Attention-Enhanced Deep Learning Framework for Automated Skin Disease Classification

Nasrin Sultana Samia Tabassum Hrity Shah Tanzeem Afsar Shahriar Arefin Zummon

ABSTRACT

Skin disease is the most common health problem worldwide. It affects millions of people and often impacts both physical health and quality of life. Early detection of skin conditions is essential to prevent complications and improve treatment outcomes. The availability of Dermatologists is limited in many regions particularly in rural and underserved areas. As a result, patients often experience delays in diagnosis and treatment. In recent times, advances in artificial intelligence, especially deep learning have shown great potential in medical image analysis and automated illness detection. This research presents deep learning based framework for automated skin disease detection using dermoscopic images and incorporates the trained model into a mobile app application for practical use. The suggested system is designed to assists users in identifying possible skin conditions in an early stage that encourages timely medical consultation. This model is based in an EfficientNetV2-S convolutional neural network. Additionally Convolutional Block Attention Module (CBAM) and Generalized Mean (GeM) pooling are used. Six common skin disease categories were trained and evaluated on a dataset. A two-stage training strategy was adopted. Experimental results reveal that the suggested model could efficiently learns visual patterns that is associated with several dermatological disorders, achieving a final classification accuracy of 91.77% and a macro F1-score of 0.8548 across the evaluated disease categories. A mobile healthcare application was constructed for real-time uses. This enables users to shoot or upload skin photos and get real-time analysis provided by the trained model.

Keywords

CBAM, Computer-aided diagnosis, Dermoscopy, EfficientNetV2, GeM pooling, Image classification, Skin disease detection.

1. INTRODUCTION

Skin diseases are the most common cause of all human illnesses which affects almost 900 million people in the world [1]. Skin diseases may cause rashes, inflammation, itchiness or other skin changes. Some skin conditions may be genetic, while lifestyle factors may cause others [2]. Due to particular constant exposure to UV radiation, pollution and physical trauma, the skin becomes sensitive and turns into pathological conditions [3]. Skin disorders are the most wide health issues globally. It affects nearly one-third of the world's population at any given time. Current statistics indicate that skin disorders are the fourth leading cause of non-fatal disease burden, with over 3,000 identified conditions ranging from minor rashes to life-threatening malignancies such as Melanoma [4]. The lesions of skin disease go beyond physical symptoms [5]. Including anxiety, depression and social isolation due to the visible nature of many impacts, patients mostly bear significant psychological distress. In underdeveloped regions, the costs of specialist dermatologists, including a severe shortage of trained professionals which create a significant barrier for effective treatment [6]. This inequality often leads to delayed diagnoses, particularly in the case of skin cancers when early detection is the single most critical factor in raising survival rates. Traditionally the determinant standard for skin disease diagnosis engages the combination of evidence inspection by the trained dermatologist, dermoscopic examination, and, while it is necessary, an invasive skin biopsy [7]. The highly effective controlled clinical environment possesses inherent limitations. Optical diagnosis is implicitly subjective and strongly dependent on the clinician's experience and the quality of lighting. Besides, the specialized equipment required for dermoscopy is mostly unavailable in rural or low-resource healthcare settings. Therefore, there is an emergency global need for an accessible, objective, and automated screening system that can provide rapid preliminary assessments without the immediate necessity of a specialist. In the current decades, the field of medical imaging has been revolutionized by the appearance of Artificial Intelligence (AI) and Deep Learning (DL). [8] Convolutional Neural Networks (CNNs) has exhibits remarkable capabilities in pattern recognition, often matching or exceeding human performance in specific diagnostic tasks. Yet many early deep learning models are computationally heavy. It requires important memory and processing power that will obstruct their deployment on mobile and edge devices [9]. This carried the development of most efficient architectures that emphasize both performance and speed. The study focuses on the implementation of a state-of-the-art diagnostic system that utilizes the EfficientNetV2-S architecture. EfficientNet means paradigm shift in neural network design through the innovative "Compound Scaling" method [10]. Unlike traditional scaling techniques that focus solely on increasing the depth (number of layers) and width (number of channels) in the network, EfficientNetV2-S uses a compound coefficient to uniformly scale depth, width and image resolution [11]. This results in a model that is significantly more compact and faster in maintaining high-level feature extraction capabilities. In skin disease classification, where subtle variations in texture, color, and border irregularity are critical, the balanced scaling of EfficientNetV2-S provides a distinct advantage in capturing fine-grained pathological features [12]. A major challenge to training robust deep learning models for dermatology is the scarcity and imbalance of high-quality clinical datasets. In a real-world setting, some skin conditions are far more common than others, leading to datasets where models may become biased toward the majority classes. To overcome this, our research incorporates the Augmentor library, an advanced data augmentation tool. To apply various transformations such as rotation, flipping and zooming in the original clinical images, we synthesized a balanced dataset of 3942 images for each of the six disease categories. The final component in this work is the integration of trained model to a portable smartphone application. To leverage the high-definition cameras and processing power of modern mobile devices, the proposed system allows users to capture an image of a skin lesion and receive an instantaneous classification. This ”pocket-dermatologist” approach serves as a decision-support system, empowering primary care providers and individuals in underserved communities to identify potentially dangerous conditions at an early stage.

2. LITERATURE REVIEW

The evolution of automated dermatological diagnostics is marked by a transition from traditional machine learning to sophisticated attention-aware deep learning frameworks[13] Earlier benchmarks established by Hameed et al. [14] utilized Support Vector Machines (SVM) with a quadratic kernel. It achieved an 83% accuracy and an F1 score of 0.81 across five classes. While Nasr-Esfahani et al. [15] introduced standard Convolutional Neural Networks (CNNs) for binary classification. Their 81% accuracy highlighted the challenges of capturing complex morphological features. Furthermore, Hajgude et al. [16] utilized an SVM-based approach for three diseases, reaching a 90.7% accuracy. However, such traditional methods often struggle with scalability as the number of disease categories increases. In contrast, our proposed work utilizes the EfficientNetV2S-CBAM-GeM architecture. It achieves a robust 91.77% classification accuracy and a macro-F1 score of 0.8548 across a much broader range of six disease categories. A defining feature of our approach is the two-stage training strategy. During the initial stage, the classification head was adapted to the dataset, with the macro-F1 score increasing from 0.4466 to 0.5846. The subsequent fine-tuning stage—where the backbone layers were unfrozen—allowed the network to learn domain-specific patterns, pushing the F1-score to its peak of 0.8548 and significantly reducing validation loss to 0.4385. The performance of our model is further enhanced by the Convolutional Block Attention Module (CBAM), which directs the network to focus on informative lesion regions while suppressing irrelevant background noise. This is complemented by Generalized Mean (GeM) pooling, ensuring the model captures hierarchical features such as texture irregularities and pigmentation differences. By reaching high accuracy and a strong F1-score across six classes without requiring manual segmentation or external metadata, our model offers a highly specialized and balanced solution for real-time dermoscopic analysis in mobile environments. Table 1 compares the proposed model with existing methods in terms of the number of classes, accuracy and F1-score. The proposed approach achieved the highest accuracy while effectively handling six disease classes.

3. MATERIALS AND METHODS

This section explains the overall methodology of the proposed deep learning-based skin disease detection system. The workflow of this study involves several stages including dataset preparation, image preprocessing, deep learning architecture and the training strategy for model optimization.

Table 1. Comparative Analysis of Skin Disease Detection Methods

Author	Method	No. of Classes	Accuracy	F1-Score
Hameed et al.	SVM (Quadratic Kernel)	5	83%	0.81
Nasr-Esfahani et al.	Standard CNN	2	81%	0.79
Hajgude et al.	SVM	3	90.7%	0.89
Proposed Work	EfficientNetV2-S-CBAM-GeM	6	91.77%	0.8548

3.1 Dataset Description

This study utilized a publicly available benchmark dataset titled ’Skin Disease Dataset’, created by Mishra and published via Kaggle [17]. The dataset serves as an open-access image resource for training deep learning models in automated dermatological diagnostics. The full dataset can be accessed directly through the official repository link: https://www.kaggle.com/datasets/pacificrm/skindiseasedataset.
This study extricates a targeted subset that consists of 3,942 images across six prominent classes are: Acne, Actinic Keratosis, Eczema, Infestation Bites, Moles and Sun/Sunlight Damage—to evaluate the proposed architecture. The total image distribution allocates 3,548 samples for training and validation, alongside 394 samples explicitly reserved for testing. The structured class breakdown is summarized in Table 2.

Table 2. Class Distribution Across Training and Testing Subsets

Class	Training Images	Testing Images
Acne	593	65
Actinic Keratosis	748	83
Eczema	1010	112
Infestation Bites	524	60
Moles	361	40
Sunlight Damage	312	34
Total	3548	394

The dataset contains a moderate level of class imbalance. For example, eczema has the highest number of images while Sunlight Damage contains comparatively fewer samples. To minimize the effect of this imbalance during model training, class-balanced sampling strategies were applied. Before being passed into the neural network all images were resized and normalized to ensure a consistent input format across the dataset.

3.2 Data Preprocessing

In this study, several preprocessing techniques were applied to prepare the dermoscopic image dataset before training the classification model. First the system resized all input images to a uniform resolution of 224×224 pixels. After resizing, image normalization was performed to stabilize the training process. Normalization adjusts the pixel intensity values so that they fall within a consistent numerical range. Data augmentation techniques were applied during the training stage. In this work, several augmentation strategies were implemented including horizontal flipping, random rotations, random cropping and scaling. All these adjust brightness and contrast. These transformations help simulate the types of variations that may occur when images are captured in real-world situations. For instance, users may capture skin images under different lighting conditions from different angles or using devices with varying

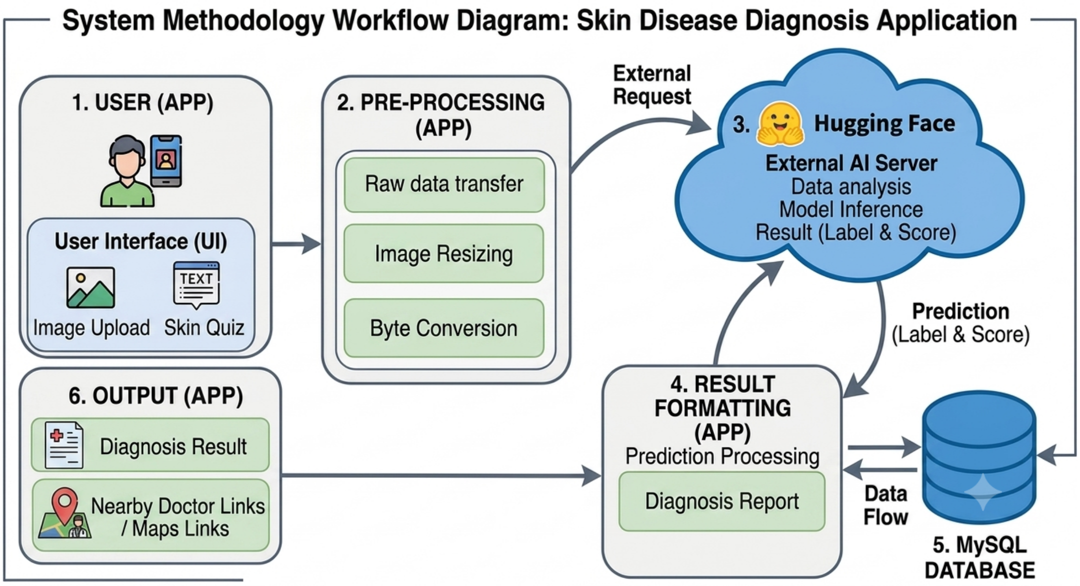

Fig. 1. Overall Workflow Diagram

camera quality. By exposing the model to these variations during training, the network learns more robust feature representations and becomes better prepared to handle real-world inputs. Finally, the dataset was divided into training and validation subsets to evaluate model performance during the training phase. A validation split of 15% was used to monitor the learning process and detect potential overfitting. Stratified sampling was applied when creating the split so that each disease category remained proportionally represented in both the training and validation sets. This approach ensures that the evaluation metrics accurately reflect the model's ability to generalize across all classes present in the dataset.

3.3 Proposed Model Architecture

The proposed skin disease classification model is built on a deep convolutional neural network architecture. This is designed to effectively capture visual patterns present in dermoscopic images.

3.3.1 EfficientNetV2 Backbone. The study proposed EfficientNetV2-S as the backbone [18]. The architecture contains multiple convolutional blocks such as fused mobile inverted bottleneck layers. This improves both training efficiency and inference speed. In this study, the EfficientNetV2-S backbone works as the primary feature extractor. It transforms the input image into high-level feature maps. These capture important visual patterns, skin texture variations, lesion boundaries, pigmentation differences and structural irregularities.

3.3.2 Convolutional Block Attention Module (CBAM). To enhance the network's ability to identify important visual features a Convolutional Block Attention Module (CBAM) was integrated into the architecture [19]. CBAM is an attention mechanism that sequentially applies two complementary operations: channel attention and spatial attention. The channel attention module identifies which feature channels contain the most useful information for classification. The network can emphasize disease-related patterns. After channel attention, the spatial attention mechanism identifies the most informative spatial regions within the feature maps. This allows the model to focus on areas where lesions are present while ignoring irrelevant background regions. By combining both attention mechanisms, CBAM improves the model's ability to capture subtle visual cues that are often present in dermatological images.

3.3.3 GeM Pooling Layer. Following the attention module, the architecture employs a Generalized Mean (GeM) pooling layer to aggregate spatial features [20]. Unlike traditional pooling methods such as average pooling or max pooling, GeM pooling introduces a learnable parameter that allows the network to adaptively control how features are aggregated. This flexibility enables the model to capture more informative feature representations.

3.3.4 Classification Layer. After feature extraction and pooling, the resulting feature vector is passed through a fully connected classification layer. This layer maps the extracted features to the six predefined disease categories. A softmax activation function is applied to convert the outputs into probability values representing the likelihood of each class. The class with the highest probability is selected as the final prediction produced by the model.

3.4 Training Strategy

This study followed a well-structured training strategy to ensure stable learning and reliable model performance. Initially a two-stage training procedure was adopted. In the first stage the backbone layers of the EfficientNet network were kept frozen. Only the newly added classification layers were trained at this stage. This approach allows the model to adjust the final layers specifically for the skin disease classification task. By doing this, the model can quickly adapt to the new dataset without disturbing the previously learned feature representations. After that, the training process entered the second stage where the backbone layers were gradually unfrozen. At this stage the entire network was allowed to update its parameters through fine-tuning. This gradual adjustment of the network improves the model's ability to capture domain-specific features relevant to dermatological image analysis. To optimize the learning process the AdamW optimization algorithm was used during training. This adaptive behavior allows the model to converge more efficiently compared to traditional optimization methods. For the classification objective, categorical cross-entropy loss was employed. This loss function is commonly used in multi-class classification problems. It measures the difference between the predicted probability distribution produced by the model and the true class labels. By minimizing this loss, the model gradually improves its predictions and learns to assign higher probabilities to the correct disease categories. Another important challenge addressed during training was class imbalance in the dataset. Some skin disease categories contain more samples than others. To reduce this bias a weighted sampling strategy was applied during training. A WeightedRandomSampler was used to ensure that samples from less frequent classes appear more often in training batches. Throughout the training process, model performance was continuously monitored using validation accuracy. The model that achieved the best validation performance was saved and later used for final evaluation on the testing dataset. This model selection strategy helps ensure that the chosen model generalizes well to unseen data rather than simply memorizing patterns from the training set.

4. RESULT ANALYSIS

This section presents the experimental evaluation of the proposed deep learning–based skin disease detection system. The performance of the model was assessed using a testing dataset. The evaluation focused on classification accuracy, confusion matrix analysis, and qualitative prediction results to provide both quantitative and practical insights into the effectiveness of the proposed system.

4.1 Training Performance

The training performance of the proposed skin disease classification model was evaluated using a two-stage training strategy. This approach was adopted to gradually adapt the pretrained EfficientNetV2-S backbone to the skin disease dataset. During the first stage the backbone network was frozen and only the newly added classification layers were trained. This stage served as a warm-up phase where the model learned to map extracted features to the six diseases. At the first six epochs, the model showed consistent improvement. The training accuracy increased from 0.3793 to 0.6075 and the validation accuracy improved from 0.4802 to 0.6087. Similarly, the macro-F1 score increased from 0.4466 to 0.5846. It indicates that the classification head was gradually adapting to the dataset. At the second stage, the backbone layers were unfrozen and the entire network was fine-tuned. This stage allowed deeper layers of the network to learn domain-specific patterns present in dermoscopic images. The model showed significant performance improvement during this phase. Validation accuracy increased from 0.7599 to 0.8752. The macro-F1 score improved to 0.8548. Validation loss was gradually reduced. The validation loss decreased from 0.7003 to 0.4385. These suggest that the model was learning increasingly discriminative feature representations. These improvements indicate that the fine-tuning stage allowed the network to better capture disease-specific visual patterns. Overall, the two-stage training strategy proved effective in improving model performance and stabilizing the learning process.

4.2 Classification Accuracy

The classification accuracy of the proposed model was evaluated using the testing dataset. Providing an explicit quantitative evaluation, the proposed network was evaluated on the completely held-out test set consisting of 394 images. The framework achieved a final overall accuracy of 91.77% and a macro F1-score of 0.8548. The use of EfficientNetV2-S as the backbone network allowed the system to extract rich hierarchical features from dermoscopic images. These features capture various dermatological characteristics such as lesion boundaries, color variations, texture irregularities, and pigmentation differences. In addition to the backbone network, the CBAM attention mechanism contributed to improved classification accuracy. It guides the model to focus on the most informative regions within the image. Instead of treating the entire image equally, the attention module emphasizes lesion areas while suppressing irrelevant background information. This is particularly important in dermoscopic images where the diseased region may occupy only a small portion of the frame. Another factor contributing to improved performance is the use of data augmentation techniques during training. Augmentation introduced variations in lighting conditions, orientation, and scale. It allows the model to learn more generalized feature representations. The overall summary metrics obtained from the test environment are formally presented in Table 3.

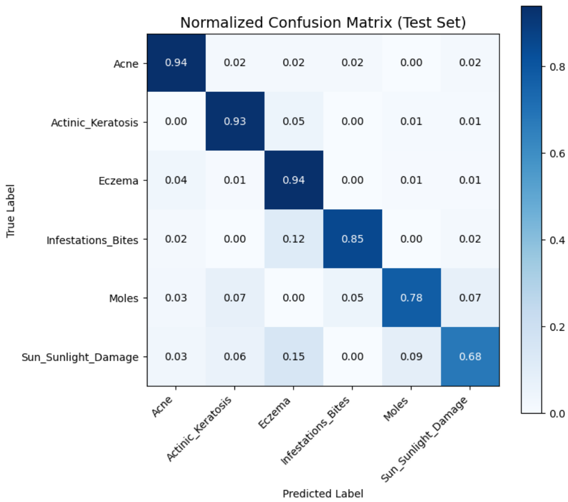

Fig. 2. Confusion Matrix Analysis

Table 3. Overall Performance Metrics on the Held-Out Test Set

Metric	Obtained Value
Overall Accuracy	91.77%
Macro Precision	0.8643
Macro Recall (Sensitivity)	0.8550
Macro F1-Score	0.8548

4.3 Prediction Confidence and Mobile Application Performance

The proposed model also generates a confidence score for each prediction. This value represents the probability of the predicted class by the softmax output layer. It indicates how certain the model is about its decision. Higher confidence values generally suggest stronger model certainty and lower values indicate that the prediction is less reliable. The system avoids presenting the result as a definitive diagnosis and instead recommends that the user consult a qualified dermatologist. Qualitative evaluation was also conducted by examining several prediction examples from the testing dataset.

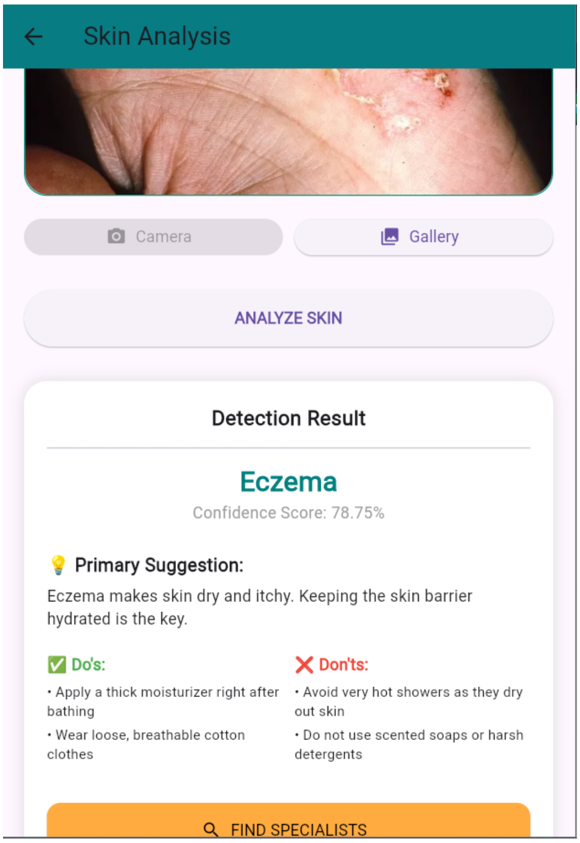

Fig. 3. Mobile App Implementation

In many cases the model successfully identified visible dermatological patterns. It also produced correct predictions with high confidence. The trained model was integrated into a mobile healthcare application to evaluate its practical usability. Users can capture or upload skin images using a smartphone. After the image is sent to an AI inference server for analysis. The model processes the image and returns the predicted disease category with the confidence score. The application then presents the result with precautionary guidance. Additionally a dermatologist recommendation feature allows users to browse doctor information. This enables users to seek professional consultation when needed.

4.4 Discussion

This research explores the use of deep learning techniques for automated skin disease classification using dermoscopic images. The primary objective is to develop a reliable system that assists users in identifying common skin conditions through a mobile application. The mentioned approach combines an EfficientNetV2-S based convolutional neural network with attention mechanisms that integrate the trained model into a practical mobile healthcare platform. The results demonstrate that the suggested model is capable of learning meaningful visual features from dermoscopic images. The training process showed steady improvement in performance as the model progressed through the two-stage training strategy. During this initial stage, where the backbone network remained frozen, the classification layers gradually adapted to the skin disease dataset. Among the key factors contributing to the model's performance is the use of EfficientNetV2-S as the feature extraction backbone. Then the system utilized the Convolutional Block Attention Module (CBAM) to enhance feature representation. The attention mechanism allows the network to focus on the most informative regions within an image. The dataset used in this study contains variations in lighting conditions, skin tones and image quality. Data augmentation was performed using techniques such as rotation, flipping and brightness adjustments. This helped the network learn more robust feature representations and improved its ability to generalize to unseen images. Another notable aspect of this research is the integration of the trained model within a mobile application environment. The addition of a dermatologist recommendation feature further enhances the usability of the system. Instead of presenting the AI prediction as a final diagnosis the application provides precautionary suggestions. It allows users to consult professional dermatologists when necessary. Despite the promising results it is important to recognize that automated skin disease detection remains a challenging problem. Dermatological images often contain subtle visual differences between conditions and even experienced clinicians may require additional clinical context for accurate diagnosis. Overall the findings of this research demonstrate that combining deep learning models with mobile healthcare technologies can provide a practical solution for preliminary skin disease screening. Future improvements include expanding the dataset with more diverse skin conditions.

5. CONCLUSION

Deep learning algorithms are used to detect various skin diseases. Skin Disease detection is important to reduce the death rate. The dermatological process to detect skin disease type is very costly compared to deep learning. In this research, an automated diagnostic system for the classification of six different skin diseases was successfully developed that leverages the EfficientNetV2-S deep learning architecture and advanced image processing techniques. By employing the Augmentor library to address data imbalance, a robust training environment was established, resulting in a classification accuracy of 91.77% and a validation accuracy of 87.52%. These findings demonstrate that the proposed model, through this unique compound scaling method, provides a highly efficient and reliable performance suitable for mobile platforms. This system has been integrated into a smartphone application, offering a portable and cost-effective screening tool. This study has the potential to function as a vital decision-support system for healthcare providers, especially in low-resource settings where access to expert dermatologists is limited. By facilitating early detection and rapid assessment, this automated approach contributes significantly to improving dermatological care and reducing the global burden of skin diseases.

6. REFERENCES

[1] Hay, R., Bendeck, S. E., Chen, S., Estrada, R., Haddix, A., McLeod, T., Mahé, A. (2006). Skin diseases. Disease Control Priorities in Developing Countries. 2nd edition.

[2] Pezzolo, E., Naldi, L. (2020). Epidemiology of major chronic inflammatory immune-related skin diseases in 2019. Expert review of clinical immunology, 16(2), 155-166.

[3] Ali, A., Khan, H., Bahadar, R., Riaz, A., Asad, M. H. H. B. (2020). The impact of airborne pollution and exposure to solar ultraviolet radiation on skin: mechanistic and physiological insight. Environmental Science and Pollution Research, 27(23), 28730-28736.

[4] Basra, M. K., Shahrukh, M. (2009). Burden of skin diseases. Expert Review of Pharmacoeconomics Outcomes Research, 9(3), 271-283.

[5] Habif, T. P., Campbell, J. L., Dinulos, J. G., Chapman, M. S., Zug, K. A. (2011). Skin disease e-book: diagnosis and treatment. Elsevier Health Sciences.

[6] Chang, A. Y., Kiprono, S. K., Maurer, T. A. (2017). Providing dermatological care in resource-limited settings: barriers and potential solutions. British Journal of Dermatology, 177(1), 247-248.

[7] Ibrahim, H., El-Taieb, M., Ahmed, A., Hamada, R., Nada, E. (2017). Dermoscopy versus skin biopsy in diagnosis of suspicious skin lesions. Al-Azhar Assiut Medical Journal, 15(4), 203-209.

[8] Lamba, R. (2025). Advances in AI for medical imaging: a review of machine and deep learning in disease detection. Procedia Computer Science, 260, 262-273.

[9] Justus, D., Brennan, J., Bonner, S., McGough, A. S. (2018, December). Predicting the computational cost of deep learning models. In 2018 IEEE international conference on big data (Big Data) (pp. 3873-3882). IEEE.

[10] Mohan, A. (2024). Enhanced multiple dense layer efficientnet. Purdue University. .

[11] Sridhar, P., Sathiya, R. R., Jayashree, P., Kaviya Sri, A., Lakshaya, L. (2024, November). EfficientNet for Enhanced Detection of Medicinal Plants: A Deep Learning Approach. In International Conference on Smart System for Integrated Computing and Communication (pp. 177-190). Cham: Springer Nature Switzerland..

[12] Ahmed, W. A., Massoud, M. A., El-Bouridy, M. E. (2024). Optimizing MRI-based medical diagnosis: Comparative analysis of EfficientNet performance with varying learning rates. Journal of the Egyptian Society of Tribology, 21(2), 105-119.

[13] Zhang, J., Zhong, F., He, K., Ji, M., Li, S., Li, C. (2023). Recent advancements and perspectives in the diagnosis of skin diseases using machine learning and deep learning: A review. Diagnostics, 13(23), 3506.

[14] Hameed N, Shabut A, Hossain MA. A Computer-aided diagnosis system for classifying prominent skin lesions using machine learning. In: 2018 10th computer science and electronic engineering (CEEC); 2018. pp. 186–91.

[15] Nasr-Esfahani E, Samavi S, Karimi N, Soroushmehr SRM, Jafari MH, Ward K, et al. Melanoma detection by analysis of clinical images using a convolutional neural network. Annu Int Conf IEEE Eng Med Biol Soc. 2016:1–4. https://ieeexplore.ieee. org/document/7590963

[16] Hajgude, M. J., Bhavsar, A., Achara, H., Khubchandani, N. (2019). Skin disease detection using image processing with data mining and deep learning. International Research Journal of Engineering and Technology (IRJET), 6(4), 4363-4366.

[17] P. K. Mishra, "Skin Disease Dataset," Kaggle Repository, 2024. [Online]. Available: `https://www.kaggle.com/datasets/pacificrm/skindiseasedataset`.

[18] Tan, M., Le, Q. (2021, July). Efficientnetv2: Smaller models and faster training. In International conference on machine learning (pp. 10096-10106). PMLR.

[19] Woo, S., Park, J., Lee, J. Y., Kweon, I. S. (2018). Cbam: Convolutional block attention module. In Proceedings of the European conference on computer vision (ECCV) (pp. 3-19).

[20] Radenović, F., Tolias, G., Chum, O. (2018). Fine-tuning CNN image retrieval with no human annotation. IEEE transactions on pattern analysis and machine intelligence, 41(7), 1655-1668.