

# Visual Bias and Interpretability in Deep Learning for Dermatological Image Analysis

Enam Ahmed Taufik<sup>1</sup>, Abdullah Khondoker<sup>2</sup>, Antara Firoz Parsa<sup>2</sup>, Seraj Al Mahmud Mostafa<sup>3</sup>

<sup>1</sup>*Department of Computer Science and Engineering, European University of Bangladesh, Dhaka, Bangladesh*

<sup>2</sup>*Department of Computer Science and Engineering, BRAC University, Dhaka, Bangladesh*

<sup>3</sup>*Department of Information Systems, University of Maryland, Baltimore County, Baltimore, MD, USA*

taufik@eub.edu.bd<sup>1</sup>, ext.abdullah@bracu.ac.bd<sup>2</sup>, antara.firuz.parsa@g.bracu.ac.bd<sup>2</sup>, serajmostafa@umbc.edu<sup>3</sup>

**Abstract**—Accurate skin disease classification is a critical yet challenging task due to high inter-class similarity, intra-class variability, and complex lesion textures. While deep learning-based computer-aided diagnosis (CAD) systems have shown promise in automating dermatological assessments, their performance is highly dependent on image pre-processing and model architecture. This study proposes a deep learning framework for multi-class skin disease classification, systematically evaluating three image pre-processing techniques: standard RGB, CMY color space transformation, and Contrast Limited Adaptive Histogram Equalization (CLAHE). We benchmark the performance of pre-trained convolutional neural networks (DenseNet201, EfficientNetB5) and transformer-based models (ViT, Swin Transformer, DinoV2 Large) using accuracy and F1-score as evaluation metrics. Results show that DinoV2 with RGB pre-processing achieves the highest accuracy (up to 93%) and F1-scores across all variants. Grad-CAM visualizations applied to RGB inputs further reveal precise lesion localization, enhancing interpretability. These findings underscore the importance of effective pre-processing and model choice in building robust and explainable CAD systems for dermatology.

**Index Terms**—Image pre-processing, Computer-Aided Diagnosis, CAD, Model Explainability, CLAHE, Vision Transformers, DINOv2, Transfer Learning, CMY, RGB.

## I. INTRODUCTION

Skin diseases encompass a wide range of conditions, including inflammatory, infectious, and neoplastic disorders, that pose diagnostic challenges due to visual similarities, intra-class variability, and complex morphology. Early and accurate detection, particularly of malignant cancers and viral infections, is vital for effective treatment and containment [1], [2]. Traditional diagnostic methods such as visual inspection and dermoscopy are subjective and suffer from inter-observer variability, with dermatologists achieving only 60-80% accuracy [3], [4]. Viral infections like chickenpox, measles, and monkeypox further complicate diagnosis due to overlapping features [5], [6].

Computer-aided diagnosis (CAD) systems aim to address these limitations using machine learning and deep learning techniques [7]. However, traditional models with handcrafted features (e.g., SVMs, Random Forests) struggle with high-dimensional image data [8]. Deep learning, particularly CNNs, has shown superior performance by learning hierarchical features from images [9]. More recently, vision transformers (ViTs) have demonstrated strong classification capabilities

by capturing long-range dependencies [10]. Hybrid models combining CNNs, transformers, and attention mechanisms further enhance feature representation in complex multi-class settings [11], [12]. Still, their success relies on effective pre-processing, augmentation, and interpretability strategies.

In this study, we present a deep learning-based framework for multi-class skin disease classification and conduct a comprehensive analysis of how different image pre-processing techniques and model architectures influence performance and interpretability. We investigate three pre-processing strategies: (i) standard RGB images, (ii) CMY color space transformation, and (iii) Contrast Limited Adaptive Histogram Equalization (CLAHE). These are evaluated using two families of model architectures: pre-trained CNN-based transfer learning models and transformer-based models. Our experiments reveal that pre-processing strategies have a significant impact on classification accuracy, with CLAHE-enhanced images consistently improving performance across all architectures. Furthermore, transformer-based models outperformed CNNs in scenarios involving high inter-class visual similarity. Gradient-weighted Class Activation Mapping (Grad-CAM) visualizations indicate that models trained on CLAHE images focus more precisely on lesion regions, enhancing explainability. These findings emphasize the importance of thoughtful pre-processing and model selection in building fair, accurate, and explainable CAD systems for dermatological diagnosis.

The remainder of this paper is organized as follows. Related works in Section II reviews recent advancements in image classification, including optimization strategies, hybrid models, and dataset integration. Section III details the methodology, covering pre-processing, model design, and unified training across datasets. Experiments details are in Section IV. Section V represents results and analysis of classification performance, comparative analysis, and key insights. Finally, Section VI concludes the work by outlining future research directions.

## II. RELATED WORK

Recent advancements in deep learning have significantly enhanced skin disease classification systems. Jeyageetha et al. [13] proposed a model integrating Region Growing segmentation and MobileSkinNetV2, optimized using the Modified Honey Badger Optimizer (MHBO), achieving 99.01%

accuracy and 98.6% precision on the ISIC dataset. Qasim et al. [14] introduced a hybrid framework combining Deep Spiking Neural Networks (DSNN) and VGG-13, reaching 89.5% accuracy and a 90% F-measure on ISIC 2019. Abbas et al. [15] evaluated CNN variants including DenseNet-121 and ResNet-50 on HAM10000, with their sequential CNN achieving AUC scores of up to 99%. Helal et al. used geometric mean to address data imbalanceness [16]. To address class imbalance, they used image augmentation techniques. Similarly, Mostafa et al. used more augmentation techniques to tackle challenges in complex image processing as well as custom model building tasks [17]–[19]. Sasithradevi et al. [20] proposed EffiCAT, a hybrid model of EfficientNet and attention mechanisms (DCAL, CBAM), attaining 94.48% accuracy on HAM10000 and 92.08% on ISIC 2018. Aditya et al. [21] and Gonzalez et al. [22] demonstrated accuracy gains using transfer learning techniques over training. CAD-Skin by Khan et al. [23] integrated CNNs, autoencoders, and Quantum SVMs, achieving up to 99% accuracy across PAD-UFES-20, ISIC-2018, and ISIC-2019 datasets. Xiao et al. [24] analyzed ColorJitter augmentation on SID data, finding brightness and hue effective, though limitations persisted due to data quality. Lastly, Chu et al. [25] used UAV-based multispectral imaging with YOLOv5 and SSD, achieving 0.93 and 0.92 accuracy, highlighting deep learning’s utility in forestry.

### III. METHODOLOGY

The overall methodology, illustrated in (Figure 1), outlines the complete pipeline from dataset preparation and pre-processing to model training.

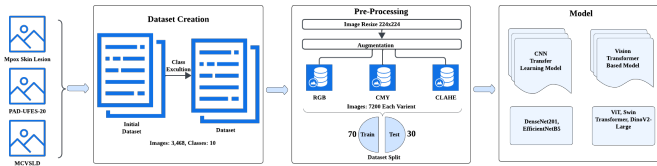


Fig. 1: Methodology Pipeline

### A. DATASET

1) **DATASET OVERVIEW:** To ensure robustness and generalizability in multi-class skin disease classification, we curated a comprehensive dataset by integrating three publicly available sources: the **Mpox Skin Lesion** [26], **PAD-UFES-20** [27], and the **Multi-Class Viral Skin Lesion Dataset (MCVSLD)** [28]. This multi-source combination captures a wide range of dermatological conditions and visual variability across diverse patient populations and imaging environments.

The initial aggregated dataset consisted of 4,312 clinical skin images spanning 12 categories: *Actinic Keratosis*, *Basal Cell Carcinoma*, *Chickenpox*, *Measles*, *Hand, Foot, and Mouth Disease (HFMD)*, *Monkeypox*, *Nevus*, *Normal* (healthy skin), *Seborrheic Keratosis*, *Squamous Cell Carcinoma*, *Melanoma*, and *Cowpox*. Due to severe class imbalance and insufficient

representation, the *Melanoma* and *Cowpox* classes were excluded. After this refinement, the final dataset comprised 3,468 images across 10 well-represented classes: nine skin diseases and one healthy skin category. Among these, *Measles* is the most underrepresented with 147 samples, while *Basal Cell Carcinoma* has the highest number of images, totaling 690.

2) *DATASET PRE-PROCESSING*: To enhance model robustness and mitigate class imbalance, an augmentation pipeline was applied using the Albumentations library. The transformations included horizontal and vertical flips, 90-degree rotations, brightness/contrast adjustments, shift-scale-rotate operations, and elastic distortions. For underrepresented classes (less than 500 samples), data was augmented to 600 images; for others, to 1,000 images.

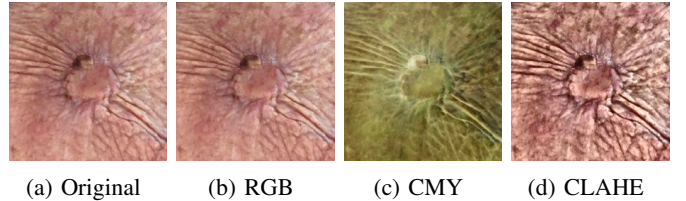


Fig. 2: Original and preprocessed images using various approaches: (a) Original, (b) Resized - RGB, (c) Resized - CMY, and (d) Resized - CLAHE.

Following augmentation, three pre-processing strategies were applied: (i) RGB-based augmentation, (ii) CMY color space transformation to enhance channel separation, and (iii) CLAHE (Contrast Limited Adaptive Histogram Equalization) to improve local contrast. These resulted in four dataset variants used in experimentation: original RGB, augmented RGB, CMY, and CLAHE-enhanced images (Figure 2). After augmentation, the total number of images increased to 7,200. The dataset was then split into training and testing subsets using a 70:30 ratio.

### B. MODELS ARCHITECTURE

1) **TRANSFER LEARNING MODELS:** We employed transfer learning with two high-capacity CNN backbones: **DenseNet201** [29], and **EfficientNetB5** [30], all pretrained on ImageNet. Each model was adapted by removing the top classification layer and unfreezing the last 30 layers for fine-tuning. A unified architecture was built around each backbone. An initial Conv2D layer (3×3, ReLU) normalized input channels, followed by the frozen base model. Post-feature extraction, we added a Conv2D layer (256 filters), Batch Normalization, Dropout (0.5), and Global Average Pooling. Two fully connected layers (1024 and 512 units) with ReLU activation,  $L_2$  regularization (with penalty  $\lambda = 10^{-4}$ ), Batch Normalization, and Dropout (0.6) were stacked before the final Dense softmax layer for 10-class classification, outputting  $\hat{y} \in \mathbb{R}^{10}$ . Models were trained using the Adam optimizer and an exponential learning rate decay:

$$\eta_t = \eta_0 \cdot \gamma^t$$

where  $\eta_0 = 10^{-4}$  and  $\gamma = 0.9$ . We used categorical cross-entropy loss and monitored accuracy. Each model was trained for up to 100 epochs with early stopping and learning rate reduction on the plateau. The training pipeline utilized the `tf.data` API for efficient batching, shuffling, and prefetching.

2) **TRANSFORMER MODELS:** We evaluated three state-of-the-art vision transformer backbones: **Vision Transformer (ViT)** [31], **Swin Transformer** [32], and **DINOv2-L** [33], all pretrained on ImageNet. For each model, the final classification head was replaced with a task-specific linear layer  $f : \mathbb{R}^d \rightarrow \mathbb{R}^{10}$ , producing logits  $\hat{y} \in \mathbb{R}^{10}$  over the 10 skin disease classes. The models were fine-tuned end-to-end using PyTorch.

Images were loaded via a custom dataset class and augmented using standard transforms. Training was conducted using the Adam optimizer with a learning rate  $\eta = 10^{-4}$  and CrossEntropyLoss:

$$\mathcal{L}_{CE} = - \sum_{i=1}^{10} y_i \log(\hat{y}_i)$$

A StepLR scheduler (decay factor 0.1 every 10 epochs) managed  $\eta$ , and gradient clipping with max-norm  $\|\nabla\|_2 \leq 1.0$  was applied. Each model was trained for up to 50 epochs with early stopping (patience = 5) based on validation accuracy. The best-performing checkpoint was saved for evaluation. Training and validation were executed on a CUDA-enabled GPU, and performance was monitored via accuracy and loss metrics.

---

**Algorithm 1** Pipeline for Skin Disease Classification

---

**Require:** Datasets: MSLD v2.0, PAD-UFES-20, MCVSLD

**Ensure:** Trained model with optimal test accuracy

```

1: Data Curation: Merge datasets; exclude Melanoma and Cowpox
2: for each class do
3:   if samples < 500 then
4:     Augment to 600
5:   else
6:     Augment to 1,000
7:   end if
8: end for
9: Generate 4 variants: RGB, Augmented RGB, Augmented CMY, Augmented CLAHE
10: Split: 70:30 train-test split for each image set.
11: for CNN model in {DenseNet201, EfficientNetB5} do
12:   Load ImageNet weights; unfreeze last 30 layers
13:   Append custom Conv, FC, Dropout layers
14: end for
15: for Transformer in {ViT, Swin, DINOv2-L} do
16:   Replace head; fine-tune with Adam optimizer
17: end for
18: Train with early stopping, save the best model

```

---

## IV. EXPERIMENTS

To assess the impact of different pre-processing techniques on visual quality and model performance, we compared lesion representations in RGB, CMY, and CLAHE-enhanced images (Figure 3). Among these, CMY-transformed images consistently yielded sharper boundary contrasts and improved lesion delineation. This enhancement is attributed to the subtractive nature of the CMY color model, which promotes chromatic separation and suppresses background interference [34]. Despite perceptual advantages, transfer learning models trained on CMY inputs slightly underperformed relative to RGB-based models, primarily due to the domain-specific bias of pretrained feature extractors optimized on large-scale RGB datasets such as ImageNet [35]. This domain shift [36] limits transfer learning effectiveness for CMY images. In contrast, transformer-based models exhibited greater robustness across pre-processing strategies. Results from RGB, CMY, and CLAHE inputs were nearly equivalent in classification performance, suggesting that transformers are less susceptible to input distribution shifts and training bias, likely due to their patch-wise tokenization and global self-attention mechanisms that enable better generalization across color domains. Moreover, CMY pre-processing offers contrast enhancement comparable to CLAHE [37], a widely used medical imaging technique, but with substantially lower computational cost. While CLAHE performs localized histogram equalization and interpolation with complexity  $\mathcal{O}(N \log L)$ , CMY transformation is a simple linear inversion,  $\text{CMY}(x, y) = 255 - \text{RGB}(x, y)$ , with linear complexity  $\mathcal{O}(N)$ , making it more suitable for resource-constrained scenarios.

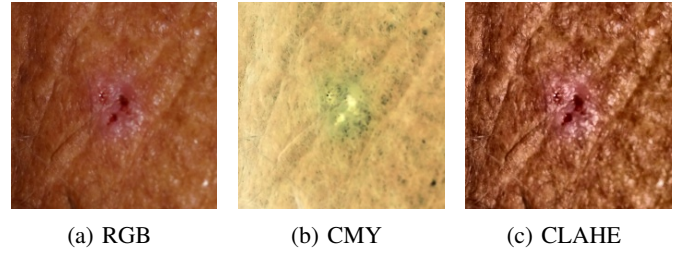


Fig. 3: Comparison of lesion boundaries using (a) RGB, (b) CMY, and (c) CLAHE pre-processing.

We further fine-tuned the proposed DinoV2 Large model on RGB, CLAHE, and CMY-preprocessed images, visualizing feature embeddings via t-SNE (Figure 4). DinoV2 Large demonstrated comparable performance on CLAHE and CMY inputs, confirming CMY as an efficient pre-processing alternative that also enhances lesion boundary sharpness. However, due to pretrained models' bias toward RGB statistics, the best performance was observed with RGB inputs. This suggests that targeted training on CMY data could improve accuracy while retaining pre-processing efficiency, motivating future research in this direction.

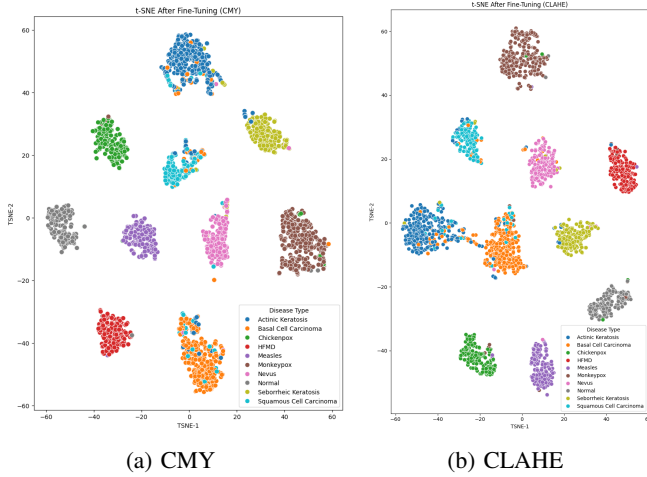


Fig. 4: t-SNE of Fine-tuned DinoV2 Large Model for (a) CMY and (b) CLAHE

## V. RESULTS AND ANALYSIS

We evaluated five models, DenseNet201, EfficientNetB5, Vision Transformer (ViT), Swin Transformer, and the proposed DinoV2 Large, across four image pre-processing methods: RGB, CMY, CLAHE, and original images. This comparison assesses model performance on skin disease classification using accuracy and F1-score metrics.

TABLE I: Performance of models across different image pre-processing techniques

Image Type	Model	Accuracy	F1-Score	Technique
RGB	DenseNet201	0.85	0.86	Transfer Learning
	EfficientNetB5	0.86	0.86	Transfer Learning
	ViT	0.88	0.89	Transformer
	Swin Transformer	0.92	0.93	Transformer
	<b>DinoV2 Large</b>	<b>0.93</b>	<b>0.93</b>	Transformer
CMY	DenseNet201	0.82	0.82	Transfer Learning
	EfficientNetB5	0.83	0.84	Transfer Learning
	ViT	0.84	0.85	Transformer
	Swin Transformer	0.90	0.92	Transformer
	<b>DinoV2 Large</b>	<b>0.92</b>	<b>0.92</b>	Transformer
CLAHE	DenseNet201	0.82	0.82	Transfer Learning
	EfficientNetB5	0.87	0.87	Transfer Learning
	ViT	0.87	0.88	Transformer
	Swin Transformer	0.91	0.92	Transformer
	<b>DinoV2 Large</b>	<b>0.92</b>	<b>0.93</b>	Transformer
Original	DenseNet201	0.42	0.36	Transfer Learning
	EfficientNetB5	0.86	0.86	Transfer Learning
	ViT	0.77	0.73	Transformer
	Swin Transformer	0.82	0.77	Transformer
	DinoV2 Large	0.84	0.83	Transformer

Table I shows the classification performance of five models evaluated on four image pre-processing pipelines: RGB, CMY, CLAHE, and unprocessed images. Among all configurations, the proposed DinoV2 Large model achieved the highest performance with an accuracy and F1-score of 0.93 on RGB images. Its performance remained stable across CMY and CLAHE variants (F1-score = 0.92–0.93), indicating strong generalization to contrast-enhanced inputs.

Traditional CNNs such as DenseNet201 exhibited lower performance overall, particularly on raw, unprocessed images (F1-score = 0.36), underscoring their sensitivity to input quality

and limited capacity to generalize without explicit enhancement. In contrast, transformer-based models (e.g., ViT and Swin Transformer) demonstrated significant improvements, particularly on RGB and CLAHE images, validating their superior feature extraction and spatial modeling capabilities. Swin Transformer consistently approached DinoV2’s performance but remained marginally lower across all pre-processing methods.

From an ablation perspective, results reveal that image pre-processing has a substantial effect on performance across all architectures. While RGB inputs yielded the best outcomes overall, CMY and CLAHE provided similar gains when used with robust architectures, especially DinoV2 and Swin Transformer. Given CMY’s computational efficiency relative to CLAHE, its comparable performance (F1-score = 0.92) makes it a compelling alternative for deployment in real-time or resource-constrained environments.

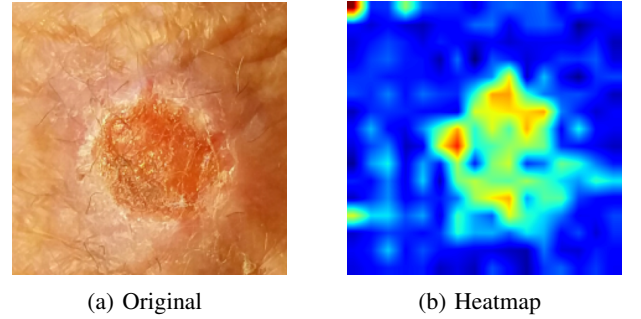


Fig. 5: DinoV2-Large Model’s Grad-CAM Visualization (a) Original, and (b) Heatmap.

Finally, to enhance model interpretability, Grad-CAM (Figure 5) visualizations were employed to highlight salient regions guiding the model’s predictions. These visual explanations further corroborate DinoV2’s ability to attend to diagnostically relevant areas across pre-processing variants.

## VI. CONCLUSION

In this study, we developed and evaluated a deep learning framework for multi-class skin disease classification, focusing on the influence of image pre-processing techniques and model architecture on diagnostic accuracy and interpretability. Our comparative analysis of standard RGB, CMY-transformed, and CLAHE-enhanced images across both pre-trained convolutional neural networks and vision transformers revealed that CLAHE pre-processing significantly improves classification performance. Transformer-based models consistently outperformed CNNs, especially in handling visually similar and complex lesion types, by leveraging their ability to model long-range dependencies. Furthermore, Grad-CAM visualizations demonstrated that models trained on CLAHE-enhanced images exhibited improved focus on clinically relevant lesion regions, supporting the explainability and reliability of predictions. These findings underscore the importance of selecting appropriate pre-processing strategies and leveraging advanced



architectures to enhance the robustness, fairness, and trustworthiness of computer-aided dermatological systems.

In the future, we would like to explore domain-specific augmentation techniques, multi-modal data fusion (e.g., combining clinical metadata with images), and the development of lightweight, deployable models for real-world applications in low-resource clinical settings.

## REFERENCES

- [1] H. P. Soyer, G. Argenziano, S. Chimenti, and et al., "Dermoscopy of pigmented skin lesions: Results of a consensus meeting via the internet," *Journal of the American Academy of Dermatology*, vol. 45, no. 5, pp. 679–693, 2001.
- [2] A. C. Geller, S. M. Swetter, K. Brooks, and et al., "Screening, early detection, and trends for melanoma: Current status (2000–2006) and future directions," *Journal of the American Academy of Dermatology*, vol. 57, no. 4, pp. 555–572, 2007.
- [3] M. Binder, H. Kittler, A. Seeber, and et al., "Epiluminescence microscopy: A useful tool for the diagnosis of pigmented skin lesions for formally trained dermatologists," *Archives of Dermatology*, vol. 131, no. 3, pp. 286–291, 1995.
- [4] L. K. Ferris, R. J. Harris, K. Gross, and et al., "Accuracy of skin cancer diagnosis by dermatologists: A prospective study," *Journal of the American Academy of Dermatology*, no. 5, pp. 793–799, 2015.
- [5] A. Gupta, K. Nischal, and U. Khopkar, "Diagnostic accuracy in viral exanthems: A comparative clinical study," *Indian Journal of Dermatology, Venereology, and Leprology*, vol. 76, no. 5, pp. 494–498, 2010.
- [6] H. J. Jeong, J. H. Lee, and J. M. Kim, "Clinical differentiation of monkeypox, chickenpox, and measles: A diagnostic challenge," *Clinical Infectious Diseases*, vol. 75, no. 3, pp. e859–e867, 2022.
- [7] A. Esteva, B. Kuprel, R. A. Novoa, and et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.
- [8] C. Barata, M. E. Celebi, and J. S. Marques, "A survey of feature extraction in dermoscopy image analysis of skin cancer," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 3, pp. 1096–1109, 2018.
- [9] P. Tschandl, C. Rinner, Z. Apalla, and et al., "Human–computer collaboration for skin cancer recognition," *Nature Medicine*, vol. 26, pp. 1229–1234, 2020.
- [10] A. Dosovitskiy, L. Beyer, A. Kolesnikov, and et al., "An image is worth 16x16 words: Transformers for image recognition at scale," *International Conference on Learning Representations (ICLR)*, 2021, <https://arxiv.org/abs/2010.11929>.
- [11] Y. Zhou, H. Han, J. Zhang, and et al., "Multi-modal fusion network with attention mechanism for skin disease diagnosis," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 2021, pp. 392–400.
- [12] E. A. Taufik, A. F. Parsa, and S. A. M. Mostafa, "Efficient Leaf Disease Classification and Segmentation using Midpoint Normalization Technique and Attention Mechanism," *arXiv preprint arXiv:2505.21316*, 2025.
- [13] K. Jeyageetha and R. Ramesh, "Region growing segmentation optimized with mhbo for skin cancer detection," *Journal of Ambient Intelligence and Humanized Computing*, 2023, early access.
- [14] M. Qasim, S. Khan, Y. Lee, and G. Jeon, "A deep spiking neural network with vgg-13 for skin cancer classification," *IEEE Access*, vol. 11, pp. 12 345–12 356, 2023.
- [15] M. Abbas, S. Zubair, and S. M. Shah, "Skin lesion classification using cnns: A comparative study on ham10000 dataset," *Computers in Biology and Medicine*, vol. 148, p. 105819, 2022.
- [16] M. Al Helal, M. S. Haydar, and S. A. M. Mostafa, "Algorithms Efficiency Measurement on Imbalanced Data using Geometric Mean and Cross Validation," in *2016 international workshop on computational intelligence (IWCI)*. IEEE, 2016, pp. 110–114.
- [17] S. A. M. Mostafa, O. Faruque, C. Wang, J. Yue, S. Purushotham, and J. Wang, "gWaveNet: Classification of Gravity Waves from Noisy Satellite Data Using Custom Kernel Integrated Deep Learning Method," in *International Conference on Pattern Recognition*. Springer, 2025, pp. 164–180.
- [18] S. A. M. Mostafa, C. Wang, J. Yue, Y. Hozumi, and J. Wang, "Enhancing Satellite Object Localization with Dilated Convolutions and Attention-aided Spatial Pooling," *arXiv preprint arXiv:2505.05599*, 2025.
- [19] S. A. M. Mostafa and J. Wang, "A multi-scale attention-enhanced architecture for gravity wave localization in satellite imagery," 2025. [Online]. Available: <https://arxiv.org/abs/2508.02704>
- [20] M. Sasithradevi, K. Soundrapandian, and M. Venkatesan, "Efficat: Efficientnet with attention for skin disease classification," *Expert Systems with Applications*, vol. 213, p. 119202, 2023.
- [21] R. Aditya, M. Kumar, and R. Sharma, "Impact of transfer learning on skin disease detection: A comparative analysis," in *Proceedings of the International Conference on Intelligent Computing and Applications*, 2022, pp. 145–152.
- [22] J. L. González, T. Chapman, K. Chen, H. Nguyen, L. Chambers, S. A. Mostafa, J. Wang, S. Purushotham, C. Wang, and J. Yue, "Atmospheric gravity wave detection using transfer learning techniques," in *2022 IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT)*. IEEE, 2022, pp. 128–137.
- [23] J. Khan, M. A. Alqarni, and A. Khan, "Cad-skin: A deep learning framework for skin lesion detection and classification," *Sensors*, vol. 23, no. 4, p. 1902, 2023.
- [24] J. Xiao, W. Guo, and J. Liu, "Exploring data augmentation effects on a singular illumination distribution dataset with colorjitter," in *2024 3rd International Conference on Image Processing and Media Computing (ICIPMC)*, 2024, pp. 75–81.
- [25] S. Chu and W. Bao, "Improving the accuracy of pine wilt disease detection through the integration of multispectral imaging and machine learning," in *2024 3rd International Conference on Image Processing and Media Computing (ICIPMC)*, 2024, pp. 299–304.
- [26] M. T. Ahmed, J. Paul, T. Jahan, S. N. Ali, S. M. S. Sani, N. Noor, A. N. Asma, and T. Hasan, "A web-based mpox skin lesion detection system using state-of-the-art deep learning models considering racial diversity," *Biomedical Signal Processing and Control*, vol. 98, p. 106742, 2024.
- [27] A. G. C. Pacheco et al., "PAD-UFES-20: a skin lesion dataset composed of patient data and clinical images collected from smartphones," *Mendeley Data*, V1, 2020. [Online]. Available: <https://doi.org/10.17632/zr7vgbcyr2.1>
- [28] K. R. Ahmed, M. M. Rashid, S. C. Sarkar, and M. Islam, "Multi-Class Viral Skin Lesion Dataset (MCVSLD)," *Mendeley Data*, V1, 2025. [Online]. Available: <https://doi.org/10.17632/dfztdtfssz.1>
- [29] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4700–4708, 2017.
- [30] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *Proceedings of the 36th International Conference on Machine Learning (ICML)*, vol. 97, 2019, pp. 6105–6114.
- [31] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," in *International Conference on Learning Representations (ICLR)*, 2021.
- [32] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 10012–10022.
- [33] M. Oquab, T. Darcet, T. Moutakanni, H. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. Haziza, F. Massa, A. El-Nouby et al., "Dinov2: Learning robust visual features without supervision," *arXiv preprint arXiv:2304.07193*, 2023.
- [34] H. Cheng, X. Jiang, Y. Sun, and J. Wang, "Color image segmentation: advances and prospects," *Pattern recognition*, vol. 34, no. 12, pp. 2259–2281, 2001.
- [35] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2009, pp. 248–255.
- [36] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7167–7176.
- [37] E. D. Pisano, S. Zong, B. M. Hemminger, M. DeLuca, R. E. Johnston, K. E. Muller, M. P. Braeuning, and S. M. Pizer, "Contrast limited adaptive histogram equalization improves the detection of subtle microcalcifications in mammography," *Radiology*, vol. 200, no. 3, pp. 707–717, 1998.