Journal of Innovative Image Processing (ISSN: 2582-4252)
www.irojournals.com/iroiip/

MG-ResViT: Dynamic Residual Learning with Contrastive Feature Optimization and PCA-Optimized Cross-Block Feature Fusion for Fine-Grained Mangrove Species Classification

Jasten Keneth D. Treceñe[1], Arnel C. Fajardo[2]

College of Computing Studies, Isabela State University – Cauayan Campus, Cauayan City, Philippines.
E-mail: [1]jastenkenneth.trecene@evsu.edu.ph, [2]acfajardo@gmail.com

Abstract

Mangrove conservation and monitoring are critically important for biodiversity. However, accurate classification remains challenging due to the morphological similarities among species. This paper proposes MG-ResViT, a novel deep learning framework that enhances mangrove species feature extraction for classification using a dynamic residual connection with spatially adaptive attention gates that capture discriminative local features, a hybrid loss that combines supervised contrastive learning and cross-entropy for optimizing feature space geometry, and PCA-optimized cross-block feature fusion for efficient multi-scale feature integration. The proposed model was evaluated using a ground-truth dataset of 3 mangrove species, composed of 1,000 images per species, which underwent preprocessing and data augmentation. Results revealed that the proposed MG-ResViT achieved an overall accuracy of 92.8% with only 6.2M parameters compared to other state-of-the-art models. Based on the results from the ablation studies conducted, the full MG-ResViT model provided excellent feature learning capability compared to the other model variants, with a high reduction in inter-class similarity (0.210) and improved in intra-class similarity (0.893). The silhouette scores also indicated that the full model has a well-defined and compact cluster

DOI: https://doi.org/10.36548/jiip.2025.2.008
Received: 02.05.2025, received in revised form: 01.06.2025, accepted: 15.06.2025, published: 24.06.2025

(0.68) compared to other model variants such as the baseline EfficientNet-B0 + CE with 0.44, + SupCon only with 0.58, and + Dynamic Residuals only with 0.65. Moreover, the comparative analysis showed MG-ResViT (92.8%) outperformed ViT-Small (91.2%), ResNet-50 (89.3%), DenseNet-121 (90.0%), and EfficientNet-B0 (88.0%) in both accuracy and computational efficiency. Thus, the proposed MG-ResViT model has the potential for a more accurate fine-grained mangrove species classification, which is important for conservation and monitoring.

Keywords: Deep Learning, Dynamic Residual Networks, Ecological Conservation, Fine-Grained Visual Recognition, Mangrove Species Classification.

1. Introduction

Mangrove forests are commonly found along tropical and subtropical coastlines and form a unique wetland ecosystem that provides habitat and food for diverse marine and terrestrial species [13], [14]. They also help with coastal protection and contribute significantly to carbon sequestration [12]. Despite the increasing recognition of their ecological significance, mangroves remain among the most threatened ecosystems in the world. Deforestation, urbanization, and other climate-related stressors have caused the mangrove ecosystem to decline over the years. In the Philippines, the mangrove forests have suffered a decline of 10% between 1990 and 2010 [25]. In the Eastern Visayas region, mangrove forests also experienced a massive decline, with a loss of 8,800 ha after Super Typhoon Haiyan in 2013. Although various mangrove rehabilitation initiatives have been initiated, many were unsustainable because of poor species selection and monitoring [26]. In the local context, mangrove rehabilitations have faced significant challenges due to incorrect species selection, improper planting strategies, and the absence of experts in mangrove species identification among the local government and residents [26]. Additionally, three true mangrove species are most commonly found and thrive in the area: *Avicennia alba, Rhizophora apiculata, and Sonneratia alba* which makes them critical to local conservation efforts. Furthermore, there is a lack of accessible and ground-truth mangrove datasets available, as most existing research relies only on remote sensing and synthetic data.

Accurate species-level monitoring is essential for effective conservation; however, manual field surveys are labour-intensive and fundamentally subjective [12]. Furthermore, classification remains challenging due to the high morphological similarities of mangroves [12], phenotypic variation due to environmental factors, and overlapping structural characteristics [2]. Although there have been various advances in deep learning, it is still challenging for current methods, particularly for mangrove species classification.

Recent studies reveal a number of limitations pertaining tothe classification of mangrove species. Convolutional Neural Networks (CNNs), such asResNet [20], MobileNet [3], and EfficientNet [3], are not able to capture the fine visual discriminations of mangrove species when differentiating among morphologically similar groups. According to Wang et al., even the best model could only achieve an accuracy score between 82% and 85% with datasets of mangrove species due to high intra-class variation and low inter-class variance [19]. Additionally, most existing networks with residual connections have fixed skip paths [21] that do not appropriately address the spatial heterogeneity of the diagnostic features of mangrove leaves and bark. While contrastive learning has found success in standard recognition, as noted in the study by Wang et al., few empirical studies on the application of mangrove classification, especially for fine-grained data, have been conducted [19]. Moreover, current methods continue to rely solely on cross-entropy [1].

Recent work by Liu et al. introduced the use of attention mechanisms for plant species classification [10]. However, it remainschallenging to address gradient instability in the deeper layers. Additionally, other CNN-based models have struggled to accurately focus on the spatial locations of plants, hindering the extraction of relevant features for classification tasks [10]. Similarly, Devarajan et al. proposed a hybrid CNN–Transformer architecture designed for emotion recognition from EEG signals [5]. However, their method requires computational resources that are impractical for field deployment. In the specific context of mangrove classification, Tan et al. achieved moderate performance through data augmentation [16], while Li et al. demonstrated the benefits of transfer learning [9]. Nevertheless, these approaches face difficulties in explicitly optimizing both feature discriminability and computational efficiency which is an essential requirement for conservation applications in resource-limited settings.

To address these gaps, this study proposed MG-ResViT, an architecture that integrates dynamic residual connections with spatially adaptive attention gates, a hybrid loss function combining supervised contrastive learning with cross-entropy, and a computationally efficient cross-block feature fusion module with PCA-based compression to reduce feature redundancy while maintaining model compactness and capturing multi-scale features. To the best of our knowledge, this is the first study to apply this hybrid CNN-Transformer with SupCon and dynamic residuals to real-world and field-captured images of mangrove species. This work is structured as follows: related works are discussed in Section 2, the proposed methodology in Section 3, the results in Section 4, and the conclusion in Section 5

1.1 Key Contributions of the Study

- For a comprehensive contextual understanding of mangrove image patterns, the proposed MG-ResViT combines EfficientNet-B0 for local spatial feature extraction and a Transformer module for global dependencies.
- The use of patch-adaptive dynamic residual blocks, a novel residual design where channel-wise attention scores control the dynamic fusion between identity mapping and convolutional transformation, helped improve gradient flow and enhanced local adaptability across varied image patches. This method represents a new form of spatial attention-enhanced skip connections.
- The integration of dual-head optimization with supervised contrastive learning featuresa dual-branch output design that enables simultaneous optimization for classification and representation learning. Supervised contrastive loss and cross-entropy loss are jointly applied to enhance intra-class compactness and inter-class separability in the learned feature space.
- A lightweight PCA-based cross-block feature fusion strategy was proposed to compress and integrate multi-scale features from the CNN and Transformer branches. This approach improved feature reuse while maintaining discriminative power with reduced dimensionality.
- The proposed architecture was evaluated on field-collected images of mangrove species. Extensive ablation studies and comparisons with state-of-the-art models were performed to validate the effectiveness of the proposed method for the challenging fine-grained mangrove species.

1.2 Objectives of the Study

The primary objective of this study is to design a novel hybrid architecture to address the challenges of fine-grained mangrove species classification. Specifically, this study seeks to:

- Develop MG-ResViT, a hybrid architecture that combines an EfficientNet-B0 backbone, Transformers, and dynamic residual connections with spatially adaptive attention gates for enhanced feature extraction.

- Optimize both local discriminative learning and global class separation by integrating a hybrid loss function that combines supervised contrastive learning (SupCon) and cross-entropy loss.

- Compress multi-scale features and reduce feature redundancy by implementing PCA-based cross-block feature fusion.

- Evaluate the performance of the model using standard evaluation metrics, compare the proposed model to other state-of-the-art models, and perform ablation studies to evaluate the individual contributions of the model variants to the overall performance.

2. Related Works

Recent advancements in dynamic neural networks have revolutionized adaptive feature extraction. However, their potential for ecological applications remains underexplored. Alzubaidi et al. introduced a learnable routing mechanism to improve computational efficiency, achieving a 40% reduction in FLOPs for ImageNet classification through dynamic branch selection [2]. Nevertheless, their evaluation excluded fine-grained ecological datasets, where spatial feature importance varies significantly in mangroves, as leaf veins and bark textures differ in diagnostic value. In the study by Schlemper et al., the authors produced satisfactory performance with their proposed attentional-gated skip connections for medical image classification; however, their method requires high computational resources [15]. Their approach also assumes that feature relevance within object boundaries is uniform and struggles to account for scenarios where key diagnostic traits occupy only a small portion of the image. Although their approach is effective for medical images with similar structures, this is a common situation when distinguishing species with similar morphological traits.

The study by Khosla et al. demonstrated that the supervised contrastive loss (SupCon) outperforms well-established alternatives based on cross-entropy loss for object classification in general, as its strength lies in augmenting the separability of features from objects [7].

SupCon performed marginally better than cross-entropy by 1% on datasets such as CIFAR-10 and ImageNet [7]. Contrastingly, SupCon is used for creating a compact intra-class cluster through better inter-class separation by optimizing feature space geometry. Its constraint depends largely on batch statistics for sampling positive and negative pairs. Accordingly, obtaining an estimate of gradients has considerable scope for bias because batch sizes are not large enough to represent all classes, which would later affect long-tailed mangrove datasets.

Furthermore, there is still a challenge to the standard application of supervised contrastive learning, wherein there is no principled way to balance the preservation of global structure and the extraction of local discriminative features. Moreover, traditional SupCon treats all features equally, which overlooks the different spatial regions that contribute variably to species discrimination and the fact that feature importance shifts across network depth. These limitations motivated the approach to combine the feature separation capabilities of SupCon with the stable classification performance of cross-entropy loss. This integration introduces dynamic feature weighting to account for spatial and depth-wise variations in feature importance. Additionally, the introduction of cross-stage partial connections has demonstrated the benefits of selective feature reuse in deep networks [11].

Several studies have been conducted on mangrove species classification. Zhang et al. employed a data augmentation strategy and achieved an overall accuracy of 92.1% [23]. However, these methods treat all image regions equally during feature extraction and fail to explicitly optimize the feature space geometry. Furthermore, various hybrid models have been proposed for image classification. Devarajan et al. proposed a hybrid CNN-Transformer architecture and achieved an 87.00% classification accuracy compared to AlexNet (83.50%), VGG-16 (85.00%), ResNet-50 (85.50%), GoogleNet (85.00%), and MobileNetV2 (86.00%) [5]. A hybrid identification method was proposed that combines the time-frequency threshold of the mangrove index with a random forest binary classifier and achieves an overall accuracy of 92.86% [6].

Based on the previous studies, various limitations were found in fine-grained recognition and mangrove classifications. Wang et al. produced only a low classification accuracy between 82% and 85% when trained on mangrove species because of high intra-class variation and morphological similarities [18]. Xu et al. cited that most existing residual connections have skip paths that do not appropriately address the spatial heterogeneity of the diagnostic features of mangrove leaves [21]. The attentional-gated connections of Schlemper et al. assume uniform relevance and struggle with partially occluded or local features [15]. Moreover, the supervised contrastive loss of Khosla et al. suffers from batch-wise sampling bias, which becomes less effective in spatially complex tasks [7]. Furthermore, transfer learning and data augmentation methods used in the studies of Li et al. [9] and Tan et al. [16] face difficulties in optimizing both feature discriminability and computational efficiency. Although the proposed hybrid CNN-Transformer model of Devarajan et al. [5] is accurate, it is computationally heavy and may be challenging for field deployment. These limitations formed the basis of the proposed MG-ResViT, which integrates adaptive residuals, transformer modules, dual-head optimization using contrastive learning, and a feature fusion strategy in a lightweight and effective architecture. The present work advances the field by simultaneously addressing both limitations through principled architectural innovations.

3. Materials and Methods

This study proposed a novel hybrid deep learning model for mangrove species feature extraction and classification, which integrates a CNN backbone with dynamic residual learning and supervised contrastive learning. The proposed model was evaluated using a real-world dataset of mangrove leaf images from three species. The full pipeline includes data preparation, proposed model architecture design, loss formulation, training strategy, and evaluation. The entire model simulation and performance evaluation were conducted using Python 3.10.5 and Jupyter Notebook on a workstation equipped with an NVIDIA GeForce RTX 3050 GPU, an AMD Ryzen 5 processor, and 16 GB of RAM.

3.1 Data Preparation and Preprocessing

The data used in this study were gathered from the mangrove sites in the Municipality of Tanauan, Leyte, Philippines. Three (3) mangrove species were used in the study: *Avicennia alba*, *Rhizophora apiculata*, and *Sonneratia alba*. These are the common species found in the area. The dataset used for the analysis was a combination of ground truth and augmented data using common augmentation techniques. The ground-truth images were collected through field sampling under varying environmental conditions, such as different lighting, distances, and angles, to ensure a diverse dataset that captured the natural variability in the appearance of mangroves. The Field Guide to Philippine Mangroves [14] was utilized, and collaboration with mangrove experts ensured accurate labeling of the dataset. Moreover, a custom PyTorch class was used to load and label the images. For the augmentation, 500 images per species were generated using controlled variations in appearance while preserving the key morphological features of mangrove leaves. In total, the dataset is composed of 3,000 images, with 1,000 images per species. Figure 1 shows the three mangrove species used in the study.

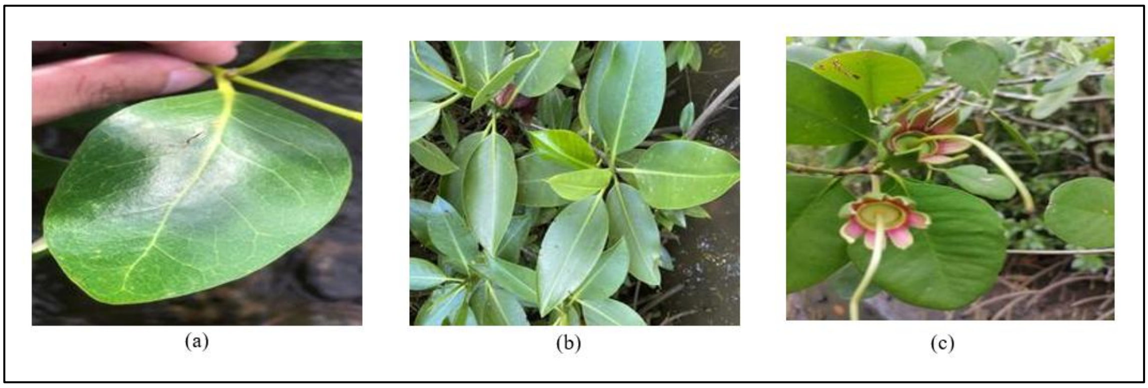

Figure 1. Sample of the Mangrove Species Dataset Used in the Study: (a) *Avicennia Alba*, (b) *Rhizophora Apiculata,* and (c) *Sonneratia Alba*

The images were pre-processed using a series of transformations to ensure consistency across the dataset. All images were resized to a standard resolution of 224x224 pixels, normalized using the ImageNet mean and standard deviation, and transformed into PyTorch tensors. Additionally, the dataset was divided into 80% for training and 20% for validation using random sampling to maintain class balance.

3.2 Proposed Method

The proposed MG-ResViT model is a hybrid architecture that integrates EfficientNet-B0 as the backbone, dynamic residual blocks, and dual-head output. EfficientNet-B0 was utilized because of its superior trade-off between accuracy and computational efficiency. Unlike traditional CNNs, EfficientNet-B0 uses compound scaling to optimize network depth, width, and resolution, making it suitable for ecological applications with constrained computing resources. Moreover, the model utilized its architecture as a feature extractor. This backbone generated a 1280-dimensional spatial feature map from the input image pre-trained on ImageNet. This ensures that there is an effective low- and mid-level representation of learning.

To adaptively fuse the original features with their transformations, four patch-adaptive dynamic residual blocks were introduced. Each block utilized an attention mechanism to compute a channel-wise importance score (α) that controls the mix between identity mapping and convolutional transformation. It is computed using Equation 1.

$$Output = (1 - \alpha) \cdot Conv(x) + \alpha \cdot x \quad (1)$$

Furthermore, when $\alpha = 1$, the block will prioritize the original features and perform identity mapping, subsequently preserving the spatial integrity. On the other hand, if $\alpha = 0$, the block will rely more on the transformed features which will allow for a flexible mix that enables the network to dynamically decide on how much transformation is needed for each of the channels. The proposed adaptive fusion mechanism helps in handling variations across patches in an image. It enables the model to preserve essential spatial structure and structural cues when necessary. Moreover, it also allows selective enhancement of discriminative and task-relevant features. This dynamic behavior makes the model more robust and context-aware and improves its ability to focus on meaningful patterns and reduce overfitting to irrelevant details.

A transformer module was integrated into the model based on the Vision Transformer (ViT) architecture. After the extraction of features using the EfficientNet-B0 and dynamic residual blocks, the resulting feature map $F \in \mathbb{R}^{H \times W \times C}$ was divided into a sequence of N non-overlapping patches. Each of these non-overlapping patches has the size of $p \times p$, and is flattened into vectors $x_i = \in \mathbb{R}^{P}$, where $P = p \times p \times C$. Each patch embedding is linearly projected and enriched with positional encodings as given by:

$$z_i = Ex_i + p_i, \quad for\ i = 1,2, \dots, N \quad (2)$$

Where $E \in \mathbb{R}^{D \times P}$ is a learnable projection matrix and $p_i \in \mathbb{R}^{D}$ is the positional encoding vector. The set of patch embeddings $Z = \{z_1, z_2, \dots, z_N\}$, was then passed through a stack of Transformer encoder layers. Each layer consisted of multi-head self-attention (MHSA), layer normalization, and a feed-forward MLP with residual connections. It is given as follows:

$$Z' = MHSA\big(LN(Z)\big) + Z \quad (3)$$

$$Z" = MLP(LN(Z')) + Z' \quad (4)$$

This process allows the model to learn long-range dependencies across spatial regions and complements the local spatial features extracted by the CNN. Moreover, the enhanced features of the Transformer were subsequently fused with convolutional branch using PCA-based cross-block fusion, which allowed the model to combine both global and local information before proceeding to the classification and contrastive learning tasks.

The dual-head output architecture served two distinct learning objectives using a shared feature representation: the classification head for supervised learning, and the projection head for self-supervised or semi-supervised learning via contrastive objectives. After the feature extractor, a 1280-dimensional pooled feature vector $f \in \mathbb{R}^{1280}$ was obtained. This shared feature vector was simultaneously passed through two separate heads.

The classification head is a simple fully connected (dense) layer that maps the 1280-dimensional feature vector to the number of output classes C. It was used to compute the cross-entropy loss during the supervised training, which is computed using Eq. 5, 6, and 7,

$$z_{cls} = W_c f + b_c \ \ where\ W_c \in \mathbb{R}^{c \times 1280}, b_c \in \mathbb{R}^c \tag{5}$$

$$\hat{y} = Softmax(z_{cls}) \tag{6}$$

$$L_{cls} = -\sum_{i=1}^{C} y_1 \log(\hat{y}_i) \tag{7}$$

Where $\hat{y}$ is the predicted class probabilities, and y is the ground-truth one-hot vector.

The projection head (Contrastive Learning Task) is a multi-layer perceptron used for contrastive learning. It maps the same feature vector into a 256-dimensional embedding space where contrastive loss was applied to encourage semantically similar samples to be closer. Let the projection head be defined as in Eq. 8.

$$z_{proj} = MLP(f) = W_2 \cdot ReLU(W_1 f + b_1) + b_2 \tag{8}$$

The resulting embedding $Z_{proj} \in \mathbb{R}^{256}$ is normalized and used for contrastive loss. Contrastive loss was applied to $\hat{Z}_{proj}$ to learn invariant representations using:

$$L_{contrast} = -\log \frac{\exp(sim(\hat{Z}_i, \hat{Z}_j)/\tau)}{\sum_{k=1}^{2N} 1_{[k \neq 1]} \exp(sim(\hat{Z}_i, \hat{Z}_k)/\tau)} \tag{9}$$

Cosine similarity was used for the contrastive loss as defined in Equation 9. Both feature vectors $\hat{Z}_i$ and $\hat{Z}_j$ are L2-normalized to ensure that $sim(\hat{Z}_i, \hat{Z}_j) \in [-1, 1]$. This normalization bounded the similarity values and prevented the numerical instability of the softmax denominator. The sharpness of the similarity distribution was controlled using a positive scalar which is the temperature parameter $\tau \in (0, 1]$. As based on the previous studies on contrastive learning [28], this study empirically set the τ to 0.07 to ensure smooth convergence and stable gradients during training. Figure 2. Architecture of MG-ResViT. The Model Combines Dynamic Residual Blocks with Patch-Wise Attention and Cross-Block Feature Fusion. Dual optimization Heads Leverage Both SupCon (contrastive) and CE (Classification) Losses for Improved Feature Separability and Accuracy

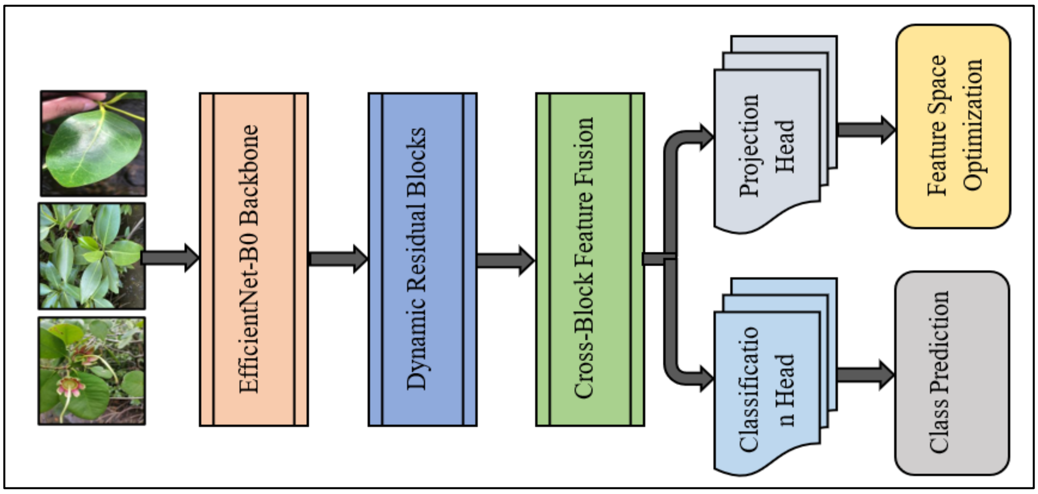

Figure 2. Architecture Diagram of the Proposed Model

3.2.1 Loss Function and Optimization

A multi-objective loss function was employed to train the model; it combined categorical classification and representation learning. It is a linear combination of cross-entropy loss and supervised contrastive loss. For the cross-entropy loss, it penalizes incorrect predictions by comparing logits with ground truth labels. It is computed using,

$$L_{CE} = \sum y \log(\hat{y}) \tag{10}$$

For the supervised contrastive loss (SupCon), this loss encourages samples from the same class to cluster in the feature space while separating those from different classes using,

$$L_{CSL} = -\frac{1}{N} \sum_{i \in I} \frac{1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(sim(z_i, z_p)/\tau)}{\sum_{a \in A(i)} \exp(sim(z_i, z_a)/\tau)} \tag{11}$$

The total loss was also computed using, $L_{total} = L_{CE} + \alpha \cdot L_{CSL}$, with $\alpha = 0.5$ which balanced the supervised classification and contrastive representation learning. Furthermore, the optimization was performed using the Adam optimizer with a learning rate of 1×10^{-3} and L2 regularization (1×10^{-5}). Gradients were clipped at a norm of 1.0 to ensure stable training.

3.2.2 Training Protocol

The proposed MG-ResViT model was trained over 50 epochs using the standard supervised learning paradigm, implemented in PyTorch. The training process was carefully designed to optimize representation learning through dual-loss objectives. The dataset was iteratively processed in mini-batches, with each batch undergoing a complete forward and backward pass through the architecture. During each epoch, forward propagation was performed by passing each input batch through the EfficientNet-B0 feature extractor, resulting in a high-dimensional feature map. This feature map was subsequently refined through a series of four dynamic Residual Attention Blocks to adaptively emphasize salient spatial features while preserving residual connections. The refined feature representation was then fed into two parallel branches: a projection head, used for supervised contrastive learning, and a classification head, responsible for final class prediction.

The model computed the cross-entropy loss based on the predicted logits and ground-truth labels. Simultaneously, the supervised contrastive (SupCon) loss was calculated from the projected feature embeddings for intra-class compactness and inter-class separability in the feature space. The total loss was obtained as a weighted summation of both loss components. Subsequently, the gradients of the total loss concerning the model parameters were computed via backpropagation. Gradient clipping was applied with a maximum norm of 1.0 to ensure numerical stability and prevent gradient explosion. To regularize the network and mitigate overfitting, the model parameters were then updated using the Adam optimizer, which was configured with a learning rate of 0.001 and a weight decay of 1e-5.

The training loss, which comprised both the cross-entropy and SupCon components was monitored and logged every 10 batches to track learning progress and detect early signs of overfitting or instability. To compute the validation loss and accuracy, a validation loop was executed at the end of every epoch. This was applied to ensure the generalization performance of the model throughout the training. The proposed training protocol is based on the demand for fine-grained image classification tasks, such as for mangrove species. The subtle inter-class variance of the fine-grained mangrove species dataset required robust and semantically aware feature embeddings. Below is the algorithm for the training protocol of the proposed model.

MG-ResViT Training Protocol

Algorithm 1: MG-ResViT Training Protocol

Input:

Dataset D with images X and labels y

Hyperparameters: epochs=50, batch_size=32, lr=0.001, weight_decay=1e-5

Model components: Backbone (EfficientNet-B0), DynamicResBlocks (x4), ProjectionHead, ClassificationHead

Output:

Trained MG-ResViT model with optimized parameters θ

1. Initialize:

- Model θ ← MG-ResViT()
- Optimizer ← Adam (θ, lr, weight_decay)
- Loss weights: λ_ce = 1.0, λ_supcon = 0.5
- Gradient clipping threshold: max_norm = 1.0

2. for epoch in 1...50 do:

3. for batch (X_batch, y_batch) in D.train_loader do:

4. features ← Backbone(X_batch) - EfficientNet-B0 extractor

5. refined_features ← DynamicResBlocks(features) - Residual attention

6. projections ← ProjectionHead(refined_features)

7. logits ← ClassificationHead(refined_features)

8. L_ce ← CrossEntropyLoss(logits, y_batch)

9. L_supcon ← SupConLoss(projections, y_batch)

10. total_loss ← λ_ce ∗ L_ce + λ_supcon ∗ L_supcon

11. $\nabla\theta$← Backpropagate(total_loss)

12. ClipGradients($\nabla\theta$, max_norm)

13. Optimizer.step(θ)

14. if batch % 10 == 0:

15. LogTrainingLoss(L_ce, L_supcon)

16. val_loss, val_acc ← Evaluate(D.val_loader)

17: LogValidationMetrics(val_loss, val_acc)

3.2.3 Evaluation Matrix

3.2.3.1 Convergence, Gradient, and Accuracy Analysis

In this study, the training and validation loss trends were analysed to evaluate the convergence and overfitting risks of the proposed model. Average gradient magnitude (AGM) was used to evaluate the change in image intensity of the model variants. This metric was used to quantify the magnitude of parameter updates during backpropagation. This is important to ensure that there is learning particularly in deep networks. Moreover, accuracy, precision, recall, and f1-score were also used to evaluate the classification performance of the proposed model.

Accuracy is the measure of the overall percentage of correctly classified samples in the dataset. It is mathematically defined as:

$$Accuracy = \frac{Number\ of\ correct\ predicitons\ (TP+TN)}{Total\ number\ of\ predictions\ (TP+TN+FP+FN)} \tag{12}$$

Precision is measured by the proportion of correctly classified positive observations to total predicted positive observations. It is given by:

$$Precision = \frac{TP}{TP+FP} \tag{13}$$

Recall, the true positive rate or the proportion of all actual positive predictions that were correctly classified as positives. It is computed using:

$$Recall = \frac{correctly\ classified\ actual\ positives\ (TP)}{all\ actual\ positives\ (TP+FN)} \tag{14}$$

F1-Score is an evaluation metric used that considers both precision and recall and is defined as:

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{15}$$

Where TP are True Positives, TN are True Negatives, FP are False Positives, and FN are False Negatives.

Average predictive entropy was also computed across all test samples to assess the confidence and uncertainty in the classification predictions of the model. It was used to quantify how uncertain the model is regarding its predictions where lower entropy indicates more confident and well-calibrated outputs [29]. For the predicted probability distribution $p = [p_1, p_2, .., p_c]$ over C classes, the entropy $H(p)$ for a single sample is computed as:

$$H(p) = -\sum_{i=1}^{C} p_i \log(p_i) \tag{16}$$

Where p_i is the predicted probability of class i and the logarithm is typically a natural log (base e). The average predictive entropy across N test samples is then defined as:

$$\mathrm{H} = \frac{1}{N} \sum_{n=1}^{N} H\left(p^{(n)}\right) = -\frac{1}{N} \sum_{n=1}^{N} \sum_{i=1}^{C} p_i^{(n)} \log(p_i^{(n)}) \tag{17}$$

3.2.3.2 Intraclass and Interclass Similarity

This study also used intraclass and interclass similarity to measure feature compactness and feature separability, respectively. The intra-class similarity is measured by how closely the feature representations are clustered within the same class [8]. Cosine similarity was used to compute the intra-class similarity. It indicates that there is a tighter feature clustering and better representation learning if the value is closer to 1. On the other hand, inter-class similarity is used to quantify the degree of feature overlap between different species [22]. Cosine similarity was also used to compute the similarity between feature vectors of different mangrove species. Cosine similarity is calculated using a mathematical formula that involves the dot product of the two vectors and their magnitudes. The formula is expressed in Equation 18.

$$similarity\ (A, B) = \cos(\theta) = \frac{A \cdot B}{||\, A\, ||\ ||\, B||} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2 \sum_{i=1}^{n} B_i^2}} \tag{18}$$

3.2.3.3 Feature Extraction Quality Metrics and Visualization

The silhouette score was also computed to measure the similarity of data points to their clusters compared to other clusters [24]. The mean silhouette score across all the data points in the dataset provided an overall measure of feature cluster quality. It is interpreted that a higher mean score indicates well-defined and compact clusters. On the other hand, a lower mean score suggests overlapping or poorly separated clusters [24]. It is computed using Equation 19.

$$s(i) = \frac{b(i)-a(i)}{larger\ of\ b(i)\ and\ a(i)} \tag{19}$$

Where $a(i)$ is the average distance from a point i to all other data points in its cluster, and $b(i)$ is the average distance from a point i to all points in the nearest neighboring cluster.

Moreover, the average Intra-Class Distance and average Inter-Class Distance were also computed to evaluate the quality and effectiveness of the models. Intra-Class Distance is calculated as the average or maximum distance between all pairs of data points within the same cluster [17]. The intra-class distance score is interpreted as a smaller value indicating that data points are more similar and tightly packed data points are within the cluster. On the other hand, Inter-Class Distance is calculated as the distance between data points in different clusters [4]. The larger the inter-class distance score, the more distinct and well-separated the clusters are from each other.

4. Results and Discussion

The proposed model for mangrove species classification was evaluated using an 80-20 split of the dataset, with 80% of the data allocated for training and 20% for validation. This split ensured that the model had sufficient data to learn patterns while maintaining a separate test set for unbiased performance evaluation. Figure 3 shows the training and validation loss and accuracy curves of MG-ResViT over 50 epochs. Results show that there is stable convergence with validation loss closely tracking training loss decreasing steadily. The plot suggests effective learning without overfitting, and the model generalized well to unseen data and demonstrated strong optimization.

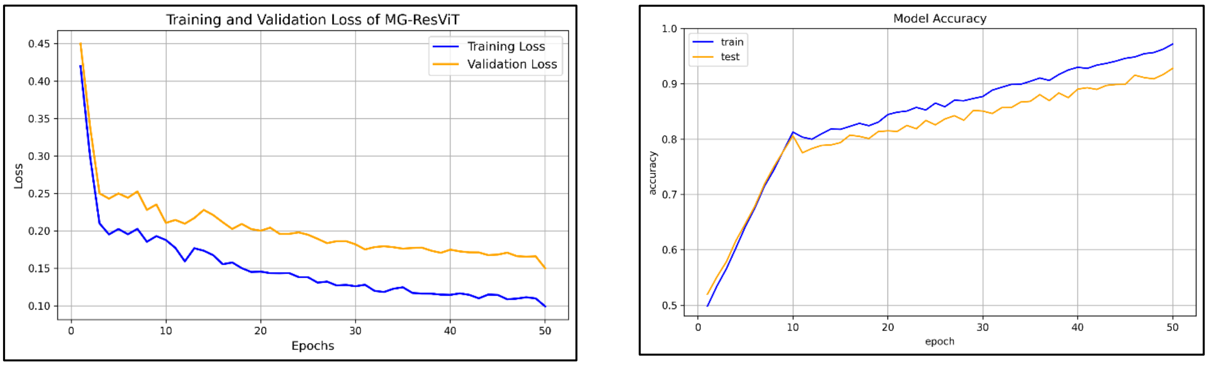

Figure 3. Proposed MG-ResViT Model Validation and Loss Curve

Moreover, the accuracy plot reveals a rapid rise in the early epochs and stabilized near 90% by epoch 30. Although minor fluctuations occurred during training, there is an overall upward trend. The final accuracy confirmed the discriminative power of the model for mangrove species classification. In addition, fluctuations are reflected due to minor batch-wise variability [27].

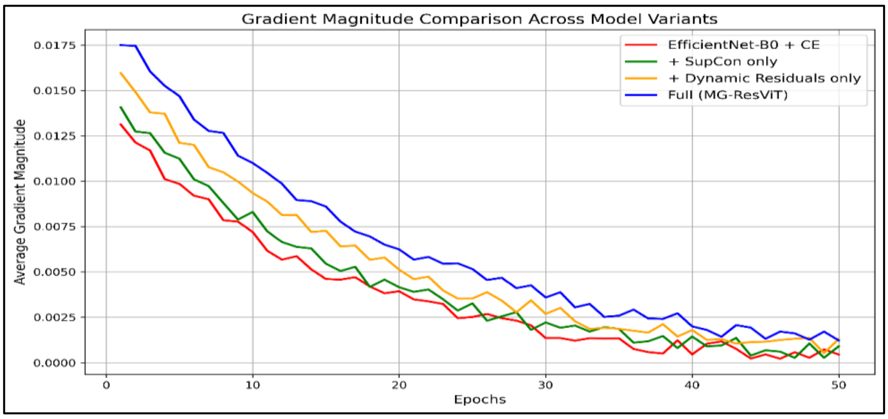

Figure 4. Average Gradient Magnitude of the Different Model Variants

Figure 4 shows a comparison of the average gradient magnitudes across different model variants over 50 training epochs. Four model variants were simulated for the ablation study, EfficientNet-B0 with cross-entropy (CE) loss (red), the same model with supervised contrastive loss (SupCon only, green), the model with dynamic residuals only (yellow), and the full hybrid model combining all components (MG-ResViT, blue). Based on the results, the baseline shows the most rapid reduction in gradient magnitude, which indicates faster convergence, while the full model maintains a higher gradient value for a longer period. Furthermore, the data show that the full MG-ResViT model preserved more gradient flow and learned more actively during training. In addition, the full model avoided early stagnation which led to improved generalization.

Table 1. Ablation Study of Different Model Variants

Model Variant	Train Acc	Val Acc	Avg Intra-Class Distance	Avg Inter-Class Distance	Silhouette Score	Entropy
EfficientNet-B0 + CE (baseline)	94.2%	88.0%	0.523	1.003	0.44	0.48
+ SupCon only	95.1%	89.7%	0.417	1.204	0.58	0.40
+ Dynamic Residuals only	95.8%	90.1%	0.409	1.264	0.56	0.38
Full (MG-ResViT)	97.2%	92.8%	0.327	1.512	0.68	0.33

Table 1 shows the systematic ablation study comparing the performance based on the different variants of the proposed model. The baseline EfficientNet-B0 with cross-entropy (CE) loss achieved a validation accuracy of 88.0%. The introduction of supervised contrastive learning (+ SupCon) alone yielded a 1.7% validation accuracy improvement (89.7%) and was accompanied by a 20.3% reduction in intra-class distance from the baseline 0.523 to + SupCon 0.417. Moreover, there was also a increase of 20.0% increase in inter-class distance from the baseline model (1.003) to the +SupCon model (1.204). The results show that SupCon is effective in enhancing feature space separation, particularly for discriminating morphologically similar mangrove species.

Additionally, the dynamic residual blocks (+ Dynamic Residuals only) contributed a further increase in validation accuracy to 90.1%, the intra-class distance was reduced by 1.9% (from 0.417 to 0.409). This consistent improvement suggests that the adaptive skip connections better preserved the discriminative local features, such as leaf vein patterns, which are important for fine-grained species classification. Additionally, the complete model MG-ResViT integrated both components achieved the highest performance across all metrics with a 92.8% validation accuracy, 0.327 intra-class distance, and 1.512 inter-class distance. The improvement in the silhouette score from the baseline model 0.44 to the full MG-ResViT 0.68 demonstrates the enhanced ability of the model to produce compact and well-separated clusters in the feature space.

Furthermore, the full MG-ResViT model achieved the lowest average predictive entropy of 0.33 compared to each architectural enhancement, which had 0.40 for the + SupCon-only variant, and 0.38 for the + Dynamic Residuals-only variant. The results show that the full model variant has better-calibrated class probabilities and the most confident predictions. Additionally, the results present progressively reduced entropy, which further validates the contribution of improved class compactness and feature separability.

The results revealed three key insights, first, SupCon contributed significantly to feature space organization, second, dynamic residuals provided finer intra-class feature refinement, and third, the combined approach yielded complementary benefits beyond additive improvements. Therefore, the results quantitively validate that the full MG-ResViT configurations achieved an optimal balance between global feature separation and local discriminative ability for mangrove species classification.

Table 2. Ablation Study for Intra-Class Similarity Scores Per Species Across Model Variants

Model Variant	**Species A (*Rhizophora*)**	**Species B (*Avicennia*)**	**Species C (*Sonneratia*)**	**Mean Intra-Class Similarity**
EfficientNet-B0 + CE (Baseline)	0.71	0.69	0.68	0.693
+ SupCon only	0.84	0.82	0.83	0.830
+Dynamic Residuals only	0.78	0.76	0.75	0.763
Full (MG-ResViT)	0.91	0.88	0.89	0.893

Table 2 presents the intra-class similarity per species across model variants. The baseline model EfficientNet-B0 + CE achieved a mean intra-class similarity score of 0.693 and showed moderate similarity scores across species, Rhizophora (0.71), Avicennia (0.69), and Sonneratia (0.68). On the other hand, the introduction of + SupCon alone improved the intra-class similarity with a mean score of 0.830. Moreover, it also boosted the similarity by 18.3% to 19.7% across species with scores, Rhizophora (0.84), Avicennia (0.82), and Sonneratia (0.83). In addition, the integration of the + Dynamic Residuals model variant yielded a mean intra-class similarity score of 0.763, whereas the full MG-ResViT model variant achieved a high improvement of 28.2% compared to the baseline model. MG-ResViT attained a mean intra-class similarity score of 0.893 which is higher than that of other model variants. The results show that there is a balanced performance across species. Moreover, SupCon produced the most significant gains by learning robust embeddings, while the dynamic residuals provided species-specific refinement.

Table 3. Ablation Study for Inter-class Similarity Scores Per Species Across Model Variants

Model Variant	Avicennia alba vs Rhizophora apiculata	Avicennia alba vs Sonneratia alba	Rhizophora apiculata vs Sonneratia alba	Mean Inter-Class Similarity
EfficientNet-B0 + CE (Baseline)	0.45	0.43	0.47	0.450
+ SupCon only	0.32	0.30	0.33	0.317
+ Dynamic Residuals only	0.37	0.36	0.39	0.373
Full (MG-ResViT)	0.21	0.19	0.23	0.210

The inter-class similarity scores per species across model variants given in Table 3 depict the superior ability of the proposed MG-ResViT model to distinguish between mangrove species. The baseline model (EfficientNet-B0 + CE) produced a mean inter-class similarity score of 0.450 and showed high feature overlap between species pairs, *Avicennia alba* vs *Rhizophora apiculate* (0.45), *Avicennia alba* vs *Sonneratia alba* (0.43), and *Rhizophora apiculata* vs *Sonneratia alba* (0.47). This reflects the challenge of separating morphologically similar taxa. The + SupCon model variant achieved a 29.6% reduction in mean similarity from the baseline of 0.450 to 0.317. The introduction of contrastive learning proved its ability to separate features of the species in embedding space. Moreover, there is an additional 15.1% reduction over the baseline provided by the integration of dynamic residuals with a mean inter-class similarity score of 0.373. The full MG-ResViT model achieved a separation mean of 0.210 while species *Avicennia alba* and *Sonneratia alba* had the lowest similarity of 0.19, these species are most often confused by field experts. These findings confirm that SupCon has established strong decision boundaries, dynamic residuals have refined the local discriminative features, and the combined approach has reduced misclassification risk.

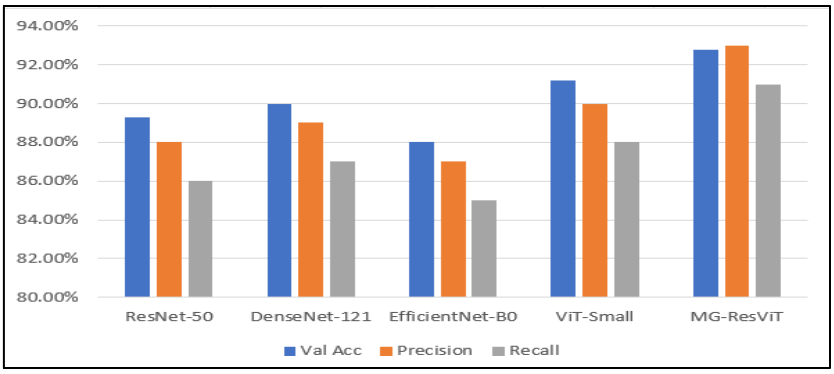

Figure 5. Comparison of the Validation Accuracy, Precision, and Recall of the Proposed MG-ResViT and the Different SOTA Models.

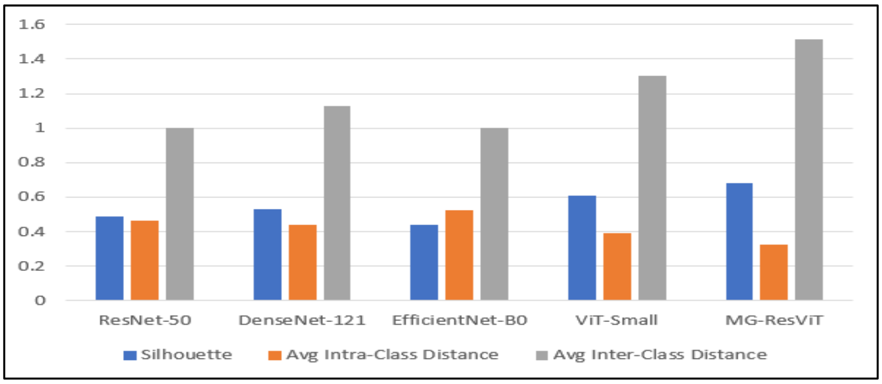

Figure 6. Comparison of the Silhouette Score, Average Intra-class Distance, and Average Inter-Class Distance of the Proposed MG-ResViT and the Different SOTA Models.

As shown in Figure 5, the proposed MG-ResViT model outperformed other models in terms of its accuracy, precision, and recall. The comparison of the silhouette scores, average intra-class distance, and average inter-class distance depicted in Figure 6 represents the feature learning capabilities of the proposed MG-ResViT model compared to other models.

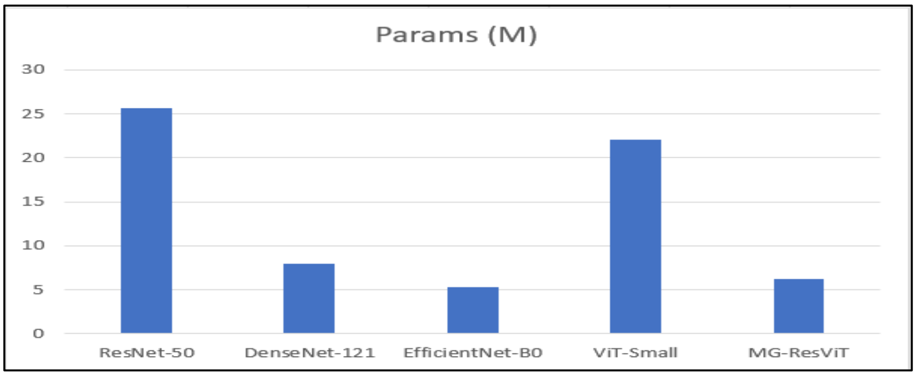

Figure 7. Comparison of the Number of Parameters of the Proposed MG-ResViT and the Different SOTA Models.

The comparison of the number of parameters of the different models presented in Figure 7 illustrates that EfficientNet-B0 has the lowest number of parameters, followed by the proposed MG-ResViT and DenseNet-121 models. ResNet-50 produced the highest number of parameters followed by ViT-Small.

Table 4. Comparison with SOTA Models

Model	Val Acc	Precision	Recall	Params (M)	Silhouette	Avg Intra-Class Distance	Avg Inter-Class Distance	Entropy
ResNet-50	89.3%	0.88	0.86	25.6	0.49	0.463	1.003	0.45
DenseNet-121	90.0%	0.89	0.87	7.98	0.53	0.441	1.126	0.41
EfficientNet-B0	88.0%	0.87	0.85	5.3	0.44	0.523	1.003	0.48
ViT-Small	91.2%	0.90	0.88	22.1	0.61	0.392	1.304	0.36
MG-ResViT	92.8%	0.93	0.91	6.2	0.68	0.327	1.512	0.33

As shown in Table 4, the proposed MG-ResViT has established new performance standards across all key metrics compared to other architectures. All models were simulated on similar datasets and the same workstation. In terms of accuracy and efficiency, the proposed model outperformed ViT-Small with 91.2% accuracy and 92.8%, respectively, while using 72% fewer parameters (MG-ResViT: 6.2M vs. ViT-Small: 22.1M). Additionally, the proposed model (92.8%) surpassed the accuracy of DenseNet-121 (90.0%) and had 22% fewer parameters (MG-ResViT: 6.2M vs. DenseNet-121: 7.98M). Moreover, MG-ResViT achieved higher accuracy than EffecientNet-B0 (88.0%) despite having more parameters (see Table 4). These findings demonstrate the value of the proposed architectural modifications.

For feature space performance comparison, the proposed model provided the highest silhouette score of 0.68 compared to other SOTA models. The high silhouette score indicated a better-defined and more distinct cluster. The proposed model also achieved optimal intra-class and inter-class ratios, the intra-class distance (0.327) is lower than that of ViT-Small (0.392), DenseNet-121 (0.441), ResNet-50 (0.463), and EfficientNet-B0 (0.523). Relatively, the inter-class distance of MG-ResViT (1.512) exceeded ViT-Small (1.304), DenseNet-121 (1.126), ResNet-50 (1.003), and EfficientNet-B0 (1.003).

Furthermore, the 6.2M parameter count demonstrated efficient feature learning without sacrificing discriminative power. The 0.68 silhouette score confirmed the excellent cluster separation that is important for handling phenotypic plasticity within species, hybrid specimens, and non-ideal field imaging conditions. Notably, the 1.512 inter-class distance of MG-ResViT provided a new benchmark, proving its effectiveness for fine-grained separation of visually similar species like *Rhizophora apiculata* and *Sonneratia alba*, juvenile versus mature leaves, and stress-affected specimens. For the average predicted entropy, MG-ResViT produced the lowest value of 0.33 compared to other models, ResNet-50 (0.45), DenseNet-121 (0.41), EfficientNet-B0 (0.48), and ViT-Small (0.36). The results indicate that the proposed model has the highest prediction confidence and feature compactness among all models.

5. Conclusion

This study proposed MG-ResViT, a novel deep learning model that established new state-of-the-art performance for feature extraction of mangrove species for classification by applying a dynamic residual connection with spatially-adaptive attention gates, hybrid supervised contrastive and cross-entropy loss optimization, and PCA-optimized cross-block feature fusion. The comprehensive experiments demonstrated that MG-ResViT outperformed existing approaches across all evaluation metrics, achieving a validation accuracy of 92.8% with exceptional feature space organization (Silhouette Score = 0.68). The reduction of both inter-class and intra-class similarity of the proposed model proved its effectiveness in handling fine-grained visual differences that challenge both automated systems and human experts. These advancements are particularly significant for ecological conservation applications, wherein accurate species classification under variable field conditions is crucial for biodiversity monitoring and habitat protection.

The proposed MG-ResViT was trained and evaluated on a dataset with only three mangrove species from a single geographic region. Although the model performed well compared to other SOTA models, its generalizability to broader ecological conditions may be explored. Moreover, the integration of both supervised contrastive loss and cross-entropy loss in the dual-head setup produced an additional training complexity. Also, the current model was simulated only on RGB imagery. Additional data modalities can be incorporated to further enhance classification performance such as hyperspectral imagery, NDVI, or time-series patterns. Lastly, future work can be conducted to further improve the model through more advanced data augmentation techniques, model quantization for edge deployment, and more hyperparameter tuning. The model has potential for extension to other fine-grained species classification tasks.

References

[1] Aljundi, Rahaf, Yash Patel, Milan Sulc, Nikolay Chumerin, and Daniel Olmeda Reino. "Contrastive classification and representation learning with probabilistic interpretation." In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 37, no. 6, 2023, 6675-6683.

[2] Alzubaidi, Laith, Jinglan Zhang, Amjad J. Humaidi, Ayad Al-Dujaili, Ye Duan, Omran Al-Shamma, José Santamaría, Mohammed A. Fadhel, Muthana Al-Amidie, and Laith Farhan. "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions." Journal of big Data 8 (2021): 1-74.

[3] Bauravindah, Achmad, and Dhomas Hatta Fudholi. "Lightweight models for real-time steganalysis: A Comparison of MobileNet, ShuffleNet, and EfficientNet." Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi) 8, no. 6 (2024): 737-747.

[4] Binu Jose, A., and Pranesh Das. "A multi-objective approach for inter-cluster and intra-cluster distance analysis for numeric data." In Soft Computing: Theories and Applications: Proceedings of SoCTA 2021, Singapore: Springer Nature Singapore, 2022, 319-332.

[5] Devarajan, Kasthuri, Suresh Ponnan, and Sundresan Perumal. "Hybrid CNN-transformer architecture for enhanced EEG-based emotion recognition: capturing local and global dependencies with self-attention mechanisms." Discover Computing 28, no. 1 (2025): 1-25.

[6] Jian, Zhuokai, Bin Ai, Jiali Zeng, and Yuchao Sun. "A hybrid mangrove identification method by combining the time-frequency threshold of the Mangrove Index with a random forest binary classifier." IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (2024).

[7] Khosla, Prannay, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. "Supervised contrastive learning." Advances in neural information processing systems 33 (2020): 18661-18673.

[8] Li, Cong, Gong Cheng, Guangxing Wang, Peicheng Zhou, and Junwei Han. "Instance-aware distillation for efficient object detection in remote sensing images." IEEE Transactions on Geoscience and Remote Sensing 61 (2023): 1-11.

[9] Li, Yuyang, Bolin Fu, Xidong Sun, Donglin Fan, Yeqiao Wang, Hongchang He, Ertao Gao, Wen He, and Yuefeng Yao. "Comparison of different transfer learning methods for classification of mangrove communities using MCCUNet and UAV multispectral images." Remote Sensing 14, no. 21 (2022): 5533.

[10] Liu, Tonglai, Xuanzhou Chen, Wanzhen Zhang, Xuekai Gao, Liqiong Lu, and Shuangyin Liu. "Early Plant Classification Model Based on Dual Attention Mechanism and Multi-Scale Module." AgriEngineering 7, no. 3 (2025): 66.

[11] Lu, Yan-Feng, Qian Yu, Jing-Wen Gao, Yi Li, Jun-Cheng Zou, and Hong Qiao. "Cross stage partial connections based weighted bi-directional feature pyramid and enhanced spatial transformation network for robust object detection." Neurocomputing 513 (2022): 70-82.

[12] Mittermeier, Russell A., Will R. Turner, Frank W. Larsen, Thomas M. Brooks, and Claude Gascon. "Global biodiversity conservation: the critical role of hotspots." Biodiversity hotspots: distribution and protection of conservation priority areas (2011): 3-22.

[13] Nelson, James A., Justin Lesser, W. Ryan James, David P. Behringer, Victoria Furka, and Jennifer C. Doerr. "Food web response to foundation species change in a coastal ecosystem." Food Webs 21 (2019): e00125.

[14] Primavera, Jurgenne H., Daniel A. Friess, Hanneke Van Lavieren, and Shing Yip Lee. "The mangrove ecosystem." World seas: an environmental evaluation (2019): 1-34.

[15] Schlemper, Jo, Ozan Oktay, Michiel Schaap, Mattias Heinrich, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. "Attention gated networks: Learning to leverage salient regions in medical images." Medical image analysis 53 (2019): 197-207.

[16] Tan, Linlin, and Haishan Wu. "Artificial Intelligence Mangrove Monitoring System Based on Deep Learning and Sentinel-2 Satellite Data in the UAE (2017-2024)." arXiv preprint arXiv:2411.11918 (2024).

[17] Venkataramanan, Aishwarya, Martin Laviale, Cécile Figus, Philippe Usseglio-Polatera, and Cédric Pradalier. "Tackling inter-class similarity and intra-class variance for microscopic image-based classification." In International conference on computer vision systems, Cham: Springer International Publishing, 2021, 93-103.

[18] Wang, Dezhi, Bo Wan, Penghua Qiu, Yanjun Su, Qinghua Guo, and Xincai Wu. "Artificial mangrove species mapping using pléiades-1: An evaluation of pixel-based and object-based classifications with selected machine learning algorithms." Remote Sensing 10, no. 2 (2018): 294.

[19] Wang, Zihu, Yu Wang, Zhuotong Chen, Hanbin Hu, and Peng Li. "Contrastive learning with consistent representations." arXiv preprint arXiv:2302.01541 (2023).

[20] Wen, Lei, Zikai Xiao, Xiaoting Xu, and Bin Liu. "Disaster Recognition and Classification Based on Improved ResNet-50 Neural Network." Applied Sciences 15, no. 9 (2025): 5143.

[21] Xu, Guoping, Xiaxia Wang, Xinglong Wu, Xuesong Leng, and Yongchao Xu. "Development of skip connection in deep neural networks for computer vision and medical image analysis: A survey." arXiv preprint arXiv:2405.01725 (2024).

[22] Yang, Xiaoran, Shuhan Yu, and Wenxi Xu. "Enhanced Convolutional Neural Networks for Improved Image Classification." arXiv preprint arXiv:2502.00663 (2025).

[23] Zhang, Hanwen, Shan Wei, Xindan Liang, Yiping Chen, and Hongsheng Zhang. "Scale effects in mangrove mapping from ultra-high-resolution remote sensing imagery." International Journal of Applied Earth Observation and Geoinformation 136 (2025): 104310.

[24] Rotem, Oded, Tamar Schwartz, Ron Maor, Yishay Tauber, Maya Tsarfati Shapiro, Marcos Meseguer, Daniella Gilboa, Daniel S. Seidman, and Assaf Zaritsky. "Visual interpretability of image-based classification models by generative latent space disentanglement applied to in vitro fertilization." Nature communications 15, no. 1 (2024): 7390.

[25] Gerona-Daga, Maria Elisa B., and Severino G. Salmo III. "A systematic review of mangrove restoration studies in Southeast Asia: Challenges and opportunities for the United Nation's Decade on Ecosystem Restoration." Frontiers in Marine Science 9 (2022): 987737.

[26] Camacho, Leni D., Dixon T. Gevaña, Lorena L. Sabino, Clarissa D. Ruzol, Josephine E. Garcia, April Charmaine D. Camacho, Thaung Naing Oo et al. "Sustainable mangrove rehabilitation: Lessons and insights from community-based management in the Philippines and Myanmar." APN Science Bulletin (2020).

[27] Kandel, Ibrahem, and Mauro Castelli. "The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset." ICT express 6, no. 4 (2020): 312-315.

[28] Hou, Pengyue, and Xingyu Li. "Improving Contrastive Learning of Sentence Embeddings with Focal InfoNCE." In Findings of the Association for Computational Linguistics: EMNLP 2023, 2023, 4757-4762.

[29] Milanés-Hermosilla, Daily, Rafael Trujillo-Codorniú, Saddid Lamar-Carbonell, Roberto Sagaró-Zamora, Jorge Jadid Tamayo-Pacheco, John Jairo Villarejo-Mayor, and Denis Delisle-Rodriguez. "Robust motor imagery tasks classification approach using bayesian neural network." Sensors 23, no. 2 (2023): 703.