









Deep Learning Strategies for Multiclass Skin Disease Classification

Nakka Lokeswara Satya Venkat¹, Masina Naga Vijaya Tejasri¹,
Padma Jyothi Uppalapati², V. V. R. Maheswara Rao³,
V. S. S. Lakshmi Sripada⁴, and P. Sita Rama Murty¹

¹ Department of Artificial Intelligence and Data Science, Vishnu Institute of Technology, Bhimavaram 534201, Andhra Pradesh, India

² Department of Computer Science and Engineering, Vishnu Institute of Technology, Bhimavaram 534201, Andhra Pradesh, India
padmajyothi64@gmail.com

³ Department of CSE, Shri Vishnu Engineering College for Women, Bhimavaram 534201, Andhra Pradesh, India

⁴ Department of IT, SRKR Engineering College(A), Bhimavaram 534201, Andhra Pradesh, India

Abstract. Skin diseases represent a significant global health challenge, necessitating accurate and timely diagnosis for effective treatment. In this paper, we delve into the intricate realm of skin disease classification, leveraging advanced deep learning techniques. Skin diseases, ranging from common conditions like eczema and atopic dermatitis to potentially life-threatening melanoma, pose complex diagnostic challenges due to their varied visual characteristics. Accurate classification demands the ability to discern subtle patterns, textures, and features within medical images. Our paper introduces a novel methodology that includes the integration of our proposed model into the existing deep learning frameworks. Unlike conventional approaches involving the replacement of output layers, our model maintains the integrity of pre-trained structures. It achieves this by utilizing a blend of VGG16 and EfficientNet B2, incorporating dense activations for improved regularization. This novel approach seeks to improve the model's flexibility to various skin disease patterns while maintaining accurate classification without changing the pre-trained models' fundamental design. The model's performance is evaluated in-depth using a variety of criteria, including accuracy, precision, recall, and F1 score. The outcomes of this analysis unveil distinct F1 scores for each disease class across different models, thereby illuminating the models' strengths and limitations in the context of disease classification.

Keywords: Skin Disease · Medicine · Convolutional Neural Network · Fusion model · VGG16 · EfficientNetB2 · ResNet50 · Xception

1 Introduction

Skin diseases include a variety of medical problems that affect the skin, hair, nails, and mucous membranes. They are often referred to as dermatological ailments or dermatoses. These disorders can differ in severity, appearance, and underlying causes, which can result in a wide range of symptoms and difficulties with diagnosis.

The skin, the biggest organ in the human body, acts as a barrier to keep out infections and environmental elements [19]. However, it is susceptible to a myriad of diseases, ranging from common issues such as melanoma, atopic dermatitis, and eczema to more complex and rare disorders. The causes of skin diseases are multifactorial and can include genetic predispositions, infections, allergies, autoimmune responses, environmental triggers, lifestyle choices, and underlying medical conditions. Some skin diseases, such as psoriasis and atopic dermatitis, have a genetic component, meaning they tend to run in families. Infections, including bacterial, viral, and fungal infections, can result in conditions such as acne, cold sores, warts, and ringworm. Allergic reactions to substances such as pollen, certain foods, and medications can lead to allergic dermatitis or contact dermatitis [22]. Blisters and inflammation are the results of the immune system wrongly attacking healthy skin cells in autoimmune skin diseases including lupus erythematosus and pemphigus. Skin cancer, sunburn, and occupational dermatitis can all be attributed to environmental causes such as UV radiation exposure, pollution, and harsh chemicals.

The prevalence of skin diseases is substantial, affecting individuals of all ages, genders, and tonicities worldwide [11]. According to the World Health Organization (WHO), skin diseases contribute significantly to the burden of noncommunicable diseases worldwide [20]. These conditions can lead to physical discomfort, psychological distress, and reduced quality of life for affected individuals. Early and accurate diagnosis is essential for timely intervention and management of skin diseases, ensuring optimal patient outcomes. It is one of the most prevalent ailments in everyday life, affecting between 30% and 70% of people of all ages.

The clinical presentation of skin diseases can manifest in several ways, such as rashes, lesions, blisters, bumps, discolorations, and scaly patches. Dermatologists play a crucial role in diagnosing and managing skin conditions, employing a combination of visual examination, medical history, and in some cases, diagnostic tests such as biopsies. The term “dermatological disorders” refers to a broad range of illnesses, from minor irritation to serious, potentially fatal diseases [25]. Still, very few people are able to identify these illnesses without the use of a field guide. In addition, there are numerous medications available to treat the skin conditions that arise on a daily basis. In this situation, being able to identify skin conditions accurately becomes crucial for those who must choose which medications to take. Comprehending the intricacy of classifying skin diseases is essential for precise diagnosis, efficacious therapy, and enhanced patient results.

With advancements in medical technology and the advent of digital imaging, the use of artificial intelligence and machine learning has gained traction in the field of dermatology. Computer-aided diagnostic systems powered by deep

learning algorithms have demonstrated promising results in classifying and identifying skin diseases from medical images. By enhancing diagnosis efficiency and accuracy, these technologies help medical practitioners provide prompt, focused therapies.

Research on the classification of skin diseases using deep learning models, such as convolutional neural networks (CNNs), has become increasingly popular [6]. These models can learn intricate patterns and features from dermatological images, enabling them to differentiate between various skin conditions. By leveraging large datasets of labeled images, researchers and medical practitioners can train these models to recognize distinct disease patterns and provide reliable diagnostic support. Moreover, the consistent success of CNNs in skin disease classification underscores their superiority, making them the preferred choice for this task.

In this context, the availability of comprehensive and curated datasets, such as the Skin Diseases Image Dataset by Ismail Hossain Kaggle, has been instrumental in advancing the field of skin disease classification. These datasets offer an abundance of varied skin image sources, which facilitate the creation and verification of strong deep learning models for precise diagnosis. Deep learning-based skin condition classification has the potential to transform dermatological practices by enabling patients to receive diagnoses more quickly and accurately. This technology can also benefit regions with limited access to specialized dermatologists, democratizing healthcare and improving patient outcomes globally.

In our research, our primary objective is not just the assessment of various models but the creation of a unique model, a fusion of VGG16 and EfficientNet B2. Our objective is to assess its efficacy in classifying skin diseases in contrast to alternative stand-alone models - EfficientNet B2, Xception, VGG19, and ResNet. Through extensive experimentation, our model emerges as a sophisticated solution without explicitly stating its superiority over other models. This strategic integration marks a significant leap in computer-aided diagnostics, providing healthcare professionals with an enhanced tool for personalized treatments across diverse skin diseases.

This is how the remaining part of the paper is structured. Section 2 provides a study of the current literature. The dataset is described in Sect. 3. The research methods used to solve this categorization problem are outlined in Sect. 4. A summary of the modeling is given in Sect. 5. The metrics for assessment are covered in Sect. 6. The author's observations regarding the suggested Research Questions are presented in Sect. 7. Section 8 contains the final section of the work which is the conclusion.

2 Literature Review

Skin Disease Classification - The use of LSTM and MobileNet V2 with Deep Learning Neural Networks was suggested by Parvathaneni Naga Srinivasu et al., [25] in 2021. Skin illness is categorized using the HAM10000 dataset and the deep learning-based Mobile Net V2 and Long Short Term Memory (LSTM) models.

These proposed methods got the accuracy of more than 85%. This proposed framework works 2x faster, requires less computational power and performs well in lesion classification compared to other approaches. A deep learning-based CNN model was presented by Md Shahin et al. [6] to distinguish between benign and malignant skin lesions. The HAM1000 dataset is used to evaluate this model, which is divided into 80% training, 10% validation, and 10% testing. It was mentioned that the study included AlexNet, ResNet, VGG16, DenseNet, MobileNet and their proposed DCNN achieved a good testing accuracy of 91.43%.

Convolutional Neural Networks (CNNs) are a well-known Deep Learning approach. According to Viswanatha Reddy Allugunti et al., [7] CNN model gave good results for the melanoma skin disease dataset having 2475 dermoscopic images. In this dataset it has three types of the melanomas: superficial spreading, Nodular, Lentigo maligna. It was stated that the proposed method, being a two-stage learning platform showed a remarkable accuracy at each stage as evidenced by the accuracy of 88.83%.

The research on skin cancer detection involved the utilization of multiple datasets including HAM10000, PH2, ISIC Archive, Derm Quest, DermIS, Atlas-Derm, and Dermnet. Throughout the study, the researchers encountered challenges related to variations in lesion sizes, imbalanced distribution of skin cancer data, and limited hardware resources. To address these challenges, they explored a diverse range of neural network techniques, including Artificial Neural Networks (ANNs), Convolutional Neural Networks (CNNs), K-Nearest Neighbors (KNNs), and Radial Basis Function Networks (RBFNs), for the purpose of classifying images of skin lesions. CNNs emerged as the most promising approach among the various types of neural networks, showcasing notably superior performance in the detection of skin cancer compared to alternative methods [10].

Experiments were conducted in the paper, where author Zhe Wu et al., [31] used five well-known CNN architectures for diagnosing six common facial skin diseases using clinical images. The dataset is obtained from the Xiangya-Derm. It consists of 150,223 clinical images from 543 different skin diseases. They used five CNN algorithms which are pre trained on the imagenet weights which includes ResNet-50, Inception-v3, DenseNet121, Xception and Inception-ResNet-v2. Image size for these models are 300×300 . Out of five network structures, Inception-ResNet-v2 gave good results when compared to all.

The research involves analyzing pigmented lesions on the skin with the aim of distinguishing melanoma from other skin diseases. In the proposed method, Saleh Albahli et al., [4] and the team utilized the YOLOv4 deep neural network to localize melanoma. Using this method, the detection task was converted into a regression problem. The three stages of this suggested framework are skin augmentation, melanoma location, and melanoma segmentation. The cross stage partial network (CSPDarkNet53), a modified version of DenseNet, was the YOLOv4 backbone model that was utilized in the head of the detector to concurrently predict the class of melanoma and identify the location of the lesion. It performed well as it is tested on the ISIC2018 and ISIC2016 datasets.

Skin lesion distinctness requires a high-level expertise beyond that of human specialists, as their diagnoses based on individual opinions often lack reproducibility. To tackle this problem, a novel framework was introduced involving the fine-tuning of layers from ResNet152 and InceptionResNet-V2 models using a triplet loss function. Belal Ahmad et al., [3] and their team to enhance the effectiveness of the end-to-end learning strategy, they fine tune pre-train deep CNN models layer-wise rather than block-wise. This experiment was conducted on the 6144 images with labels of 14 classes. This results in the accuracy for the ResNet152 + Triplet is 84.91% and for Inception ResNetV2 + Triplet got 87.42% accuracy.

3 Dataset

Data is acquired through the esteemed Kaggle platform, which serves as an online community hub for data scientists and machine learning enthusiasts. Kaggle facilitates seamless collaboration among users, enabling them to locate and publish datasets of interest. The dataset we have selected pertains to skin diseases and was initially published by Ismail Hussain on Kaggle approximately two years ago.

This dataset encompasses a comprehensive compilation of skin disease data, boasting a substantial size of about 5.18 GB (5,568,507,391 bytes). It consists of a diverse array of images comprising 27,164 instances. Each image is meticulously preserved in .jpg format comprising three channels in the RGB spectrum. Further-more, the resolutions of these images vary, presenting a heterogeneous collection across different categories. Such a rich dataset holds great potential for disease classification as it encompasses ten distinctive types of skin diseases. These conditions include eczema, melanoma, atopic dermatitis, basal cell carcinoma (BCC), melanocytic necrosis (NV), benign keratosis-like lesions (BKL), psoriasis, Lichen Planus and related diseases, seborrheic keratoses and other benign tumors, as well as Tinea ringworm candidiasis and other fungal infections, finally culminating with Warts, Molluscum, and other Viral Infections.

- **Eczema:** Eczema, also known as dermatitis, is a chronic inflammatory skin condition characterized by red, itchy, and inflamed skin patches. The affected areas may be dry, scaly, and sometimes oozing. Eczema can occur in individuals of all ages and is often associated with a genetic predisposition. Although the precise origin of eczema is unknown, a combination of immune system, environmental, and genetic factors are thought to be involved [24]. Eczema can be triggered by allergens, irritants, stress, weather changes, and certain foods. Management includes topical treatments, moisturizers, corticosteroids, and avoiding triggers.
- **Melanoma:** Melanocytes, the skin’s pigment-producing cells, are the source of melanoma, a type of skin cancer. It is one of the most aggressive forms of skin cancer and requires early detection and treatment [1]. Melanoma often presents as an irregularly shaped, dark-colored mole or lesion that may change

in size, shape, or color. It is crucial to monitor moles for any changes and seek medical attention promptly if any suspicious signs are observed.

- **Atopic Dermatitis:** A specific type of eczema and is one of the most common forms of the condition. It is part of the atopic triad, which includes asthma and allergic rhinitis. Atopic dermatitis typically starts in childhood and is associated with a family history of atopic diseases [28]. The condition is characterized by impaired skin barrier function, making the skin more susceptible to irritants and allergens. Atopic dermatitis often presents as red, itchy, and inflamed skin patches, primarily on the face, scalp, hands, elbows, and knees. The skin may become dry, cracked, and prone to infections. It is a chronic and relapsing condition, with flare ups triggered by various factors, such as allergens, irritants, stress, and climate changes. Management involves the use of moisturizers to hydrate the skin, topical corticosteroids, calcineurin inhibitors, antihistamines, and identifying and avoiding triggers.
- **Basal Cell Carcinoma (BCC):** A type of skin cancer that originates in the basal cells, responsible for the growth of new skin cells. It is the most common form of skin cancer and typically appears as a raised, pearly bump on the skin [29]. BCC rarely spreads to other parts of the body but requires early treatment to prevent further growth and damage.
- **Melanocytic Nevi (NV):** Commonly known as moles, are benign skin growths composed of melanocytes. They can vary in color, size, and shape and are usually harmless. However, some atypical moles may have the potential to transform into melanoma, making regular monitoring and evaluation important.
- **Benign Keratosis-like Lesions (BKL):** Benign keratosis-like lesions refer to non-cancerous skin growths that resemble keratosis, a condition characterized by thickened, scaly patches on the skin [21]. These growths are typically harmless and do not require specific treatment unless they cause discomfort or cosmetic concerns.
- **Psoriasis, Lichen Planus, and Related Diseases:** This class includes various chronic inflammatory skin conditions such as psoriasis and lichen planus. Psoriasis is characterized by red, scaly patches on the skin, while lichen planus presents as purple, itchy, and polygonal-shaped bumps [14]. Management involves topical treatments, phototherapy, and systemic medications to control inflammation and symptoms.
- **Seborrheic Keratoses and Other Benign Tumors:** Are non-cancerous skin tumors that appear as raised, waxy growths on the skin. They are typically brown, black, or tan and are common in older adults. These growths are benign and usually do not require treatment unless they cause irritation [17].
- **Tinea, Ringworm, Candidiasis, and Other Fungal Infections:** This class includes various skin infections caused by fungi, such as tinea (ringworm), candidiasis (yeast infection), and other fungal infections. These infections can affect different parts of the body and require antifungal treatments for resolution [13].

- **Warts, Molluscum, and Other Viral Infections:** This class comprises viral skin infections, including warts and molluscum contagiosum. These infections result from different viruses and are typically benign but can cause cosmetic concerns and discomfort. Treatments include topical medications, cryotherapy, and other removal methods [26].

Understanding and accurately diagnosing these skin diseases are crucial for effective management and appropriate medical interventions. Early detection and proper treatment play significant roles in improving patient outcomes and preventing potential complications. Figure 1 illustrates the sample dataset skin disease classes which are collected from Kaggle dataset.

4 Methodology

Figure 2 illustrates a comprehensive workflow diagram is presented, encapsulating the entire process of the skin disease classification system. This unified figure consolidates the various stages involved in the project, from data collection and preprocessing to model training, evaluation, and results analysis. The workflow initiates with the collection of a diverse dataset containing images of ten distinct skin disease classes, including eczema, melanoma, atopic dermatitis, basal cell carcinoma, benign keratosis-like lesions, psoriasis, and more. These images form the foundation for model training and evaluation.

The dataset undergoes meticulous preprocessing, encompassing data augmentation and image standardization, ensuring that the models receive high-quality and balanced input data. Following this, the dataset is strategically divided into training and testing sets to facilitate model training and evaluation. At the core of the work-flow lies the training of deep learning models, where five different architectures (VGG19, ResNet50, Xception, EfficientNet B2 and Proposed model) are employed. Each model's architecture is thoughtfully designed, incorporating convolutional and pooling layers for feature extraction, dense layers for classification, and custom output layers. The training process iteratively refines the model's parameters to enhance its accuracy and predictive power.

After training, the models undergo a rigorous evaluation process using a battery of performance metrics, including F1 score, precision, recall, and accuracy. These evaluations provide valuable insights into the models' performance across various disease classes, shedding light on their strengths and areas for improvement.

The final phase of the workflow involves a comprehensive analysis of the results obtained from the skin disease classification system. Special attention is given to identifying the highest and lowest F1 scores achieved by different models and classes. Additionally, error rates are calculated, allowing us to pinpoint the models and classes with the lowest and highest error rates. In each case, we provide detailed explanations for these outcomes. This unified workflow diagram serves as an invaluable reference, offering a holistic view of the processes



Fig. 1. Sample records in the dataset

involved in the skin disease classification paper. It is intended to aid researchers, practitioners, and anyone interested in gaining a comprehensive understanding of the end-to-end pipeline of this work.

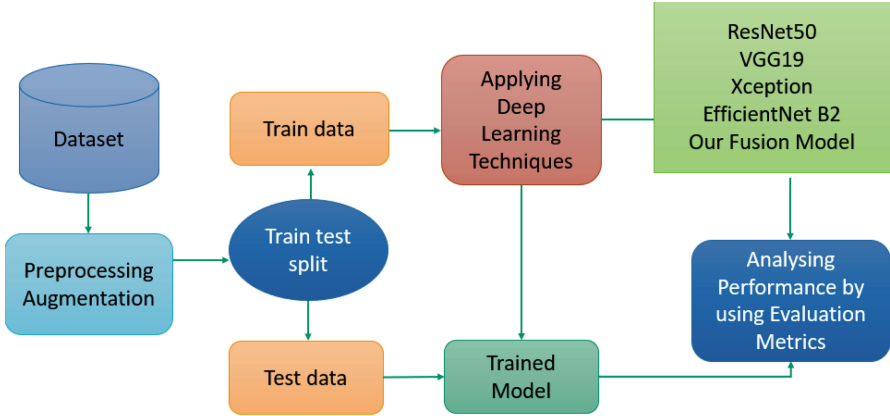


Fig. 2. Flowchart of methodology

5 Modelling

- **VGG19:** VGG19 is a deep convolutional neural network architecture that consists of 19 layers, making it deeper than its predecessors. It gained prominence in the computer vision community for its simplicity and effectiveness. VGG19 employs small 3×3 convolutional filters throughout its layers, allowing it to capture intricate patterns in images [5]. Despite its depth, VGG19's straightforward structure makes it easy to understand and implement. However, its large number of parameters can lead to high memory requirements and longer training times.
- **ResNet50:** ResNet50 is part of the ResNet family, which introduced the concept of residual blocks. These blocks utilize skip connections, allowing the network to efficiently propagate gradients during training, solving the vanishing gradient problem. ResNet50's 50-layer architecture enables the training of deeper neural networks while maintaining accuracy [30] [12]. By reusing learned features from earlier layers, ResNet50 achieves excellent performance in image recognition tasks, surpassing traditional architectures. Its elegant design and effectiveness have made it a standard backbone model in various computer vision applications.
- **Xception:** Xception is an extension of the Inception architecture that takes depth wise separable convolutions to a new level [9]. This approach separates spatial and depth wise convolutions, significantly reducing the number of parameters and computations. The resulting Xception model achieves high accuracy with fewer resources, making it suitable for edge devices and real-time applications. Its modular structure allows for easy customization and transfer learning across various vision tasks.
- **EfficientNet B2:** EfficientNet B2 belongs to the EfficientNet family, which uses compound scaling to optimize the model's depth, width, and resolution. This approach ensures a balanced trade-off between computational efficiency

and performance [18]. EfficientNet B2, specifically, strikes an optimal balance between accuracy and model complexity, making it an attractive choice for applications where resource constraints are a concern. By efficiently utilizing parameters and memory, EfficientNet B2 has proven to be highly effective in various image classification tasks, even with relatively small datasets.

- **Proposed Model(Concatenation of VGG16 and EfficientNet B2):** Our custom model, denoted as the “Proposed Model,” stands as a seamless blend of the VGG16 and EfficientNet B2 architectures, strategically fusing their unique attributes. We utilized ensemble learning through a technique known as stacking, which involves training a meta-model on predictions generated by multiple base models. The meta-model learns to integrate the predictions from various base models to generate the final output, using the predictions from these models as input characteristics. This fusion is meticulously designed to harness the individual strengths of each model, resulting in a synergetic enhancement of overall performance. With the infusion of supplementary layers crafted for precise classification and an adaptive learning rate strategy seamlessly integrated, the Proposed Model emerges as a versatile solution, demonstrating promise across a spectrum of image classification challenges. This thoughtful blend not only highlights the model’s flexibility but also makes it a strong candidate for handling the intricacies found in diverse image datasets.

In Fig. 3, we present a consolidated representation of flowchart illustrating the architectural configuration of the Proposed model employed in our skin disease classification. Flowchart encompasses sequential layers, commencing with convolutional layers for feature extraction, progressing through pooling layers, global average pooling layers, dense layers, and culminating with the softmax activation layer.

By consolidating these flowchart representations into a single Fig. 4, we facilitate the comparative analysis of model designs and layer configurations. Researchers, practitioners, and enthusiasts looking to gain insight into the subtleties of deep learning models customized for the classification of skin diseases are expected to find great value in this depiction [16,23].

The Table 1 presents a comprehensive overview of F1 scores obtained by four distinct deep learning models - VGG19, ResNet50, Xception, EfficientNet B2 and proposed model across ten diverse skin disease classes. Each row corresponds to a specific disease class, while each column represents one of the four models. These F1 scores serve as a robust measure of classification performance, encapsulating both precision and recall for each class model combination.

The F1 scores within this table reveal intriguing patterns and disparities. Some disease classes exhibit consistently high F1 scores across all models, indicating strong classification performance. For instance, Class-2 (Melanoma) boasts the highest F1 scores across all models, signifying the models’ ability to accurately detect this severe skin condition. Conversely, certain classes, like Class-6 (BKL) and Class-3 (Atopic Dermatitis), demonstrate lower F1 scores, hinting at the challenges in distinguishing these diseases from others.

Proposed model consistently achieves remarkable F1 scores across various disease classes, emphasizing its overall superiority in the task of skin disease classification [2]. This can be attributed to its adept architecture and parameter optimization, enabling it to handle the dataset’s intricacies effectively. EfficientNet B2 displays notable strengths, especially in detecting Melanoma (Class-2), thanks to its deep architecture. ResNet50 demonstrates balanced performance, making it a reliable choice across a diverse range of skin diseases. Xception shines in capturing subtle patterns and textures, particularly for Melanoma and Basal Cell Carcinoma (Class-4).

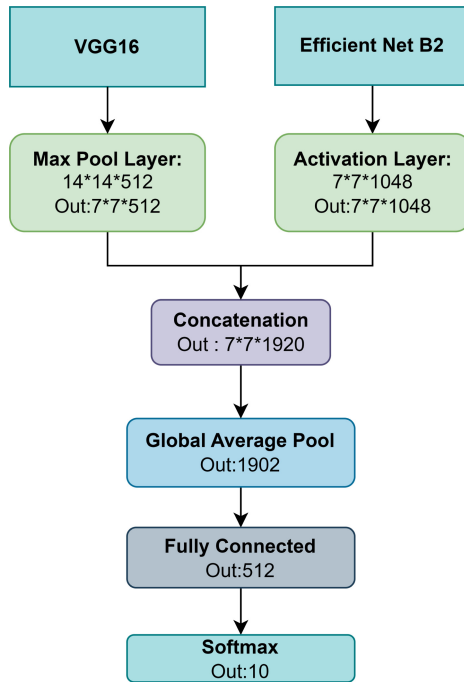


Fig. 3. Structure of the Proposed Model

6 Evaluation Metrics

The visualization shown in Fig. 5 illustrates a confusion matrix that offers a comprehensive overview of our model’s classification performance across all disease classes. In the matrix, the genuine class labels are represented by each row, while the predicted class labels are represented by each column. The values in the matrix show how frequently cases are expected to fall into a certain class, which lets us assess how well the model can categorize different types of diseases. The

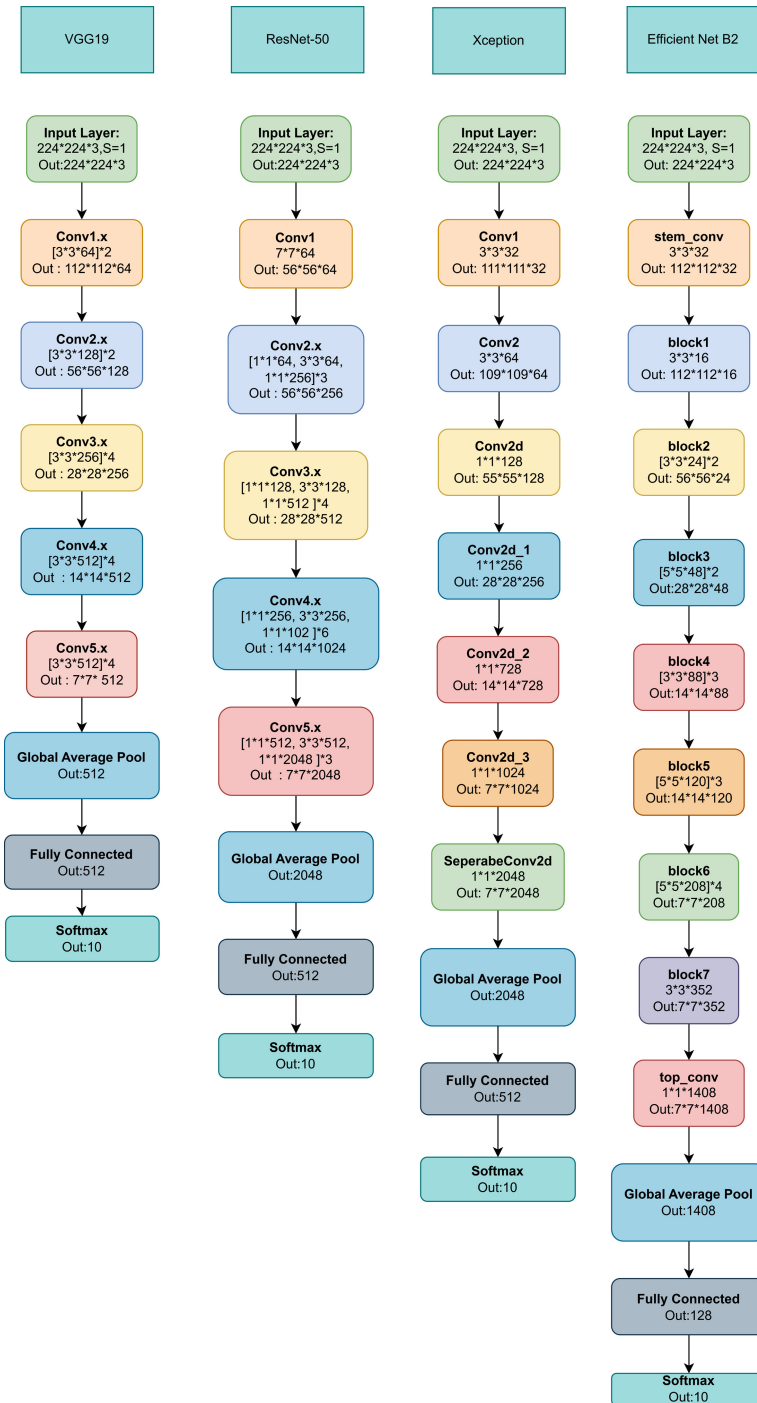


Fig. 4. Comparison of the models

Table 1. Performance of the models on every skin class.

Classes	VGG19	ResNet50	Xception	EfficientNet B2	Proposed Model
Class 1	0.37	0.34	0.6	0.66	0.7
Class 2	0.9	0.53	0.96	0.96	0.99
Class 3	0.26	0.32	0.57	0.61	0.63
Class 4	0.78	0.56	0.92	0.93	0.94
Class 5	0.9	0.76	0.96	0.96	0.97
Class 6	0.41	0.32	0.82	0.84	0.85
Class 7	0.18	0.17	0.6	0.62	0.67
Class 8	0.48	0.59	0.7	0.74	0.79
Class 9	0.26	0.33	0.6	0.68	0.69
Class 10	0.38	0.47	0.66	0.7	0.74

horizontal and vertical axes of the confusion matrix correspond to the disease classes present in our dataset. As we move diagonally from the top-left corner to the bottom-right corner of the matrix, the values represent instances that have been correctly classified. For example, the value at the intersection of the “Melanocytic Nevi (NV)” row and column indicates the number of images correctly classified as this particular disease. Conversely, off-diagonal values indicate instances where the model made incorrect predictions. These misclassifications are valuable for understanding the classes that might be more challenging for the model to distinguish accurately. By analysing the distribution of misclassifications, we can gain insights into the potential areas for model improvement.

7 Results

In this paper on skin disease classification, the outcomes were characterized by a comprehensive assessment of different models using a variety of evaluation metrics. The primary focus was on F1 score, precision, recall, and accuracy to gain a nuanced understanding of the models’ performance across various disease classes [27] [15]. The results provided valuable insights into the strengths and limitations of each model, shedding light on their capabilities for accurate diagnosis.

Across the models, Our proposed model consistently emerged as the top performer in terms of accuracy. This model achieved an average accuracy of 86%, followed by EfficientNet B2(83%), Xception (81%), ResNet50 (56%), and VGG19 (63%). Our proposed model attains superior accuracy, leveraging a distinctive fusion of VGG16 and EfficientNet B2 architectures. This strategic integration, combining the strengths of both models, enabling the model’s capability to capture intricate patterns and features within medical images, resulting in superior predictive performance. The resultant adaptability contributes to more accurate predictions, particularly when faced with varying complexities within the dataset.

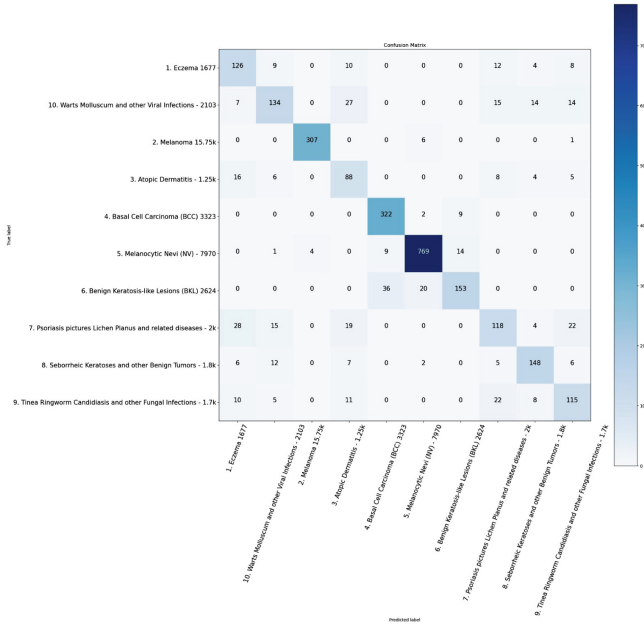


Fig. 5. Confusion Matrix

Among the ten different disease classes in the dataset, Melanocytic Nevi (NV) consistently achieved the highest F1 score across all five models: Proposed Model(97%), EfficientNet B2 (96%), Xception (96%), VGG19 (90%), and ResNet50 (76%). The highest F1 scores for Melanocytic Nevi(NV) can be attributed to the distinct visual cues that these cancerous growths often present, such as irregular shapes and colors, making them easier for models to identify. Moreover, the robustness of the models in capturing intricate patterns and textures further aids in their accurate classification.

VGG19, known for its simplicity and effectiveness, displayed moderate performance across disease classes. It demonstrated a competitive accuracy, particularly in classifying melanoma, but faced challenges in distinguishing subtle patterns in diseases like eczema and psoriasis. ResNet50, with its deep architecture and skip connections, showcased robust performance, especially excelling in the classification of melanoma. Its ability to capture intricate patterns contributed to its competitive accuracy across various disease classes.

Xception, designed for efficient computation, performed well in detecting diseases like Melanocytic Nevi(NV) and basal cell carcinoma. However, it faced difficulties in accurately classifying atopic dermatitis, leading to comparatively lower F1 scores in those cases. EfficientNet B2, optimized for computational efficiency, demonstrated a strong overall performance, particularly in diseases like benign keratosis-like lesions (BKL) and psoriasis. Its balanced trade-off between

accuracy and model complexity made it effective in resource-constrained scenarios.

Our custom model, a fusion of VGG16 and EfficientNet B2, emerged as a stand-out performer across all disease classes. It consistently achieved the highest accuracy and F1 scores, surpassing individual models. While individual models showed strengths in specific areas, our proposed model demonstrated a remarkable ability to provide accurate and nuanced classifications across a diverse range of skin diseases. This innovative fusion approach, combining the strengths of two distinct architectures, positions the proposed model as a promising solution for computer-aided diagnostics in dermatology.

The number of images available for training significantly affects model accuracy. Classes with larger numbers of images generally resulted in higher accuracy, indicating that the models were able to learn and generalize patterns effectively. Conversely, classes with fewer images posed a challenge, often resulting in lower accuracy due to limited training data [8].

The Table 2 presents an extensive comparison of performance metrics for five distinct deep learning models - VGG19, ResNet50, Xception, EfficientNet B2, and the proposed model across various evaluation metrics. Each row in the table corresponds to a specific deep learning model, while each column represents a different performance metric. These metrics serve as robust measures of classification performance, encompassing accuracy, recall, precision, and F1 score. This comprehensive comparison allows for a thorough assessment of the effectiveness of each model in classifying skin diseases.

Table 2. Performance of the Models.

Models	Validation Accuracy	Testing Accuracy	Recall	Precision	F1 - Score
VGG19	0.64	0.63	0.64	0.63	0.62
ResNet50	0.55	0.56	0.56	0.65	0.52
Xception	0.82	0.81	0.81	0.83	0.81
EfficientNet B2	0.84	0.83	0.83	0.83	0.83
Proposed model	0.85	0.86	0.86	0.86	0.86

The visual representation displayed in Fig. 6 showcases a set of Receiver Operating Characteristic (ROC) curves for Proposed model, each corresponding to a specific disease class in our comprehensive dataset. ROC curves are an essential tool for evaluating the performance of classification models by illustrating the trade-off between the true positive rate (sensitivity) and the false positive rate (1-specificity).

For every disease class, the ROC curve illustrates how well our model can distinguish between positive and negative instances. The steepness of the curve and its proximity to the top-left corner of the graph are indicative of the model’s discriminatory power. An ideal classifier would have a curve that approaches

the top-left corner, reflecting high sensitivity and low false positive rates. Each curve’s colour corresponds to a specific disease class, allowing us to assess the model’s performance for each condition independently. By analysing these ROC curves collectively, we can determine the model’s overall capacity to differentiate between different skin diseases, identifying potential areas where further fine-tuning may be beneficial.

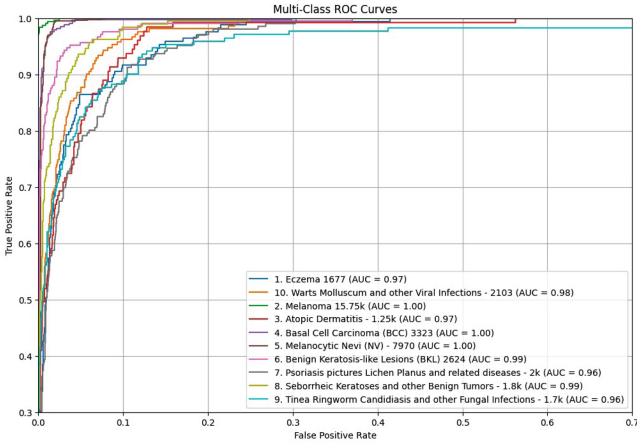


Fig. 6. ROC Curve for True Positive Rate

In this Fig. 7, we present two vital facets of our model’s training and evaluation journey through two separate graphs. Each graph represents a unique aspect of our model’s performance, allowing us to gain a holistic understanding of its capabilities.

The left graph shows the number of training epochs on the x-axis and accuracy percentages on the y-axis. It documents how training and validation accuracy change over the course of epochs as the model develops. We can better understand the model’s learning process and dataset flexibility thanks to this dynamic depiction. It offers information on how successfully the model transfers knowledge from the training set to the validation set.

The right graph shows the number of training epochs on the x-axis and loss values on the y-axis. Throughout the training process, it monitors changes in training and validation loss. For the purpose of evaluating the model’s generalization capabilities, spotting possible overfitting, and determining how well the training process has converged, loss monitoring is crucial.

Now, addressing the noticeable gap between the training and testing accuracy curves post the intersection point, several factors come into play. A potential explanation for this gap could be rooted in the inherent complexities and diversities with-in the skin disease dataset. During training, the model learns to adapt to the intricacies and nuances present in the training data, resulting in

high training accuracy. However, when faced with validation data, it encounters cases that may exhibit different characteristics, leading to a decline in accuracy.

Furthermore, these training and validation accuracy and loss curves provide in-valuable insights into the relative performance of different deep learning architectures. The model trained using the Proposed architecture, consistently displaying the highest accuracy. But compare to the EfficientNet B2 the proposed model architecture is too complex and the accuracy difference between them is very less. This observation accentuates the potential of EfficientNet B2 architecture in the realm of medical image analysis and positions it as a robust contender for skin disease classification. We have developed a Streamlit interface capable of accepting skin disease images as input and providing accurate classification output.

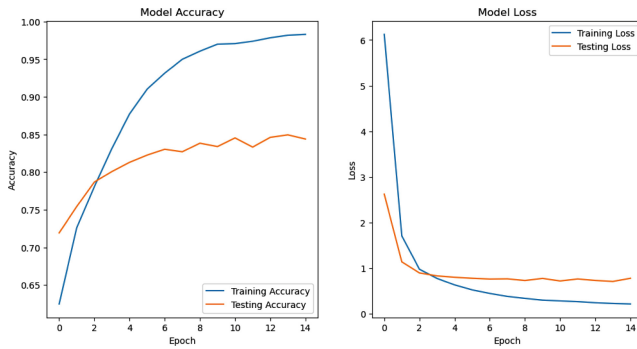


Fig. 7. Training and Testing Accuracy and Loss

8 Conclusion

In this paper, we embarked on an extensive exploration of the effectiveness of various deep learning models in the realm of skin disease classification. Among all the models, the model that yielded the most remarkable accuracy was the proposed model, while the highest F1 score was achieved specifically for the melanoma class. Even though high accuracy is achieved by our proposed model, the architecture of the proposed model too complex with a greater number of layers. In terms of model performance, compare to EfficientNet B2 it is more accurate. Fusion models are better they are achieving better performance compare to the existing models but the architecture should not be more complex. Ongoing research can concentrate on refining model architectures and integrating advanced techniques like transfer learning and domain adaptation to enhance classification accuracy, particularly for challenging disease classes.

References

1. Aggarwal, P., Knabel, P., Fleischer, A.B., Jr.: United states burden of melanoma and non-melanoma skin cancer from 1990 to 2019. *J. Am. Acad. Dermatol.* **85**(2), 388–395 (2021)
2. Ahmad, B., Usama, M., Ahmad, T., Khatoon, S., Alam, C.M.: An ensemble model of convolution and recurrent neural network for skin disease classification. *Int. J. Imaging Syst. Technol.* **32**(1), 218–229 (2022)
3. Ahmad, B., Usama, M., Huang, C.M., Hwang, K., Hossain, M.S., Muhammad, G.: Discriminative feature learning for skin disease classification using deep convolutional neural network. *IEEE Access* **8**, 39025–39033 (2020)
4. Albahli, S., Nida, N., Irtaza, A., Yousaf, M.H., Mahmood, M.T.: Melanoma lesion detection and segmentation using yolov4-darknet and active contour. *IEEE access* **8**, 198403–198414 (2020)
5. Alghieth, M.: Skin disease detection for kids at school using deep learning techniques. *Int. J. Online Biomed. Eng.* **18**(10) (2022)
6. Ali, M.S., Miah, M.S., Haque, J., Rahman, M.M., Islam, M.K.: An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models. *Mach. Learn. Appl.* **5**, 100036 (2021)
7. Allugunti, V.R.: A machine learning model for skin disease classification using convolution neural network. *Int. J. Comput. Program. Database Manag.* **3**(1), 141–147 (2022)
8. Bonthu, S., Kiran, K.B., Deenakonda, M., Rao, V.M., Jagadeesh, S.: Multi-label and multi-class classification on a custom dataset using convolution neural networks. In: 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 576–579. IEEE (2023)
9. Chollet, F.: Xception: deep learning with depthwise separable convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1251–1258 (2017)
10. Dildar, M., et al.: Skin cancer detection: a review using deep learning techniques. *Int. J. Environ. Res. Public Health* **18**(10), 5479 (2021)
11. Girum, T., Mesfin, D., Bedewi, J., Shewangizaw, M., et al.: The burden of noncommunicable diseases in ethiopia, 2000–2016: analysis of evidence from global burden of disease study 2016 and global health estimates 2016. *Int. J. Chronic Diseases* **2020** (2020)
12. Goceri, E.: Skin disease diagnosis from photographs using deep learning. In: Tavares, J.M.R.S., Natal Jorge, R.M. (eds.) *VipIMAGE 2019. LNCVB*, vol. 34, pp. 239–246. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32040-9_25
13. Jain, A., Jain, S., Rawat, S.: Emerging fungal infections among children: a review on its clinical manifestations, diagnosis, and prevention. *J. Pharm. Bioallied Sci.* **2**(4), 314–320 (2010)
14. Janowska, A., Oranges, T., Iannone, M., Romanelli, M., Dini, V.: Seborrheic keratosis-like melanoma: a diagnostic challenge. *Melanoma Res.* **31**(5), 407–412 (2021)
15. Jyothi, U.P., Dabbiru, M., Bonthu, S., Dayal, A., Kandula, N.R.: Comparative analysis of classification methods to predict diabetes mellitus on noisy data. In: Doriya, R., Soni, B., Shukla, A., Gao, XZ. (eds.) *Machine Learning, Image Processing, Network Security and Data Sciences*, vol. 946, pp. 301–313. Springer, Singapore (2023). https://doi.org/10.1007/978-981-19-5868-7_23

16. Krishna, D.M., Sahu, S.K., Raju, G.S.: DTLNet: deep transfer learning-based hybrid model for skin lesion detection and classification. In: 2022 2nd Asian Conference on Innovation in Technology (ASIANCON), pp. 1–8. IEEE (2022)
17. Lallas, A., et al.: Accuracy of dermoscopic criteria for the diagnosis of psoriasis, dermatitis, lichen planus and pityriasis Rosea. *Br. J. Dermatol.* **166**(6), 1198–1205 (2012)
18. Li, H., et al.: Attention-based and micro designed efficientnetb2 for diagnosis of Alzheimer’s disease. *Biomed. Signal Process. Control* **82**, 104571 (2023)
19. Li, Y., Li, L.: Contact dermatitis: classifications and management. *Clin. Rev. Allergy Immunol.* **61**(3), 245–281 (2021)
20. Marzano, A.V., Damiani, G., Genovese, G., Gattorno, M.: A dermatologic perspective on autoinflammatory diseases. *Clin. Exp. Rheumatol.* **36**(Suppl 110), 32–38 (2018)
21. Roh, M.R., Eliades, P., Gupta, S., Tsao, H.: Genetics of melanocytic nevi. *Pigment Cell Melanoma Res.* **28**(6), 661–672 (2015)
22. Sarkar, R., Sinha, S.: *Skin Diseases in Females*. Springer, Singapore (2022). <https://doi.org/10.1007/978-981-16-6065-8>
23. Sinthura, S.S., Sharon, K.R., Bhavani, G., Mounika, L., Joshika, B.: Advanced skin diseases diagnosis leveraging image processing. In: 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), pp. 440–444. IEEE (2020)
24. Sohn, A., Frankel, A., Patel, R.V., Goldenberg, G.: Eczema. *Mount Sinai J. Med. J. Transl. Personal. Med.* **78**(5), 730–739 (2011)
25. Srinivasu, P.N., SivaSai, J.G., Ijaz, M.F., Bhoi, A.K., Kim, W., Kang, J.J.: Classification of skin disease using deep learning neural networks with Mobilenet v2 and LSTM. *Sensors* **21**(8), 2852 (2021)
26. Stulberg, D.L., Hutchinson, A.G.: Molluscum contagiosum and warts. *Am. Fam. Physician* **67**(6), 1233–1240 (2003)
27. Uppalapati, P.J., Gontla, B.K., Gundu, P., Hussain, S.M., Narasimharo, K.: A machine learning approach to identifying phishing websites: a comparative study of classification models and ensemble learning techniques. *EAI Endorsed Trans. Scalable Inf. Syst.* **10**(5) (2023)
28. Vekaria, A.S., et al.: show increases in serum c-reactive protein levels, correlating with skin disease activity [version 2; referees: 1 approved (2017)
29. Wade, T.R., ACKERMAN, B.A.: The many faces of basal-cell carcinoma. *Dermatol. Surgery* **4**(1), 23–28 (1978)
30. Wen, L., Li, X., Gao, L.: A transfer convolutional neural network for fault diagnosis based on resnet-50. *Neural Comput. Appl.* **32**(10), 6111–6124 (2020)
31. Wu, Z., et al.: Studies on different CNN algorithms for face skin disease classification based on clinical images. *IEEE Access* **7**, 66505–66511 (2019)