.16
tp_sp	15532.59	15787.44	11401.48	10430.17	-7988.1	16059.07	13267.97	5949.38	13745.19	1120.39	15787.16
tp_d2m	15367.23	16013.91	13333.57	12199.87	-2350.55	16219.41	13217.43	7363.11	14508.7	792.58	16013.1
tp_t2m	15287.74	15520.17	10035.36	9959.64	2507.33	15776.61	13471.65	5040.49	12876.23	1242.74	15508.93
tp_tcwv	10048.65	10903.97	5894.28	9154.21	10915.99	10917.71	10915.12	4533.76	8936.78	4265.23	10917.34
sst_sp	−1325.32	49.5	−533.79	350.22	49.5	22.17	-33.72	513.67	164.86	407.73	578.4
sst_d2m	−1345	50.76	−1083.49	302.86	50.73	22.95	-28.83	253.09	168.63	242.25	469.35
sst_t2m	−1203.35	162.73	−618.14	448.95	152.65	150.95	128.43	460.89	264.38	428.89	618.16
sst_tcwv	417.41	1791.89	626.53	2026.01	1791.88	1761.61	1705.82	1295.29	1915.3	1284.37	2078.17
sp_d2m	4586.21	4653.68	−1841.18	5989.32	4628.09	4820.51	4820.53	6154.26	1396.57	2264.35	5246.3
sp_t2m	6477.65	−567.87	−169.25	7643.56	6717.94	6866.83	6866.86	7055.27	3117.15	3747.81	7662.74
sp_tcwv	1936.88	2854.23	−962.18	3013.48	2882.62	2925.1	2925.1	1694.33	1773.39	1231.72	3024.19
d2m_t2m	4554.14	4680.8	−1491.79	5721.31	4682.9	4821.86	4821.89	5031.73	1714.52	2267.49	5699.35
d2m_tcwv	2156.32	2732.28	−1547.55	3246.14	2877.41	2992.38	2992.41	1674.31	1356.45	953.43	3228.64
t2m_tcwv	2370.72	3286.09	−354.74	3385.12	3343	3365.94	3365.95	1835.68	2346.46	1532.67	3412.36

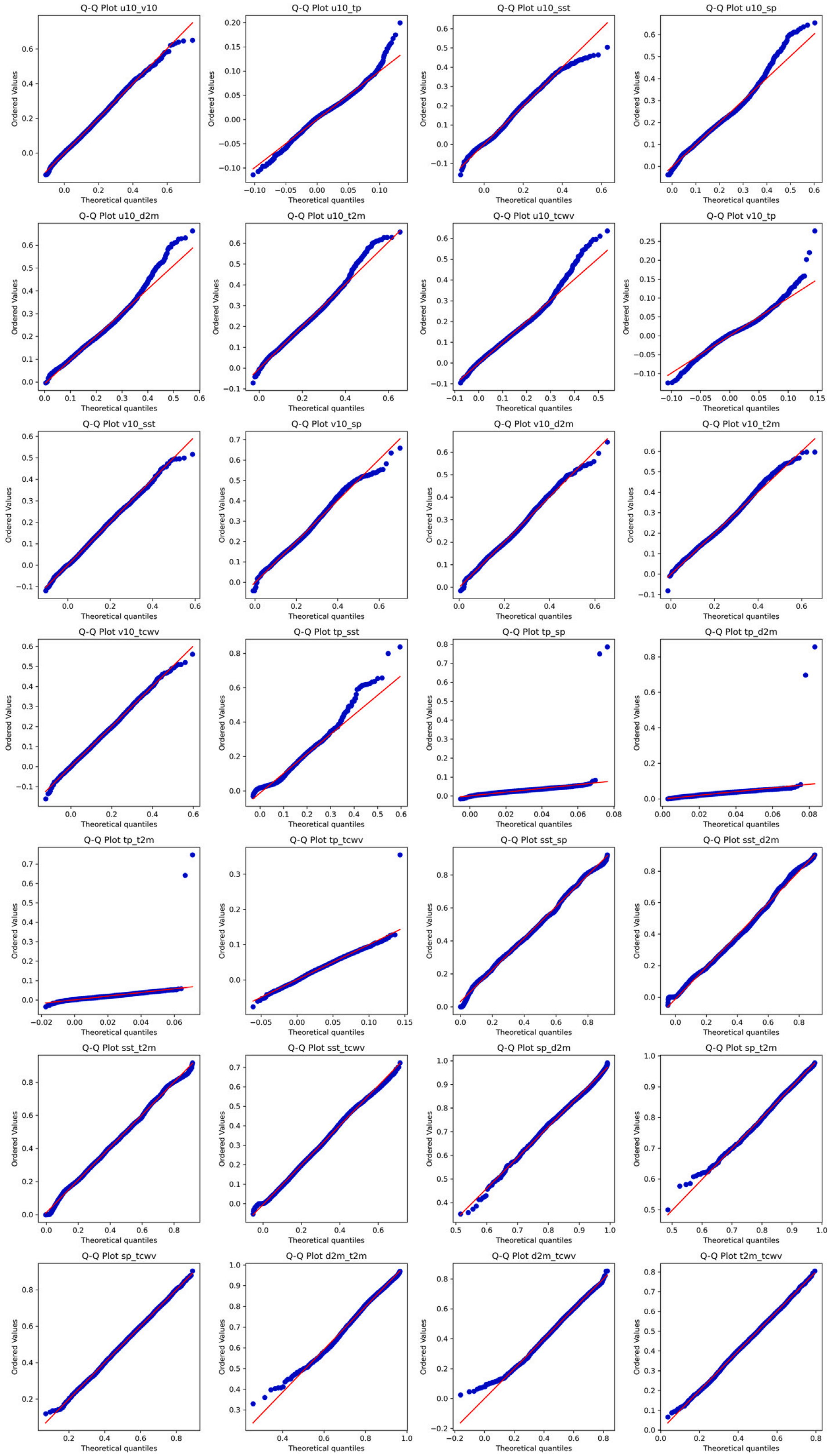

Fig. A.1. Q-Q plots comparing the empirical SSIM values with the theoretical Beta distribution across feature pairs.

Appendix B. Detailed computational complexity of the proposed SGA-Net

In this Section the computational complexity of the proposed model is constructed. The high-level architecture of the proposed SGA-Net is presented in Fig. B.2 (also Fig. 2 of the main manuscript), and it is visible that it consists of a max-pool branch, a min-pool branch, an avg-pool branch, two no-pool branches, several Spatial Attention (SA), Multi-Head Self-Attention (MHSA) in Global Feature Extractor (GFE) blocks, which are dominated by convolutional layers, and a Channel Attention (CA) block. The asymptotic complexity of each core module is discussed as follows.

Convolutional layer. For the first convolutional layer across every branch with input dimensions (H, W, C), kernel size K, and O_F number of output filters, the computational complexity is

$$\mathcal{O}\left(HW\,K^2\,C\,O_F\right). \tag{23}$$

For the successive convolutional operations, the input channel C is the same as the number of output filters of the previous layer. So, the computational complexity is

$$\mathcal{O}\left(HW\,K^2\,O_F^2\right). \tag{24}$$

Pooling layer. For a single pooling layer with input dimensions (H, W, C), the computational complexity is

$$\mathcal{O}(HWC). \tag{25}$$

SA Block. The spatial attention module involves two channel-wise pooling and a convolutional layer with 2 channel dimension, 1 filter size, and 7 kernel size, leading to a complexity of

$$\mathcal{O}(HWC + HWK^2). \tag{26}$$

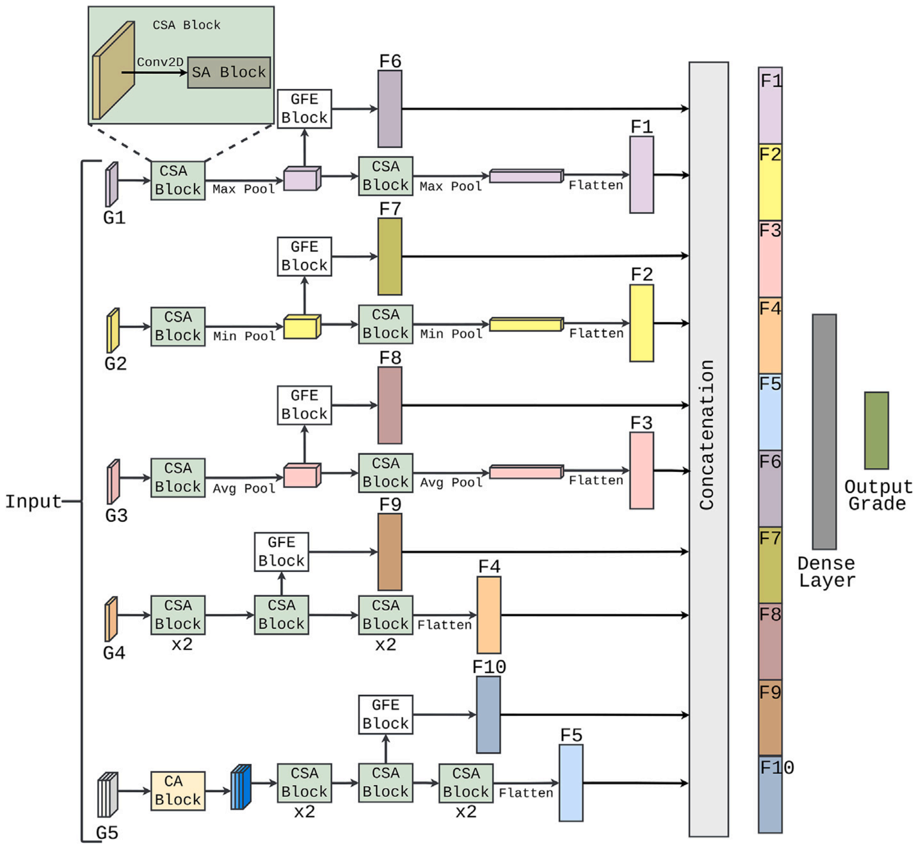

Fig. B.2. Proposed SGA-Net Model: The CSA Block, which comprises of 2D convolution and the SA Block, the CA Block, and the novel GFE Block, are its feature extractor components.

CA Block. The SE block performs global average pooling and two fully connected transformations. Thus, the complexity will be

$$\mathcal{O}(HWC + C^2/r), \tag{27}$$

where r is the reduction ratio.

MHSA Block. After reshaping the feature map into (HW, C), the self-attention module computes the query, key, and value projections followed by attention operations. The overall complexity is

$$\mathcal{O}(HW\,C^2) + \mathcal{O}\left((HW)^2 C\right). \tag{28}$$

Concatenation and dense layers. Feature concatenation across branches has a linear cost,

$$\mathcal{O}(d), \tag{29}$$

where d is the dimensionality of the concatenated feature vector. The dense layers require

$$\mathcal{O}(d \cdot 50), \tag{30}$$

Overall training complexity. Let N be the number of training samples, B be the batch size, and E be the number of epochs. The G_5 block contains five convolutional layers, one CA module, five SA modules, and one GFE block consisting of a MHSA module, and it is followed by convolution and pooling operations. Accordingly, the overall computational complexity is

$$\begin{aligned}\mathcal{O}\Bigg(\frac{N}{B}\,E\Big[&(HWK^2CO_F) + 5(HWK^2O_F^2) + 5(HWC + HWK^2)\\ &+ (HWC + C^2/r) + (H^2W^2C) + (HW\,C^2) + (HWC) + (50d)\Big]\Bigg)\end{aligned} \tag{31}$$

In conventional CNN architectures, convolutional kernel sizes and channel dimensions are typically small, and the number of neurons in the dense layers is generally treated as a constant with respect to the input resolution, making their contribution to the overall complexity negligible. Similarly, the reduction ratio r is a constant parameter used to reduce feature dimensionality. Consequently, the computational complexity is primarily governed by the input spatial dimensions H and W. From the derived Eq. 31, the dominant asymptotic term is $\left(H^2W^2C\right)$, originating from the Multi-Head Self-Attention (MHSA) operation within the GFE block. This term scales quadratically with spatial resolution, thereby overshadowing all remaining components, which scale linearly with spatial size (HW). Therefore, the overall model complexity is dominated by:

$$\mathcal{O}\left(H^2W^2C\right) \tag{32}$$

Appendix C. Class-wise density map visualization using Grad-CAM of Grade 1, 3, 4 and 6

This Section presents the remaining grade's Grad-CAM visualizations, which display the attention patterns of the baseline models and the proposed SGA-Net and are compared across multiple meteorological features. A detailed analysis and comparative discussion of the visualizations in this Appendix C and Section 5.6 of the manuscript is provided below. As stated in the main manuscript, each of these density maps has been annotated with (x) and the feature maps are annotated with (.) to indicate cyclone eye.

For wind speed (ws) analysis, [13] demonstrates the ability to extract meaningful spatial patterns for cyclone Grades 0 (main manuscript Fig. 15), 1, 2 (main manuscript Fig. 16), 6, and 7 (main manuscript Fig. 18); however, consistent tail and center shifts are observed across all grades, and the method fails to capture the relevant structures for Grades 3, 4, and 5 (main manuscript Fig. 17). In contrast, the approaches proposed by [11] and [12] exhibit complete failure across all grades, showing no reliable localization of wind-speed features. [3] either focuses on spatially irrelevant regions or entirely misses the features for lower grades (up to Grade 3). Its performance improves with increasing cyclone intensity, yielding more coherent pattern extraction for higher grades. The proposed method provides substantially more robust localization. The local branch accurately captures wind-speed patterns with no discernible tail shifts, particularly up to Grade 5, although minor center shifts are present. For Grades 6 and 7, the global branch exhibits superior focus, demonstrating that the hybrid local–global feature integration enables enhanced performance at higher cyclone intensities.

For total precipitation (tp), [13] performs poorly across all cyclone grades, failing to extract any meaningful spatial features. In contrast, the methods of [11] and [12] demonstrate comparatively stronger capability, capturing precipitation patterns to a reasonable extent despite the intrinsic low contrast of the data. The approach by [3] achieves only partial success, with reliable localization limited to Grades 1 and 3. In the proposed method, the local branch does not accurately capture precipitation structures for Grades 0–3, primarily because the cyclone eye lacks a distinct spatial signature at these early intensities; however, the global branch is able to identify broader spatial relationships to a certain degree. For these lower grades, [11] and [12] perform comparatively better. As cyclone intensity increases and the eye becomes more clearly defined (Grade 4 onward), the proposed method demonstrates precise localization of the eye region and consistent focus on relevant precipitation patterns, even under low-contrast conditions. In contrast, the focus of [11] and [12] becomes increasingly diffuse at higher intensities, indicating less reliable feature localization.

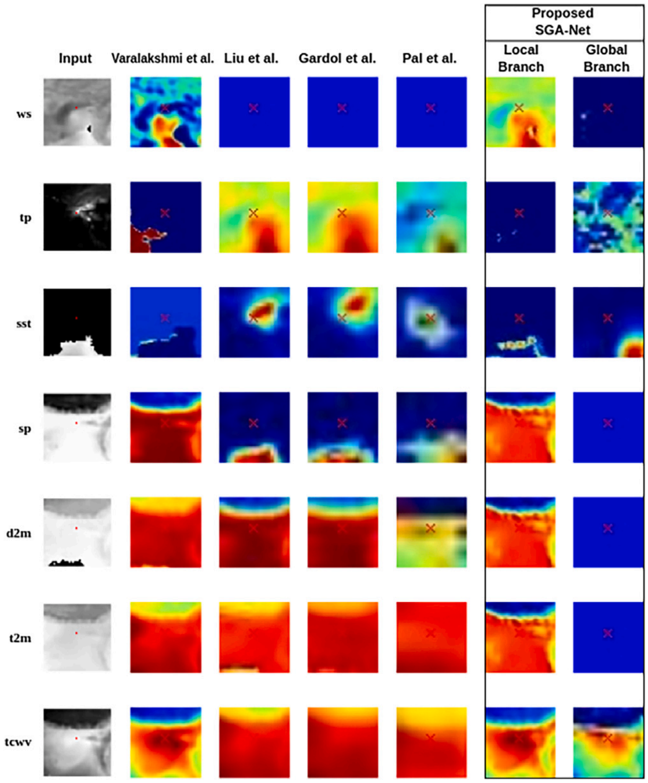

Fig. C.3. Visualization of Grade 1 attention patterns of SGA-Net and SOTA using GradCAM.

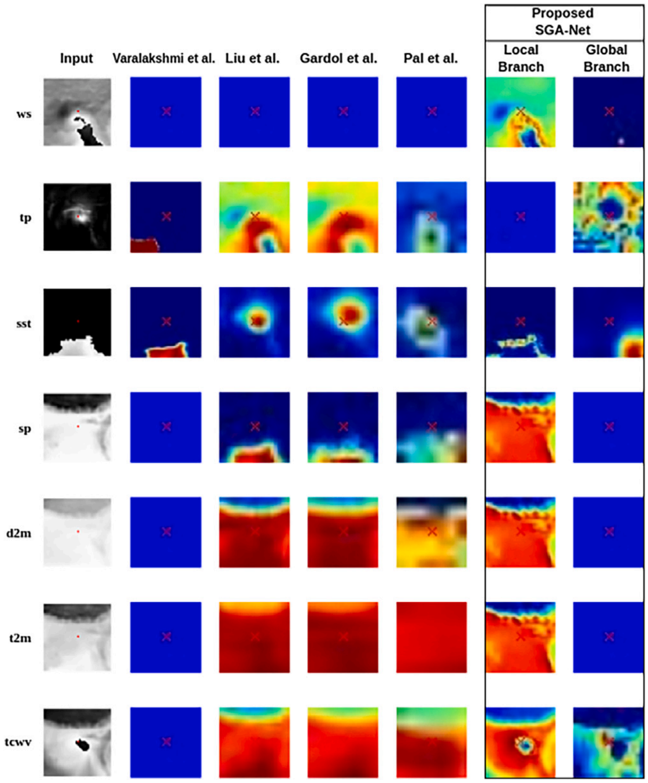

Fig. C.4. Visualization of Grade 3 attention patterns of SGA-Net and SOTA using GradCAM.

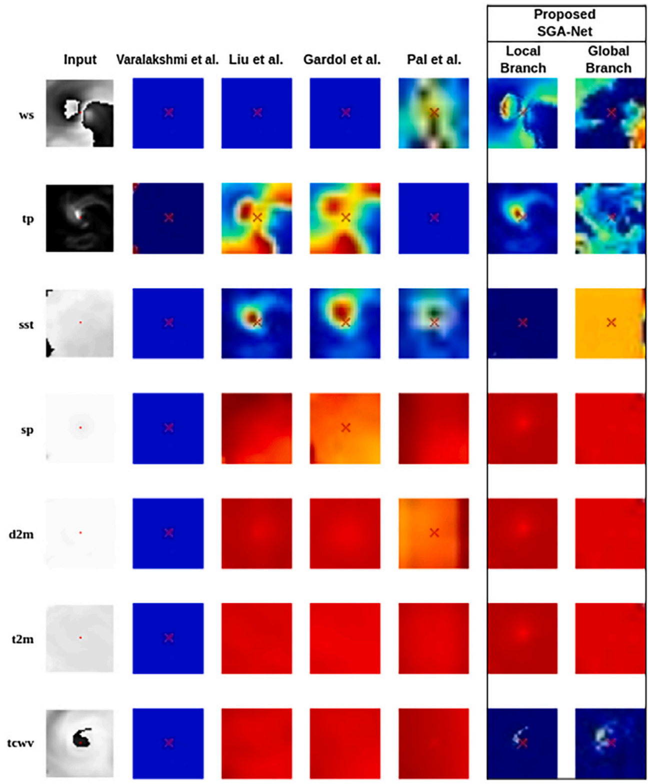

Fig. C.5. Visualization of Grade 4 attention patterns of SGA-Net and SOTA using GradCAM.

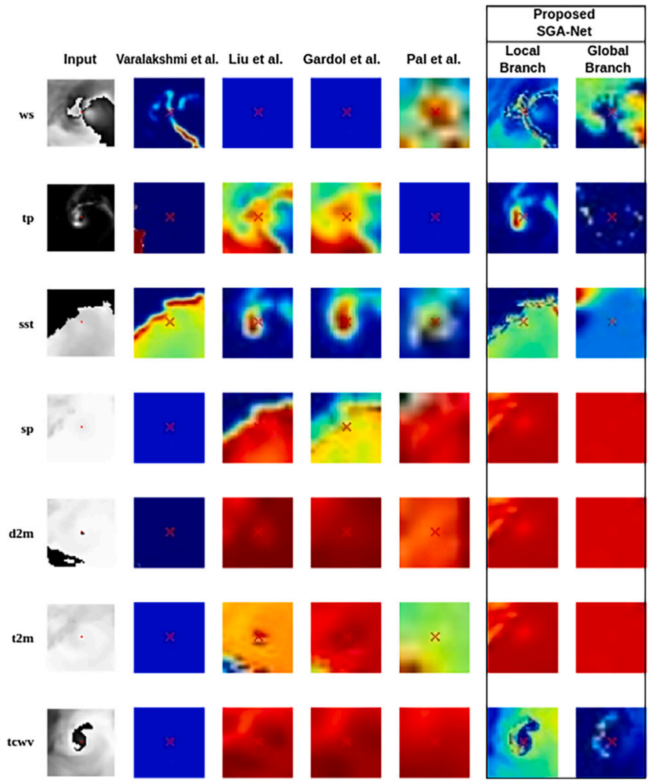

Fig. C.6. Visualization of Grade 6 attention patterns of SGA-Net and SOTA using GradCAM.

For sea surface temperature (sst), the proposed model assigns meaningful attention to relevant regions primarily at higher cyclone grades through its local branch. In contrast, the focus produced by [11], [12], and [3] is largely misaligned, failing to highlight physically relevant areas. [13] exhibits comparatively better performance in a few isolated instances but lacks consistency across grades. It is important to note that for SST, the spatial domain contains only oceanic information, with land regions masked. Consequently, none of the methods—including the proposed model—can accurately pinpoint the cyclone eye location based solely on SST-derived features.

For sea pressure (sp), dew point temperature at 2 m ($d2m$), and air temperature at 2 m ($t2m$), the proposed model produces a unified density map, reflecting the shared feature extractor used for these variables. Visualizations from the local branch consistently reveal a slightly elevated intensity at the cyclone center across all grades, corresponding to the region occupied by the cyclone eye. In contrast, existing SOTA approaches employ separate frameworks for each variable; however, in every case, the proposed model's local branch yields more coherent and physically meaningful density maps.

For total columnar water vapor ($tcwv$), the methods proposed by [11], [12], and [3] fail to capture the relevant spatial structures across all cyclone grades. [13] shows comparatively better performance at lower grades but does not generalize effectively to higher intensities. In contrast, the proposed model consistently identifies the water vapor distribution for Grades 0–3 through both its local and global branches. For Grades 4 and 5, both branches predominantly emphasize the cyclone center, while for Grades above 5, the global branch maintains precise focus on the eye region, and the local branch captures the broader structural patterns surrounding the core.

The Grad-CAM visualizations provide interpretability across multiple meteorological variables and demonstrate that the proposed SGA-Net achieves superior feature localization and physically meaningful attention distributions compared with existing SOTA methods in most scenarios. The baseline models frequently exhibit diffuse or inconsistent attention, often failing to concentrate on the critical cyclonic regions. In contrast, SGA-Net's local–global fusion architecture yields a more structured and interpretable attention mechanism. This complementary focus enables the model to integrate localized intensity information with broader contextual cues, resulting in a more comprehensive and physically coherent representation of cyclone dynamics.

Appendix D. Theoretical ground behind setting the SSIM threshold of 0.8 for grouping features

The threshold in this study was chosen based on empirical distribution characteristics of inter-feature structural similarity rather than an arbitrary assumption. The choice of SSIM threshold was made from pairwise SSIM values, which are shown in Table D.2. It demonstrates that a majority of these pairs have low levels of similarity in structure (SSIM < 0.4), suggesting that these are largely independent structures in space. Only an exceedingly rare number of feature pairs have consistently higher levels of SSIM measure (> 0.8), which indicates greater structural similarity between features, whereas lower values reflect weak or negligible similarity [44]. This naturally motivates the choice of a higher threshold, i.e., 0.8 a conservative threshold to identify structurally similar features.

Table D.2
Mean SSIM values of every feature pair.

Feature_Pair	**u10_v10**	**u10_tp**	**u10_sst**	**u10_sp**	**u10_d2m**	**u10_t2m**	**u10_tcwv**
SSIM	0.1469	0.0181	0.1057	0.1906	0.1985	0.196	0.127
Feature_Pair	**v10_tp**	**v10_sst**	**v10_sp**	**v10_d2m**	**v10_t2m**	**v10_tcwv**	**tp_sst**
SSIM	0.0124	0.1091	0.209	0.2097	0.2145	0.143	0.0936
Feature_Pair	**tp_sp**	**tp_d2m**	**tp_t2m**	**tp_tcwv**	**sst_sp**	**sst_d2m**	**sst_t2m**
SSIM	0.0238	0.0268	0.0155	0.0385	0.4088	0.4052	0.4143
Feature_Pair	**sst_tcwv**	**sp_d2m**	**sp_t2m**	**sp_tcwv**	**d2m_t2m**	**d2m_tcwv**	**t2m_tcwv**
SSIM	0.2719	**0.8805**	**0.8804**	0.5647	**0.8233**	0.5242	0.4607

Data availability

The authors do not have permission to share data.

References

[1] W.M. Gray, Tropical Cyclone Genesis, (Ph.D. thesis), Colorado State University. Libraries, 1975.
[2] J.C.L. Chan, The Physics of tropical cyclone motion, Annu. Rev. Fluid Mech. 37 (1) (2005) 99–128.
[3] S. Pal, U. Das, O. Bandyopadhyay, SSN: a novel CNN-based architecture for classification of tropical cyclone images from INSAT-3d, IEEE Trans. Geosci. Remote Sens. 62 (2024) 1–8.
[4] S. Rehman, M. Sahana, P. Kumar, R. Ahmed, H. Sajjad, Assessing hazards induced vulnerability in coastal districts of India using site-specific indicators: an integrated approach, GeoJournal 86 (5) (2021) 2245–2266.
[5] M. Mondal, S. Haldar, A. Biswas, S. Mandal, S. Bhattacharya, S. Paul, Modeling cyclone-induced multi-hazard risk assessment using analytical hierarchical processing and GIS for coastal West Bengal, India, Reg. Stud. Mar. Sci. 44 (2021) 101779.
[6] V.F. Dvorak, Tropical Cyclone Intensity Analysis Using Satellite Data, vol. 11, US Department of Commerce, National Oceanic and Atmospheric Administration, 1984.
[7] M.F. Piñeros, E.A. Ritchie, J.S. Tyo, Objective measures of tropical cyclone structure and intensity change from remotely sensed infrared image data, IEEE Trans. Geosci. Remote Sens. 46 (11) (2008) 3574–3580.
[8] D.C. Herndon, C. Velden, K. Brueske, R. Wacker, B. Kabat, Upgrades to the UW-CIMSS AMSU-based tropical cyclone intensity estimation algorithm, 2004.

[9] A. Ghosh, Data Science and Cases in Sustainability: Pattern Recognition and Machine Learning, Mathematics for Sustainable Developments, Springer Nature Singapore, 2026, https://books.google.co.in/books?id=_Ghf0QEACAAJ.
[10] B.-F. Chen, Y.-T. Kuo, T.-S. Huang, A deep learning ensemble approach for predicting tropical cyclone rapid intensification, Atmos. Sci. Lett. 24 (5) (2023) e1151.
[11] Y. Liu, E. Racah, J. Correa, A. Khosrowshahi, D. Lavers, K. Kunkel, M. Wehner, W. Collins, et al., Application of deep convolutional neural networks for detecting extreme weather in climate datasets, arXiv preprint arXiv:1605.01156, 2016.
[12] S. Gardoll, O. Boucher, Classification of tropical cyclone containing images using a convolutional neural network: performance and sensitivity to the learning dataset, Geosci. Model Dev. 15 (18) (2022) 7051–7073.
[13] P. Varalakshmi, N. Vasumathi, R. Venkatesan, Tropical cyclone prediction based on multi-model fusion across Indian coastal region, Prog. Oceanogr. 193 (2021) 102557.
[14] T.-H. Do, S.-T. Phan, D.-T. Du, D.-Q. Dang, K.-H. Mai, L.R. Hole, Multi-heat keypoint incorporation in deep learning model to tropical cyclone centering and intensity classifying from geostationary satellite images, npj Clim. Atmos. Sci. 15 (2025).
[15] Z. Zong, G. Song, Y. Liu, Detrs with collaborative hybrid assignments training, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 6748–6758.
[16] Z. Zhang, X. Yang, L. Shi, B. Wang, Z. Du, F. Zhang, R. Liu, A neural network framework for fine-grained tropical cyclone intensity prediction, Knowl.-Based Syst. 241 (2022) 108195.
[17] G. Xu, Y. Li, C. Ma, X. Li, Y. Ye, Q. Lin, Z. Huang, S. Chen, TFG-net: tropical cyclone intensity estimation from a fine-grained perspective with the graph convolution neural network, Eng. Appl. Artif. Intell. 118 (2023) 105673.
[18] Y. Tian, W. Zhou, P.K.Y. Cheung, Z. Liu, Vision transformer for extracting tropical cyclone intensity from satellite images, Adv. Atmos. Sci. 42 (1) (2025) 79–93.
[19] W. Tian, X. Li, J. Fan, H. Zhao, RTCC-net: tropical cyclone generation classification model based on multi-source information fusion, Sci. Rep. 16 (2026) 1.
[20] Z. Zhang, X. Yang, X. Wang, B. Wang, C. Wang, Z. Du, A neural network with spatiotemporal encoding module for tropical cyclone intensity estimation from infrared satellite image, Knowl.-Based Syst. 258 (2022) 110005.
[21] Y. Wang, J. Wang, Y. Chang, K. Cai, S. Li, Y. Dong, Graph-theoretical investigation of trajectory dynamics and size characteristics in tropical cyclones, Nat. Hazards (2025) 1–18.
[22] K.R. Knapp, M.C. Kruk, D.H. Levinson, H.J. Diamond, C.J. Neumann, The International best track archive for Climate Stewardship (IBTrACS) unifying tropical cyclone data, Bull. Am. Meteorol. Soc. 91 (3) (2010) 363–376.
[23] H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, et al., The ERA5 global reanalysis, Q. J. R. Meteorol. Soc. 146 (730) (2020) 1999–2049.
[24] M. Ranzato, Y.-L. Boureau, Y. Cun, et al., Sparse feature learning for deep belief networks, Adv. Neural Inf. Process. Syst. 20 (2007).
[25] Y. LeCun, B. Boser, J. Denker, D. Henderson, R. Howard, W. Hubbard, L. Jackel, Handwritten digit recognition with a back-propagation network, Adv. Neural Inf. Process. Syst. 2 (1989).
[26] N.-I. Galanis, P. Vafiadis, K.-G. Mirzaev, G.A. Papakostas, Convolutional neural networks: a roundup and benchmark of their pooling layer variants, Algorithms 15 (11) (2022) 391.
[27] D. Yu, H. Wang, P. Chen, Z. Wei, Mixed pooling for convolutional neural networks, in: International Conference on Rough Sets and Knowledge Technology, Springer, 2014, pp. 364–375.
[28] K. He, X. Zhang, S. Ren, J. Sun, Spatial pyramid pooling in deep convolutional networks for visual recognition, IEEE Trans. Pattern Anal. Mach. Intell. 37 (9) (2015) 1904–1916.
[29] P. Sermanet, S. Chintala, Y. LeCun, Convolutional neural networks applied to house numbers digit classification, in: Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), IEEE, 2012, pp. 3288–3291.
[30] L. Dong, L. He, M. Mao, G. Kong, X. Wu, Q. Zhang, X. Cao, E. Izquierdo, CUNet: a compact unsupervised network for image classification, IEEE Trans. Multimedia 20 (8) (2017) 2012–2021.
[31] M. Sun, Z. Song, X. Jiang, J. Pan, Y. Pang, Learning pooling for convolutional neural network, Neurocomputing 224 (2017) 96–104.
[32] C. Ozdemir, Y. Dogan, Y. Kaya, A new local pooling approach for convolutional neural network: local binary pattern, Multimed. Tools Appl. 83 (12) (2024) 34137–34151.
[33] J.-Y. Zhou, C.-H. Yang, X.-L. Wang, S.-Y. Cao, A soft sensor modeling framework embedded with domain knowledge based on spatio-temporal deep LSTM for process industry, Eng. Appl. Artif. Intell. 126 (2023) 106847.
[34] M.-H. Guo, T.-X. Xu, J.-J. Liu, Z.-N. Liu, P.-T. Jiang, T.-J. Mu, S.-H. Zhang, R.R. Martin, M.-M. Cheng, S.-M. Hu, Attention mechanisms in computer vision: a survey, Comput. Vis. Media 8 (3) (2022) 331–368.
[35] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.
[36] Z. Gao, J. Xie, Q. Wang, P. Li, Global second-order pooling convolutional networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 3024–3033.
[37] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, Q. Hu, ECA-net: efficient channel attention for deep convolutional neural networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 11534–11542.
[38] Z. Qin, P. Zhang, F. Wu, X. Li, FcaNet: frequency channel attention networks, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 783–792.
[39] G. Strang, The discrete cosine transform, SIAM Rev. 41 (1) (1999) 135–147.
[40] S. Woo, J. Park, J.-Y. Lee, I.S. Kweon, CBAM: convolutional block attention module, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 3–19.
[41] M. Jaderberg, K. Simonyan, A. Zisserman, et al., Spatial transformer networks, Adv. Neural Inf. Process. Syst. 28 (2015).
[42] H. Yin, A. Vahdat, J.M. Alvarez, A. Mallya, J. Kautz, P. Molchanov, A-ViT: adaptive tokens for efficient vision transformer, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 10809–10818.
[43] A.M. Reza, Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement, J. VLSI Signal Process. Syst. Signal Image Video Technol. 38 (1) (2004) 35–44.
[44] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.
[45] A.K. Gupta, S. Nadarajah, Handbook of Beta Distribution and Its Applications, CRC Press, 2004.
[46] C. Villani, The Wasserstein distances, in: Optimal Transport: Old and New, Springer, 2009, pp. 93–111.
[47] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: visual explanations from deep networks via gradient-based localization, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2017, pp. 618–626.
[48] W. Duan, J. Yuan, X. Duan, D. Feng, Seasonal variation of tropical cyclone genesis and the related large-scale environments: comparison between the Bay of Bengal and Arabian Sea sub-basins, Atmosphere 12 (12) (2021) 1593.
[49] X.-T. Fan, Y. Li, A.M. Lyu, L.-S. Liu, Statistical and comparative analysis of tropical cyclone activity over the Arabian Sea and Bay of Bengal (1977–2018), J. Trop. Meteorol. 26 (3) (2020) 441–452.
[50] S. Koley, Augmenting efficacy of global climate model forecasts: machine learning appraisal of remote sensing data, Int. J. Eng. Trends Technol. 72 (6) (2024) 442–502.