frontiers in Environmental Science

TYPE Original Research
PUBLISHED 04 May 2026
DOI 10.3389/fenvs.2026.1797545

OPEN ACCESS

EDITED BY
Nir Krakauer,
City College of New York (CUNY),
United States

REVIEWED BY
Amit Pimpalkar,
Ramdeobaba University, India
Peerapol Khunarsa,
Uttaradit Rajabhat University, Thailand

*CORRESPONDENCE
Ateeq Ur Rehman,
202411144@gachon.ac.kr

RECEIVED 27 January 2026
REVISED 01 March 2026
ACCEPTED 16 March 2026
PUBLISHED 04 May 2026

CITATION
Shandilya G, Gupta S, Saudagar AKJ, Ikram S, Rehman AU, De la Torre Díez I, Mohamed HG, Casanova RP, Castilla ÁK and Kaur U (2026) Enhanced weather classification using xception with SENet and attention mechanisms.
Front. Environ. Sci. 14:1797545.
doi: 10.3389/fenvs.2026.1797545

Enhanced weather classification using xception with SENet and attention mechanisms

Gunjan Shandilya[1], Sheifali Gupta[1], Abdul Khader Jilani Saudagar[2], Sunnia Ikram[3], Ateeq Ur Rehman[4]*, Isabel De la Torre Díez[5], Heba G. Mohamed[6], Ramón Pali Casanova[7,8,9], Ángel Kuc Castilla[7,8,10,11] and Upinder Kaur[12]

[1]Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab, India, [2]Information Systems Department, College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia, [3]Department of Software Engineering, The Islamia University Bahawalpur, Bahawalpur, Pakistan, [4]School of Computing, Gachon University, Seongnam-si, Republic of Korea, [5]Department of Signal Theory and Communications and Telematics Engineering, University of Valladolid, Valladolid, Spain, [6]Department of Electrical Engineering, College of Engineering, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia, [7]Universidad Europea del Atlántico, Santander, Spain, [8]Department of Projects, Universidad Internacional Iberoamericana, Arecibo, PR, United States, [9]Department of Projects, Universidade Internacional do Cuanza, Cuito, Bié, Angola, [10]Department of Projects, Fundación Universitaria Internacional de Colombia, Bogotá, Colombia, [11]Department of Projects, Universidad de La Romana, La Romana, Dominican Republic, [12]School of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab, India

Introduction: Weather classification plays a crucial role in applications such as environmental monitoring, disaster management, and smart city infrastructure. Accurate and efficient classification of weather conditions from images remains a challenging task due to variations in illumination, texture, and atmospheric conditions.

Methods: This study proposes an efficient deep learning framework for multi-class weather classification by integrating the Xception architecture with Squeeze-and-Excitation (SE) blocks and a spatial attention mechanism. Transfer learning with pre-trained ImageNet weights was employed, and a comparative analysis was conducted using EfficientNet-B3, ResNet152V2, and Xception architectures. The proposed enhanced Xception model incorporates channel-wise recalibration and spatial feature refinement to improve representational capability. The model was trained and evaluated on the Multi-Class Weather Dataset (MWD), which consists of 1,125 images categorized into four classes: sunshine, cloudy, rain, and sunrise. To ensure robustness and generalization, 5-fold cross-validation, statistical significance testing, calibration analysis, and robustness evaluation under image perturbations were performed.

Results: The proposed model achieved a classification accuracy of 99.06% on the test set. Additionally, it attained a macro precision of 98.3%, macro recall of 97.7%, and macro F1-score of 98.0%. The model demonstrated strong generalization capability and robustness under varying perturbation conditions, with only moderate computational overhead.

Discussion: The integration of SE blocks and spatial attention significantly enhances feature representation by emphasizing informative channels and spatial regions. Compared to baseline architectures, the proposed framework

shows superior performance in terms of accuracy and robustness. These results indicate that the model is well-suited for real-world weather classification applications, particularly in intelligent environmental monitoring systems.

KEYWORDS

attention mechanisms, cloudy, deep learning (DL), image classification, rain, squeeze and excitation (SE), sunrise, sunshine

1 Introduction

Weather classification at exact levels plays an essential role in industries such as disaster relief and aviation together with urban development and agriculture. Truthful and prompt weather predictions decrease possible hazards and help organizations manage resources optimally while improving their decision processes (Schultz et al., 2019). Historical weather forecasting has used numerical models for analyzing meteorological factors that include pressure and temperature with humidity and wind speed measurements. The advancement of computer vision technology and DL methods introduced image-based weather categorization as a new research field. Visual data obtained from distant locations and urban areas has the potential to generate real-time weather monitoring information through this emerging research field according to (Gdeeb, 2024; Goswami, 2020). CNNs have proven to be an effective DL method which shows excellent results when performing image classifications in recent years. The development of DL technology has created rising interest for implementing DL solutions to classify weather through image analysis. One key challenge in weather classification with images originates from the visual aspects that appear identically across different weather conditions (Nalluri et al., 2019). Repeated grey skies frequently occur both during cloudy conditions and rain while bright light appears whenever the sun rises or sets. Weather classification algorithms need to detect small visual cues when analyzing weather types since these conditions share identical visual elements.

Transfer learning (TL) serves as the method employed to handle these limitations. It allows pre-trained ImageNet models to adapt their functionality for analyzing limited datasets such as weather classifications (Kalkan et al., 2022). The employment of pre-trained models leads to speeding up learning processes and enhancing photo weather identification accuracy rates. The main priority of this research project involves classifying four distinct weather patterns, including cloudy weather and rain and sunshine, along with morning sun.

Using SENet with the most successful TL model brings about additional gains in classification precision. The feature extraction process receives optimization through SENet because these models incorporate attention mechanisms which help the model to concentrate on images' most essential parts (Ghaleb et al., 2022; Gladh and Sahlin, 2021). The model learns essential picture elements through this attention process which makes it effective for class distinction under natural conditions. This work aims to show that attention methods such as SENet and transfer learning may be used together to greatly increase the robustness and accuracy of weather classification models. The work advances the field of deep learning-based weather prediction systems, which may be used in real-time for smart city infrastructure, environmental monitoring, and disaster management. It does this by concentrating on image-based classification. The main contributions made by this study are.

- The study conducts an extensive comparative evaluation among three prominent transfer learning models, Xception, EfficientNetB3, and ResNet152V2, identifying Xception as the most effective baseline model for weather classification based on superior performance metrics.
- For enhancing the capability for feature extraction, the selected Xception model is augmented with SE blocks, enabling the model to dynamically recalibrate channel-wise feature responses and significantly improve classification accuracy by emphasizing critical weather-related features.
- The proposed framework further integrates an attention mechanism, which selectively highlights relevant spatial regions in images. This strategic integration improves the model's robustness, effectively distinguishing visually similar weather conditions, particularly in scenarios with subtle or challenging patterns.
- The study demonstrates the significance of combining the Xception architecture, Squeeze-and-Excitation (SE) blocks, and the attention mechanism, creating a unified, robust framework. This synergistic integration capitalizes on each technique's strengths, substantially improving the model's capability to accurately classify weather conditions, even with limited datasets and subtle visual differences among classes.

The identification of the weather based on image analysis faces challenges due to the overlapping images of various kinds of weather, including the differentiation between rain and clouds, and the differentiation between sunrise and natural exposure to the sun. The unclear patterns in the images regarding the weather lead to prediction errors, which affect important government decisions regarding farming and environmental monitoring, and disaster management preparedness. This study presents an innovative approach that combines the use of Xception, SENet, and attention mechanisms, since these components provide optimized localization and feature representation to address the challenges. The research applies a transfer learning method for enhancing model performance while working on a limited number of weather images in order to achieve accurate classification results.

The paper follows a structured plan that includes a thorough review of recent findings in weather image classification through Section 2. Section 3 details the research design by presenting an illustration of the model structures along with data preprocessing methods and the combination of SENet and attention components. A performance evaluation with existing approaches accompanies

TABLE 1 Comparative analysis of existing models in literature with their drawbacks.

Ref./Year	Model used	Dataset used	Classes	Performance%	Drawbacks
(Al-Haija et al., 2022)/ 2022	ResNet50	RFS (3,300 images: rain, fog, snow)	3	Accuracy - 80.7	Moderate accuracy, cannot generalize well to more weather types
(Schultz et al., 2021)/ 2022	ResNet50	DAWM 2020 + MCWCD 2018 (1,656 images)	6	Accuracy-98.48, Precision-98.51, Recall-98.41	No feature enhancement modules like SE or attention, Limited scalability
(Ranjith Ku et al., 2024)/2024	Multi-Class SVM + Grid Search	GitHub weather samples (400)	2 (snowfall types)	Improved from 83 to 91	Poor feature learning, small dataset
(Li et al., 2023)/2023	Vision Transformer (ViT) + MLP	Segmented weather dataset	14	Recognition rate-92.83	High training cost; needs large datasets; overfitting risk on small data
(Mittal and Sangwan, 2023)/2023	InceptionV3 + Logistic Regression	Multi-class weather dataset (1,125)	4	Accuracy-97.77	Simple decision layer, it lacks deep semantic understanding
(Gupta and Goel, 2023)/2023	DenseNet-161 + Logistic Regression	MWD (Multi-class Weather Dataset)	4	Accuracy-98.2, F1-99.10	High model complexity, lacks explainability
(Samo et al., 2023)/ 2023	ViT + Focal Loss	Road Weather Dataset (3,560 images)	8	Accuracy-92, F1-81.22	Data imbalance not fully solved; performance degrades on subtle weather cues
(Jovanovic et al., 2024)/2024	XGBoost (for solar events), RNN (sunspots) + PSO tuning	WEAPD Dataset	11	XGBoost-98.1, RNN-0.84	Not suitable for image classification, it is focused on time series predictions
(Papadimitriou et al., 2023)/2023	CNN	Weather image dataset	4	Accuracy-98	Limited capability in handling overlapping weather patterns
(Tian and Kim, 2023)/2023	Faster R-CNN + FPN + ResNet101	Foggy vehicle image dataset	-	Detection Accuracy-96.83	Designed for object detection in fog, it is not optimized for weather condition classification

the experimental results Section 4. Section 5 ends the research by offering future research guidelines.

2 Literature review

Several studies have been conducted by employing machine learning techniques to keep a close check on the weather image classification to prevent any natural disaster. In their study, authors of (Al-Haija et al., 2022) employed various DL models, out of which ResNet50 has outperformed other models with an accuracy of 80.7%. They developed their customized RFS dataset of 3,300 images, these images were a combination of rain, fog, and snow images. The authors of (Schultz et al., 2021) employed ResNet50 to classify the weather images into six categories by using two datasets DAWM 2020 + MCWCD2018 which have 1,656 images. They attained an accuracy of 98.48%, precision of 98.51% and recall of 98.41%.

This study of (Sun et al., 2024) assesses the effect of weather conditions on the severity of driver injuries using a novel technique termed the Multi-Class-Weather phenomenon, which integrates ResNet9, multinomial logic, and self-learning methodologies. It makes use of deep supervised learning techniques to investigate the relationship between driving injuries and the weather. The findings indicate a clear connection between fog conditions and more severe injuries. The model demonstrates capability to reduce overfitting and simultaneously reaches 87% accuracy performance. The research by Ranjith Ku et al. (2024) investigates the value of snowfall predictions for the agricultural sector and aviation industry and climate organization reporting. The method brings fresh analysis to snowfall classification through study of three meteorological elements including precipitation and wind speed together with temperature measurement. Performance evaluation takes place by removing 400 items from GitHub before dividing them into training and testing datasets with training chairs consisting of 70% of items and testing positions containing 30% of items. When it came to classifying different types of snowfall according to variables including wind speed, gust, minimum and maximum temperature, and precipitation, the model's initial accuracy was 83%. The accuracy of the model increased to 91% when the SVM configuration was optimized using hyper-parameter tuning methods like grid search. This research (Li et al., 2023) adopts a multi-layer transformer method, which uses self-attention features to determine similarity between separated image segments. An MLP classifier performs classification duties on the newly created weighted positions. Research findings demonstrate that the Vision Transformers (ViT)-based DL algorithm enhances prediction accuracy to 92.83% while dealing with 14 separate weather conditions. To optimize performance in less time, the research by Mittal and Sangwan (2023) suggests a transfer learning-based system for categorizing weather photos utilizing characteristics acquired from pre-trained deep CNN models. Larger, better-quality data sets enhance model accuracy, and the Spark platform allows for the scalability of this framework when working with larger datasets. An amalgamation of the InceptionV3 framework and a Logistic Regression classifier yields the highest results, with a maximum accuracy of 97.77%, according to extensive trials conducted on a meteorological image dataset. The

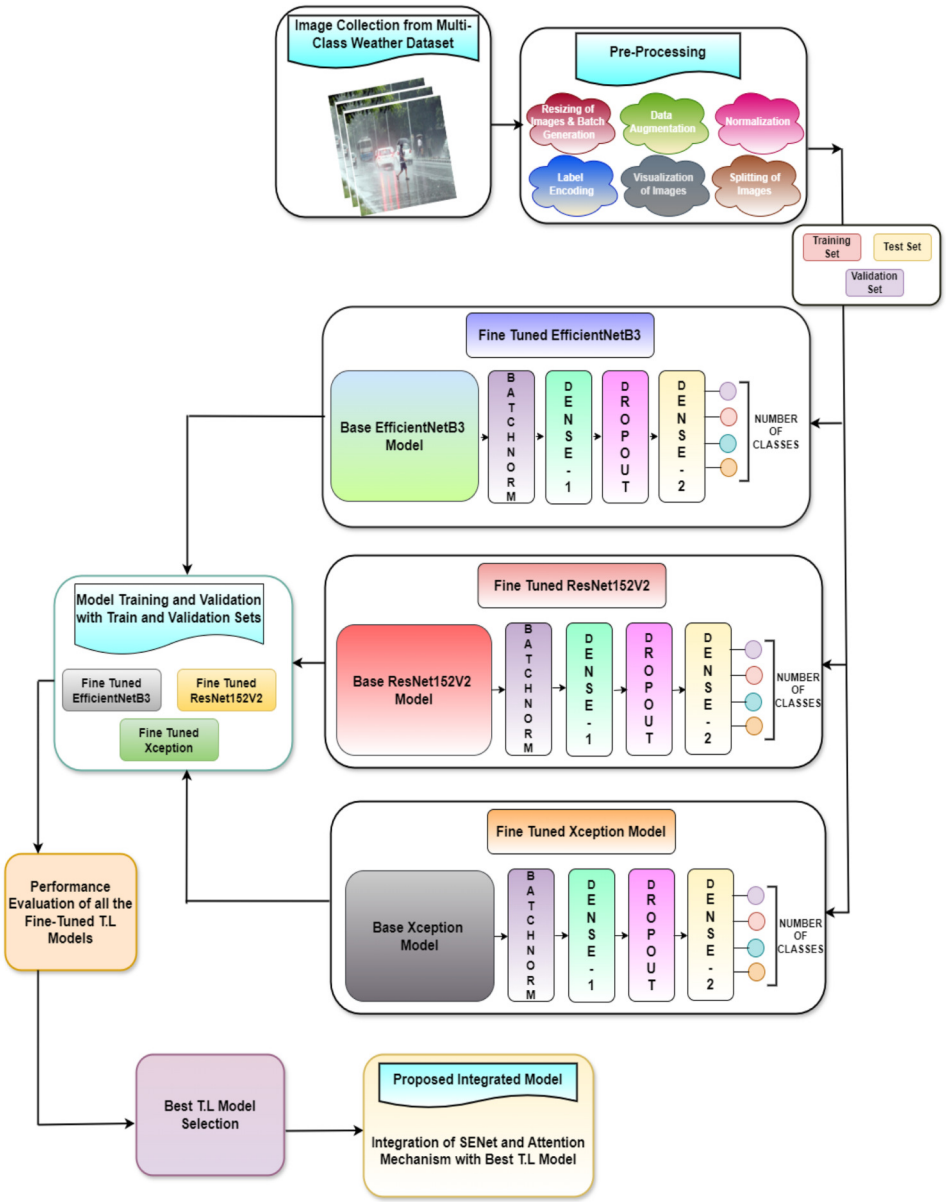

FIGURE 1
Proposed methodology.

study by (Gupta and Goel, 2023) investigates the application of TL and DL to the classification of various weather kinds. It makes use of the four weather categories found in the Multi-class Weather Dataset (MWD). To deal with the problem of low-labeled data, the study extracts detailed features from the photos and fine-tunes the models using pre-trained models. The classifier for the weather categories is logistic regression. With an astounding F1 score of 99.10%, flawless precision, and 98.2% accuracy, DenseNet-161 outperformed the other models. The authors of (Sharma et al., 2023) have designed a DL model for identifying and categorizing meteorological situations. It mixes CNNs with SVMs. Ten thousand pics representing five different weather environments were used to test the model. It outperformed previous models in multi-class classification with a strong total accuracy of 97.24%. A thorough examination of the variables influencing the model's performance is also included in the study. To solve data imbalances and concentrate on hard-to-learn weather situations, the research of Samo et al. (2023) investigates the use of focused loss. Additionally, it looks into how modern vision transformers use an attention mechanism to modify dynamic weight pixel-by-pixel. Vision transformers achieve 92% validation accuracy and an F1-score of 81.22%, outperforming CNNs despite the imbalance in the dataset. They altered a dataset of 3,560 photos for eight distinct classifications.

In the study of Jovanovic et al. (2024), the authors evaluated the performance of XGBoost in the classification of solar events and RNNs in the prediction of sunspots with the inclusion of the modified particle swarm optimization algorithm in the process. The RNN model obtained the best R^2 value of 0.840,448, while

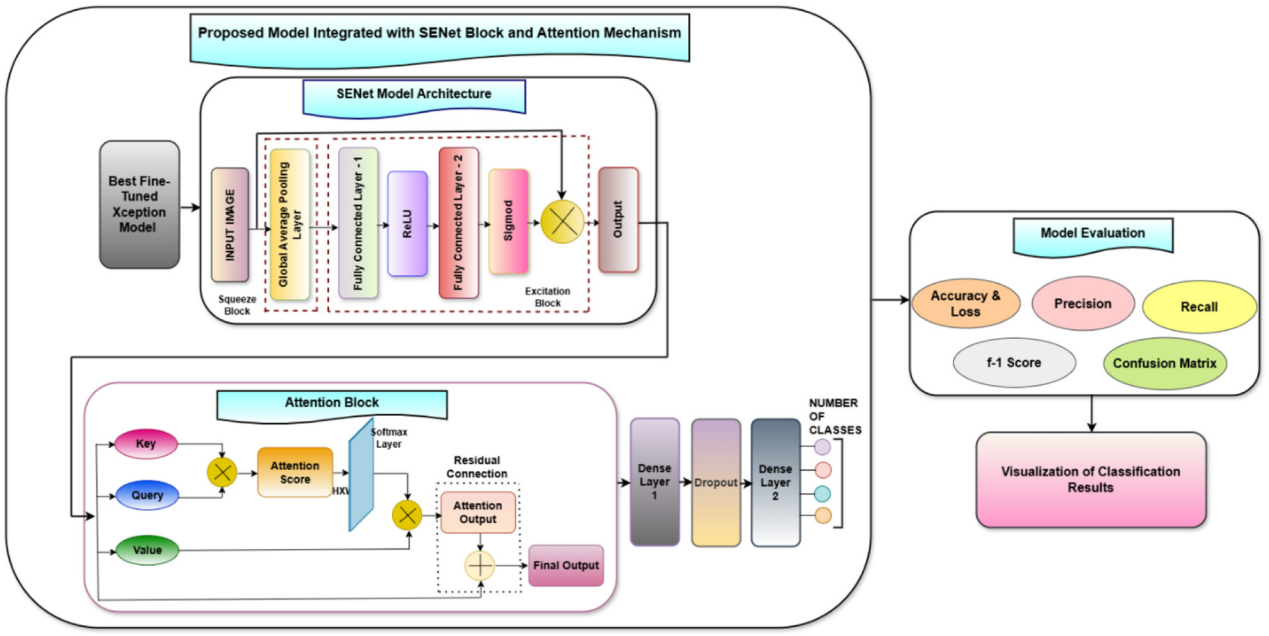

FIGURE 2
Proposed integrated model architecture.

the XGBoost model obtained an accuracy of 0.981,565, outperforming the results of the other models. In the study of Zheng et al. (2023), the authors proposed the use of the ConvLSTM-LSTM model in the prediction of wind speed in urban regions based on the spatial-temporal relationship of the historical meteorological data obtained from the weather station. This model outperformed the CNN-LSTM model by 22.12%, 22.80%, and 12.24%, and the ConvLSTM-FC model by 14.84%, 15.04%, and 7.51% in R^2, respectively, in the prediction of wind speed, showing its potential in the prediction of abrupt weather changes. In the study of Shrivastava et al. (2023), the authors used the time series data in the forecasting of the temperature in New Delhi, comparing the performance of the MPR and DNN models in the prediction, with the DNNM-3 model showing an accuracy of 96.4% in the forecasting process. In the study of Venkatachalam et al. (2023), the authors proposed the use of the Transductive LSTM (T-LSTM) model in the prediction of weather with the inclusion of the HHWD and Jena Climate datasets, showing an accuracy of 98.2% in the prediction process based on the RMSE, loss, and MAE performance metrics used in the model. Finally, the study of (Papadimitriou et al., 2023) proposes a CNN-based weather image classification method for cloudy, sunny, wet, and snowy days with 98% accuracy and better than existing approaches.

The article of (Tian and Kim, 2023) presents an enhanced Faster R-CNN model for identifying vehicles in foggy circumstances. By categorizing and preprocessing photos with haze, detection accuracy was increased. The initial VGG16 framework was updated with ResNet101, and a Feature Pyramid Network (FPN) was included to enhance the feature extraction for smaller objects particularly. The modified model attained a detection accuracy of 96.83%.

Recent developments in the field of atmospheric science and AI research also show the increasing importance of deep learning in weather analysis. For example, research in (Zhao et al., 2025) shows the effects of remote thermal forcing, like the Tibetan Plateau heating effect, on global weather patterns (Kang et al., 2018). This also shows the importance of learning long-range dependencies in weather analysis, which we have attempted to address in this paper by incorporating attention mechanisms in the proposed model. Research in typhoon prediction (Xu et al., 2025) and plateau vortices (Zhao et al., 2024) also show the importance of AI in weather analysis for better prediction accuracy at the local level, which we have attempted to address by incorporating adaptive feature recalibration in the proposed model. Research in belief shift clustering (Xu et al., 2025) and hybrid models (Zhang et al., 2023) also shows the importance of dynamic learning of features in weather analysis, which we have attempted to address by incorporating hierarchical weather feature extraction in the proposed model. All this research provides a platform for developing robust classification models incorporating spatial and channel-wise features (Elhoseiny et al., 2015). Table 1 shows a comparison analysis of the key models, techniques, datasets, performance, and limitations.

Though the SE blocks and attention mechanisms have been studied individually in the past, the novelty in the proposed framework comes in the form of task-specific integration and validation. Unlike other existing techniques in the field of weather classification, which have been based on conventional CNNs and transformers without dual-level feature refinement, the proposed approach incorporates the benefits of SE blocks and attention mechanisms within a task-specific and fine-tuned Xception backbone to address the problem of visually similar classes such as Shine and Sunrise. The proposed architecture allows the network to refine the features in the channel and

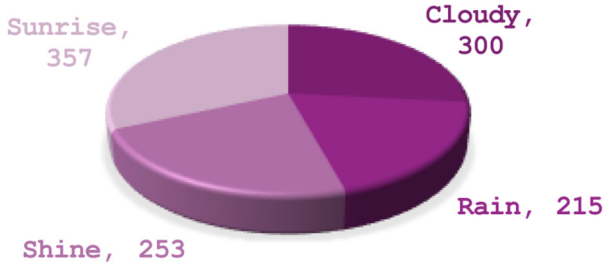

FIGURE 3
Class wise image data.

spatial dimensions, resulting in better weighting of feature importance and the discrimination of subtle differences in the patterns of the weather classes. Moreover, the proposed approach has been validated using an ablation study, resulting in incremental improvements in the performance of the network with the addition of SE blocks and attention mechanisms. Unlike other metrics, the proposed approach has been validated in terms of macro-F1 scores and standard deviation to assess the stability and generalization capabilities of the network.

3 Proposed framework

The focus of this study is to classify images of weather into four classes, which are sunshine, cloudy, rain, and shine. In order to accomplish this, three different TL models, Xception (Chollet, 2017), EfficientNetB3 (Tan and Le, 2019), and ResNet152V2 (He et al., 2016), have been used. In selecting these three models, it was strategically decided to choose the three best-performing models for the purpose of this study. ResNet152V2 is a residual connection-based deep learning architecture, which is best for training deep networks because it overcomes the vanishing gradients problem. In addition, EfficientNetB3 is a lightweight version of the EfficientNet architecture, which is best for environments where computational resources are scarce. On the other hand, the Xception model is based on depthwise separable convolutions, which reduces the number of parameters and computational complexity without affecting the accuracy.

FIGURE 4
Sample images from dataset (Chollet, 2017).

FIGURE 5
Comparison of preprocessing techniques: **(a)** Original image, **(b)** Histogram Equalization, and **(c)** CLAHE-enhanced image.

Specifically, MobileNet emphasizes computational efficiency but at the expense of depth and performance, whereas DenseNet has higher memory usage because of densely connected layers, which could be a drawback for using the model on resource-constrained devices. Moreover, the lower-tier EfficientNet models, such as EfficientNetB0 to B1, do not offer the depth and representation capacity of EfficientNetB3, which is more effective in accurately distinguishing visually similar weather conditions. Figure 1 depicts the proposed framework for the weather classification research. The methodology proposed in Figures 1, 2 is an all-embracing framework proposed specifically for multi-class weather condition classification. The proposed method uses various state-of-the-art DL methods and attention mechanisms. Initially, images from a multi-class weather dataset are gathered; in each of the acquired images, particular weather conditions are classified as rain, sunshine, sunrise, and cloudy. It provides a good basis for building a strong model for classification because a heterogeneous and representative dataset will be as important as the model's good generalization capability across different weather conditions. Before actual model training, many necessary preprocessing steps will be required on the dataset so that the input data is in the best possible format for training the models. The images are first resized to a uniform dimension of 224 × 224 pixels to meet the input requirements of the pre-trained models, and normalization was performed by scaling the pixel values to accelerate convergence during the training of the neural network. The dataset is therefore split into training and validation/test splits to incorporate objective evaluation of the models' performance. Label encoding will also be done using categorical labels, such as weather conditions, and turning them into numerical values that will interpret them for the model.

In the subsequent stage, three pre-trained CNNs, namely, EfficientNetB3, ResNet152V2, and Xception, were utilized. Transfer learning relies on the deep feature representations learned from large-scale datasets such as ImageNet to reduce the amount of data and computational power required during training. Each model is fine-tuned with its original top layers removed and replaced by a new task-specific layer, created specifically for weather classification. New layers comprise batch normalization, dense and dropout layers. The batch

TABLE 2 **Data augmentation parameters.**

Transformation	Parameter value	Purpose
Horizontal Flip	True	Simulate viewpoint variation
Vertical Flip	True	Improve spatial invariance
Rotation	±20°	Simulate camera tilt
Zoom	0.9–1.1	Simulate distance variation
Brightness	0.85–1.15	Simulate lighting variation

Normalization layer normalizes the input for each layer so that the training can be made stable and speedy. The first dense layer learned complex representations over the extracted features, while the second dense layer will output the final predictions over the number of weather classes. With a dropout rate of 0.45, this layer has the effect of preventing overfitting, where, in each step of training, a portion of the neurons is randomly deactivated, which strengthens the generalization capability of the model.

Each fine-tuned model is executed on the training dataset and validated on the validation set, then evaluation metrics like accuracy, precision, recall, and F1-score are calculated, alongside a generated confusion matrix for an evaluation of their misclassification rate. The comparison among these three fine-tuned models enables the best architecture.

The best-performing model is optimized from the transferred learning stage with the added Squeeze-and-excitation Block (SENet) (Hu et al., 2018) and Attention Block (Vaswani et al., 2017). Figure 2 gives the design of this integrated TL model. The attention mechanism is what makes this model focus further, more intensively on particular regions of the weather images, emphasizing the important features at the same time while reducing noise and irrelevant information. This in turn increases the sensitivity of the model towards weather conditions, where subtle features like fog or clouds during complex scenes are hard to detect.

SENet has been applied after the best-performing model's output, followed by the batch normalization layer, which has normalized the feature maps and avoids overfitting with

TABLE 3 CLASS-WISE dataset distribution before and after augmentation.

Class	Total images	Validation	Test	Training images (before augmentation)	Training images after augmentation
Cloudy	300	60	18	222	1,110
Rain	215	43	21	151	755
Shine	253	51	25	177	885
Sunrise	357	53	26	278	1,390
Total	1,125	207	90	828	4,140

TABLE 4 Evaluation of transfer learning models with key features and selection justification

Model	Parameters (millions)	Input size	Top-1 accuracy (%)	MACs (GFLOPs)	Complexity	Key features
MobileNetV2	3.4	224 × 224	71.8	0.3	Low	Highly efficient, suitable for mobile applications
DenseNet121	8	224 × 224	74.9	2.9	Moderate	Dense connectivity, good feature reuse
EfficientNetB0	5.3	224 × 224	77.1	0.39	Low	Compound scaling, efficient for small datasets
EfficientNetB1	7.8	240 × 240	79.1	0.7	Moderate	Improved depth over B0
EfficientNetB3	12	300 × 300	81.6	1.8	Moderate	Balanced depth and efficiency
ResNet50	25.6	224 × 224	76	4.1	High	Deep residual learning, good generalization
ResNet152V2	60.4	224 × 224	78.3	11.5	Very High	Very deep, high capacity, needs more resources
Xception	22.9	299 × 299	79	8.4	Moderate	Depthwise separable convolutions, high efficiency

preserved essential findings. The SENet block output is used as input to the attention block. Then, two dense layers are used, along with a dropout layer. This integrated SENet model is tested using the test dataset, which in turn proves the actual efficiency of the model. The model's prediction capabilities are determined based on the different parameters of the model's efficiency. The confusion matrix proves the classification mistakes the model has made at any given weather condition. This provides a detailed idea of the model's strengths and weaknesses. Finally, the results obtained, which include the classification, are visually compared to find the model's categories, which have been correctly predicted. All of these steps are part of the methodology's workflow, which provides highly precise and accurate solutions in the classification of weather conditions. The evaluation of the integrated SENet model is done using the test dataset, which in turn proves the actual efficiency of the model. The model's prediction capabilities are determined based on the different parameters of the model's efficiency. The confusion matrix proves the classification mistakes the model has made at any given weather condition. This provides a detailed idea of the model's strengths and weaknesses. Finally, the results obtained, which include the classification, are visually compared to find the model's categories, which have been correctly predicted. All of these steps are part of the methodology's workflow, which provides highly precise and accurate solutions in the classification of weather conditions.

3.1 Dataset employed

For conducting the classification task efficiently, the dataset plays a very crucial task. Here in this study, the Multi Class Weather dataset (MWD) (Ajayi, 2018) has been employed for effective model training and testing. The dataset provides a framework for outside weather study by extracting numerous attributes for distinguishing distinct weather situations like sunrise, shine, rain and cloudy. The Multi-Class Weather dataset used within this research includes 1,125 images spread into four weather classes. The directory structure of the dataset was verified before conducting the experiments. The number of images per class was counted, and exploratory data analysis was conducted to examine the imbalance in the distribution of data and determine representative samples. Samples and statistics of class frequency were generated to verify the integrity of the dataset before modeling.

1. Sunrise: This class includes images that portray the atmospheric conditions during the early hours of the day,

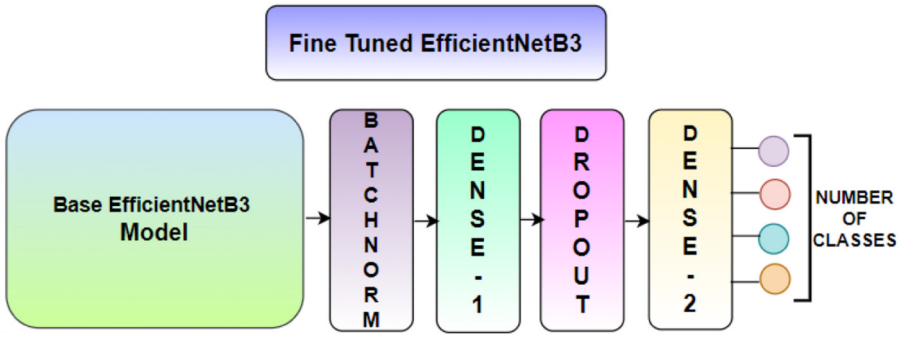

FIGURE 6
Architecture of fine-tuned EfficientNetB3.

with softer light, warmer tones, and a gradual landscape illumination.

2. Shine: In this class, the images are mostly with the sky not obscuring direct sunlight. Such images have relatively high brightness and contrast with hardly any cloudiness and deep shadows.
3. Rain: This class is more complicated because the images are of rainy conditions, possibly with raindrops, reflections on the water surface, and darkened skies. Visibility during rainy conditions raises many problems for successful classification.
4. Cloudy: This class consists of images dominated by overcast skies, which scatter or obscure sunlight with clouds. Due to the variation in cloud densities, kinds, and textures, it is an interesting category for model differentiation.

Figure 3 gives the dataset's image description class-wise and Figure 4 shows some of the sample images for every weather class. These images were captured while maintaining the natural settings and circumstances.

3.2 Preprocessing

To effectively prepare the data set for the training of the model, the preprocessing operations shown in Figure 5 are crucial to the process. Each of the operations plays a vital role in ensuring that the input data to the model is not only standardized and dependable, but also capable of being recognized effectively by the model (Tan and Le, 2019). Since the images used in representing the weather conditions are usually not of the same resolution, the need to resize all the images to a standard dimension, i.e., 224 × 224 pixels, arises, since the pre-trained models require images of the same dimension for effective training and testing. To effectively utilize the memory and training capabilities of the model, the process of batch generation is used, allowing the model to process images in batches, i.e., in smaller numbers. After the image resizing step, data augmentation methods were used on the training data alone. The data augmentation methods used were horizontal flipping (probability p = 0.5), vertical flipping (probability p = 0.5), rotation by a random angle between ±20°, zooming between 0.9 and 1.1, and brightness adjustment by a random amount between ±15%.

TABLE 5 Parameter table for fine-tuned EFFICIENTNETB3 model.

Layer type	Output shape	Parameters
EfficientNetB3	None, 1,536	10,783,535
Batch Normalization	None, 1,536	6,144
Dense 1, (256 units + ReLU)	None, 256	393,472
Dropout (0.45)	None, 256	0
Dense 2 (4 units, SoftMax)	None, 4	1,028
Total Parameters:11,183,882		
Trainable Parameters: 11,142,730		
Non-Trainable Parameters: 41,152		

Subsequently, the normalization of the pixel values takes place, and the pixel values are scaled to the desired range, i.e., between 0 and 1. This helps to prevent the possibility of numerical instability and allows the network to converge more quickly to the desired results. Next, the process of label encoding takes place, wherein the categorical nature of the weather labels will be converted to a numerical form. Typically, one-hot encoding will be used to achieve the desired results in the context of multi-class classification problems. An image visualization step will be carried out to ensure that the augmentation and other image processing steps have been done correctly. The dataset was strictly separated into training, validation, and test sets before the application of the data augmentation techniques. Data augmentation was performed on the training set to increase the diversity of the images, and the same was not performed on the validation and test sets to avoid any form of data leakage and to obtain unbiased results.

To improve the generalization capability and prevent overfitting caused by the limited amount of data, data augmentation was used for the training set only after the dataset was partitioned into 70:20:10. The dataset had 828 training images, 207 validation images, and 90 test images. After data augmentation, the training set was enlarged to 4,140 samples, while the validation and test sets remained the same to ensure unbiased evaluation. The data augmentation strategies were chosen to mimic various environmental and acquisition changes without altering the

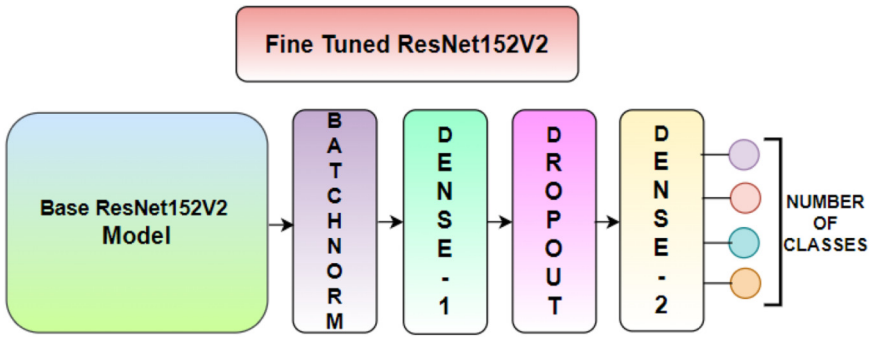

FIGURE 7
Architecture of fine-tuned ResNet152V2 model.

TABLE 6 RESNET152V2 parameter description.

Layer type	Output shape	Parameters
ResNet152V2	None, 2048	58,331,648
Batch Normalization	None, 2048	8,192
Dense 1, (256 units + ReLU)	None, 256	524,544
Dropout	None, 256	0
Dense 2 (4 units, SoftMax)	None, 4	1,028
Total Parameters: 58,865,412		
Trainable Parameters: 58,717,572		
Non-Trainable Parameters: 147,840		

semantic meanings of weather patterns. The detailed augmentation configuration is summarized in Table 2. Specifically, horizontal and vertical flipping were applied, random rotation was limited to ±20°, zoom range was set between 0.9 and 1.1. Brightness variation was implemented using a multiplicative intensity factor within the range [0.85, 1.15], corresponding to ±15% variation. The ranges were selected to create moderate geometric and photometric variation while maintaining structural weather conditions. The augmentations were performed using an on-the-fly approach during training, meaning that augmentations were generated on a batch-by-batch basis during training rather than being stored in a database. This means that this process did not require additional memory storage or parameters to be learned by the model; it simply required a small amount of additional computation during training.

In order to take into consideration, the changes in exposure and contrast that are often observed in weather-related images, further image enhancement techniques were included in the preprocessing stage. These techniques are Histogram Equalization (HE) and Contrast Limited Adaptive Histogram Equalization (CLAHE). These techniques are used to further enhance image contrast. Histogram Equalization enhances image contrast by distributing intensity values across an image (Dasu et al., 2026). Figure 5 displays the histogram equalization and CLAHE visualization.

In Contrast Limited Adaptive Histogram Equalization method is used to divide an image into small regions or tiles. The image contrast is enhanced in these regions without allowing excessive enhancement by specifying a predefined limit. The representative examples of the image and the enhanced versions are shown in Figure 5. The addition of these preprocessing techniques enhances feature visibility and facilitates better feature extraction.

Dataset distribution for each class has been illustrated in Table 3, for before and after data augmentation. The dataset is moderately imbalanced, with the highest number of samples belonging to class 'Sunrise' and the least number of samples belonging to class 'Rain'. The dataset was divided in a proportionate ratio of 70:20:10, yielding 828, 207, and 90 samples for training, validation, and testing, respectively. The data was split using stratified random sampling, with a fixed seed (seed = 42) to ensure reproducibility. We split the data directly using Scikit-learn's train_test_split function, and we stored the split indices so that our experiments can be exactly replicated. Data augmentation was only done on the training set in a proportionate manner, increasing the number of training samples five-fold from 828 to 4,140, while maintaining the original class distribution. The validation and test sets were not altered.

The augmentation probabilities were explicitly implemented to ensure consistency between the methodological description and the executable code, and all experiments were re-run after verification whereas zoom transformation was implemented explicitly using a scaling interval of [0.9, 1.1] to ensure precise consistency between the methodological description and the augmentation configuration and Brightness variation was implemented using a multiplicative intensity factor within the range [0.85, 1.15], corresponding to ±15% variation.

3.3 Selection of best TL models

The selection of EfficientNetB3, ResNet152V2, and Xception was based on a balance between representation capability and both efficiency and computation. Although EfficientNetB0 and MobileNetV2 are highly efficient, they might not be sufficiently deep to learn complex visual patterns in weather scenes. EfficientNetB3 provides a good compromise between efficiency and scalability, and it can be used in a mid-sized dataset. ResNet152V2 is computationally expensive, but it is a high-capacity model due to its depth. Xception provides a good compromise between accuracy and complexity using depth-wise separable convolutions and can be used as a baseline model and later

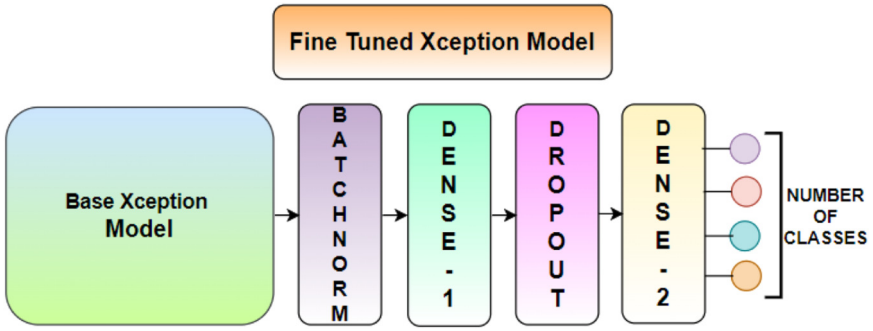

FIGURE 8
Fine-tuned xception model architecture.

TABLE 7 Parameter description of fine-tuned Xception model.

Layer type	Output shape	Parameters
Xception	None, 2048	20,861,480
Batch Normalization	None, 2048	8,192
Dense 1, (256 units + ReLU)	None, 256	524,544
Dropout	None, 256	0
Dense 2 (4 units, SoftMax)	None, 4	1,028
Total Parameters:21,395,244		
Trainable Parameters: 21,336,244		
Non-Trainable Parameters: 58,624		

enhanced using SE and attention mechanisms. A comparison of the transfer learning models is shown in Table 4.

3.4 Fine-tuned EfficientNetB3 model

This segment explains the optimized EfficientNetB3 model architecture that was utilized in the weather classification task (Figure 6). The EfficientNetB3 model was pre-trained on the ImageNet dataset to function as a feature extraction model that can extract high-level features from input images. The fully connected layers of the model were modified to suit the specific classification task. Additionally, the model utilized the Batch Normalization technique after the convolutional layers to improve training stability by eliminating gradient-related problems. The features extracted from the input images by the EfficientNetB3 model were then passed to the fully connected Dense layer with ReLU activation to extract task-specific features. The features were then passed to the dropout layer to prevent overfitting. Finally, the features were passed to the Dense layer with the SoftMax activation function to generate probabilities that represent the various classes of weather conditions. This model maintained the computational efficiency of the EfficientNetB3 model while improving its adaptability to the specific dataset.

Table. 5 gives the parameter table for the fine-tuned EfficientNetB3 framework. The model architecture consists of several key layers. The EfficientNetB3 base is the pre-trained model that extracts the features. Its total number of parameters is 10, 783, 535 and, therefore, generates a feature vector of shape 1,536. To stabilize and speed up training, the basic output is normalized through a Batch Normalization layer containing 6,144 parameters. Another layer is the Dense Layer, adding 393,472 parameters. It has 256 units and ReLU activation. This fully connected layer captures much more complex relationships within the feature set. Thereafter a Dropout Layer is used with a rate of 0.45. It adds no trainable parameters; it randomly deactivates 45% of the neurons during training time to avoid overfitting. At last, this layer is used to create output with the assistance of a SoftMax activation function that calculates class probabilities and a Dense Layer with 4 units, representing the 4 classes. There are 1,028 parameters added to the model by this layer. The total number of parameters in the model is 11, 184, 179.

3.5 Fine-Tuned ResNet152V2 model

The optimized ResNet152V2 architecture has been shown in Figure 7. The ResNet152V2 architecture is a deep convolutional neural network with 152 layers. It uses residual connections to allow the gradients to flow effectively, enabling the training of a deep neural network. The pre-trained image classification architecture on the ImageNet dataset was used as the backbone architecture. It was fine-tuned to classify the images into various weather classes. The extracted features were then passed through a Dense layer with ReLU activation to learn the features for the classification task. The features were then passed through a dropout layer to prevent overfitting. Finally, the features were passed through a Dense layer with SoftMax activation to predict the probabilities of the images belonging to the various weather classes.

The parameter configuration of the fine-tuned ResNet152V2 framework is presented in Table 6. The framework consists of a pre-trained ResNet152V2 backbone, which provides a 2048-dimensional feature vector. A Batch Normalization layer is used for the normalization of the features. A fully connected Dense layer with a reduced dimension of 256 features is used for task-specific learning. A dropout layer has been added to avoid overfitting during the training process. The learned features are then transformed into the final outputs using a SoftMax activation function in the final Dense layer. The total number of parameters

TABLE 8 Parameter description of proposed model.

Layer type	Output shape	Parameters
Xception	None, 2048	20,861,480
Batch Normalization	None, 2048	8,192
Dense 1, (256 units + ReLU)	None, 256	524,544
Dropout	None, 256	0
Dense 2 (4 units, SoftMax)	None, 4	1,028
Total Parameters: 21,395,244		
Trainable Parameters: 21,336,244		
Non-Trainable Parameters: 58,624		

in the framework is around 58.8 million. The additional layers have a small number of parameters. The framework effectively utilizes the potential of the backbone network for the final task of multi-class weather classification.

3.6 Fine-Tuned Xception Model

The optimized Xception network architecture is shown in Figure 8 that is applied for multi-class weather classification. The network is composed of a pre-trained Xception model that acts as a feature extractor, providing a 2048-dimensional representation of the input images. For the network to learn and adapt to the classification task, some extra layers were added to the pre-trained model. A Batch Normalization layer is applied to normalize the features and improve training performance. Then, the features are passed through a Dense layer of 256 units and ReLU activation to learn the features, followed by a Dropout layer to avoid overfitting. Finally, the features are passed through another Dense layer of four units and SoftMax activation to obtain the probability distributions of the four weather classes.

Table 7 summarizes the architecture of the model, as well as the number of parameters within each layer and the output shapes. The base Xception model has 20,861,480 parameters, with most being trainable and outputting feature maps of shape (None, 2048). These characteristics are passed through 8,192 parameters using the Batch Normalization layer. Using the ReLU activation function in the Dense layer with 256 units, which has 524,544 parameters and reduces the dimensionality of the feature map. Dropout does not introduce any new parameters, but it will help to avoid overfitting. The model consists of 21,395,244 parameters altogether, where 21,336,620 are trainable, meaning they can be updated in the course of training the model. The remaining 58,624 parameters are not updated during training; usually, inherited from the frozen Batch Normalization layers or the pre-trained Xception model.

3.7 Proposed model integrated with SENet block and attention mechanism

The proposed architecture utilizes a fine-tuned version of an Xception model with an SENet block and self-attention mechanism (Shelke et al., 2025b) to improve the accuracy of weather classification. As illustrated in Figure 2, the input image is first passed through the

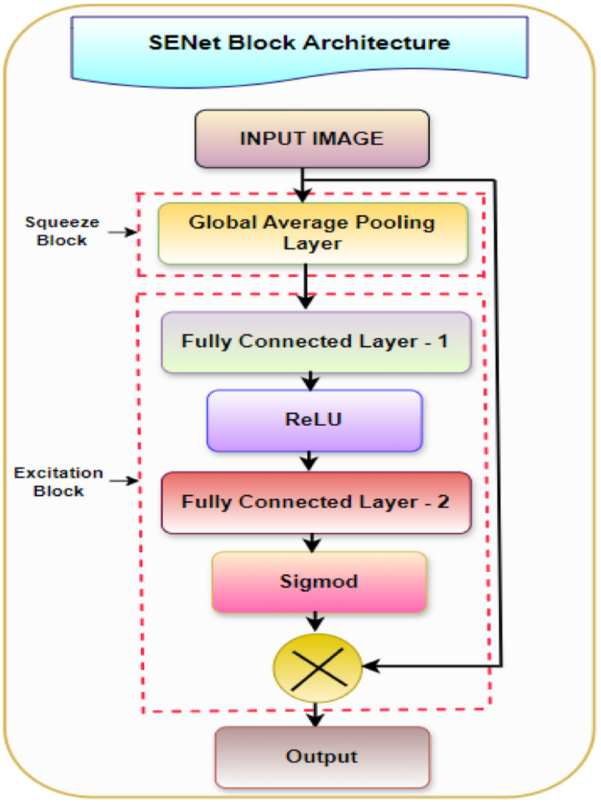

FIGURE 9
Architecture of the SENet model.

fine-tuned version of the Xception model to efficiently extract spatial and semantic features using depth-wise separable convolutions as defined in Equation 1. Then, the high-level feature map is further improved using an SENet block, which is followed by a self-attention mechanism to capture key spatial features. Finally, global average pooling and fully connected layers with dropout are used to classify the input images while avoiding overfitting:

$$F \in R^{h \times w \times c} \tag{1}$$

Here, the height, width, and number of channels are given by h, w, c, respectively. The extracted feature map F is then passed to the Squeeze-and-Excitation (SE) block. The first step in this block is the squeeze operation, where global spatial information is compressed into a channel descriptor using Global Average Pooling (GAP) as defined in Equation 2:

$$z_c = \frac{1}{h \times w} \sum_{i=1}^{h} \sum_{j=1}^{w} F_c(i, j), \forall c \in \{1, 2, \ldots, C\} \tag{2}$$

This results in a compact vector $z \in R^c$, which is passed through two fully connected (dense) layers in the excitation operation as given by Equation 3:

$$s = \sigma(W_2 \delta(W_1 z)) \tag{3}$$

Here, $W_1 \in R^{\frac{c}{r} \times c}$, and $W_2 \in R^{c \times \frac{c}{r}}$ are the weights of the dense layers, δ represents the ReLU activation, and σ denotes the sigmoid activation. The intermediate reduction ratio r (typically 16) helps

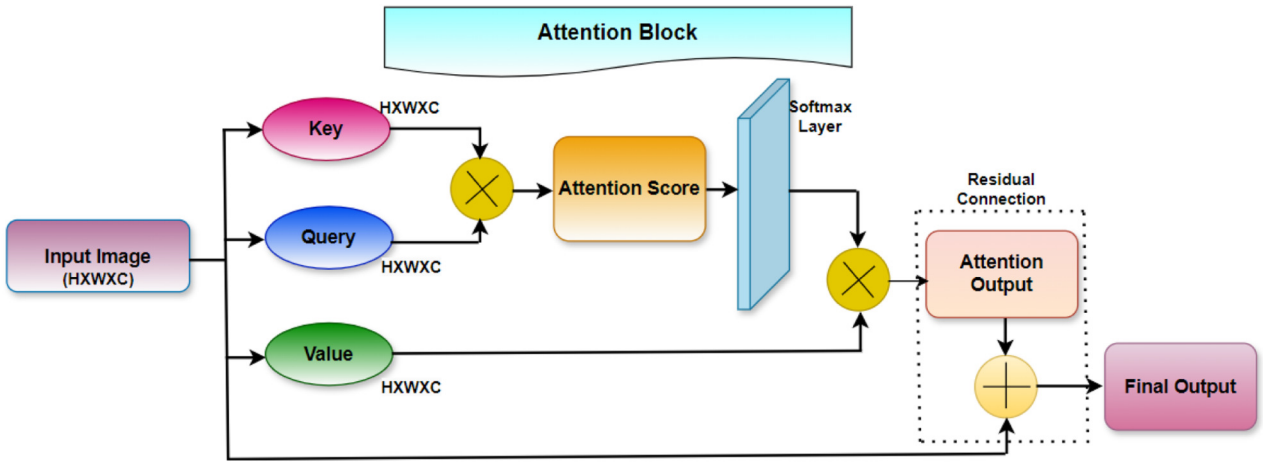

FIGURE 10
Design of attention mechanism.

TABLE 9 Optimized fine-tuned EFFICIENTNETB3 model accuracy and loss.

Epoch	Training loss (TL)	Training accuracy (TA)	Validation loss (VL)	Validation accuracy (VA)
1.	1.57	0.4178	1.52	0.4561
5.	0.44	0.8322	0.31	0.9037
10.	0.21	0.9210	0.16	0.9556
15.	0.17	0.9566	0.14	0.9485
20.	0.10	0.9633	0.13	0.9630
25.	0.06	0.9822	0.12	0.9704
30	0.05	0.9833	0.11	0.9639

control model complexity. The resulting weight vector s is then used to recalibrate the original feature maps through channel-wise multiplication as given by Equation 4.

$$\tilde{F}_c = s_c \cdot F_c \tag{4}$$

This operation emphasizes the most informative channels while suppressing less relevant ones, enhancing the model's sensitivity to critical weather patterns. Following the SE block, the Attention Block is introduced to further refine the spatial relevance of the feature map. The recalibrated feature map $F^{\sim}$ is linearly projected into three separate matrices Query (Q), Key (K), and Value (V) as given in Equation 5.

$$Q = \tilde{F}W^Q, K = \tilde{F}W^K, V = \tilde{F}W^V \tag{5}$$

where W^Q, W^K, and W^V are learnable projection weights. The scaled dot-product attention mechanism is then used to compute attention scores as can be seen from Equation 6:

$$Attention(Q, K, V) = SoftMax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{6}$$

Here, d_k represents the dimension of the key vectors and is used for scaling to prevent overly large values in the SoftMax computation. The attention output is then added to the recalibrated features *via* a residual connection to retain original spatial information. The final output O is flattened and passed through a series of fully connected layers for classification. This includes one Dense layer with ReLU activation function, a dropout layer and final output Dense layer with SoftMax activation function given by Equation 7:

$$\hat{y}_i = \frac{\exp(z_i)}{\sum_{j=1}^{c}\exp(z_j)}, i = 1, 2, \ldots, C \tag{7}$$

where $\hat{y}_i$ *is* the probability of the input image belonging to class i, and C is the total number of weather categories.

Through the combined effects of fine-tuned transfer learning, channel recalibration through SE, and contextual feature enhancement through attention, the proposed model is capable of focusing on critical features to achieve high accuracy in the classification of visually similar and context-sensitive weather conditions. The novelty of the proposed model does not lie in the introduction of new architectural components but rather in the systematic integration and extensive evaluation of the Xception backbone model with Squeeze-and-Excitation blocks and spatial

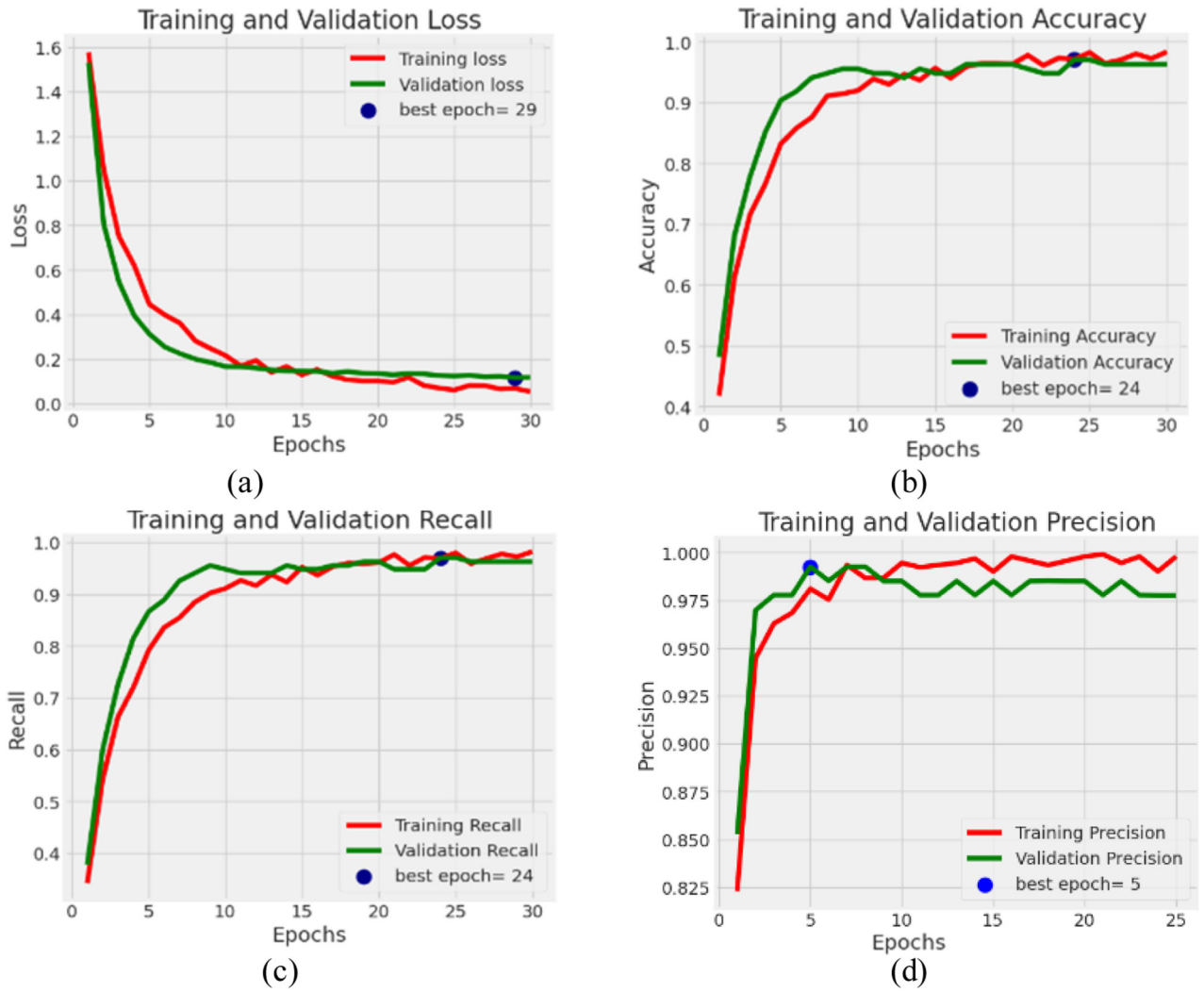

FIGURE 11
Graphical presentation of fine-tuned EfficientNetB3 **(a)** loss **(b)** accuracy **(c)** recall **(d)** precision.

attention mechanisms within a unified framework for weather image classification. Although various studies have attempted to explore the proposed model components individually and in limited combinations, the novelty of the proposed model is in the systematic exploration of the complementary effects of the proposed model components for feature refinement and spatial discrimination through extensive ablation and validation. More specifically, the use of SE blocks helps in adaptive channel-wise feature recalibration, which improves the discriminability of weather-related attributes. Additionally, the use of spatial attention helps in better visual cue localization, allowing for better differentiation between visually ambiguous weather conditions, such as fog, haze, and clouds. Apart from the architectural design, the study also provides a detailed evaluation of robustness, statistical significance, and explainability, making it a more comprehensive evaluation compared to other recent works in the similar domain. Table 8 describes the parameters used in the proposed model.

From Table 6 it can be seen that the number of parameters in the proposed architecture is about 27 million, and hence this architecture falls under the category of moderate complexity. This architecture is not light-weight but is more efficient than very deep models such as ResNet152V2 (~60M parameters), while providing better performance in terms of classification accuracy.

3.7.1 SENet block

The overall architecture of the SENet block that enhances feature representation through channel-wise recalibration has been shown in Figure 9. The input feature map is initially fed into a Global Average Pooling layer that integrates spatial information into a single descriptor for every channel. This is then processed by a fully connected layer with ReLU activation that reduces the feature dimension by a factor of 1/16.

The feature dimensions are thus reduced and captures compact dependencies between different channels. This reduced feature map is then mapped back to the original dimension using another fully connected layer followed by a sigmoid activation function that

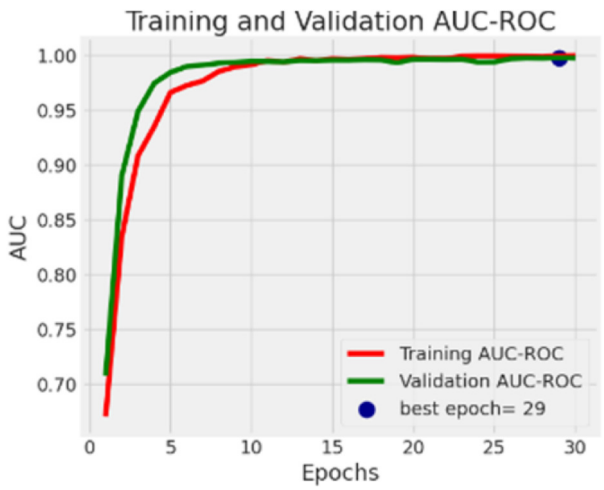

FIGURE 12
Fine-tuned EfficientNetB3 model AUC-ROC.

generates attention weights between 0 and 1 for every pair of channels. This attention map is then element-wise multiplied with the original input feature map that enhances informative features and suppresses irrelevant ones. This enhances overall classification performance.

3.7.2 Attention block

The proposed Attention Block, shown in Figure 10, incorporates a scaled dot-product self-attention mechanism to address the long-range spatial dependencies in the extracted feature maps. The self-attention mechanism allows the network to concentrate on the most informative spatial locations of the input feature map. Let the input feature map be represented by Equation 8 as:

$$X \in R^{H \times W \times C} \tag{8}$$

The feature map is reshaped into a two-dimensional representation by using Equation 9.

$$X' \in R^{N \times C} \tag{9}$$

where N = H × W. Three learnable linear transformations are applied to generate the Query (Q), Key (K), and Value (V) matrices, which are defined by the Equations 10–12.

$$Q = X'W_Q \tag{10}$$

$$K = X'W_K \tag{11}$$

$$V = X'W_V \tag{12}$$

where Equations 13–15 give the values for W_Q, W_K and W_V.

$$W_k \in R^{(C \times d_k)} \tag{13}$$

$$W_Q \in R^{(C \times d_q)} \tag{14}$$

$$W_V \in R^{(C \times d_v)} \tag{15}$$

In this implementation, the projection dimensions are defined by Equation 16:

$$d_k = d_v = \frac{c}{8} \tag{16}$$

The attention weights are computed using the scaled dot-product attention formulation as given in Equation 17:

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{17}$$

The SoftMax function normalizes the attention scores across spatial positions, assigning higher weights to more informative regions of the feature map. The resulting attention output is reshaped back to H × W × d_v and projected to match the original channel dimension. A residual connection is then applied as defined in Equation 18:

$$Y = X + Attention(X) \tag{18}$$

This residual formulation preserves original feature information while enhancing context-aware spatial representation. The attention mechanism improves the network's ability to capture complex spatial dependencies, which is particularly beneficial for distinguishing visually overlapping weather categories.

4 Results and discussion

This section discusses the model performance results for all the above-mentioned models. These results were recorded during the testing of the models on the test set.

4.1 Hyperparameters and evaluation metrics

In the current research, transfer learning was used with pre-trained weights for the Xception and ResNet152V2 models.

TABLE 10 Model performance matrix for fine-tuned EFFICIENTNETB3 model.

Classes	Precision	Recall	F-1 score	Accuracy	Cohen's Kappa
Cloudy	0.916	0.912	0.914	0.965	0.943
Rain	0.933	1.00	0.966		
Shine	1.00	0.920	0.960		
Sunrise	0.945	0.970	0.957		
Macro Average Performance	0.948	0.950	0.949	0.965	0.943

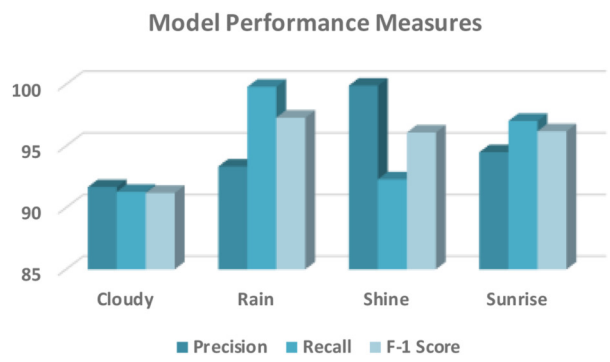

FIGURE 13
Graphical presentation of performance measures for fine-tuned EfficientNetB3.

During the training process, the higher layers were fine-tuned while the convolutional layers were frozen or had a lower learning rate. This is done to avoid overfitting the model. The choice of hyperparameters is an essential task for optimal model performance. Hyperparameters were determined experimentally for the model. The model was optimized iteratively. The model was optimized with the Adam optimizer and an initial learning rate of 0.0001. This is because the Adam optimizer has adaptive learning rates and converges quickly. To avoid overfitting of the model, the dropout rate was set to 0.5 between the dense layers. This randomly disables half of the neurons in the model to improve the model's ability to generalize. The training was performed for 30 epochs, using a batch size of 32 images, allowing efficient utilization of computational resources while ensuring model stability and convergence. SoftMax activation function has been used in the last output layer. Several performance metrics were utilized to comprehensively evaluate the effectiveness of the proposed weather classification model:

Accuracy measures the ratio of correctly predicted instances to the total number of predictions as given by Equation 19.

$$Accuracy = (TP + TN)/(TP + FP + TN + FN) \quad (19)$$

Precision, as given by Equation 20, reflects the proportion of positive identifications that were actually correct, crucial for assessing class-wise prediction reliability.

$$Precision = TP/(TP + FP) \quad (20)$$

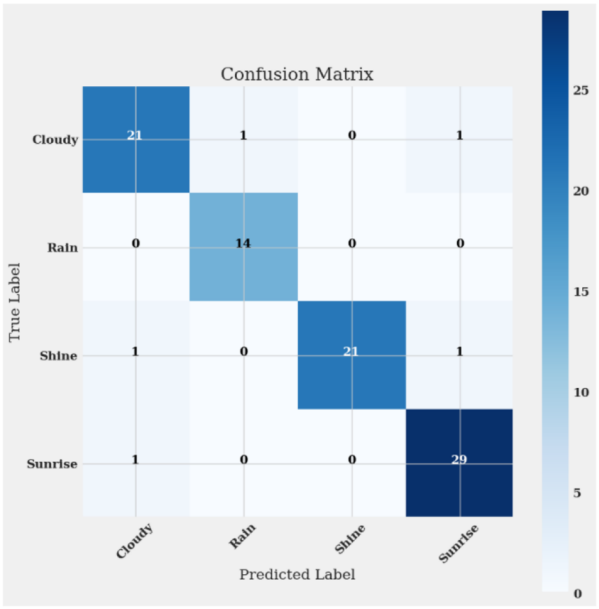

FIGURE 14
Confusion matrix for fine-tuned EfficientNetB3.

Recall, as given by Equation 21, indicates the model's capability to identify all relevant

$$Recall = TP/(TP + FN) \quad (21)$$

F1-score combines precision and recall into a harmonic mean, balancing both false positives and false negatives as given by Equation 22.

$$F1\ Score = 2\ (Precision^{*}Recall)/(Precision + Recall) \quad (22)$$

AUC-ROC (Area Under Curve - Receiver Operating Characteristics) summarizes a model's ability to distinguish between classes at various threshold settings, with values closer to 1.0 indicating superior performance. Confusion Matrix provides a comprehensive visualization of the model's performance, clearly showing true positive (TP), false positive (FP), true negative (TN), and false negative (FN) predictions across all classes.

4.2 Fine-tuned EfficientNetB3 model results

In this section, the classification results of the EfficientNetB3 model have been shown. Table 7 gives the

TABLE 11 Accuracy and loss measurement for fine-tuned RESNET152V2.

Epoch	Training loss (TL)	Training accuracy (TA)	Validation loss (VL)	Validation accuracy (VA)
1.	1.03	0.6056	1.62	0.4881
5.	0.13	0.9701	0.18	0.9255
10.	0.09	0.9810	0.10	0.9704
15.	0.04	0.9901	0.09	0.9771
20.	0.05	0.9856	0.09	0.9778
25.	0.02	0.9967	0.07	0.9704
30	0.01	0.9989	0.07	0.9769

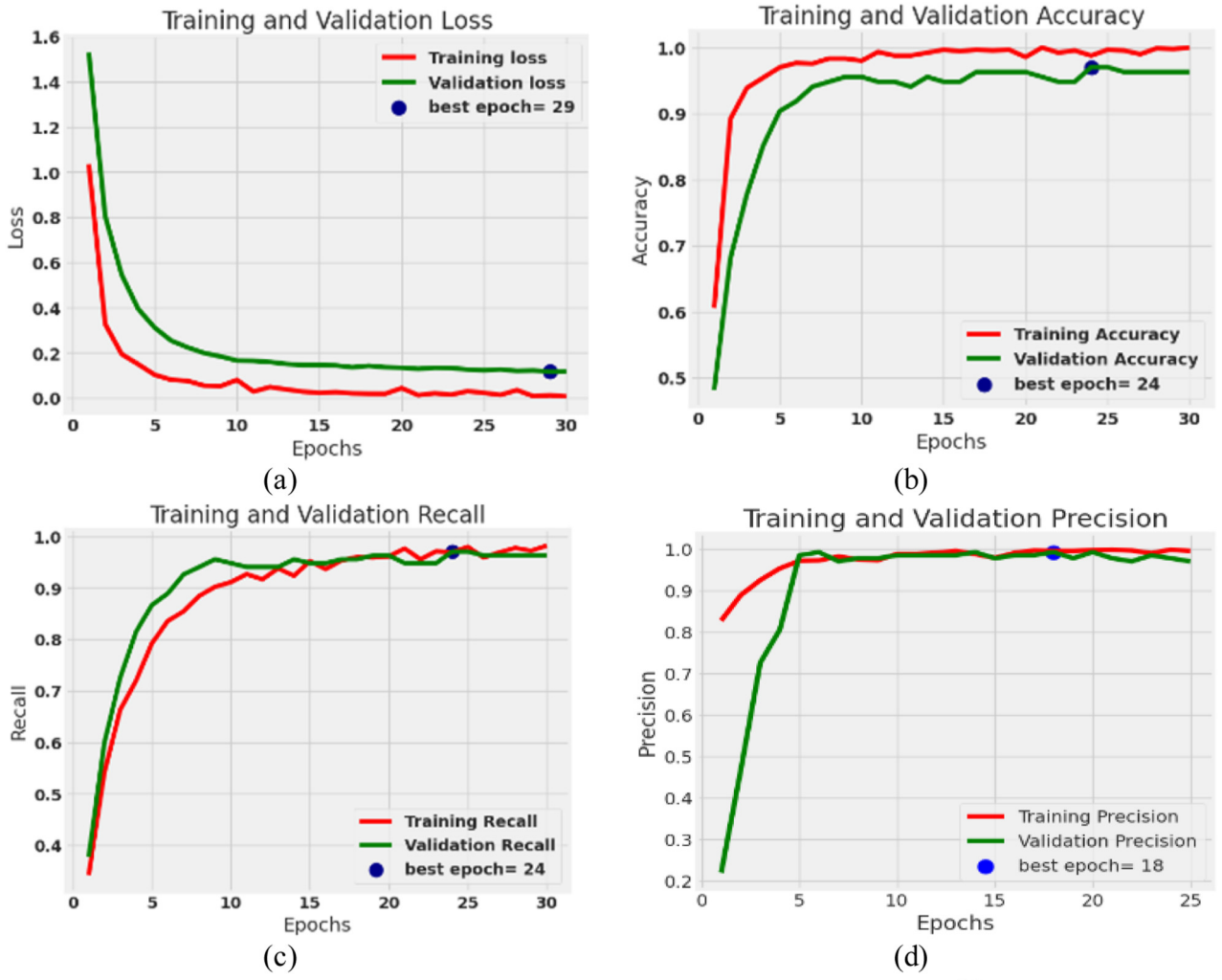

FIGURE 15
Graphical Presentation of the fine-tuned ResNet152V2 model **(a)** Loss **(b)** Accuracy **(c)** Recall **(d)** Precision.

model's accuracy and loss development measured during training and validation of the model. From Table 9, it can be seen that the model's performance improves throughout 30 epochs according to the training and validation criteria. In the beginning, epoch 1, T.A stands at 0.417 and V.A at 0.456, and both the TL is 1.57 while the V.L is 1.52; this is to be expected since the model is still learning from the data. But by epoch 5, the improvement is very good, T.A went up to 0.8322 and the V.A to 0.9037, and both training and validation loss plummeted. From epoch 10, it is observed that the model learned well and sufficiently generalized new data as it had a T.A of 0.9210 and a V.A of 0.9556 alongside some further value reductions in the loss value. From epoch 15 to epoch 20, further improvement in both measures of training and validation was encountered. This resulted in the training reaching an accuracy of up to 0.9633 while the validation is at a peak at 0.9639. It shows that the model is not overfitting but has rather nicely stuck to the complexity of the data.

Figure 11 shows the training dynamics of the proposed model for 30 epochs. The training accuracy (TA) was steadily improving with the highest accuracy of 98.33% at the 30th epoch, whereas the

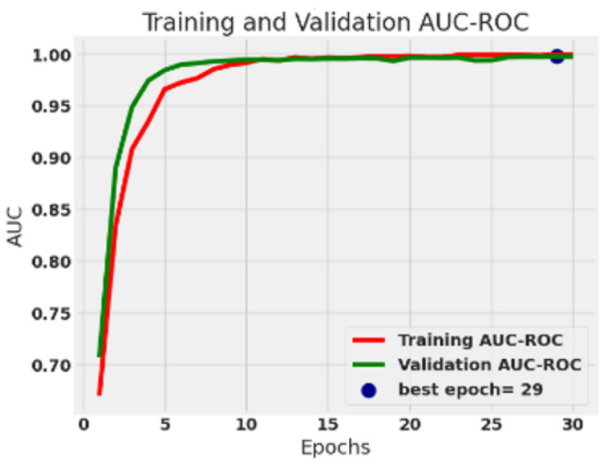

FIGURE 16
AUC-ROC of the fine-tuned ResNet152V2 model.

TABLE 12 Performance parameter analysis for fine-tuned RESNET152V2.

Classes	Precision	Recall	F-1 score	Accuracy	Cohen's kappa
Cloudy	0.923	0.965	0.945	0.976	0.964
Rain	0.989	0.988	0.988		
Shine	0.920	0.925	0.922		
Sunrise	0.972	0.932	0.952		
Macro Average Performance	0.951	0.952	0.951	0.976	0.964

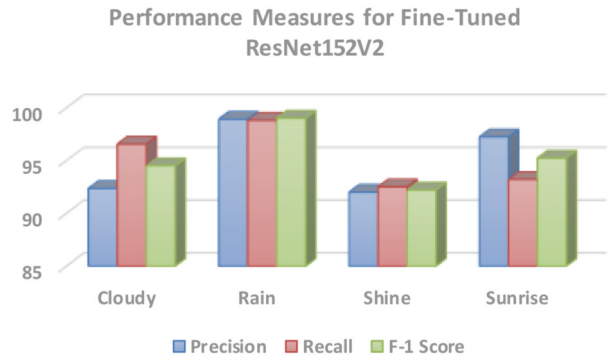

FIGURE 17
Performance Measurement Presentation of the fine-tuned ResNet152V2.

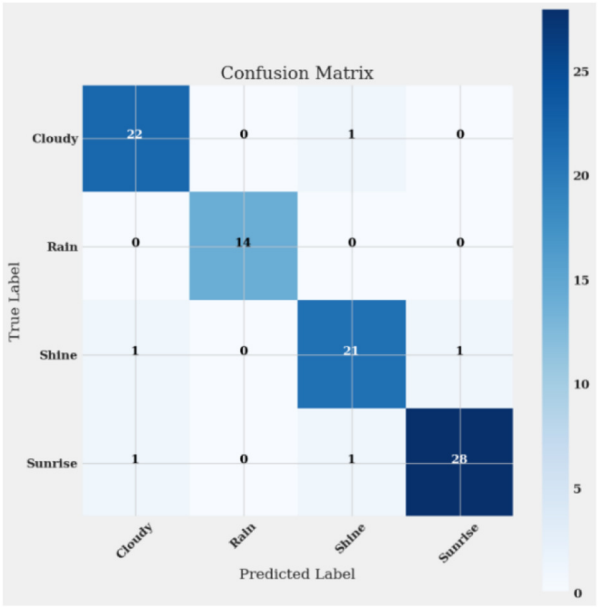

FIGURE 18
Fine-tuned ResNet152V2 confusion matrix.

validation accuracy (VA) remained relatively constant at 96%–97%, showing excellent generalization with minimal deviation between training and validation accuracy curves. However, a small fluctuation in the validation accuracy curve was observed at approximately the 25th epoch with a 1% deviation. However, this does not signify any overfitting as the deviation between the training accuracy and the validation accuracy curves remains small. As presented in Figure 11a, the training and validation losses start at a high value of 1.6. However, the losses reduce sharply for the initial epochs. The losses continue to reduce steadily and then remain constant after 25 epochs. The minimum validation loss was observed at the 29th epoch. Figure 11b shows the accuracy curves for the training and validation accuracy. The training accuracy and the validation accuracy increase sharply for the initial epochs. The accuracy then stabilizes at a high value of 95%. The highest validation accuracy was observed at the 24th epoch. Figure 11c shows the recall curves for the training and validation accuracy. The training accuracy and the validation accuracy increase sharply for the initial epochs. The accuracy then stabilizes at a high value of 1.0.

Figure 11d plots the model's precision, which is the fraction of true positive predictions among all positive predictions. The red curve indicates the training precision, and the green curve follows validation precision.

For both curves, there are steady increases in the first epochs until the precision values reach almost 1.0. This means that the model is fairly accurate with its positive predictions. The best validation precision is depicted at epoch 5, and it has been underlined by the blue dot. For all epochs, across the entire graph, the training precision slightly exceeds the validation precision. Figure 12 shows the AUC-ROC performance for the training epochs. The AUC-ROC metric assesses the classification's capacity to differentiate between classes. The closer the value to 1.0, the better the separability. The training (red) and validation (green) curves sharply increase at the start of the epochs, reaching a state of near-perfect classification. The validation curve surpasses the training curve in the later epochs. The highest validation performance is at epoch 29, as shown by the blue marker.

Table 10 shows the classification performance metrics for the test set. The precision for the Cloudy class was 0.9167, while the recall was 0.9128. The Rain class had a precision of 0.9334 and a perfect recall of 1.0. The Shine class had a perfect precision of 1.0 but a slightly lower recall of 0.92. The Sunrise class had a high precision of 0.945 and a high recall of 0.9703. The model have showed overall accuracy of 0.965 and cohen's kappa of 0.943. The classification performance metrics for the test set are as follows.

Figure 13 displays the model's performance measures graphically. From where these values can easily be analyzed.

Figure 14's confusion matrix provides a clear overview of the model's performance across four weather categories: cloudy,

TABLE 13 Accuracy and loss development for fine-tuned Xception model.

Epoch	Training loss (TL)	Training accuracy (TA)	Validation loss (VL)	Validation accuracy (VA)
1.	0.92	0.6501	1.06	0.685
5.	0.09	0.9789	0.32	0.9810
10.	0.06	0.9845	0.07	0.9853
15.	0.04	0.9933	0.05	0.9815
20.	0.02	0.9978	0.05	0.9810
25.	0.02	0.9979	0.05	0.9817
30	0.02	0.9980	0.04	0.9820

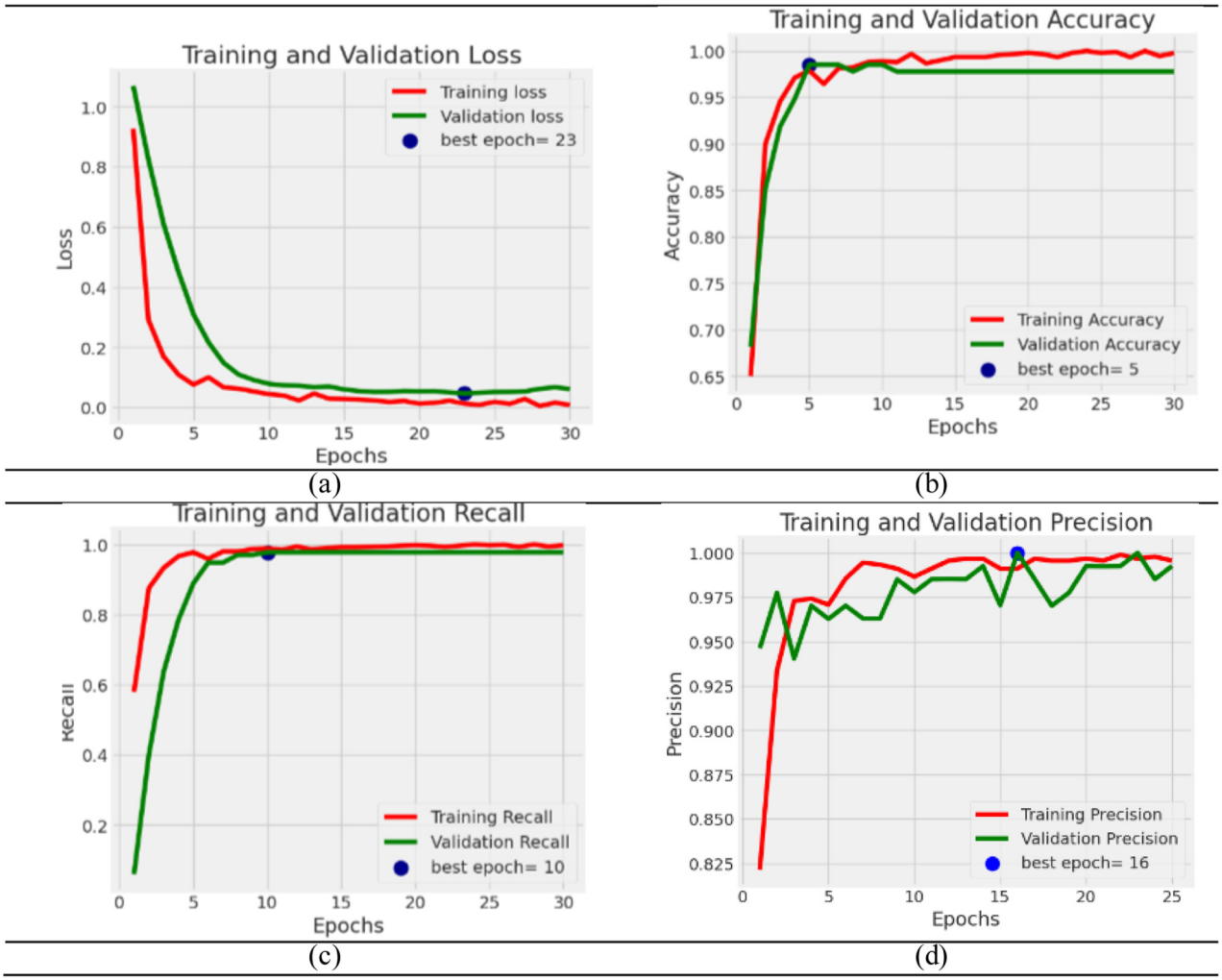

FIGURE 19
Performance measurement for fine-tuned xception model **(a)** training and validation loss **(b)** training and validation accuracy **(c)** training and validation recall **(d)** training and validation precision.

rainy, shine, and sunrise. It shows the number of correct and incorrect predictions for each class, with correct predictions shown on the diagonal and errors appearing in the off-diagonal cells.

For instance, the model was able to correctly predict 21 instances of the "Cloudy" class but incorrectly predicted 1 instance of "Sunrise" and another instance of "Shine." On the other hand, all 14 instances of the "Rain" class were correctly predicted with no

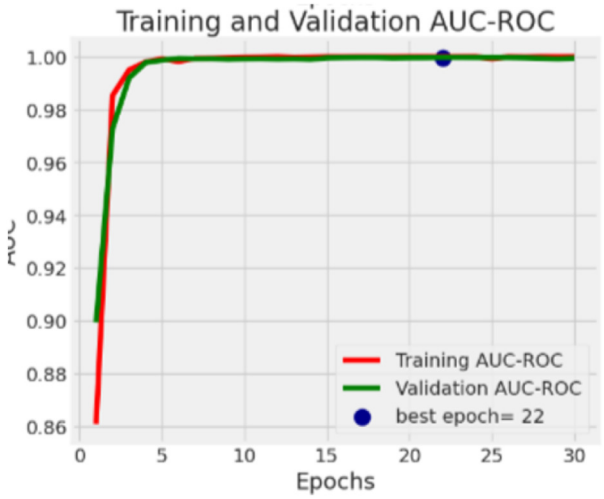

FIGURE 20
Model AUC-ROC curve for fine-tuned xception model.

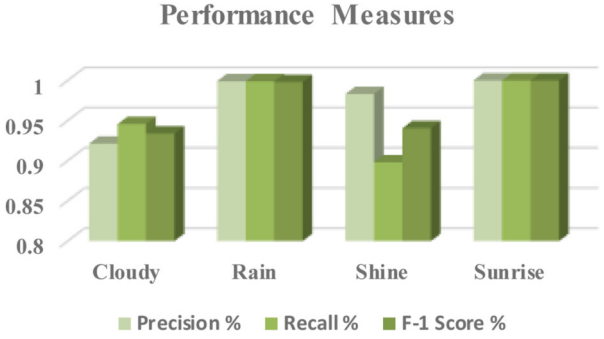

FIGURE 21
Fine-tuned xception model performance measures presentation.

errors. The "Shine" class saw 21 instances correctly predicted with only 1 instance incorrectly predicted as "Cloudy." On the other hand, the "Sunrise" class saw 29 instances correctly predicted with 1 instance incorrectly predicted as "Cloudy." Based on the above instances correctly predicted and incorrectly predicted by the model, it can be analyzed that the model has correctly predicted 96.59% of the instances with an overall accuracy rate of 96.59%. It is very clear that the model has done extremely well in correctly forecasting instances of Rain and Shine. The high F1 score is an indication that the model is striking a balance between recall and precision in all the instances. This is an indication that the model is a reasonably accurate classifier of weather patterns.

4.3 Fine-Tuned ResNet152V2 model results

The epoch-wise Table 11 provides detailed metrics of training and validation loss and accuracy at various times. In the case of epoch 1, the TA with a loss of 1.03 is 0.6056, whereas the VA with a higher loss of 1.62 is 0.4881. At epoch 5, where TA is at 0.9701 and VA at 0.9255, the model shows improvement with lower losses. At epoch 10, the accuracy rises to 0.9810 for the training and 0.9704 for the validation while training and validation losses significantly decrease. The model continues in this trend; at epoch 15, V.A rises to 0.9771 and TL decreases to 0.04 with 0.9901 accuracies. Epoch 25 approaches perfect TA at 0.9967 while the VA remains steady at 0.9704 as training progresses. At epoch 30, the loss on the training set is very low at 0.01, and the final accuracy on the training set stands at 0.9989. The VA, meanwhile, remains steady at 0.9669 with a loss of 0.07.

Figure 15a displays the model loss. Here, the training and validation loss against epochs is intended to prove that the model decreased its loss to as low as 30 epochs. The training loss at the outset is roughly 1.5, but by the last epoch, it sharply decreases and then stabilizes at 0.01. It exemplifies just how well the model can minimize errors on a training set. The VL also begins at 1.6 but has a comparable steep drop before leveling out to 0.07 by epoch 30. The model is generalizing well to new, unseen data without overfitting; the best VL epoch is 29, represented by a blue dot. The TA of the model grows rapidly as depicted in Figure 15b. It reaches nearly perfect by epoch 10 and remains stable. In addition to stability, VA peaks at 0.97-0.98, and 24 is identified as the best epoch. The difference between training and validation runs' accuracy is strongly indicative that the model is well-generalized and not subject to overfitting.

Figures 15c,d show the trend of the recall and precision of the model over the course of 30 epochs. The recall of the model, both for the training set and the validation set, increases rapidly in the initial epochs, with the recall exceeding 0.95 by the 10th epoch.

The maximum recall for the validation set occurs at epoch 24, after which the recall remains almost constant with minor fluctuations. Similarly, the precision of the model increases rapidly in the initial epochs, exceeding a precision of 0.95 by the 10th epoch. The maximum precision for the validation set occurs at

TABLE 14 FINE-TUNED Xception performance parameter analysis.

Classes	Precision	Recall	F-1 score	Accuracy	Cohen's kappa
Cloudy	0.921	0.945	0.933	0.982	0.975
Rain	0.999	0.999	0.999		
Shine	0.983	0.898	0.940		
Sunrise	1.0	1.0	1.0		
Macro Average Performance	0.975	0.960	0.968	0.982	0.975

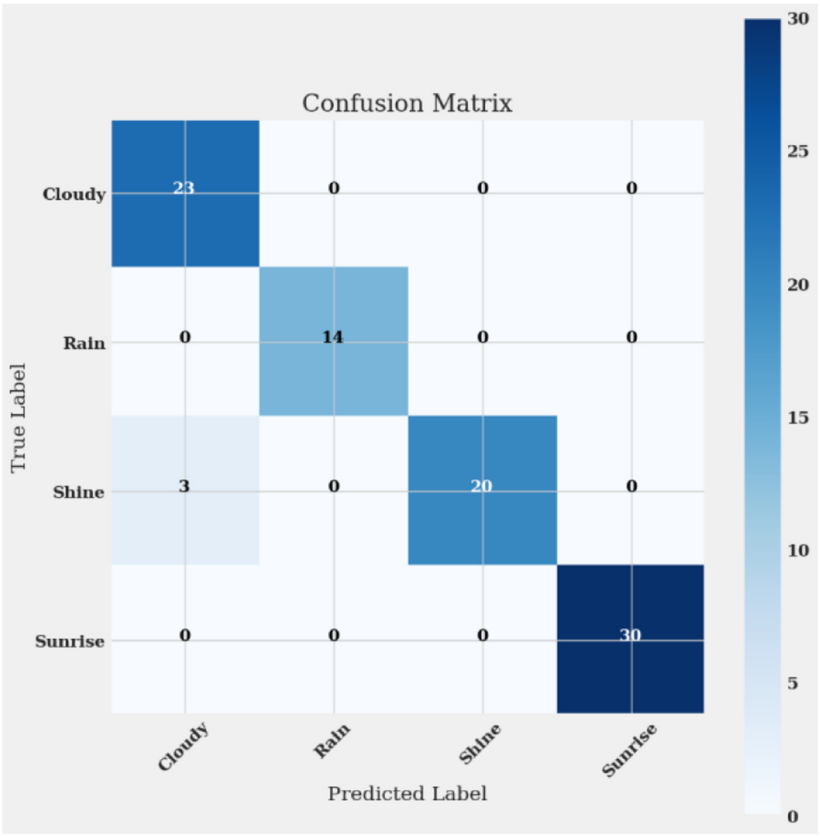

FIGURE 22
Confusion matrix for fine-tuned xception model.

TABLE 15 Accuracy and loss development for integrated Xception-senet model.

Epoch	Training loss (TL)	Training accuracy (TA)	Validation loss (VL)	Validation accuracy (VA)
1.	1.21	0.5844	1.59	0.4519
5.	0.14	0.9577	0.17	0.9332
10.	0.09	0.9812	0.10	0.9723
15.	0.04	0.9943	0.09	0.9789
20.	0.03	0.9954	0.08	0.9801
25.	0.02	0.9955	0.07	0.9812
30	0.01	0.9979	0.04	0.9906

epoch 18, after which the precision remains high for all epochs. Figure 16 Model's AUC-ROC Graphical Presentation of the fine-tuned ResNet152V2. Table 12 shows the class-wise performance metrics obtained during the evaluation phase. The performance of the Cloudy class shows high precision and recall, thus ensuring a balanced F1 score. The performance of the Rain class shows near-perfect results, with precision and recall values close to 0.99, thus ensuring reliable results. The precision and recall values for the Shine class are well above 0.92, thus ensuring reliable and consistent results. Overall accuracy and Cohen's kappa for fine-tuned ResNet152V2 are 0.976 and 0.964 respectively. The performance of the classes shows consistent performance, with high precision, recall, and F1 scores, thus ensuring the effectiveness of the proposed model in multi-class weather classification.

Figure 17 graphically shows the model's performance measures; from this, the values of these parameters can be visualized.

Confusion matrix achieved during the model's testing phase has been shown in Figure 18. From the figure, it can be analyzed that 22 images for the Cloudy class were correctly classified, and 1 image was misclassified. For the Rain class, 14 images were correctly classified. 21 images for the Shine class were correctly classified, and 28 images for the Sunrise class were correctly detected.

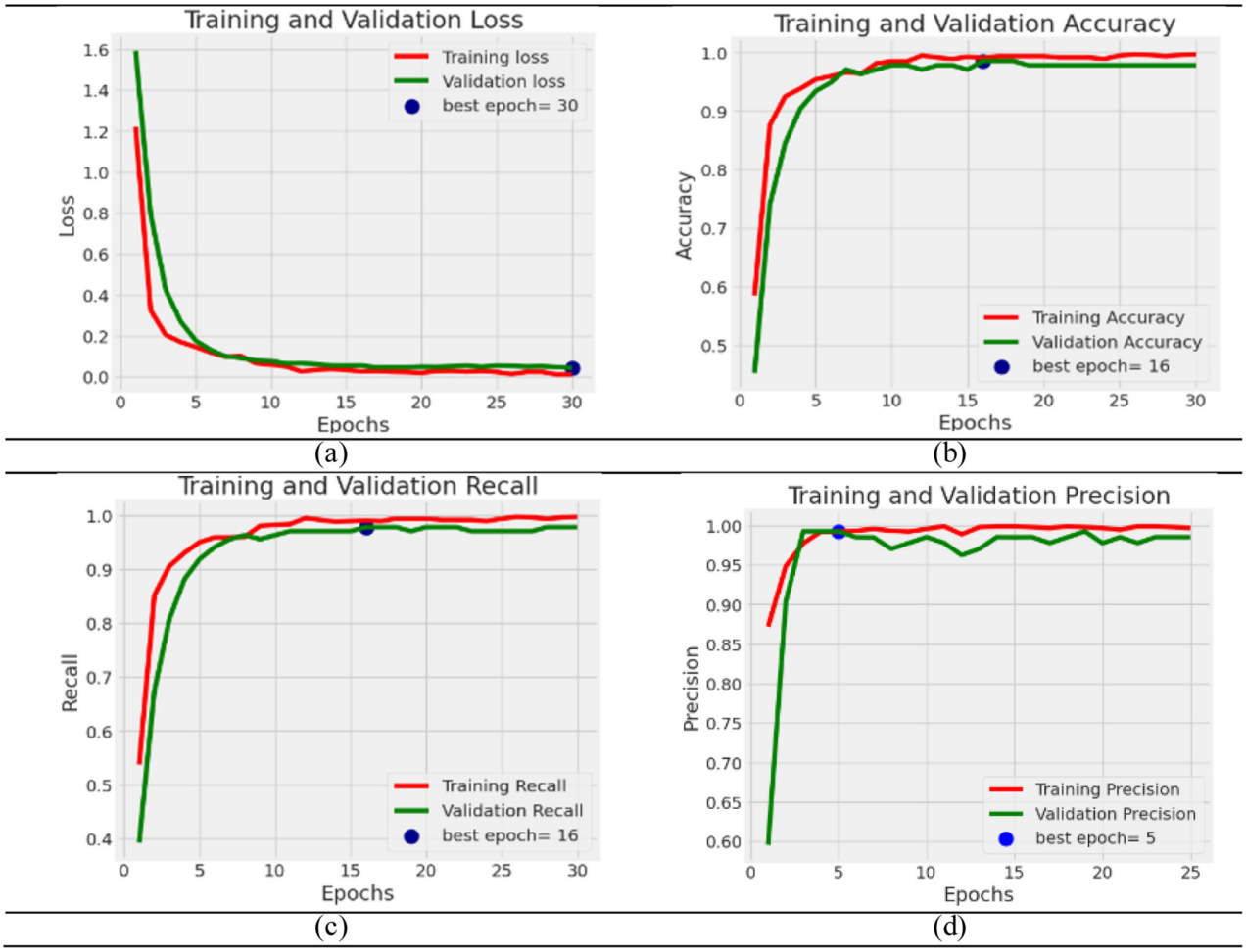

FIGURE 23
Accuracy and loss measurement for integrated Xception-SENet model **(a)** training and validation loss **(b)** training and validation accuracy **(c)** recall **(d)** precision.

4.4 Fine-Tuned Xception Model results

This section discusses the training and validation performance of the refined Xception model in terms of epochs. Table 13 shows the accuracy and loss of the model on each epoch during training and validation. The model starts with low accuracy and high loss because of the initialization of model weights. However, the model shows rapid improvement in performance in the initial few epochs. Both training and validation accuracy reach above 97% by the 5th epoch. The loss of the model also reduces substantially in these initial epochs. After the 10th epoch, validation accuracy is around 98.5%, along with a significant reduction in validation loss. However, after the 15th epoch, the training and validation performance of the model converges. Validation accuracy is around 98%, while validation loss is consistently low. Training accuracy also increases in later epochs compared to the initial epochs. However, validation performance does not increase in later epochs. Thus, the results show that the model generalizes well in the initial epochs and does not suffer from overfitting in later epochs.

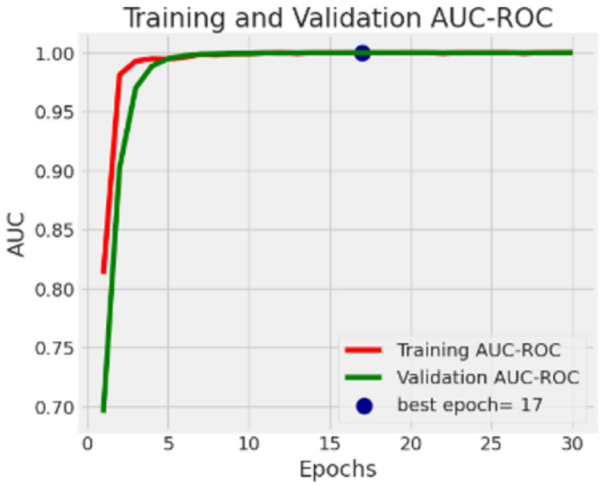

FIGURE 24
Integrated Xception-SENet model AUC-ROC curve.

TABLE 16 Integrated Xception-senet model's performance parameter analysis.

Classes	Precision	Recall	F-1 score	Accuracy	Cohen's kappa
Cloudy	0.942	1.0	0.971	0.9906	0.986
Rain	1.0	1.0	1.0		
Shine	1.0	0.911	0.955		
Sunrise	0.990	1.0	0.995		
Macro Average Performance	0.983	0.977	0.980		

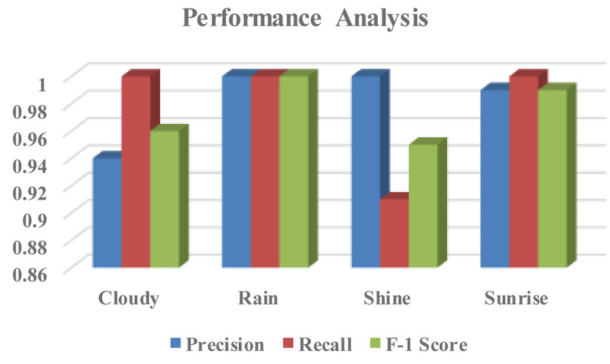

FIGURE 25
Presentation of performance analysis for integrated Xception-SENet model.

Figures 19, 20 display the graphical view of the model accuracy and loss curves along with the precision, recall and AUC-ROC for training and validation. The model loss is displayed in Figure 19a.

Figure 19A demonstrates that at first, both losses declined swiftly, with the TL dropping more quickly and stabilizing about epoch 10. Additionally, VL declines but stabilizes sooner, showing that the model has strong generality. The graphical representation of model accuracy is presented in Figure 19b. Both accuracies improve significantly, with VA stabilizing at 0.98 and TA attaining virtually 0.99 by epoch 10. The validation accuracy remains steady with additional training beyond epoch 5, suggesting little overfitting, whereas training accuracy increases marginally. Figure 19c displays the model's validation and training recall. By epoch 10, training and validation recalls have stabilized at 0.98 and are almost comparable, indicating that the model finds relevant positive events well. With early convergence at epochs 5–10 and consistent validation performance throughout, the model exhibits remarkable overall performance with no observable overfitting. Figure 19d displays the model's precision graph. The image, which displays that both the training and validation curves stabilize around 0.98, demonstrates that the model maintains a high degree of precision throughout epochs. The model reached peak precision early in the training phase, as indicated by the blue dot at epoch 16, which is the best epoch or the epoch with the highest precision.

Figure 20 illustrates the AUC-ROC performance of the fine-tuned Xception model. The similarity between the training and validation performance curves suggests that the learning is stable and that there is little overfitting in the model. The AUC-ROC values are close to 1, which implies that the classes are well separated from each other, and the best validation performance is obtained around epoch 22. Table 14 presents the performance metrics of the proposed model in detail.

The overall performance of the proposed model is high. The Rain and Sunrise classes are classified almost perfectly, and the Cloudy and Shine classes also have good performance metrics, although lower than the first two classes. For instance, the Shine class is classified with high precision and lower recall values, which implies that fewer instances are missed in classification. The overall accuracy is high, i.e., 0,982 and Cohen's kappa score of 0.975, which ensures the reliability of the model in multi-class classification of weather types. Figure 21 presents the performance results in a graphical manner.

Figure 22 displays the confusion matrix for the fine-tuned Xception model. From Figure 22 it can be seen that for the cloud, the classification model was perfect because all 23 instances of the class were classified under the Cloudy class without any instance being misclassified. The same is true for the Rain class where all 14 instances were correctly classified, meaning ideal performance was achieved.

In the case of Shine, the model got all 20 instances right but 3 went wrongly into the "Cloudy" class hence a slightly low recall value. Finally, in the Sunrise class, the model's performance was as follows: the model gave a perfect result for this class by accurately classifying 30 instances. In total, the confusion matrix shows that the model works well in differentiating between various weather conditions, and misclassification is the least between 'Shine' and 'Cloudy'.

4.5 Integrated Xception-SENet Attention model results

This part of the study discusses the performance evaluation of the proposed integrated fine-tuned Xception-SENet model. During the testing of the model, various hyperparameters have been recorded which are represented graphically. Table 15 gives the epoch-wise accuracy and loss development achieved during the training and validation of the proposed model. From the table, it can be seen that at Epoch 1, the model begins with relatively low accuracy about 0.5844 for training and 0.4519 for validation, as well as large loss values for both training of 1.21 and validation 1.59. Since the model is still learning and has not identified the patterns in the data, this is to be expected. A discernible improvement is seen by Epoch 5, when VA reaches 0.9332 and TA soars to 0.9577. The VL dramatically decreases to 0.17, indicating that the model is picking

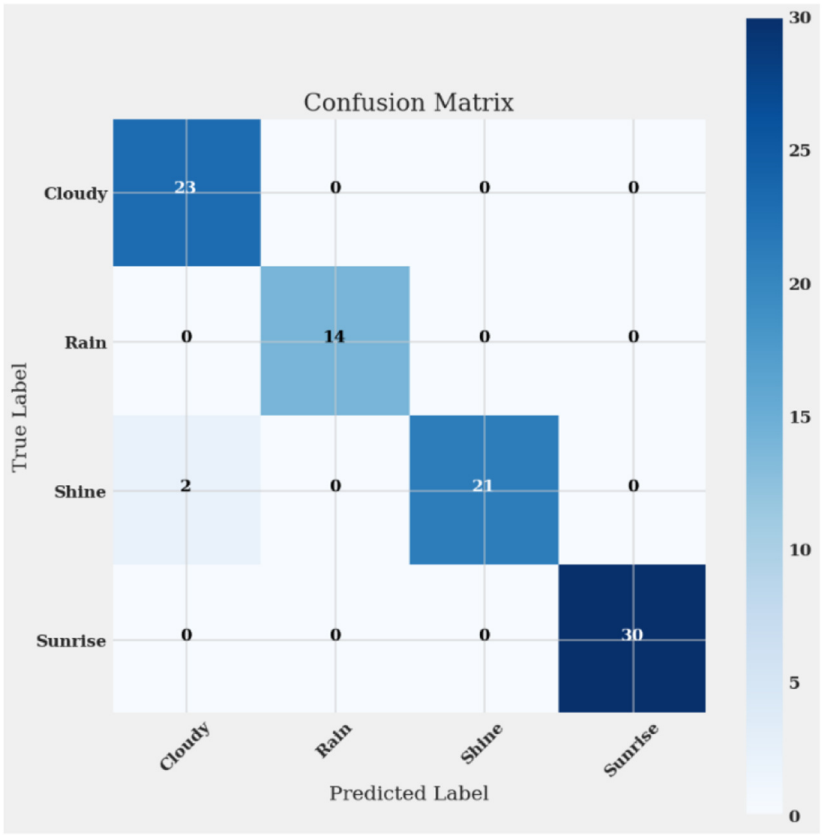

FIGURE 26
Confusion matrix for integrated Xception-SENet.

TABLE 17 Average performance parameter comparative analysis of the models.

Models	Accuracy	Precision	Recall	F-1 score
MobileNetV2	0.951	0.935	0.931	0.932
DenseNet121	0.958	0.949	0.941	0.945
EfficientNetB0	0.961	0.951	0.942	0.946
EfficientNetB3	0.965	0.948	0.950	0.949
ResNet152V2	0.976	0.951	0.952	0.951
Xception	0.982	0.975	0.960	0.968
Proposed Xception-SENet Attention	0.990	0.983	0.977	0.980

up speed. The model keeps improving at epoch 10, achieving 0.9723 accuracy on the validation set and 0.9812 accuracy on the training set. Additionally, loss values significantly drop, suggesting that the model is learning effectively. The performance stabilizes around epoch 15, with VA ranging from 0.98 to 0.99 and TA surpassing 0.99. By epoch 30, the VL has decreased to 0.04 and the TL has decreased to just 0.01; this indicates that the model is now highly tuned and operating effectively on unknown data.

Figure 23 shows the training and validation performance curves for the proposed model over 30 epochs. As shown in Figure 23a, the training and validation losses decline significantly in the initial epochs and then gradually decrease over subsequent epochs. This indicates that the training and validation losses are minimized for all epochs. Figure 23b shows the training and validation accuracy curves for the proposed model. As shown in the figure, training and validation accuracies increase significantly in the initial epochs and then increase gradually over subsequent epochs. This indicates that training and validation accuracies are maximized for all epochs. Figure 23c shows the training and validation recall curves for the proposed model. As shown in the figure, training and validation recalls increase significantly in the initial epochs and then increase gradually over subsequent epochs. This indicates that training and

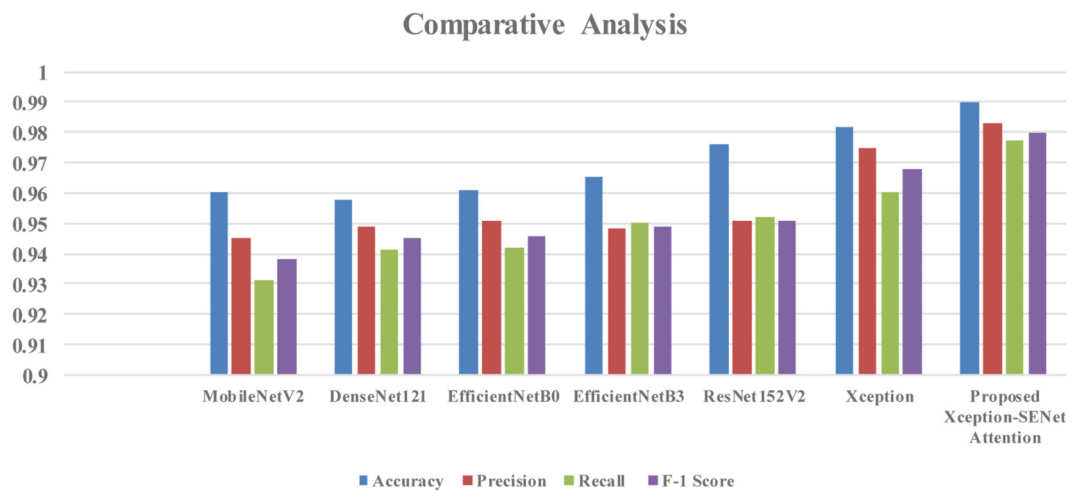

FIGURE 27
Comparative analysis of the models.

TABLE 18 Results for 5-FOLD cross-validation.

Fold	Accuracy	Macro precision	Macro recall	Macro F-1
Fold 1	0.988	0.984	0.983	0.984
Fold 2	0.992	0.987	0.986	0.987
Fold 3	0.989	0.985	0.982	0.983
Fold 4	0.991	0.986	0.985	0.986
Fold 5	0.990	0.986	0.984	0.985
Mean ± Std	0.990 ± 0.004	0.986 ± 0.001	0.984 ± 0.002	0.985 ± 0.002

TABLE 19 Quantitative ablation analysis.

Model composition	Accuracy	Precision	Recall	F-1	Parameters (millions)	FLOPs (GFLOPs)
Xception (baseline)	0.968	0.965	0.964	0.964	21.39	8.4
Xception + SE	0.979	0.978	0.977	0.977	21.67	8.6
Xception + Attention	0.982	0.981	0.980	0.980	22.10	9.1
Xception + SE + Attention (Proposed)	0.990	0.987	0.986	0.986	22.38	9.3

validation recalls are maximized for all epochs. The validation recall reaches a peak in the mid-to-late training phase, which indicates that optimal class detection is achieved for all epochs.

The model precision for training and validation sets throughout various epochs is shown in Figure 23d. Precision gauges how accurate positive predictions are, which is crucial given how expensive false positives may be. As can be seen, the model becomes quite effective at identifying true positives with few false positives as precision increases quickly over the first few epochs and stabilizes at 1.0. Epoch 5 has the highest validation precision.

The AUC-ROC for the training and validation sets throughout several epochs is shown in Figure 24. Excellent performance is indicated by a rating near 1.0. The graph illustrates how the AUC-ROC rises quickly in the initial epochs before stabilizing and approaching a perfect score. The classification report of Table 16 provides valuable insights into how well the model performs across different weather conditions.

For "Cloudy," the model has a precision of 0.942, a perfect recall of 1.0, and an F1-score of 0.971. This means it is generally good at recognizing cloudy instances, though it might sometimes label other conditions as cloudy. When it comes to "Rain," the model is flawless, achieving a precision, recall, and F1-score of 1.0, meaning it correctly identifies all rainy instances without any mistakes. For "Shine," it scores a precision of 1.0 and a recall of 0.911, leading to an F1-score of 0.955. This indicates that while the model accurately labels most "Shine" instances, it occasionally misses a few. In

TABLE 20 Statistical analysis of model variants.

Model A	Model B	Paired t-test (p)	Wilcoxon (p)	Significant (α = 0.05)
Baseline	SENet	0.8139	1.0000	No
Baseline	Attention	0.1082	0.1875	No
Baseline	Proposed	0.0086	0.0625	Yes*
SENet	Attention	0.1960	0.3125	No
SENet	Proposed	**0.0233**	0.0625	Yes*
Attention	Proposed	0.1079	0.1875	No

classifying "Sunrise", the model is nearly perfect, with a precision of 0.990, a recall of 1.0, and an F1-score of 0.995. Overall, the model's accuracy is 0.9906 and Cohen's kappa score of 0.986, demonstrating its strong ability to differentiate between the various weather conditions. Figure 25 shows the graphical view of the various performance parameters. From the graph, these parameters can easily be analyzed for the class Sunrise and Rain, the model has shown very good performance.

Confusion matrix of Figure 26 displays the model's performance for each weather situation in terms of actual versus projected labels. For instance, the model accurately identified 30 occurrences of "Sunrise," 14 occurrences of "Rain," and 23 occurrences of "Cloudy." In two instances, though, "Shine" was incorrectly classified as "Cloudy." Although there are a few mistakes in some categories, the matrix shows good performance overall.

An error analysis was performed using the confusion matrix of the proposed Xception SE-Net Attention model. As depicted in Figure 26, the overall misclassification rate is very low, as only two misclassified instances exist for the entire test set. The confusion matrix reveals the misclassification only between the Shine and Cloudy classes, where two instances of the Shine class are misclassified as instances of the Cloudy class. The confusion arises due to the visual similarity between bright overcast conditions and partially sunny days, where the cloudiness may dominate the visibility of clear illuminance conditions. The transitional weather conditions will always present some level of overlap in the visual characteristics, making them harder to distinguish using static image-based features. The Rain and Sunrise classes present a perfect classification performance for the test set.

4.6 Comparative analysis of the proposed xception-SENet-attention framework with TL models

Table 17 shows the comparative performance analysis of different transfer learning (TL) models for weather classification. Lightweight models such as MobileNet V2 and DenseNet121 have a reasonable accuracy but slightly less precision and F1-score. The EfficientNet models have shown promising performance by using a compound scaling technique. The performance of the models has been enhanced in the case of the ResNet152 V2 because of the deep structure of the network. Among the baseline models, the Xception model has shown the best performance by using the efficient technique of depth-wise separable convolution.

```
ROBUSTNESS TESTING PROTOCOL
Gaussian σ: 0.1
JPEG Quality: 30
Blur Kernel: 15x15
Perturbations applied ONLY during testing.
Model weights were NOT retrained.
```

FIGURE 28
Robustness testing protocol and perturbation settings applied during evaluation.

The proposed Xception–SENet–Attention model has shown the highest performance in all the metrics. The graphical representation of the comparative analysis of the models has been shown in Figure 27. The figure shows the performance gains of the proposed model.

4.7 Cross validation analysis for 5-fold

To evaluate the robustness and generalization capability of the proposed model beyond a single train-test split, a 5-fold cross-validation experiment was conducted on the training dataset. The dataset was partitioned into five equal subsets, where in each iteration, four folds were used for training and one-fold for validation. This process was repeated five times, ensuring that each fold served as the validation set exactly once. The performance metrics were averaged across all folds to obtain a more reliable and unbiased estimate of the model's generalization ability. The 5-fold cross-validation was implemented using stratified sampling with a fixed random seed (seed = 42), and fold-wise training logs are provided in the supplementary executable notebook to ensure reproducibility. Table 18 presents the 5-Fold cross validation results. These results demonstrate consistently high performance with low variability across folds, indicating that the model is stable and not overly dependent on a particular data partition.

4.8 Ablation analysis

An ablation study was conducted to systematically confirm the effectiveness of the proposed architectural design with four configuring results as given in Table 19 as (i) Xception Baseline, (ii) Xception +SE, (iii) Xception + Attention, and (iv)

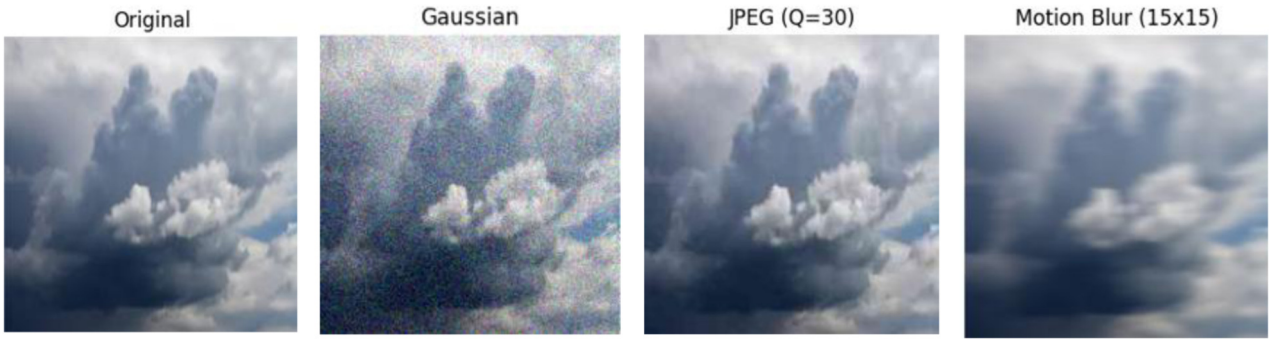

FIGURE 29
Sample perturbations applied during robustness testing (Gaussian noise σ = 0.1, JPEG quality = 30 quality = 30, Motion blur 15 × 15).

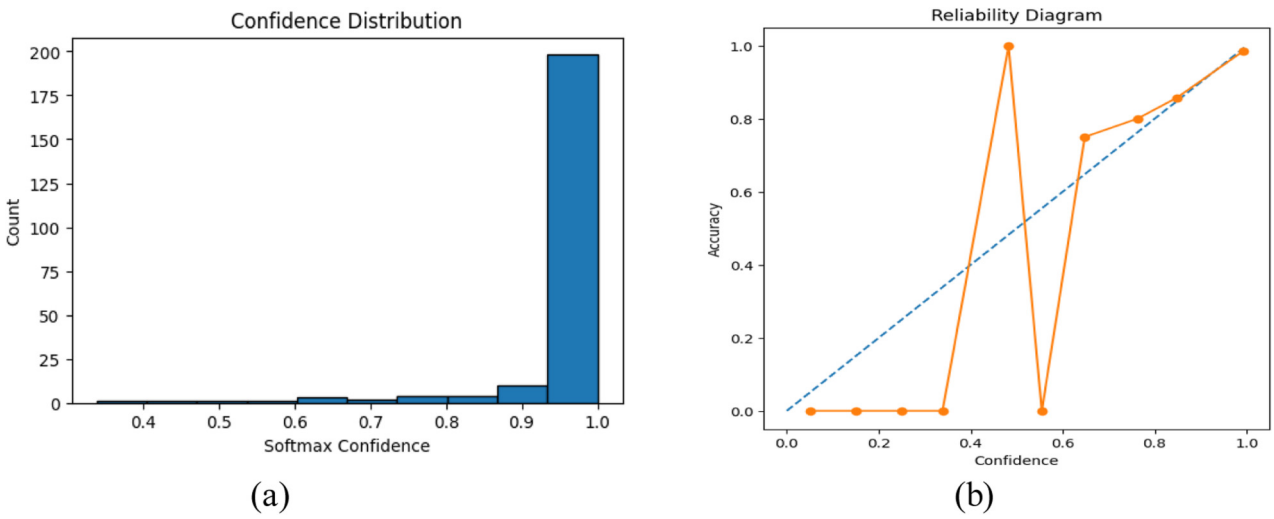

FIGURE 30
Calibration analysis of the proposed model showing **(a)** Distribution of Prediction Confidence Scores. **(b)** Reliability Diagram of the Proposed Model.

Xception +SE + Attention. As indicated in Table 20, the Xception model at the base level attained the accuracy of 96.8 with precision of 0.965 and recall of 0.964 and F1-score of 0.964. The accuracy rose to 97.9 when the SE module was included, and it rose with all the measures (Precision: 0.978, Recall: 0.977, F1: 0.977). This validates the use of channel-wise feature recalibration to improve discriminative ability. Likewise, incorporation of spatial attention mechanism also showed further enhancement and delivered an accuracy of 98.2 percent, a precision of 0.981, recall of 0.980, and F1-score of 0.980 that showed improvement in localization of spatial features. The most integrated model (Xception +SE + Attention) was the most powerful in all metrics as it has an accuracy of 99.0, precision of 0.987, recall of 0.986, and F1-score of 0.986. This is a 2.2% absolute enhancement in accuracy, which is complementary to channel and spatial attention mechanisms. Even though the new model comes with a minor increase in parameters (21.39M–22.38M) and FLOPs (8.4–9.3 GFLOPs), the computational overhead of the new model is moderate, as compared to the significant performance improvements. In general, the ablation findings are a clear indication that both SE and attention modules are positive and synergistic to the performance of overall classification.

4.9 Statistical validation analysis of model variants

Statistical validation comparison of model variants using paired t-test and Wilcoxon signed-rank test across 5-fold cross-validation has been presented in Table 20. The results indicate that the proposed model achieves statistically significant improvements over the Baseline ($p = 0.0086$) and SENet-enhanced variant ($p = 0.0233$) under the paired t-test at a significance level of $\alpha = 0.05$. However, comparisons involving Baseline vs. SENet, Baseline vs. Attention, SENet vs. Attention, and Attention vs. Proposed do not demonstrate statistical significance ($p > 0.05$), suggesting that individual architectural components do not independently yield substantial performance gains. Although the Wilcoxon test reports marginal significance ($p = 0.0625$) for comparisons involving the proposed model, this slight discrepancy is likely

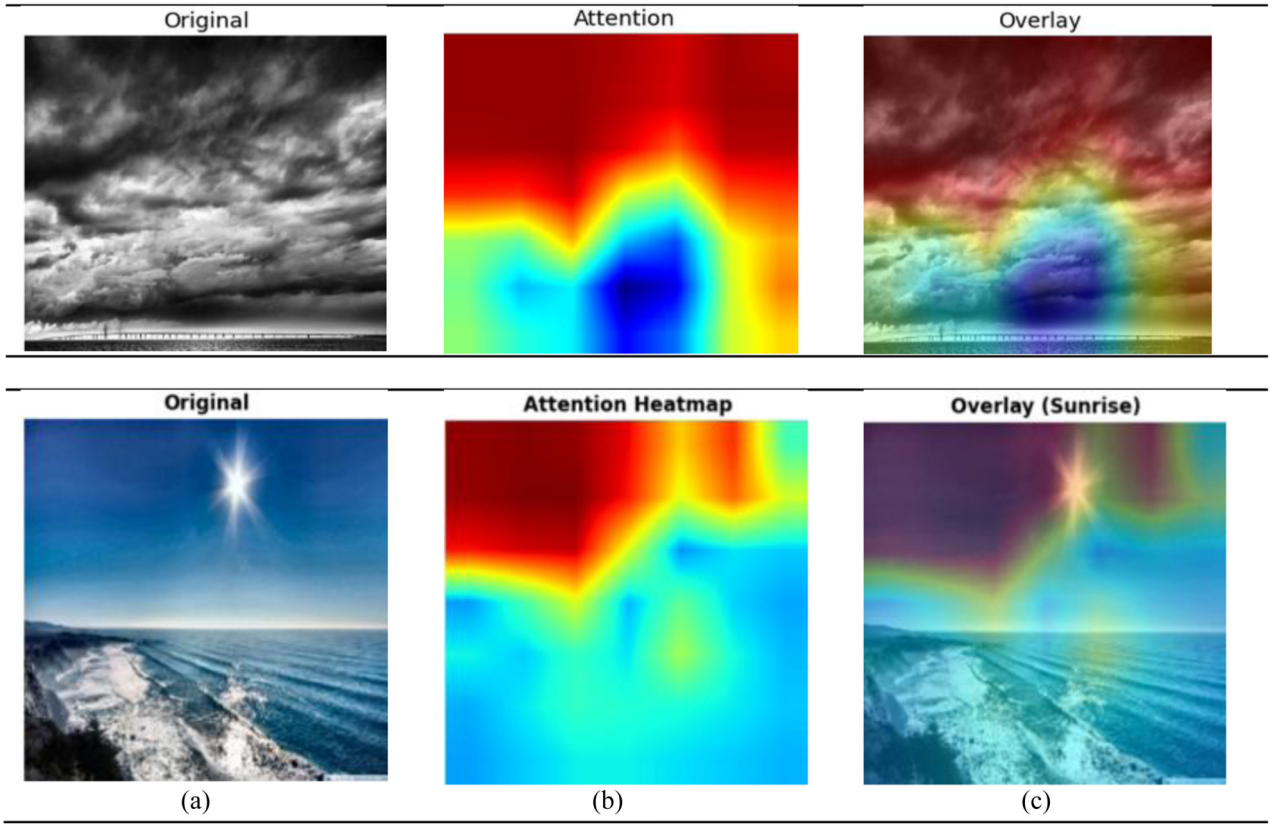

FIGURE 31
Spatial attention visualization for a sample image: **(a)** original image, **(b)** Attention heatmap, **(c)** Attention overlay.

attributable to the limited number of cross-validation folds. Overall, the statistical analysis confirms that the proposed integrated architecture provides a meaningful and statistically supported performance enhancement compared to its individual components.

4.10 Robustness evaluation under image perturbations

To test the robustness of the proposed model under realistic degradations, further robustness experiments were carried out by introducing controlled perturbations on the test data set. The aim was to test whether the proposed model still retains its classification performance under realistic degradations. For this purpose, three different types of perturbations were applied on the test data set: first, a Gaussian noise with a standard deviation $\sigma = 0.1$ was added; second, a JPEG compression with a quality factor of 30 was applied; and third, a motion blur with a kernel size of 15×15 was applied on the test data set. In this, perturbations were applied only on the test data set, and the model was not re-trained on these perturbed data sets. The testing protocol and parameter settings are summarized in Figure 28.

Some representative examples of the applied distortion are shown in Figure 29. As it can be seen, the distortion causes some quality degradation in the image. The above evaluation of robustness also supports the generalization of the proposed framework. Despite the noticeable degradation in image quality, the model demonstrates stable classification behavior with only moderate variations in performance, suggesting that the Xception–SENet–Attention framework does not purely depend on perfect image clarity.

Although performance decreases under severe perturbations, the accuracy remains above 90% across all tested conditions, indicating reasonable robustness of the proposed architecture. This stability test also supports the generalization potential of the model beyond clean image testing.

4.11 Model calibration and confidence analysis

To further test the reliability of the probabilistic results obtained by the proposed model, a calibration test was carried out through a reliability diagram and confidence distribution plot. As shown in Figure 30a, the histogram of the confidence distribution shows that the majority of the predictions are concentrated in the high-confidence interval of 0.95–1.0. This shows that the model makes strongly confident predictions for the majority of the samples. As the test accuracy of 99% has been achieved, this high concentration of confident predictions near 1.0 shows that the class separability is high and that the features have been learned in a highly discriminative manner by the Xception + SENet + Attention architecture.

To verify that these high levels of confidence are justified, the performance of the model on the task of calibration was also

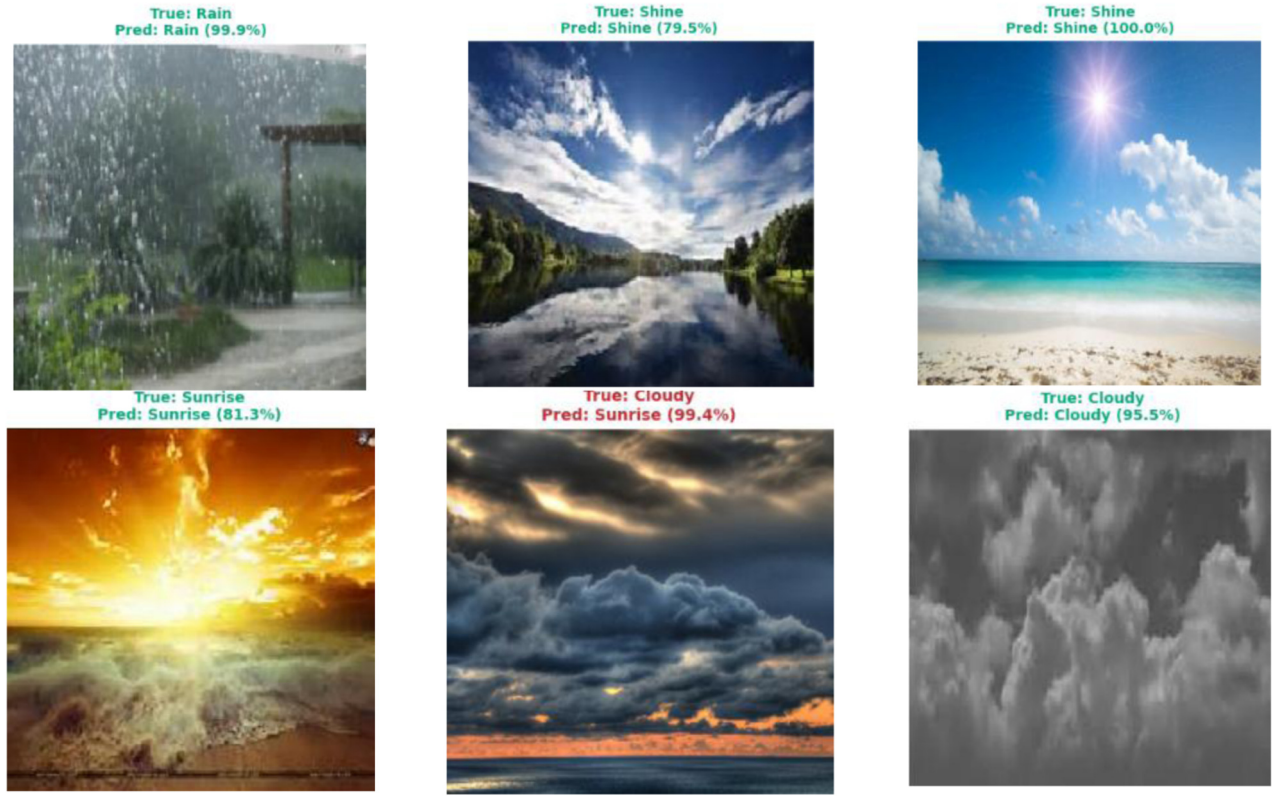

FIGURE 32
Visualization of the classification results.

evaluated using the reliability diagram and Expected Calibration Error (ECE) metric. The reliability diagram shown in Figure 30b depicts that the curve follows the ideal diagonal line closely, which means that the confidence of the predictions made by the model and the empirical accuracy of the predictions are in close agreement. The ECE value of 0.0183 is also low, which means that the difference between the confidence of the predictions and the true accuracy of the predictions is low and below 2%. This shows that the model not only performs with high accuracy but also does not suffer from overconfidence, which is often the problem with deep neural networks and models in general.

4.12 Model interpretability and attention visualization

To further understand the decision-making process of the proposed architecture, spatial attention heatmaps and overlay images for the test samples are created. The attention heatmaps indicate the regions of higher relational importance for the self-attention mechanism. As can be seen from Figure 31a, the model primarily focuses on dense clouds in cloudy images, darker clouds in rain images, and horizon illumination in images of sunrise. These visualizations again affirm that the model uses relevant meteorological features instead of background noise, thereby enhancing its interpretability. In Figure 31b attention heatmap highlights the areas of high-intensity lighting around the sun and the water, indicating the relevance of the classification decision to semantic features such as weather. Figure 31c shows the attention overlays; the overlay confirms that the model is effectively capturing discriminative features such as sunlight intensity, colour gradients in the sky, and illumination patterns reflected on the water surface.

4.13 Visualization of classification results

This section displays various classified results obtained during the model's testing on the unseen test set data. Figure 32 displays the classification results with actual and predicted classes for the images.

5 State of art comparison

This section compares the performance of the proposed Integrated Xception-SENet Attention Model against current research on classifying weather images into four categories: Sunrise, Shine, Rain, and Cloudy. Table 21 outlines this comparison, showing that the proposed model outperforms several existing models, including VGG16, ResNet50, Spiking Neural Networks (SNN), FCN-LSTM and a Hybrid model. Notably, it achieved an impressive accuracy of 99.06%, with an overall precision of 98.35%, recall of 97.75%, and F1 score of 96.09%. These outstanding results highlight the model's ability to classify weather images with enhanced efficiency and accuracy. The closest competitor to the Integrated Xception-SENet Attention Model is

TABLE 21 Comparison analysis of sota models.

Reference/ Published	Technique/Model used	Dataset/Number of images	Total classes	Performance parameters %
(Ghaleb et al., 2022)/2021	Hybrid CNN + SVM	Customized weather dataset/5,000	4	Accuracy – 94 Precision – 93 Recall – 95
(Gladh and Sahlin, 2021)/ 2022	VGG16	Swimcat/784 Hyta/32	2	Accuracy - 91.4 Accuracy – 89
(Schultz et al., 2021)/2022	ResNet50	DAWM 2020 + MCWCD 2018/1,656	6	Accuracy - 98.48 Precision - 98.51 Recall - 98.41
(Mittal and Sangwan, 2023)/ 2023	InceptionV3+LR	MWD/1,125	4	Accuracy - 97.77 Precision - 97.74 Recall - 97.75 F-1 Score - 97.74
(Jovanovic et al., 2024)/2024	Meta-based RF fusion	WEAPD/6,877	11	Accuracy – 96 Precision – 96 Recall – 95 F-1 Score – 95.5
(Guzel et al., 2024)/2024	ConvNextSmall	Multi-class weather image classification dataset/2,543	10	Accuracy – 97.27 Precision – 97.11 Recall – 96.95 F-1 Score – 97.03
(Shelke et al., 2025a)/2025	FCN-LSTM	MWD/1,125	4	Accuracy – 96.88 MSE – 7.11
Proposed Xception-SENet Attention Model		MWD/1,125	4	Accuracy – 99.06 Precision – 98.35 Recall – 97.75 F-1 Score – 96.09

ResNet50, which reached an accuracy of 98.48%. Another strong contender is the InceptionV3+LR model, which achieved an accuracy of 97.7%, precision of 97.74%, recall of 97.75%, and an F1 score of 97.74%. Overall, the Integrated Xception-SENet Attention Model stands out as the top performer among the models evaluated.

6 Conclusion, limitation and future work

In the present study, a novel integrated framework of Xception, SENet, and Attention has been proposed for multi-class weather image classification problems. Comparative evaluation of the proposed framework with other models, such as EfficientNetB3 and ResNet152V2, revealed that Xception provides the most suitable baseline model for the task. Additionally, the integration of SENet and attention further improves the performance of the classifier in distinguishing visually similar weather classes. The evaluation of the proposed model, including 5-fold cross-validation, statistical significance, calibration, and robustness, indicates the stability and generalization ability of the model. The proposed framework provides high performance in weather image classification with moderate computational complexity. Despite the fact that the dataset contains four categories of weather and a limited number of samples, the results have proven that the integration of dual-level feature refinement and transfer learning is a powerful and reliable method for weather image classification. Though the dataset is slightly class-imbalanced, the macro-averaged metrics were to be reported to make the classes have equal weights. This discourages the artificial inflating of the performance measures by the majority-class dominance. The confusion matrix also supports the fact that there is consistent classification performance even in the minority classes.

Although the empirical outcomes are strong, some weaknesses can be identified. First, the dataset is rather small and is restricted to four weather types, which can be a limitation of the range of environmental variations that people can come across in practice. Despite the use of cross-validation, augmentation and robustness testing to reduce the risks of overfitting, testing on larger and more heterogeneous datasets would provide further support to the claims of generalization. Second, the effort is based only on the features of the static images, and it does not use the temporal and multi-modal features, which can be useful in the complex weather. Lastly, in practice deployment conditions can also add more variation, like intense lighting, sensor noise or domain shifts that were not well covered in the existing data.

Future research can extend this work by exploring additional weather conditions, such as snow, hail, and thunderstorms, to create a more comprehensive classification system. Additionally, integrating real-time data from IoT devices and satellite imagery could enable on-the-fly classification, making the system highly adaptive for dynamic applications like autonomous vehicles and smart agriculture.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

GS: Conceptualization, Data curation, Methodology, Writing – original draft. SG: Data curation, Formal Analysis, Resources, Validation, Visualization, Writing – review and editing. AS: Investigation, Methodology, Project administration, Resources, Writing – review and editing. SI: Conceptualization, Data curation, Formal Analysis, Methodology, Writing – review and editing. AR: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Methodology, Writing – original draft. ID: Conceptualization, Data curation, Formal Analysis, Writing – review and editing. HM: Data curation, Formal Analysis, Funding acquisition, Writing – original draft. RC: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Methodology, Writing – review and editing. ÁC: Conceptualization, Formal Analysis, Funding acquisition, Methodology, Writing – review and editing. UK: Formal Analysis, Methodology, Writing – original draft, Writing – review and editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2026R140), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

References

Ajayi, G. (2018). *Multi-class weather dataset for image classification*. Florida, Johannesburg: Mendeley Data, V1. doi:10.17632/4drtyfjtfy.1

Al-Haija, Q. A., Gharaibeh, M., and Odeh, A. (2022). Detection in adverse weather conditions for autonomous vehicles *via* deep learning. *AI (Basel)* 3 (2), 303–317. doi:10.3390/ai3020019

Chollet, F. (2017). "Xception: deep learning with depthwise separable convolutions," in *Proc. IEEE conf. Comput. Vis. Pattern Recognit. (CVPR), Honolulu, HI, USA*, 1251–1258.

Dasu, B. T., Reddy, M. V., Kumar, K. V., Chithaluru, P., Ahmed, N., and Abd Elminaam, D. S. (2026). A self-attention driven multi-scale object detection framework for adverse weather in smart cities. *Sci. Rep.* 16, 1992. doi:10.1038/s41598-025-31660-4

Elhoseiny, M., Huang, S., and Elgammal, A. (2015). "Weather classification with deep convolutional neural networks," in *Proc. Int. Conf. Image Process. (ICIP)*, 3349–3353. doi:10.1109/ICIP.2015.7351424

Gdeeb, R. T. (2024). Weather classification using meta-based random forest fusion of transfer learning models. *Int. J. Adv. Intell. Inf.* 10 (2), 186. doi:10.26555/ijain.v10i2.1264

Ghaleb, M., Moushier, H., Shedeed, H., and Tolba, M. (2022). Weather classification using fusion of convolutional neural networks and traditional classification methods. *Int. J. Intell. Comput. Inf. Sci.* 0 (0), 1–13. doi:10.21608/ijicis.2022.117060.1156

Gladh, M., and Sahlin, D. (2021). *Image synthesis using CycleGAN to augment imbalanced data for multi-class weather classification*. Dissertation. Norrköping, Sweden: Linköping University.

Goswami, S. (2020). "Towards effective categorization of weather images using deep convolutional architecture," in *2020 international conference on industry 4.0 technology (I4Tech)*, 15, 76–79. doi:10.1109/I4Tech48345.2020.9102678

Gupta, A., and Goel, S. (2023). "Deep learning empowered weather image classification for accurate analysis," in *2023 IEEE 2nd international conference on industrial electronics: developments & applications (ICIDeA)*, 440, 567–572. doi:10.1109/ICIDeA59866.2023.10295216

Guzel, M., Kalkan, M., Bostanci, E., Acici, K., and Asuroglu, T. (2024). Cloud type classification using deep learning with cloud images. *PeerJ Comput. Sci.* 9 (e1779), e1779. doi:10.7717/peerj-cs.1779

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Identity mappings in deep residual networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV), Amsterdam, the Netherlands*, 630–645.

Hu, J., Shen, L., and Sun, G. (2018). "Squeeze-and-excitation networks," in *Proc. IEEE conf. Comput. Vis. Pattern Recognit. (CVPR), Salt Lake City, UT, USA*, 7132–7141.

Jovanovic, L., Bacanin, N., Simic, V., Mani, J., Zivkovic, M., and Sarac, M. (2024). Optimizing machine learning for space weather forecasting and event classification using modified metaheuristics. *Soft Comput.* 28 (7–8), 6383–6402. doi:10.1007/s00500-023-09496-9

Kalkan, M., Bostancı, G. E., Güzel, M. S., Kalkan, B., Özsarı, Ş., Soysal, Ö., et al. (2022). Cloudy/clear weather classification using deep learning techniques with cloud images. *Comput. Electr. Eng.* 102, 108271. doi:10.1016/j.compeleceng.2022.108271

Kang, L. W., Chou, K. L., and Fu, R. H. (2018). Deep learning-based weather image recognition. *Proc. Int. Symp. Comput. Consum. Control (IS3C)*, 384–387. doi:10.1109/IS3C.2018.00103

Li, S., Tian, W., Wu, X., Tan, C., and Cui, L. (2023). "Classification of multi-class weather image data," in *2023 9th international symposium on System security, safety, and reliability (ISSSR)*, 57, 203–207. doi:10.1109/ISSSR58837.2023.00038

Mittal, S., and Sangwan, O. P. (2023). Classifying weather images using deep neural networks for large scale datasets. *Int. J. Adv. Comput. Sci. Appl.* 14 (1). doi:10.14569/ijacsa.2023.0140136

Nalluri, S., Ramasubbareddy, S., and Kannayaram, G. (2019). Weather prediction using clustering strategies in machine learning. *J. Comput. Theor. Nanosci.* 16 (5), 1977–1981. doi:10.1166/jctn.2019.7835

Papadimitriou, O., Kanavos, A., Mylonas, P., and Maragoudakis, M. (2023). "Advancing weather image classification using deep convolutional neural networks," in *2023 18th international workshop on semantic and social media adaptation & personalization (SMAP)18th international workshop on semantic and social media adaptation & personalization (SMAP 2023)*, 157, 1–6. doi:10.1109/SMAP59435.2023.10255190

Ranjith Kumar, M., Sudeesh Kumar, V., Tharun Kaarthik, G. K., Chatiyode, V., Srinivasan Anusha, J., and Revathi, P. (2024). "Snowfall forecasting with enhanced accuracy: leveraging multi-class SVM for meteorological predictions," in *IFIP advances in information and communication technology* (Cham: Springer Nature Switzerland), 312–328.

Samo, M., Mafeni Mase, J. M., and Figueredo, G. (2023). Deep learning with attention mechanisms for road weather detection. *Sensors (Basel)* 23 (2), 798. doi:10.3390/s23020798

Schultz, M., Reitmann, S., and Alam, S. (2019). "Classification of weather impacts on airport operations," in *2019 Winter Simulation Conference (WSC)*. doi:10.1109/WSC40007.2019.9004915

Schultz, M., Reitmann, S., and Alam, S. (2021). Predictive classification and understanding of weather impact on airport performance through machine learning. *Transp. Res. Part C Emerg. Technol.* 131, 103119. doi:10.1016/j.trc.2021.103119

Sharma, G., Anand, V., Chauhan, R., Kukreti, S., and Gupta, S. (2023). "Celestial insights: an innovative convolutional neural network for high-resolution satellite remote sensing image classification," in *2023 3rd international conference on smart generation computing, communication and networking (SMART GENCON)*, 1–7. doi:10.1109/SMARTGENCON60755.2023.10442686

Shelke, N., Maurya, S., Ithape, R., Shaikh, Z., Somkunwar, R., and Pimpalkar, A. (2025a). Towards an automated weather forecasting and classification using deep learning, fully convolutional network, long short-term memory. *Int. J. Electr. Comput. Eng. (IJECE)* 15 (2), 1868–1879. doi:10.11591/ijec.v15i2.pp1868-1879

Shelke, N., Maurya, S., Ithape, R., Shaikh, Z., Somkunwar, R., and Pimpalkar, A. (2025b). Towards an automated weather forecasting and classification using deep learning, fully convolutional network, and long short-term memory. *Int. J. Electr. Comput. Eng. (IJECE)* 15 (2), 1868. doi:10.11591/ijece.v15i2.pp1868-1879

Shrivastava, V. K., Shrivastava, A., Sharma, N., Mohanty, S. N., and Pattanaik, C. R. (2023). Deep learning model for temperature prediction: an empirical study. *Model. Earth Syst. Environ.* 9 (2), 2067–2080. doi:10.1007/s40808-022-01609-x

Sun, W., Abdullah, L. N., Khalid, F. B., and Sulaiman, P. S. B. (2024). "Pioneering driver safety: evaluating weather impacts with the multi-class-weather algorithm," in *2024 7th international conference on electronics, communications, and Control Engineering (ICECC)*, 1–6. doi:10.1109/ICECC63398.2024.00008

Tan, M., and Le, Q. V. (2019). "EfficientNet: rethinking model scaling for convolutional neural networks," in *Proc. 36th Int. Conf. Mach. Learn. (ICML)* (Long Beach, CA, USA), 6105–6114.

Tian, E., and Kim, J. (2023). Improved vehicle detection using weather classification and Faster R-CNN with dark channel prior. *Electron. (Basel)* 12 (14), 3022. doi:10.3390/electronics12143022

Vaswani, A. (2017). "Attention is all you need". Long Beach, CA, USA, 5998–6008.

Venkatachalam, K., Trojovský, P., Pamucar, D., Bacanin, N., and Simic, V. (2023). DWFH: an improved data-driven deep weather forecasting hybrid model using Transductive Long Short Term Memory (T-LSTM). *Expert Syst. Appl.* 213, 119270. doi:10.1016/j.eswa.2022.119270

Xu, H., Zhao, Y., Dajun, Z., Duan, Y., and Xu, X. (2025). Exploring the typhoon intensity forecasting through integrating AI weather forecasting with regional numerical weather model. *Npj Clim. Atmos. Sci.* 8 (1), 38. doi:10.1038/s41612-025-00926-z

Zhang, Z., Liu, Z., Martin, A., and Zhou, K. (2023). BSC: belief shift clustering. *IEEE Trans. Syst. Man, Cybern. Syst.* 53 (3), 1748–1760. doi:10.1109/TSMC.2022.3205365

Zhao, Y., Lu, M., Chen, D., and Zhang, L. (2024). Understanding the weakening patterns of inner Tibetan Plateau vortices. *Environ. Res. Lett.* 19 (6), 064076. doi:10.1088/1748-9326/ad519

Zhao, Y., Lu, M., Zhang, L., and Cheng, T. F. (2025). Remote influence of Southern Tibetan Plateau heating on North Pacific atmospheric Rivers. *J. Clim.* 38 (1), 101–116. doi:10.1175/JCLI-D-23-0706.1

Zheng, L., Lu, W., and Zhou, Q. (2023). Weather image-based short-term dense wind speed forecast with a ConvLSTM-LSTM deep learning model. *Build. Environ.* 239 (110446), 110446. doi:10.1016/j.buildenv.2023.110446