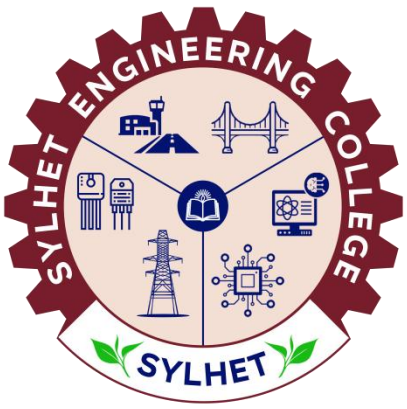
A Thesis Submitted to the Sylhet Engineering College for the Degree of
Bachelor of Science in Electrical and Electronic Engineering

A Deep Learning Approach to Build a Lightweight Skin Disease Classifier by Means of Transfer Learning.

By,
Amit Das Raj
Kriti Uchchhal Mozumder
&
Avijit Chakro Borty

Supervised by,
Salman Fazle Rabby
Assistant Professor
Department of Electrical and Electronic Engineering
Sylhet Engineering College



June, 2025
Sylhet Engineering College, Sylhet
Affiliated with
Shahjalal University of Science & Technology

The thesis titled **"A Deep Learning Approach to Build a Lightweight Skin Disease Classifier by Means of Transfer Learning"** submitted by **Amit Das Raj**, **Kriti Uchchhal Mozumder** and **Avijit Chakro Borty**; Student ID: **2019338504**, **2019338529** and **2019338556**; Session 2019-2020, to the Department of Electrical and Electronic Engineering, Sylhet Engineering College, has been accepted as satisfactory in partial fulfillment of the requirements for the Degree of Bachelor of Science in Electrical and Electronic Engineering and approved as to its style and contents.

BOARD OF EXAMINERS

Chairman

Md. Shahid Iqbal
Assistant Professor and Head,
Department of Electrical and Electronic Engineering
Sylhet Engineering College, Sylhet.

Member & Supervisor

Salman Fazle Rabby
Assistant Professor,
Department of Electrical and Electronic Engineering
Sylhet Engineering College, Sylhet.

Member

Apurbo Biswas
Assistant Professor,
Department of Electrical and Electronic Engineering
Sylhet Engineering College, Sylhet.

Member

Md. Ashraful Alam
Lecturer,
Department of Electrical and Electronic Engineering
Sylhet Engineering College, Sylhet.

Member

Mahedi Kamal Ahmed
Lecturer,
Department of Electrical and Electronic Engineering
Sylhet Engineering College, Sylhet.

Member (External)

Arif Ahammad
Assistant Professor,
Department of Electrical and Electronic Engineering
Shahjalal University of Science & Technology, Sylhet.

Acknowledgements

We begin by glorifying the Almighty Lord Krishna, the creator and sustainer of this universe. He empowered us with the strength and perseverance to complete all the academic requirements for the Bachelor of Science in Electrical and Electronic Engineering (EEE), including this thesis work. Next, it is with immense pleasure that we express our deep gratitude to our respected supervisor, **Salman Fazle Rabby**, Assistant Professor, Department of EEE, Sylhet Engineering College, for his insightful academic guidance, thoughtful suggestions, creative ideas, and constant motivation throughout our research journey. His dedicated supervision has significantly contributed to our progress and success. Moreover, We are deeply thankful to our department head, **Md. Shahid Iqbal**, faculty members **Md. Ashraful Alam** and **Mahedi Kamal Ahmed** for their unwavering support, cooperation, and encouragement during this academic pursuit. Additionally, our sincere appreciation goes to **Md. Janibul Alam Soeb**, Assistant Professor, Department of FPM (Farm Power and Machinery), Faculty of Agricultural and Engineering Technology, Sylhet Agricultural University. His expert input, valuable discussions, and consistent support have been instrumental to our thesis work.

Lastly, we owe a heartfelt debt to our beloved parents and siblings, whose love, prayers, sacrifices, and continuous support have always been our greatest strength. We are also thankful to our classmates and friends in the EEE department at Sylhet Engineering College for their companionship and encouragement throughout this journey.

Abstract

Skin diseases are common health problems around the world, and their recognition and monitoring remain major challenges for the medical industry. In severe cases, some skin diseases may even lead to skin cancer, making accurate and early detection crucial to prevent complications. While advancements in medical technology have made it possible to diagnose and detect skin diseases more quickly and accurately, their application is often limited due to high costs. Moreover, manual diagnosis by medical experts is time-consuming and subjective. Therefore, image processing techniques play a vital role in the early diagnosis of various skin diseases. In this study, an automated approach was presented for skin lesion classification using deep learning on the HAM10000 dataset. The dataset comprises dermatoscopic images of seven different types of skin lesions, including melanoma, basal cell carcinoma, and benign keratosis. ConvNeXt-Tiny model—a modern convolutional neural network architecture inspired by Vision Transformers was employed in this study. To address the inherent class imbalance in the dataset, weighted balanced class weights during training were used. The model was trained using custom focal loss with label smoothing and evaluated using accuracy, precision, recall, and F1-score. The proposed model achieved an overall accuracy of **95.83%**, with strong performance across most classes. This work demonstrates the feasibility and reliability of using deep learning models for assisting dermatologists in early skin disease detection and classification.

Keywords: *Skin disease, image processing, HAM10000, melanoma, benign keratosis, ConvNeXt-Tiny, custom focal loss, vision transformer.*

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Overview of Skin Disease

Skin diseases are a widespread health concern globally, affecting individuals of all age groups and leading to significant discomfort and healthcare challenges. These conditions involve abnormal changes in skin texture, appearance, or function due to inflammation, infections, or genetic factors. Some skin diseases may appear benign initially, but can resemble serious conditions, making accurate diagnosis essential. According to global statistics, skin diseases contribute 1.79% of the physical disabilities of humans in all countries of the world [1]. Skin problems cause suffering for about 30% to 70% of people in various countries of the world [2]

Recognizing the burden of dermatological diseases in resource-limited regions is essential for formulating an effective and long-term global strategy to address this health challenge. In many cases, skin disorders serve as visible indicators of more serious underlying systemic conditions, such as HIV and various neglected tropical diseases (NTDs) [3]. Environmental factors, UV radiation, and genetic mutations can contribute to their onset. Visual diagnosis can be difficult due to similar textures, colors, and shapes among different lesion types. This visual complexity increases the risk of misdiagnosis. Therefore, image analysis and classification techniques have gained traction in clinical support [4]. Deep learning (DL) methods, particularly those using convolutional neural networks (CNNs), offer automated, high-accuracy diagnosis through dermoscopic image analysis. Transfer learning (TL) techniques, leveraging pretrained models, help refine feature extraction and enhance classification performance on specific datasets [5].

1.2 Common Types of Skin Diseases

Among the commonly studied skin diseases are melanocytic nevi, melanoma, benign keratosis-like lesions, basal cell carcinoma, actinic keratoses, vascular lesions, and dermatofibroma. While some, like nevi and keratosis, are typically non-cancerous, others, such as melanoma or basal cell carcinoma, may require immediate clinical attention. In

fact, melanoma is considered the most common form of skin cancer, and it affects the skin surface cells known as melanocytes [6]. Seborrheic keratoses (SKs) are very common benign epidermal tumors composed of keratinocytes. They primarily emerge in adulthood, can be seen more frequently with age—especially from age 50, peaking around 60 years old [7]. Approximately 80% of non-melanoma skin cancers is basal cell carcinoma (BCC). It arises from basal keratinocytes and typically occurs on sun-exposed areas, especially the head and neck. Although BCC grows slowly and rarely metastasizes, it can cause significant local tissue destruction if left untreated [8]. In 75% cases, actinic keratosis arises on chronically sun-exposed areas. Actinic keratosis should be treated with care as it is considered the early stage of squamous cell carcinoma. For its treatment, sunscreen and protective clothing is widely used [9].

1.3 Machine Learning in Dermatology

Machine learning (ML) is revolutionizing the diagnosis and management of both cancerous and non-cancerous skin diseases by supporting faster, more accurate clinical decisions. Key applications of machine learning include:

Early Detection & Diagnosis:
ML models, especially CNNs, analyze dermoscopic images to accurately detect conditions like melanoma, eczema, psoriasis, and fungal infections. They help dermatologists by identifying ambiguous cases, tracking disease severity, and monitoring treatment responses over time.

Personalized Treatment:
By analyzing patient data and history, ML helps tailor treatment plans, predict flare-ups, and guide medication choices based on individual needs.

Education & Awareness:
ML-powered apps assist patients in symptom tracking and treatment adherence, while also supporting clinician training through annotated datasets and case simulations.

1.4 Motivation for the Work

World Health Organization (WHO) estimates that skin diseases have a significant health impact at the international level, both in population morbidity and mortality. Actinic keratosis, basal cell carcinoma, benign keratosis-like lesions, dermatofibroma, melanoma, melanocytic nevi, and vascular lesions are some of the extensive number of skin diseases which range from benign to highly malignant. Proper diagnosis on time is extremely essential, especially for dangerous forms such as melanoma, which may prove fatal if diagnosed late.

Getting a definitive diagnosis for a skin condition can take time and often varies from one doctor to another. To help solve this, the research introduces a smart, lightweight AI model that can quickly and accurately classify seven different types of skin lesions from images.

When dealing with something as aggressive as melanoma, catching it early is paramount. This model is designed to act as a reliable assistant for dermatologists, helping them spot dangerous signs with greater confidence. By making advanced diagnostics more accessible and practical for any clinic, we aim to reduce the workload on healthcare staff.

Ultimately, the goal is to provide a powerful tool that helps doctors make faster, more objective diagnoses, improving patient lives through early and consistent detection.

1.5 Objectives

The main objectives of ths thesis are:

1. To develop a pre-trained deep learning model capable of accurately classifying common skin diseases.
2. To design and implement advanced CNN architectures.
3. To evaluate model performance using key matrices such as accuracy, precission, recall.
4. To compare the efficiency of different CNN model to identify most effective architectures.

N.B.: CNN: Convolutional neual network

Chapter 2: Literature Review

2.1 Overview

Over the last few years, skin diseases diagnosis research has caught attention to the researchers worldwide. Numerous datasets that are publicly available, like ISIC, HAM10000 etc are being widely used by the researchers enabling them to explore various analytical approaches. State-of-The-Art techniques such as data partitioning, merging, clustering, and classification have been utilized to enhance the detection and treatment of skin cancer. While each of these methods presents its own set of challenges and limitations, they collectively contribute to improving clinical decision-making and supporting healthcare professionals in achieving more accurate and timely diagnoses.

Viswanatha Reddy [10] proposed a deep learning based CNN model to classify between three melanoma subtypes- lesion maligna, superficial spreading and nodular melanoma. They used dermoscopic images from the DermNet-NZ dataset. They initially used traditional classifiers like decision tree, random forests and gradient boosted tree, but achieved lower accuracy compared to their custom CNN modelwhich gained 88.83% accuracy.

Ahmed et al. [11] used pre-trained ResNet152 and InceptionResNet-V2 and fine-tunes them with a triplet loss function for classification. Their model was trained on dermoscopic images from a hospital in Wuhan. The method of using triplet loss function minimized L-2 distance of embedding space that represents same disease and maximized in case of different diseases.

Parvathaneni N. Srinivasu et al. [12] combined MobileNet V2 and LSTM and brought a hybrid deep learning framework for multi-class skin disease classification. They utilized the HAM10000 dataset of dermoscopic images. They used LSTM for augmentation to capture temporal dependencies, and a grey-level co-occurrence matrix to assess texture information. Their model achieved over 85% accuracy compared to conventional MobileNet that outperformed several benchmark architectures including VGG and standard CNN variants.

Wu et al. [13] conducted an experiment of five convolutional neural network (CNN) architectures—including Inception-ResNet-v2—on classifying six common facial skin diseases. They constructed a specialized dataset of 2,656 facial images using clinical images from a reputed hospital in China. The Inception-ResNet-v2 achieved the highest precision and recall among the tested models.

AlDera and Ben Othman [14] proposed a model in which they combined image processing and machine learning to diagnose acne, cherry angioma, melanoma, and psoriasis from skin images. Their model involves preprocessing, segmentation via Otsu's thresholding, and feature extraction using Gabor filters, entropy, and Sobel operators. They used state-of-the-art classifiers like SVM, Random Forest, and K-NN and achieved highest accuracy with SVM at 90.7%. Their study highlights the potential of hybrid feature-based classifier for effective multiclass skin disease diagnosis using a relatively small dataset.

In the study [15] led by Talha Mahboob Alam et al., An Efficient Deep Learning-Based Skin Cancer Classifier is conducted for an Imbalanced Dataset. The study explore various CNN models including AlexNet InceptionV3 RegNetY-320 for classifying skin cancer. by using high perameter RegnetY-320 they achieve high accuracy after data augmentation.

Dilshan Nandasena et al. [16] conducted a Mobile-Based Skin Disease Diagnosis System Using Convolutional Neural Networks (CNN). They have explored various models like ANN, SVM, and KNN, but CNNs consistently yielded better results in image-based skin classification tasks. MobileNet, a lightweight CNN architecture, has proven particularly effective for mobile applications due to its computational efficiency. The HAM10000 dataset is widely used for training and evaluating such systems.

Abdulmateen Adebiyi MS1 [17] introduce a Accurate Skin Lesion Classification Using Multimodal Learning on the HAM10000 and ISIC 2017 Datasets. Traditional models such as ResNet, DenseNet, and EfficientNet have demonstrated promising results, yet rely solely on visual data. Multimodal learning has emerged as a superior approach, integrating both images and clinical metadata. Studies, including this work, show that combining visual and textual data using models like ALBEF improves diagnostic performance over unimodal models. Prior work by Tschandl et al. and Tiulpin et al. further supports multimodal approaches in medical diagnostics. This evolution highlights the importance of context-aware AI in skin cancer detection.

Mohammad Fraiwan [18] introduced 'On the Automatic Detection and Classification of Skin Cancer Using Deep Transfer Learning' used 13 pretrained model like SqueezeNet GoogLeNet Inceptionv3 DarkNet-53 EfficientNet-b0 etc, found the best model accuracy using Densenet201 classify the seven-class classification using HAM10000 dataset.

Karar Ali [19] proposed an Multiclass skin cancer classification using EfficientNets – a first step towards preventing skin cancer. The study builds and experiments with the EfficientNet family of CNNs, identifying EfficientNet B4 as the best performing model. The top three layers of EfficientNet models were replaced to better suit seven-class skin cancer classification. Additional dense, batch normalization, and dropout layers were added to reduce overfitting.

Syed Inthiyaz [20] introduce Skin disease detection using deep learning,It addresses the challenges in dermatological diagnosis by applying pre-trained deep learning models such as VGG16, VGG19, and ResNet-50, combined with image preprocessing and classification via softmax and SVM, and classify Eczema Melanoma Psoriasis Healthy skin (non-diseased).The model relies on a deep neural network called ResNet-50, which has 50 layers and is especially good at avoiding common issues like vanishing gradients, the researchers used a version of ResNet-50 that was already trained on a large image dataset (ImageNet), and then fine-tuned it to recognize skin diseases more accurately.

'A skin cancer classification system for pigmented skin lesions' was developed by Jinnai et al. [21]. In this study, a total of 5846 clinical images from 3551 patients were used, covering both malignant tumors (malignant melanoma and basal cell carcinoma) and benign conditions (nevus, seborrheic keratosis, senile lentigo, and hematoma/hemangioma). To evaluate the model, a test dataset was constructed by randomly selecting 666 patients. The team trained a Faster Region-based Convolutional Neural Network (FRCNN) using VGG-16 as the backbone architecture. The model's classification performance was then compared to that of 10 board-certified dermatologists and 10 trainees. The FRCNN achieved an accuracy of 86.2% in six-class classification and 91.5% in binary (benign vs. malignant) classification, surpassing the performance of the human examiners. This work demonstrated the potential of AI to support early and accurate diagnosis of skin cancer using standard clinical photographs.

In their work [22], led by Mohamed A. Kassem et al., introduced a model aimed at achieving highly accurate classification of skin lesions. The proposed approach utilizes transfer learning combined with a pre-trained GoogleNet model. Initially, the model adopts parameters from the pre-trained network, which are subsequently fine-tuned throughout training to improve classification performance. Their system was evaluated on the ISIC 2019 dataset and achieved significant performance across eight classes of skin lesions.

In their study [23], led by Amirreza Mahbod et al., introduced a novel approach to skin lesion classification using hybrid deep neural networks. Their proposed method presents a fully automated computerized system for classifying skin lesions, leveraging optimized deep features extracted from multiple well-established convolutional neural networks (CNNs) at different abstraction levels. Specifically, three pre-trained deep models—AlexNet, VGG16, and ResNet-18—are employed as deep feature generators. The extracted features are then used to train support vector machine (SVM) classifiers.

In this study Abdulmateen Adebiyi et al. [24] , the authors advocate Comparison of Three Deep Learning Models in Accurate Classification of 770 Dermoscopy Skin Lesion Images, This study proposes a novel deep learning-based approach for classifying benign and malignant skin lesions using 770 real-world dermoscopy images sourced from the University of Missouri Healthcare. Unlike cleaner public datasets like HAM10000. Three CNN architectures—ResNet50, DenseNet121, and Inception-V3—were trained on three image sets, including versions with and without hair artifact removal. DenseNet121 performed best.

Sumit Kumar Singh et al [25] introduce Fuzzy Logic with Deep Learning for Detection of Skin Cancer presents an innovative hybrid approach that combines fuzzy logic-based segmentation with a modified YOLO deep learning model to detect melanoma from dermoscopic images. Using publicly available datasets (ISIC 2017, ISIC 2018, and PH2), the method achieves high accuracy, sensitivity, and specificity. The two-phase segmentation technique improves lesion boundary detection, while the enhanced YOLO architecture ensures real-time and precise classification. Overall, the study offers a robust and efficient framework for early skin cancer diagnosis.

Akter et al. [26] proposed a deep learning-based approach for classifying seven types of skin lesions using the HAM10000 dataset. They evaluated a custom CNN and six pretrained models—InceptionV3, Xception, DenseNet, MobileNet, ResNet-50, and VGG-16. InceptionV3 achieved the highest accuracy (90%), while ensemble stacking models performed worse, with a maximum of 78% accuracy.

Karambele et al. [27] proposed a ResNet50-based deep learning approach for classifying skin diseases using a dataset of 2,826 dermoscopic images collected from Dermnet and other sources. The study focused on five categories of skin disorders, including dermatitis, herpes, lupus, melanoma, and vascular tumors. The ResNet50 model was fine-tuned using transfer learning with ImageNet weights, and early stopping was employed to prevent overfitting. The model achieved a classification accuracy of 76.07%, with high specificity (94.03%) and balanced precision, recall, and F1-score.

Faris et al. [28] proposed a hybrid deep learning approach for classifying three types of skin conditions—eczema, scabies, and healthy skin—using a dataset of over 5,000 images sourced from DermNet and Kaggle. The authors integrated MobileNet and EfficientNetB0 combining feature maps into a unified vector of 2,304 dimensions. It achieved a training accuracy of 98%, validation accuracy of 91.93%, and test accuracy of 91.5%.

2.2 Summary

Recent studies in skin disease detection highlight the growing role of deep learning and hybrid approaches in medical image analysis. Public datasets such as ISIC and HAM10000 are widely used to train and evaluate models. Researchers have explored traditional classifiers, CNNs, transfer learning, multimodal learning, and hybrid frameworks combining CNNs with LSTM, fuzzy logic, and YOLO for improved performance. Pre-trained architectures like ResNet, DenseNet, EfficientNet, and Inception models consistently achieve high accuracy, precision, and recall. Studies demonstrate that CNN-based and lightweight models outperform conventional machine learning, supporting accurate and efficient skin disease diagnosis across multiple classes and conditions.

Chapter 3: CNN in Dermatological Imaging

3.1 Introduction to Convolutional Neural Networks (CNNs)

A Convolutional Neural Network, or CNN, is a type of AI inspired by the human brain that's excellent at analyzing images. It learns by studying vast numbers of pictures to recognize patterns. In dermatology, a CNN can look at an image of a skin lesion and make a diagnostic suggestion, much like how a dermatologist uses their experience to assess a condition. Because dermatology is so visual, this technology is a perfect fit.

This technological advancement is set to transform the field. A CNN can analyze digital images of skin lesions, providing a data-driven "second opinion" that complements a dermatologist's clinical judgment. The goal is to unite human expertise with AI's precision, leading to faster, more reliable diagnoses and ultimately, better outcomes for patients. By detecting subtle patterns that might be missed by the human eye, these systems can be particularly crucial in the early detection of skin cancers like melanoma. This synergy between human and artificial intelligence paves the way for a new era of diagnostic accuracy and personalized patient care. Essentially, CNNs act as a powerful tool to assist dermatologists, blending advanced technology with medical expertise to improve diagnostic accuracy and patient care.

3.2 Core Structures of CNN

CNNs are characterized by their ability to automatically learn hierarchical representations of data directly from raw input, making them exceptionally well-suited for tasks involving structured grid-like data such as images. The key principles that underlie CNNs include:

3.2.1 Convolution Layer

The convolutional layer is the core building block of a Convolutional Neural Network (CNN), where most computations take place. It operates using three main components: input data, a filter (or kernel), and the resulting feature map as shown in Fig. 3.1. For a color image input, the data is represented as a 3D matrix of height, width, and depth (RGB

channels). A feature detector—also known as a kernel—is used in this research to scan the input image and identify relevant features, a process referred to as convolution.

The filter is a 2D weight matrix that slides over the input, performing dot products between the filter and corresponding pixel values. A common filter size is 3x3. The result of each operation is stored in an output matrix. The filter shifts by a defined stride value until the entire image is covered, and the accumulated outputs form the feature map, also called the activation map or convolved feature.

A key principle in CNNs is parameter sharing, where filter weights remain fixed during each pass but are updated during training through backpropagation and gradient descent. Three hyperparameters crucial to the output volume are:

Number of Filters: Affects output depth; e.g., three filters generate three feature maps.
Stride: Defines how many pixels the filter moves; higher stride yields smaller outputs.
Zero-Padding: Maintains spatial dimensions when filters don't fit exactly. Types include:
Valid Padding: No padding; edges may be dropped.
Same Padding: Output size equals input size.
Full Padding: Adds zeros to enlarge the output.

After convolution, a Rectified Linear Unit (ReLU) function is applied to introduce nonlinearity.

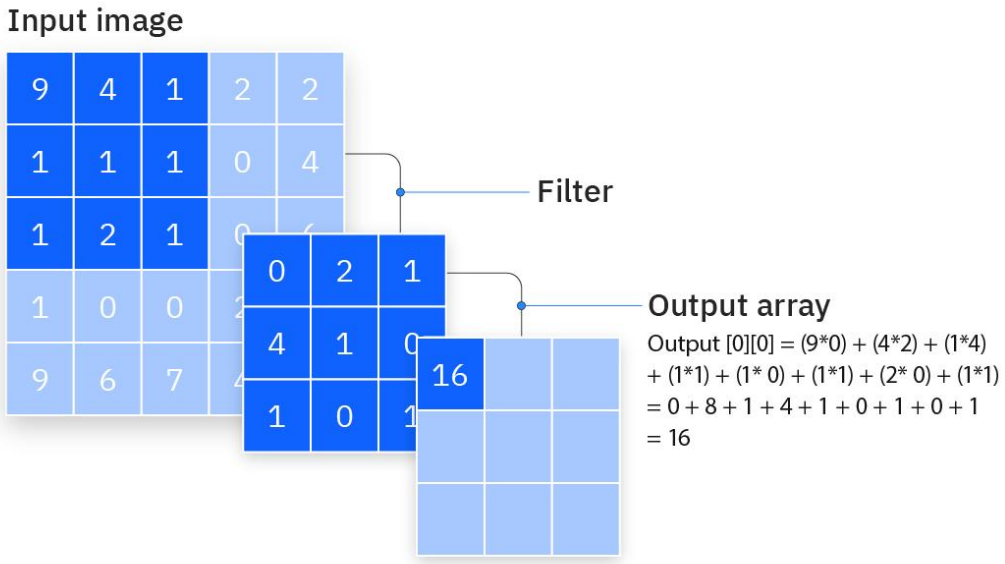


Fig. 3.1: Convolution process [30]

3.2.2 Activation Functions

Activation functions are essential in neural networks, determining whether a neuron should activate based on its input. These functions apply mathematical operations to evaluate the significance of inputs during prediction tasks. When an input is important, the neuron is activated, contributing to the network's learning process.

A node (or neuron) in an artificial neural network functions similarly to a biological neuron, receiving signals and deciding whether to activate based on their strength. In deep learning, the activation function transforms the weighted sum of inputs at a node into an output passed to the next layer or used as the final output.

Various activation functions have been used in this research, as shown in Fig. 3.2, with a primary focus on the Rectified Linear Unit (ReLU) due to its efficiency. ReLU sets all negative inputs to zero, deactivating those neurons and reducing computational load. This selective behavior enhances efficiency and enables faster convergence—typically six times faster than functions like tanh or sigmoid—without saturating in the positive range.

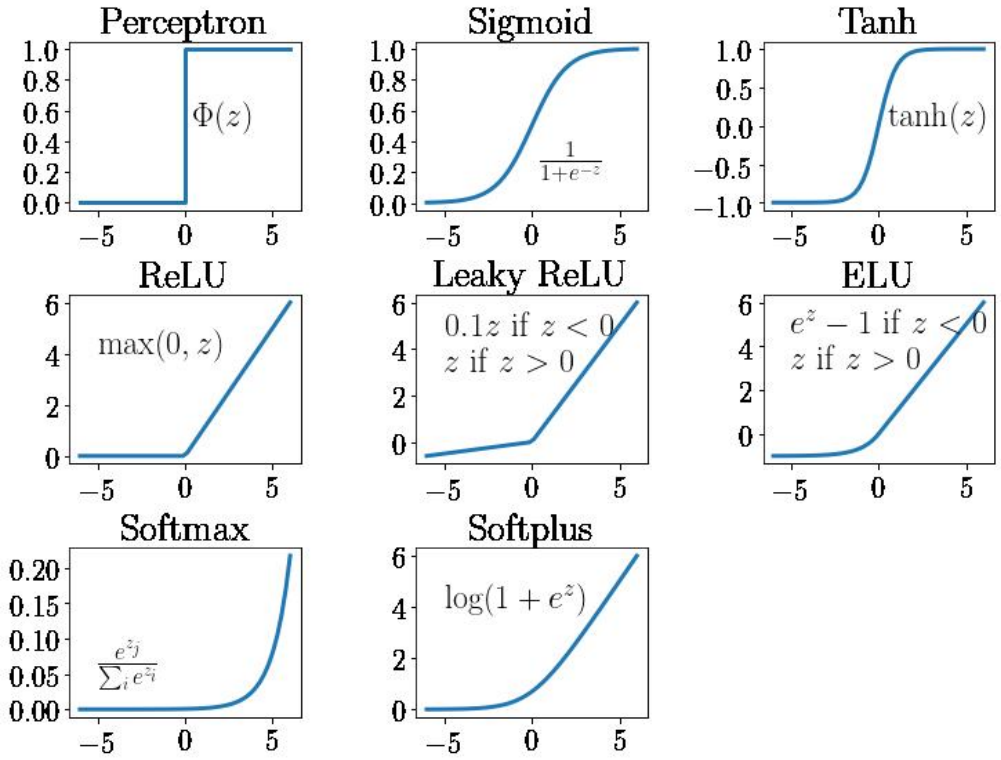


Fig. 3.2: Popular activation functions [31]

Despite these advantages, ReLU exhibits certain limitations. Specifically, it saturates in the negative region, where its gradient becomes zero. When the gradient is zero during backpropagation, weight updates do not occur, potentially hindering learning—a problem known as the "dying ReLU" phenomenon. To mitigate this issue, variants such as Leaky ReLU are introduced, which allow a small, non-zero gradient for negative inputs. Additionally, ReLU is not zero-centered, meaning that during optimization, it may follow a less direct, zigzagging path towards the optimum, possibly increasing the training time.

3.2.3 Pooling Layers

Pooling layers, also referred to as downsampling layers, are primarily responsible for dimensionality reduction, effectively decreasing the number of parameters within the input feature maps. Similar in structure to the convolutional layer, the pooling operation employs a filter that traverses the entire input. However, unlike the convolutional layer, the pooling filter does not contain learnable weights. Instead, it applies an aggregation function to the values located within its receptive field, subsequently generating the output array.

There are two principal types of pooling operations commonly utilized in CNN architectures:

Max Pooling: As illustrated in Fig. 3.3, the max pooling filter moves across the input feature map and selects the maximum pixel value within each receptive field to populate the output array. Max pooling is generally preferred over other pooling methods due to its effectiveness in preserving the most prominent features from the input.

Average Pooling: In this method, the pooling filter computes the average of all pixel values within the receptive field and assigns this average to the corresponding location in the output array as it moves across the input.

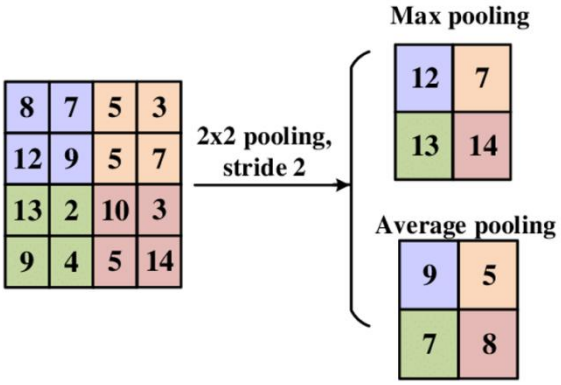


Fig. 3.3: Max pooling and Average poooling layers [32]

Pooling layers significantly reduce the model's complexity by lowering the dimensionality of feature maps, which leads to improved computational efficiency. Moreover, by simplifying the feature representations, pooling layers help to minimize the risk of overfitting, thereby enhancing the model's ability to generalize to unseen data.

3.2.4 Fully Connected Layers

Fully-connected layers, also referred to as linear layers or dense layers, establish connections between every neuron in the input layer and every neuron in the output layer. These layers are integral components in neural network architectures. As depicted in Fig. 3.4, a typical fully-connected layer example consists of four input neurons connected to eight output neurons. Fully-connected layers play a critical role in Convolutional Neural Networks (CNNs), which have demonstrated remarkable success in various computer vision tasks, including image recognition and classification.

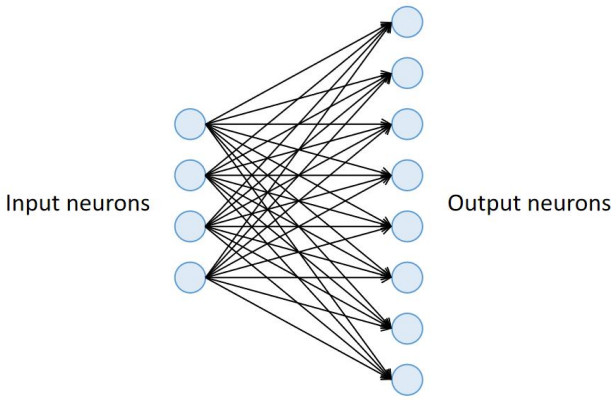


Fig. 3.4: Small fully connected layer with 4 input and 8 output neurons

In the CNN pipeline, the process initiates with convolution and pooling layers, which progressively decompose the input image into abstract feature representations and analyze these features independently. The output generated by these layers is then fed into one or more fully-connected layers. These layers synthesize the extracted features and perform high-level reasoning to drive the final classification or prediction outcome.

3.3 Popular CNN Architectures

CNN architectures are characterized by their depth, width, and connectivity patterns, which significantly influence their capacity to learn intricate patterns from data. Several seminal CNN architectures have been proposed, each offering unique insights into effective feature extraction and representation learning. Some prominent CNN architectures include:

AlexNet, introduced by Alex Krizhevsky in 2012, was a breakthrough in deep learning and significantly advanced the field of image classification. It won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) by a large margin, reducing the top-5 error rate from ~26% to 15%. As shown in Fig 3.5, AlexNet consists of eight layers (five convolutional and three fully connected), and was the first to successfully train a deep CNN using GPUs, enabling faster computations. It introduced key innovations like the ReLU activation function for faster convergence, dropout for regularization, and data augmentation for improved generalization. AlexNet's success marked the beginning of deep learning dominance in computer vision.

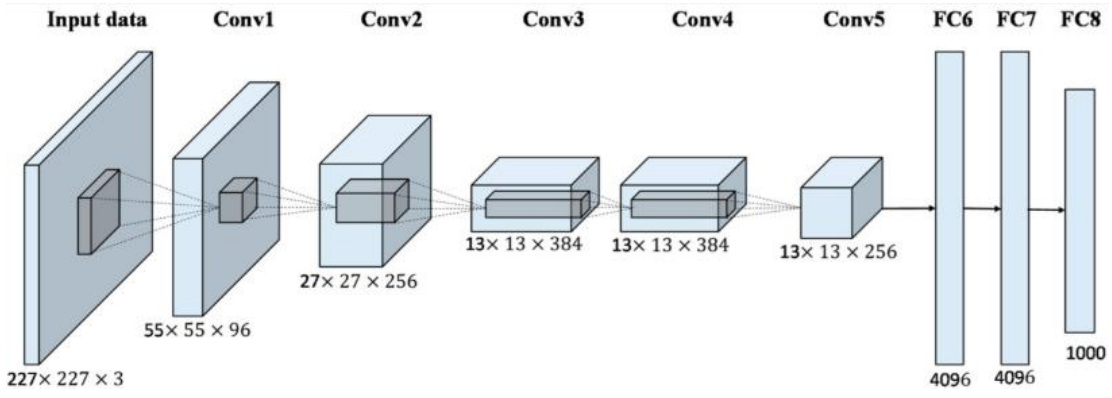


Fig. 3.5: The AlexNet architecture [33]

VGGNet, developed by the University of Oxford in 2014, is a deep CNN known for its simple and uniform design using stacked 3×3 convolutional filters. Popular versions like VGG-16 and VGG-19 demonstrated that increasing network depth improves accuracy, achieving strong results in the ImageNet Challenge. While widely used as a baseline in computer vision, its main drawback is high computational cost due to a large number of parameters.

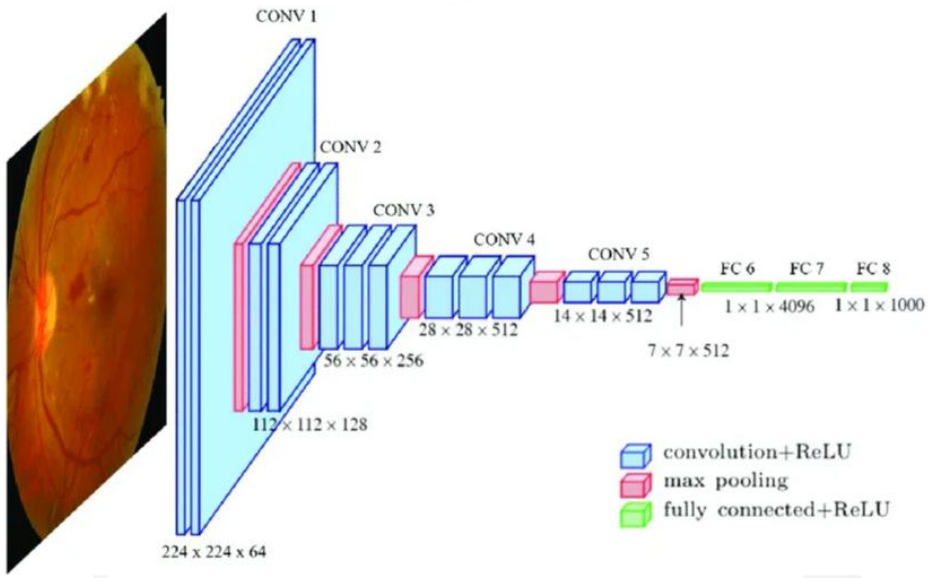


Fig. 3.6: The VGGNet architecture [34]

EfficientNet, introduced by Google AI in 2019, is a family of convolutional neural networks that achieves state-of-the-art accuracy with significantly fewer parameters and computational cost compared to previous models. Its core innovation is the compound scaling method, which uniformly scales the network's depth, width, and resolution using a fixed set of scaling coefficients. As Shown in Fig 3.7, the core building block of EfficientNet is the MBConv block, inspired by MobileNetV2. Starting with a baseline model (EfficientNet-B0) found through neural architecture search, the family includes variants up to EfficientNet-B7, each offering a trade-off between accuracy and efficiency. EfficientNet models are widely adopted in both academic research and industry due to their excellent performance and scalability, especially for tasks where computational efficiency is critical.

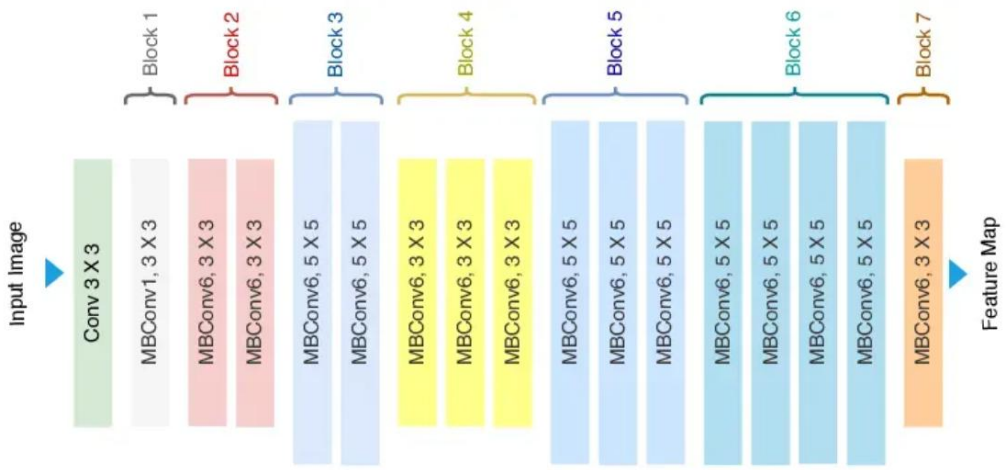


Fig. 3.7: The EfficientNet architecture [35]

ResNet (Residual Network), introduced by Microsoft Research in 2015, revolutionized deep learning by addressing the problem of vanishing gradients in very deep neural networks. Its key innovation is the residual connection or skip connection, which allows gradients to flow directly through layers by adding the input of a layer to its output, as shown in Fig 3.8. This simple yet powerful idea enables the training of extremely deep networks, such as ResNet-50, ResNet-101, and ResNet-152, with significantly improved performance. ResNet set new benchmarks in image classification (e.g., winning the ImageNet 2015 competition) and has since become a foundational architecture in computer vision tasks like object detection, segmentation, and beyond.

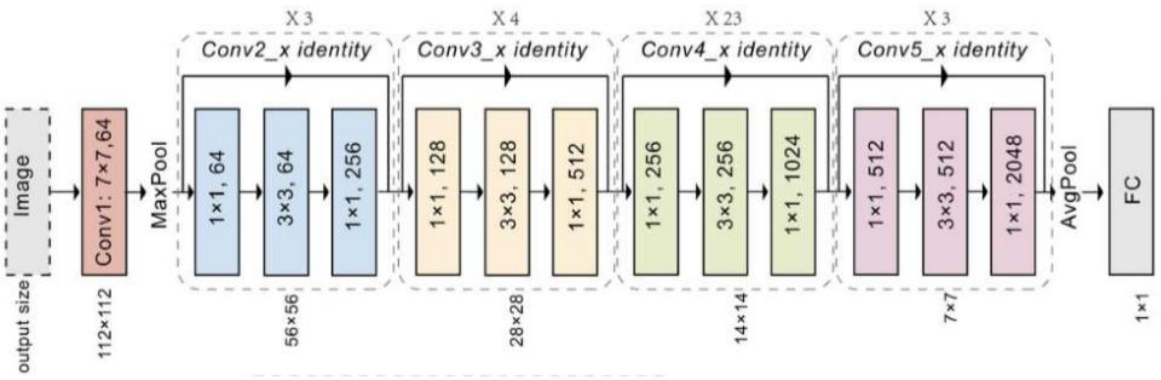


Fig. 3.8: The ResNet architecture [36]

ConvNeXt is a modern convolutional neural network architecture introduced by Facebook AI Research (FAIR) that reimagines classical ConvNet design in the era of Transformers. Inspired by the success of Vision Transformers (ViTs), ConvNeXt incorporates architectural components such as large kernel sizes, inverted bottlenecks, and LayerNorm, all while retaining the efficiency and inductive biases of CNNs, shown in Fig 3.9. It is built on top of the ResNet framework and it replaces traditional convolutions with depthwise separable convolutions and adapts the training protocols commonly used in transformer models. ConvNeXt achieves performance on par with or surpassing state-of-the-art ViTs on major benchmarks like ImageNet, demonstrating that carefully modernized convolutional networks remain highly competitive in visual recognition tasks.

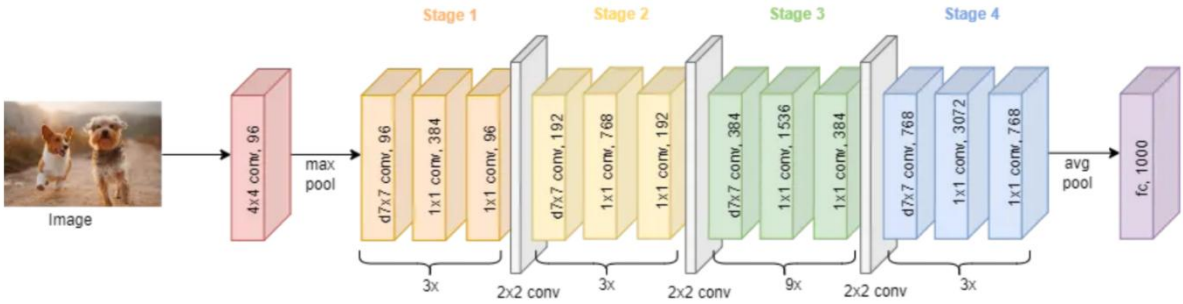


Fig. 3.9: The ConvNeXt architecture [37]

3.4 Uses of CNN in Medical Imaging

The application of Convolutional Neural Networks (CNNs) in medical image analysis has received considerable attention due to their capacity to automate and enhance diagnostic procedures. In the field of dermatology, CNNs serve a pivotal function in the classification of skin lesions, assisting clinicians with the early detection and accurate diagnosis of various skin disorders, including melanoma and other types of skin cancer.

CNNs utilize extensive datasets of labeled dermatological images to learn distinctive features that correspond to different skin conditions. By examining the textural, structural, and contextual attributes present in medical images, CNNs can accurately distinguish between benign and malignant lesions, thereby providing essential support to dermatologists during clinical evaluations and diagnostic decisions.

However, despite these promising capabilities, the integration of CNNs into dermatological image analysis faces certain challenges. These include the requirement for large, well-annotated datasets, the need for improved model interpretability, and ensuring robust generalization across diverse patient demographics and varying imaging environments.

3.5 Relevance of CNN in Dermatology

Convolutional Neural Networks (CNNs) are revolutionizing dermatology by using their advanced image analysis capabilities to interpret skin lesion images. By mimicking the human visual system, these AI models learn from vast amounts of data to identify subtle, complex patterns that may be invisible to the human eye, significantly aiding in the early and accurate detection of skin cancers like melanoma. This process provides an objective, data-driven "second opinion," which helps to reduce human error and enhance diagnostic consistency.

This technology offers crucial support for clinical decisions and can help overcome healthcare disparities by providing scalable diagnostic tools in underserved areas. The ultimate goal is not to replace the clinician but to augment their expertise, creating a powerful synergy between the dermatologist and AI. As the technology evolves, ensuring model fairness, interpretability, and data privacy remains critical for its widespread adoption and success in improving patient outcomes.

3.6 Image Preprocessing in Dermatology

Preprocessing is a fundamental step in improving the effectiveness of Convolutional Neural Networks (CNNs) for dermatological image analysis. Standard preprocessing methods include image normalization, resizing, noise reduction, and data augmentation. These techniques are essential for ensuring consistency across input images, minimizing computational complexity, and increasing the model’s resilience to variations in lighting, orientation, and background noise present in clinical images.

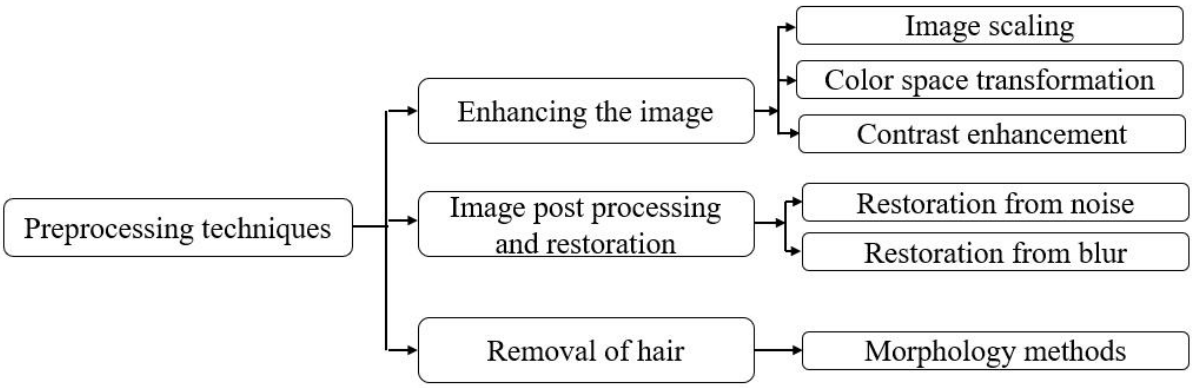


Fig. 3.10: Image preprocessing technique

In particular, data augmentation serves a vital role by artificially expanding the training dataset. Through the application of transformations such as rotation, scaling, flipping, and shifting, augmentation techniques generate diverse variations of existing images. This process enables CNNs to achieve better generalization, enhancing their capacity to accurately detect and classify a wide range of skin lesions, even in cases involving previously unseen or atypical image patterns.

3.7 Evaluation Metrics for CNN performance Analysis

Evaluation metrics are critical for determining the effectiveness of Convolutional Neural Network (CNN) architectures in the classification of skin lesions. The most commonly employed metrics include accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC-ROC). Each of these metrics offers valuable insights into various aspects of the model's classification performance, including its ability to correctly identify positive and negative cases, as well as its overall diagnostic reliability.

These evaluation criteria help assess the model's sensitivity to false positives and false negatives, which is particularly important in medical imaging where diagnostic errors can have serious consequences. The selection of appropriate metrics depends on the specific clinical objectives of the dermatological task, such as prioritizing sensitivity to detect malignant lesions or maintaining high specificity to reduce unnecessary biopsies. Balancing these metrics ensures that the CNN performs optimally in real-world diagnostic scenarios.

Chapter 4: Methodology

4.1 Overview

This study proposes a lightweight robust transfer learning based approach for multiclass skin disease classification using medical grade dermoscopic images. The proposed methodology incorporates the use of ConvNeXt-Tiny architecture pretrained on ImageNet and also finetuned on the HAM10,000 dataset. The pipeline includes data preprocessing, augmentation, balanced sampling, custom loss Formulation using Focal Loss with label smoothing and enhanced the evaluation of the model by using Test-Time Augmentation (TTA) combined with Monte Carlo Dropout. The whole dataset was divided into train = 80%, validation = 10% and test = 10% using stratified splitting mechanism.

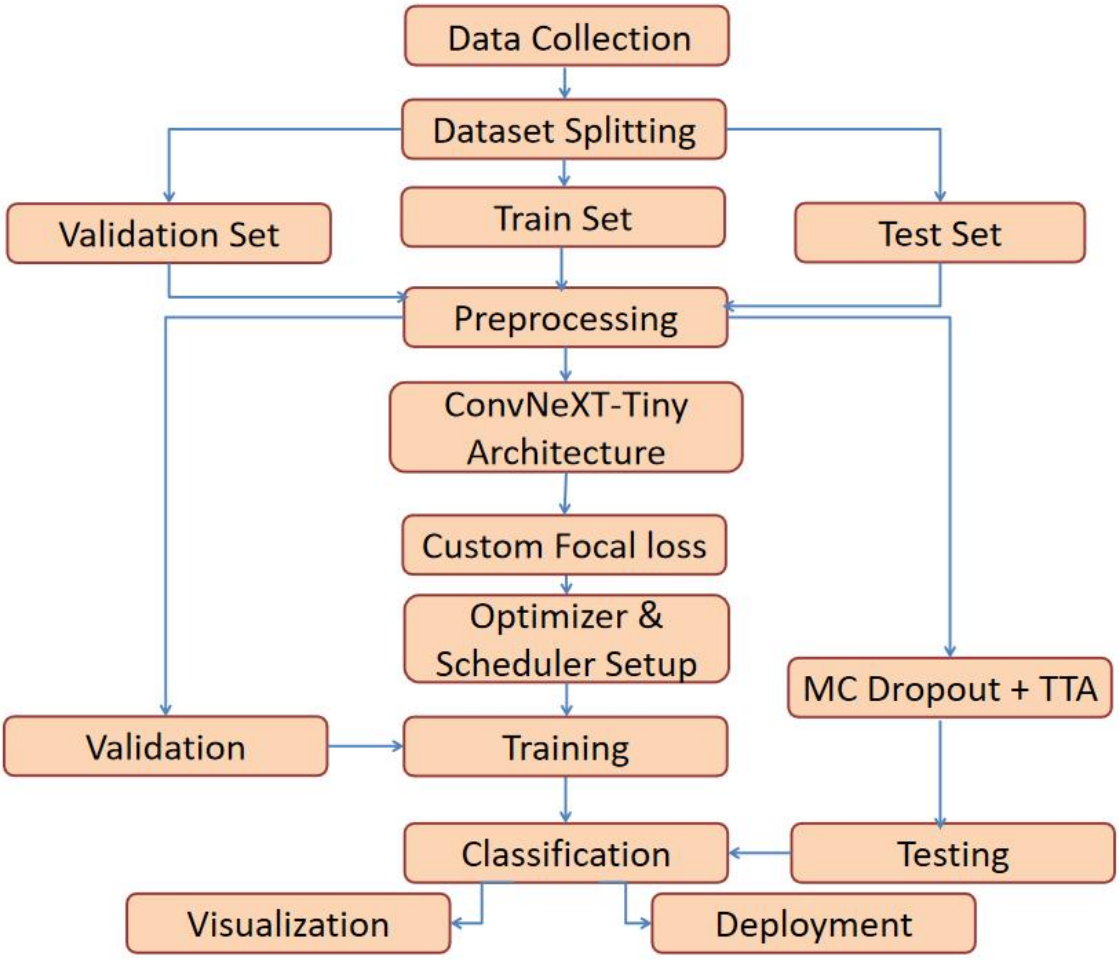


Fig. 4.1: Proposed method

Step 1: Data Collection: The HAM10,000 dataset was used for input. It contains two images files and a metadata CSV file that contains image ID, diagnosis (label) etc. It was used to merge class names with images files.

Step 2: Data Splitting: The stratified splitting method is utilized to split according to unique lesion id the dataset into train = 80%, validation = 10% and test = 10% respectively. This ensures that the distribution of each class is preserved across training, validation and test sets.

Step 3: Preprocessing: Different types of transformations are used for training, validation and test sets. Training data gets balanced sampling, hair removal on Melanoma and Actinic Keratoses ,class wise augmentation such as (random crops, flips, rotation etc) and also resize and normalization whereas the validation and test datasets are resized and normalized only..

Step 4: ConvNeXt-Tiny: The ConvNeXt-Tiny model was backbone to classify seven diseases. The model was fully fine tuned on Ham10,000 for better detection .

Step 5: Custom focal loss Optimizers and scheduler : Focal loss used for rare classes and learning scheduler and adma optimizer is used for better performance.

Step 6: Training: The augmented and balanced data of train sets are fed into the model. While training, the model learns to distinguish between skin disease classes from image patterns.

Step 7: Validation: After each epoch, data from the validation set was used to perform validation to monitor generalization performance, accuracy, loss, macro-F1 and class wise recall.

Step 8: Classification: After performing inference on input image, the raw output from the models which are described as logits were fed through an argmax operation to select the class with highest predicted probability, which converts the numerical outputs into actual class labels corresponding to the seven classes of the dataset.

Step 9: MC Dropout + TTA: Monte Carlo (MC) Dropout and Test Time Augmentation was applied to improve prediction and stability and model confidence. MC dropout keeps the dropout layers active even while evaluating the model to introduce randomness. Also TTA applies variation (flip,rotation) to each image during inference to stimulate multiple perspectives.

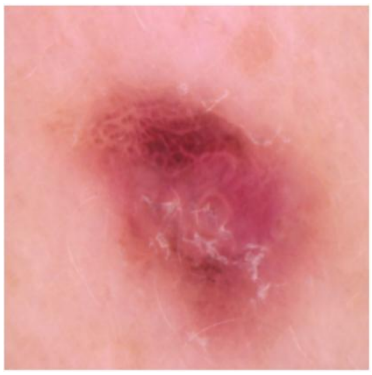
Step 10: Testing: The model is evaluated on test dataset using metrics like accuracy, macro F1, per-class recall.

Step 11: Visualization: The models performance was visualized by plotting different curves such as train-loss vs validation-loss, train-accuracy vs validation-accuracy, ROC curve, confusion matrix. The Grad-Cam visualization was also implemented where ten test images were randomly selected to visualize where the model focuses during prediction. For each image, true label, predicted label was observed and the Grad-CAM heatmap was overlaid to understand the model's decision-making and learning behavior.

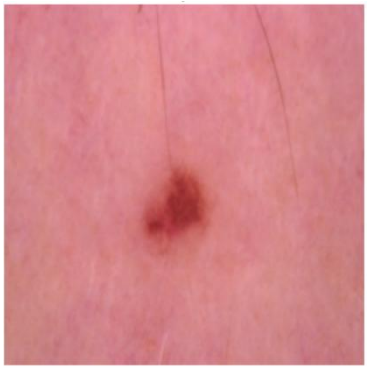
Step 12: Deployment: Model is saved and made a demo deployment using Gradio and model was tested on the dataset, ISIC 2019, achieved from Kaggle.

4.2 Dataset Used

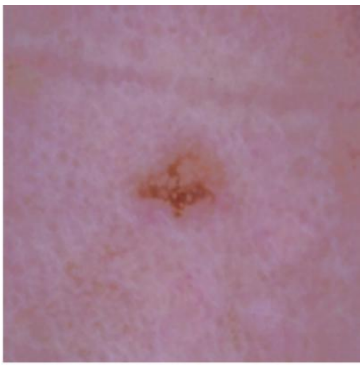
In this research, the HAM10,000 dataset [29] was used containing of 10,015 dermoscopic RGB image which covers seven distinct classes of skin diseases. The dataset was achived from Kaggle respiratory (ISIC archive). To ensure balanced evaluation and training, the data set was split into train 80%, validation 10% and test 10%. This stratified split helps maintain class distribution across all subsets and supports robust model training and evaluation. The ISIC-2019 was used for gradio interface testing.



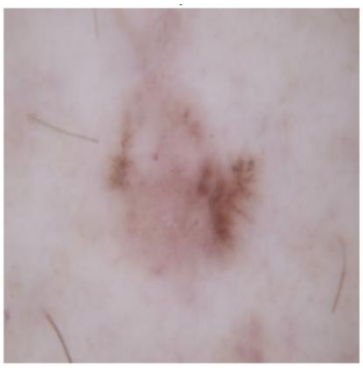
Melanoma



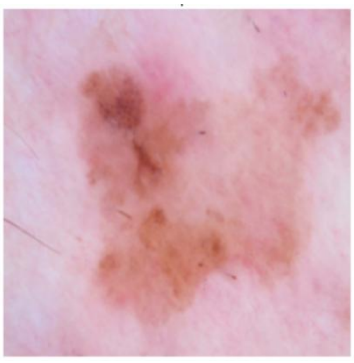
Melanocytic nevi



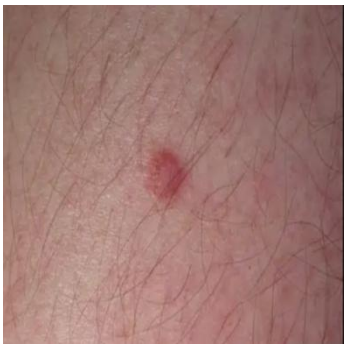
Benign keratosis



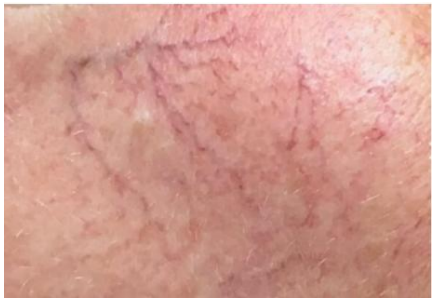
Basal cell carcinoma



Actinic keratosis



Dermatofibroma



Vascular lesions

Fig. 4.2: Samples of different skin lesions in the HAM10000 dataset [29]

4.3 Image Preprocessing and Augmentation

All dermoscopic images in the dataset were standardized by resizing them to 224×224 pixels to ensure a uniform input format compatible with the ConvNeXt-Tiny architecture. To address class imbalance and improve generalization, a balanced augmentation strategy was employed, wherein approximately 5,000 augmented samples were generated per class using a combination of random horizontal and vertical flipping, rotations of up to 15 degrees, and color jittering to simulate variation in brightness, contrast, and saturation. In addition to this general augmentation, class-specific preprocessing was applied by performing hair artifact removal exclusively on melanoma and actinic keratoses images, where occlusions from hair commonly hinder lesion visibility and can affect diagnostic clarity. For validation and test sets, only resizing and ImageNet-style normalization were applied to maintain evaluation consistency without introducing augmentation-induced variance. Furthermore, test-time augmentation (TTA) was implemented during inference by generating predictions across multiple transformed versions of each input image, and averaging the results to enhance robustness, particularly for visually ambiguous cases.

4.4 Established Methods

As part of the transfer learning approach, ResNet101, EfficientNetB0, EfficientNetB3, and DenseNet161 were employed as comparative baseline models, selected based on their prominence in prior literature. These architectures were evaluated on the same dataset to assess their effectiveness in multi-class skin lesion classification. For consistency, Focal

Loss was adopted as the loss function, along with the Adam optimizer and Softmax classifier to handle the seven output classes. The performance metrics obtained from these models are summarized in Tables 5.1, 5.2, 5.3 and 5.4

4.5 Proposed Model

The proposed methodology employs a deep learning model based on the ConvNeXt-Tiny architecture, pre-trained on the ImageNet dataset, to perform multi-class classification of skin lesions from the HAM10000 dataset. A full fine-tuning strategy was adopted, where all layers of the network were unfrozen to allow for comprehensive feature adaptation to the target domain. The model's final classifier head was replaced with a linear layer tailored to the seven specific classes of skin lesions. To enhance the model's robustness and its ability to estimate uncertainty, Monte Carlo (MC) Dropout with a dropout probability of p=0.3 was strategically injected into the model's feature blocks.

The training process was meticulously designed to address the significant class imbalance inherent in the dataset. A key innovation is the use of class-specific data augmentation, where under-represented and clinically critical classes, such as Melanoma and Actinic Keratoses, were subjected to more aggressive transformations (e.g., random vertical flips, color jitter), while majority classes received milder augmentations. Furthermore, a domain-specific pre-processing step involving an OpenCV-based hair removal algorithm was selectively applied to these same classes, which are often obscured by hair artifacts.

To guide the optimization, a sophisticated Focal Loss function was implemented with a gamma parameter of $\gamma=3$ and label smoothing of 0.05. This loss function was further enhanced with class weights calculated to balance the dataset, with additional manual up-weighting applied to the Melanoma and Actinic Keratoses classes to prioritize their correct identification. During inference, the integrated MC Dropout was leverage by performing multiple forward passes on each test image with the dropout layers activated. The final prediction is determined by averaging the softmax probabilities from these passes, yielding a more stable and reliable classification output.

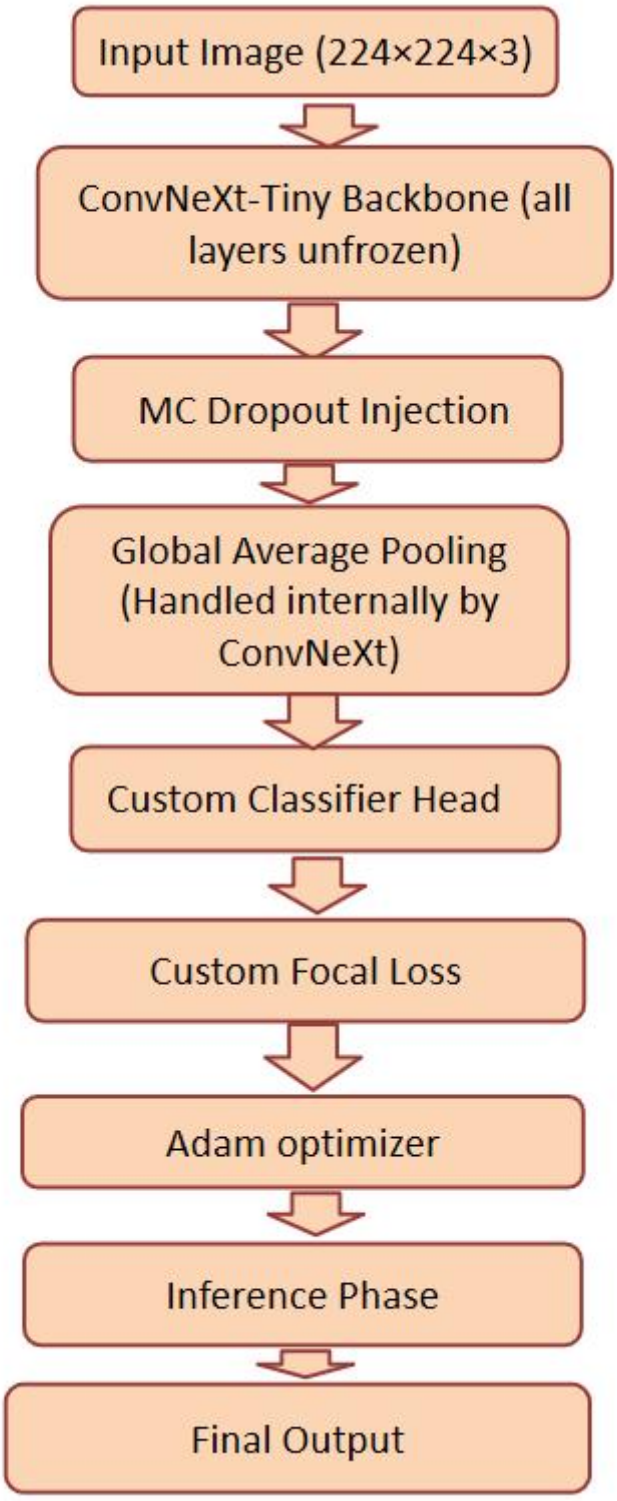


Fig. 4.3: Proposed model architecture

Input Image

Input skin lesion images are systematically prepared for the model by resizing them to a uniform resolution of 224x224 pixels with three color channels (RGB).

Backbone: ConvNeXt-Tiny

A pretrained ConvNeXt-Tiny architecture, originally trained on the ImageNet dataset, serves as the backbone of the model. This transfer learning approach leverages rich, general-purpose features to enhance lesion classification. All layers are unfrozen from the beginning, allowing full fine-tuning during training.

MC Dropout Injection

Dropout layers (p=0.3) are inserted into the final feature block and kept active during inference. This Monte Carlo (MC) Dropout method is used to model prediction uncertainty.

Global Average Pooling

The model performs global average pooling after feature extraction. This operation condenses the feature maps into a 1D tensor suitable for the classifier.

Classifier Head

The original classifier is replaced with a single, minimal linear layer. This head maps features directly to the seven output classes without any hidden layers or activations.

Loss Function: Custom Focal Loss

A custom focal loss with gamma=3 and label smoothing of 0.05 is implemented to handle class imbalance. Additionally, weights are manually boosted for classes 0 and 5.

Optimizer & Scheduling

The Adam optimizer trains the model, with a ReduceLROnPlateau scheduler adjusting the learning rate. Early stopping is used to halt training when performance stagnates.

Inference Phase (TTA + MC Dropout)

During testing, both MC Dropout and Test-Time Augmentation (TTA) are used. This process generates multiple diverse predictions for each image to improve final accuracy.

Final Output

The softmax probabilities from all TTA passes are averaged to get a final, robust prediction. The class is chosen via argmax over these averaged probabilities.

4.6 Summary of Proposed Method

This work presents a lightweight, robust pipeline for seven-class skin-lesion classification on HAM10000. Images and metadata are stratified-split 80/10/10 by lesion ID to preserve class balance across train/validation/test. Training images are resized to 224×224 and normalized; a balanced augmentation scheme expands each class (~5k samples) using flips, small rotations, and color jitter. To address class-specific occlusions, hair removal is applied only to melanoma and actinic keratoses. Validation/test sets use only resizing and normalization. At inference, test-time augmentation (TTA) averages predictions over transformed views to improve robustness.

The core model is ConvNeXt-Tiny pre-trained on ImageNet and fully fine-tuned for the target domain; the original head is replaced by a single linear layer producing seven logits. Monte Carlo (MC) Dropout (p=0.3) is injected into the feature block and kept active at test time to quantify uncertainty. Final predictions are obtained by averaging softmax probabilities from multiple MC-Dropout/TTA passes.

Training is optimized with Adam, a ReduceLROnPlateau scheduler, and early stopping. Class imbalance is handled with a custom focal loss (γ=3) plus 0.05 label smoothing, combined with class weights and additional up-weighting of melanoma and actinic keratoses to prioritize their detection. Performance is monitored on the validation set (accuracy, macro-F1, class-wise recall), and the best checkpoint is selected for testing.

For comparison, established transfer-learning baselines (ResNet101, EfficientNetB0/B3, DenseNet161) are trained under the same dataset and loss settings.

The methodology also includes interpretability and deployment steps: Grad-CAM is used to visualize model attention on dermoscopic images, and a lightweight Gradio interface is built for interactive testing on unseen images.

Chapter 5: Results and Discussion

This section presents and analyzes the outcomes obtained from various techniques applied to the task of skin disease classification. For clarity and coherence, the discussion is structured into five distinct subsections. Subsection 5.1 highlights the performance metrics achieved by the implemented pre-trained CNN models and the fine-tunes ConvNeXT-Tiny model. Subsection 5.2 illustrates the training process through visualizations of the models' loss and accuracy curves across epochs, providing insight into their learning behavior. Subsection 5.3 presents a comprehensive evaluation of the models using standard performance metrics such as confusion matrics and AUC-ROC curve, offering a detailed assessment of classification effectiveness. Subsection 5.4 presents a visual demonstration of the models robustness and prediction accuracy using Grad-cam and gradio interface. Subsection 5.5 offers a comparative analysis between the existing state-of-the-art models and the proposed model, emphasizing improvements, limitations, and potential implications for practical deployment in skin cancer detection.

5.1 Performance Results

In this study, four widely used pre-trained CNN architectures—EfficientNetB0, EfficientNetB3, ResNet101, and DenseNet161—were fine-tuned and tested on the HAM10000 dataset. Their outcomes are reported in Tables 5.1 through 5.4, showcasing precision, recall, F1-score, and overall accuracy for each class. While these models demonstrated promising performance, especially for well-represented classes such as Melanocytic Nevi, they showed limitations in detecting rarer but clinically significant classes like melanoma and actinic keratoses. To address these gaps, the proposed ConvNeXt-Tiny model was fine-tuned and evaluated separately, with results summarized in Table 5.5. This model consistently achieved higher overall accuracy compared to the earlier baselines, reflecting improved generalization and robustness. The comparative analysis across these tables not only highlights the progression in performance but also emphasizes the advantages of lightweight architectures optimized through advanced preprocessing and tailored loss functions.

Table 5.1: Performance of EfficientNetB0

	Precision	Recall	F1 score	Support
Actinic Keratosis	0.76	0.78	0.77	32
Basal Cell Carcinoma	0.82	0.96	0.88	52
Benign Keratosis	0.80	0.78	0.79	110
Dermatofibroma	1.00	1.00	1.00	11
Melanoma	0.56	0.82	0.69	112
Melanocytic Nevi	0.97	0.89	0.93	671
Vascular Lesions	0.87	0.93	0.90	14
Overall accuracy	87.52%			1002

Table 5.2: Performance of EfficientNetB3

	Precision	Recall	F1 score	Support
Actinic Keratosis	0.87	0.81	0.84	32
Basal Cell Carcinoma	0.90	0.90	0.90	52
Benign Keratosis	0.76	0.85	0.81	110
Dermatofibroma	1.00	1.00	1.00	11
Melanoma	0.55	0.85	0.66	112
Melanocytic Nevi	0.97	0.86	0.91	671
Vascular Lesions	0.65	0.93	0.76	14
Overall accuracy	86.13%			1002

Table 5.3: Performance of ResNet101

	Precision	Recall	F1 score	Support
Actinic Keratosis	0.57	0.75	0.65	32
Basal Cell Carcinoma	0.57	0.88	0.70	52
Benign Keratosis	0.70	0.62	0.66	110
Dermatofibroma	0.92	1.00	0.96	11
Melanoma	0.30	0.86	0.44	112
Melanocytic Nevi	0.98	0.62	0.76	671
Vascular Lesions	0.54	1.00	0.70	14
Overall accuracy	67.27%			1002

Table 5.4: Performance of DenseNet161

	Precision	Recall	F1 score	Support
Actinic Keratosis	0.70	0.81	0.75	32
Basal Cell Carcinoma	0.81	0.90	0.85	52
Benign Keratosis	0.87	0.74	0.80	110
Dermatofibroma	1.00	0.82	0.90	11
Melanoma	0.67	0.70	0.68	112
Melanocytic Nevi	0.94	0.94	0.94	671
Vascular Lesions	0.81	0.93	0.87	14
Overall accuracy	88.12%			1002

Table 5.5: Performance of fine-tunes ConVNeXT-Tiny

	Precision	Recall	F1 score	Support
Actinic Keratosis	0.92	0.80	0.85	15
Basal Cell Carcinoma	0.87	0.72	0.79	18
Benign Keratosis	0.88	0.84	0.86	44
Dermatofibroma	1.00	0.75	0.86	4
Melanoma	0.88	0.95	0.91	23
Melanocytic Nevi	0.98	0.98	0.98	442
Vascular Lesions	0.67	1.00	0.80	6
Overall accuracy	95.83%			552

5.2 Loss and Accuracy Curves

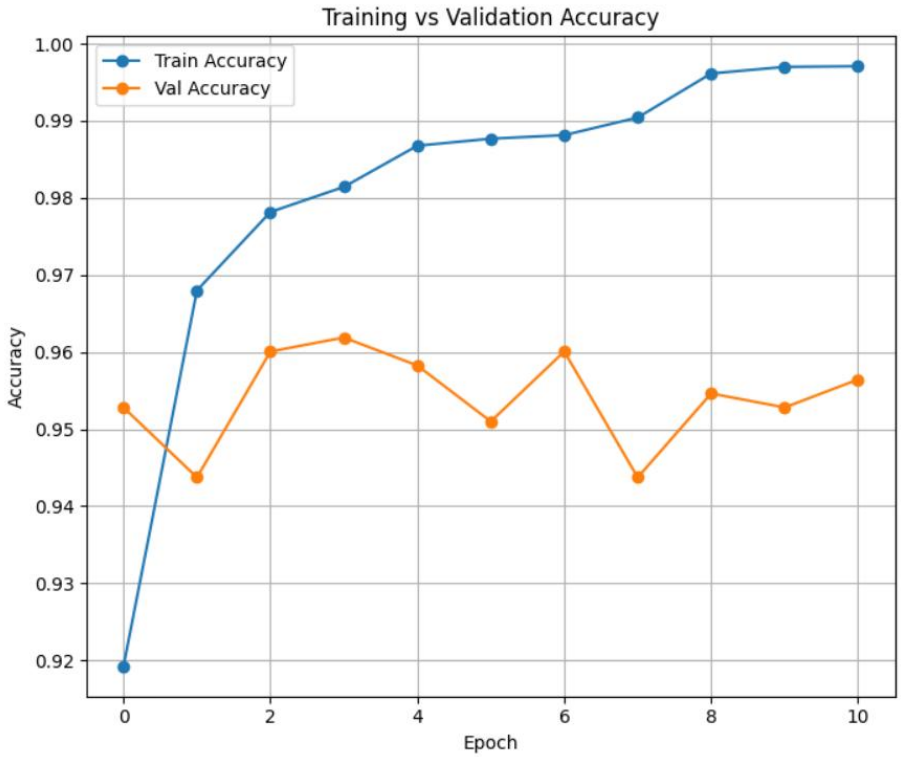


Fig. 5.1: Accuracy graph of ConvNeXT-Tiny

Fig. 5.2: Loss graph of ConvNeXT-Tiny

The accuracy and loss curves of training and validation for the proposed model hava been displayed in Fig. 5.1 and 5.2 respectively. As shown in Fig. 5.1, the training accuracy steadily increases and reaches near 100% after approximately 8 epochs, indicating that the model has effectively learned the training data. The validation accuracy also improves significantly over epochs, eventually stabilizing around 95-96%, which reflects the model's strong generalization capability on unseen data. However, a slight gap remains between the training and validation accuracy curves, suggesting mild overfitting, which is common in deep learning models trained on medical imaging datasets due to data variability.

Similarly, the loss curves in Fig. 5.2 exhibit a consistent decline for both training and validation datasets. The training loss rapidly decreases and flattens to a very low value after the initial few epochs, indicating that the model has minimized classification errors on the training set. The validation loss, while slightly higher than the training loss, also decreases and stabilizes without significant fluctuations, confirming that the model does not suffer from severe overfitting or underfitting. This trend further validates the model's robustness and its ability to generalize well to new, unseen skin disease images.

5.3 Evaluation Matrices

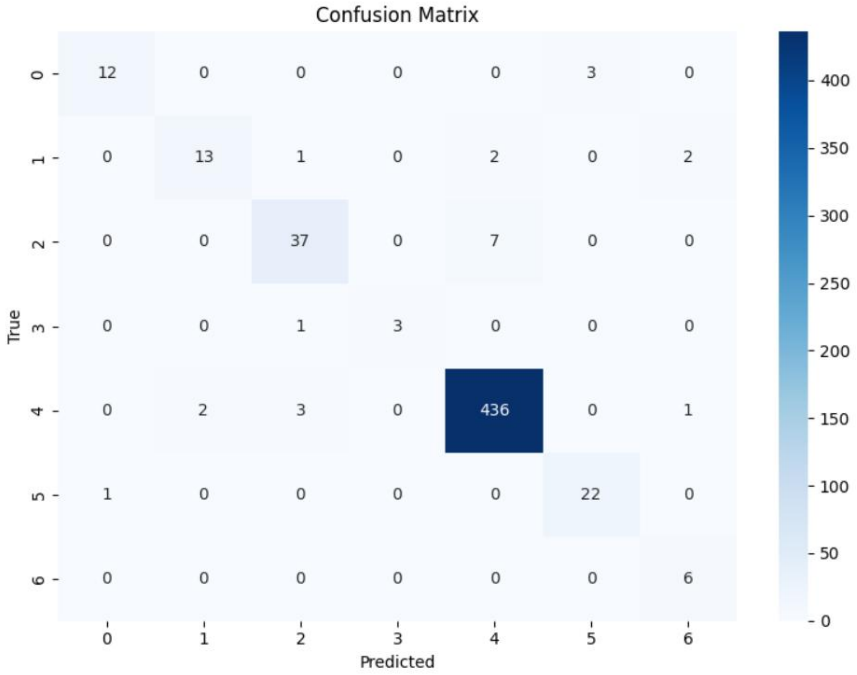


Fig. 5.3: Confusion matrix of ConvNeXT-Tiny

The confusion matrix shown in Fig. 5.3 represents the performance of the skin disease classification model on the HAM10000 dataset, which includes seven classes of skin lesions. The model demonstrates high accuracy in classifying the largest class, with 436 correct predictions for class 4, Melanocytic nevi, indicating strong performance on well-represented categories. However, the matrix also reveals challenges in less represented classes, such as class 0: Actinic keratosis and class 5: Melanoma, where some misclassifications are observed—for instance, class 0 shows three instances being predicted as class 5. Similarly, classes 1, 2, and 3 have a few misclassifications, particularly between adjacent categories, suggesting overlap in visual features among them. Overall, the model performs reliably and accurately across all classes.

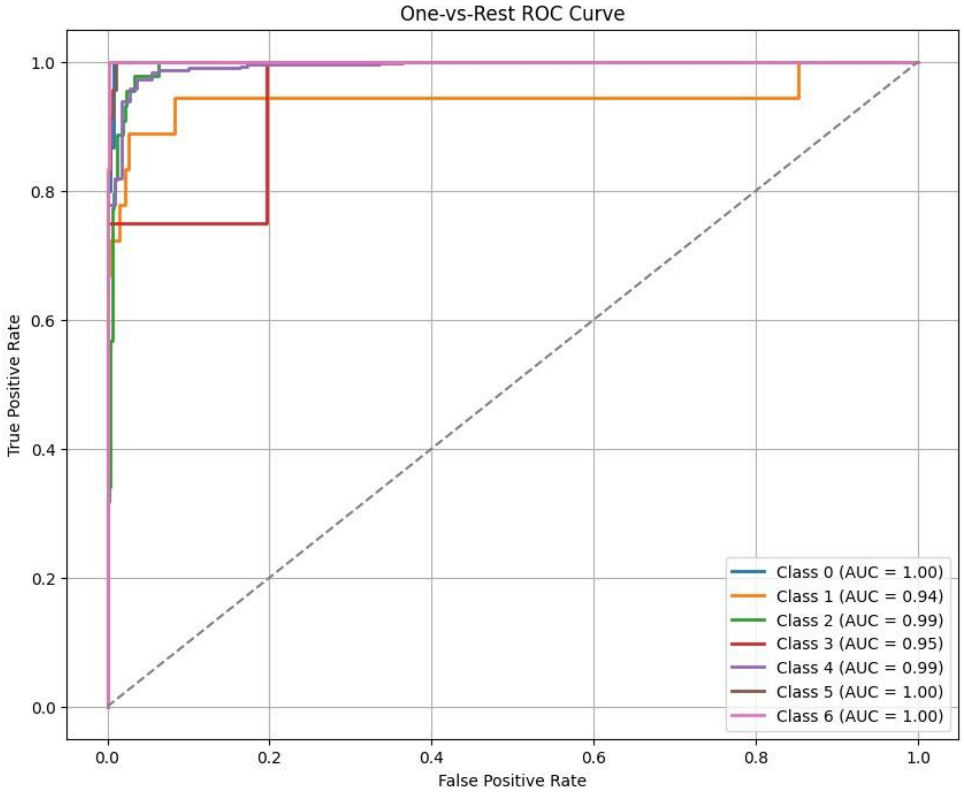


Fig. 5.4: ROC curve of ConvNeXT-Tiny

The AUC-ROC curve in Fig. 5.4 highlights the strong classification performance of the model across all seven skin lesion classes using a one-vs-rest approach. Classes 0, 5, and 6 achieved perfect AUC scores of 1.00, while Classes 2 and 4 followed closely with 0.99. Class 3 scored 0.95, and Class 1 had the lowest but still strong performance with 0.94. These high AUC values indicate that the model is highly effective at distinguishing

between different skin lesion types, with excellent overall discriminatory ability across the HAM10000 dataset.

5.4 Visualization and Deployment

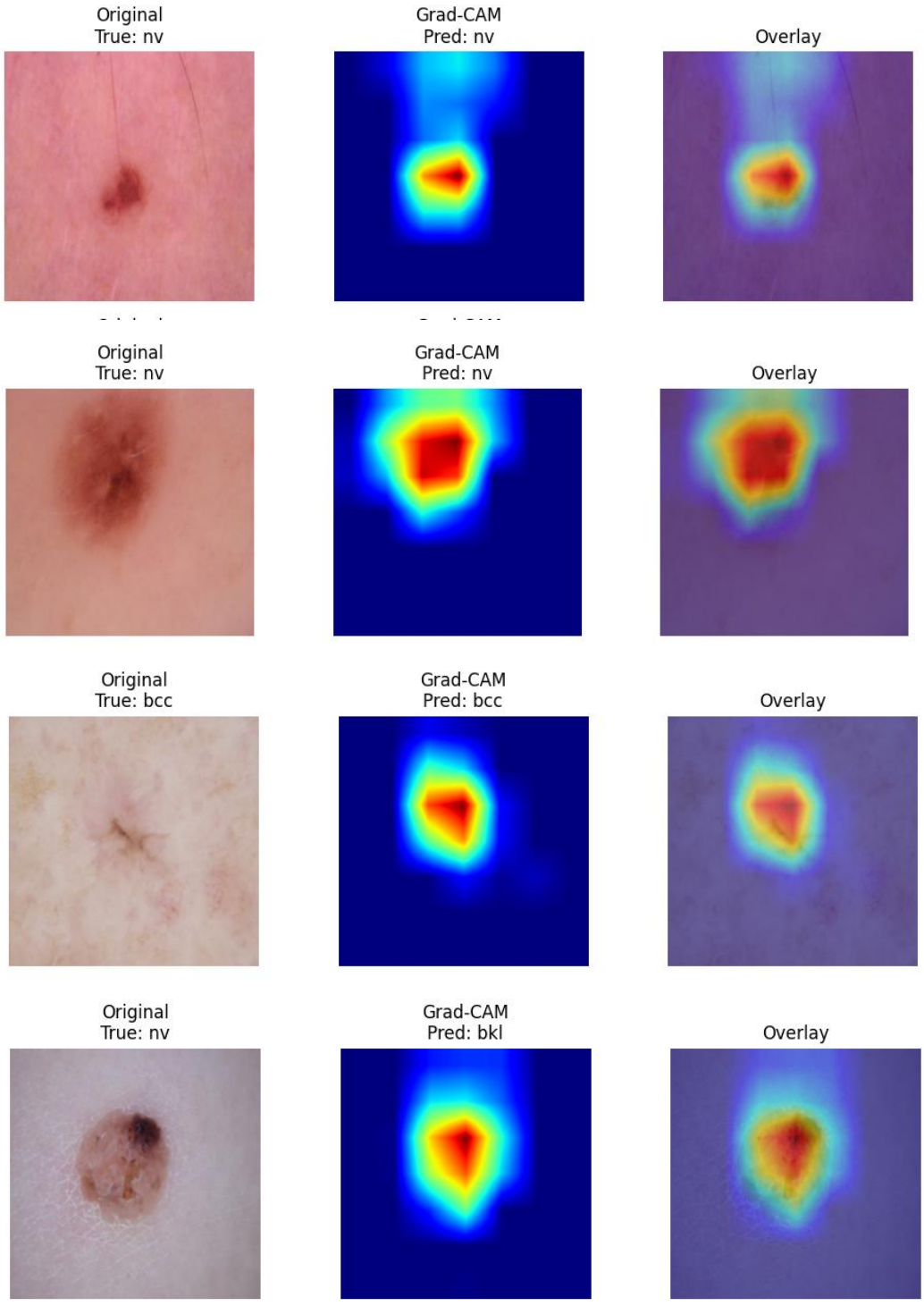


Fig. 5.5: Skin lesion visualization with Grad-CAM

To enhance the interpretability of the model's predictions, Gradient-weighted Class Activation Mapping (Grad-CAM) was applied to visualize the regions of dermoscopic images that contributed most to the classification decision. As shown in the figure, the Grad-CAM heatmaps accurately highlight the lesion areas for classes such as "nv" (nevus) and "bcc" (basal cell carcinoma), confirming that the model focuses on the relevant regions rather than surrounding skin textures. The overlay images clearly demonstrate a strong alignment between the model's attention and the actual lesion, which builds trust and reliability in the model's decision-making process—crucial for medical applications. This visualization helps clinicians or researchers understand *why* the model made a particular prediction, reducing the black-box nature of deep learning models and supporting its potential deployment in real-world diagnostic tools.

To facilitate practical usability, a lightweight Gradio interface was developed for interactive deployment of the proposed model. The interface allows users to upload dermoscopic images and instantly receive classification predictions, along with visual explanations generated through Grad-CAM heatmaps. This design makes the system accessible not only to researchers but also to healthcare professionals who may benefit from quick diagnostic support. Gradio's browser-based environment ensures platform independence, requiring no complex installation, while maintaining responsiveness and interpretability. Thus, the deployment demonstrates the feasibility of translating the trained model into a real-world decision support tool.

The Gradio interface allows users to upload a skin lesion image and receive a classification result across seven possible categories from the HAM10000 dataset. It displays the top prediction along with a confidence score, providing transparency and ease of interpretation. The deployment includes features such as Monte Carlo (MC) Dropout and Test-Time Augmentation (TTA) to improve prediction robustness and handle model uncertainty.

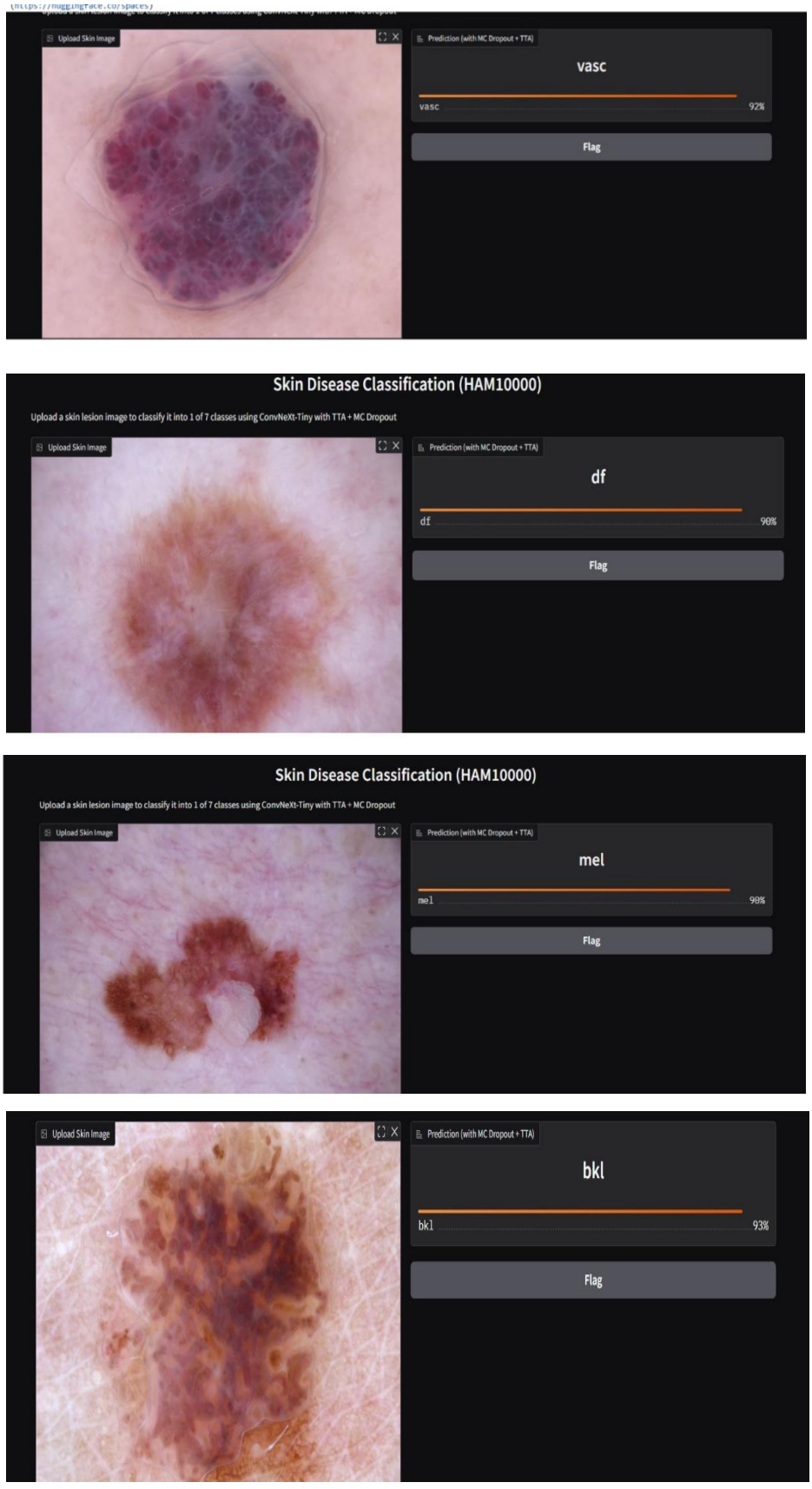


Fig. 5.6: Gradio interface deployment

5.5 Discussion

The fine-tuned ConvNeXt-Tiny network demonstrates strong potential as a lightweight tool for automated skin disease recognition, achieving an impressive overall accuracy of 95.83%. This result surpasses several deeper and more complex models while maintaining a significantly smaller size and faster performance.

The model excels at identifying common benign lesions, though it occasionally confuses visually similar conditions like benign keratosis and melanoma, a challenge that mirrors real-world dermatological difficulties. Despite this, it maintains a high true-positive rate for melanoma, highlighting its potential as a valuable screening tool.

Further analysis confirms the model's stability and reliability. The training process shows minimal overfitting, and high AUC scores indicate well-calibrated classification. Interpretability studies using Grad-CAM show that the model focuses on clinically relevant features within the lesions, increasing confidence in its diagnostic process.

Finally, live deployment results obtained via the Gradio interface corroborate offline metrics. Real-time tests on previously unseen dermoscopic images consistently produced accurate top-1 predictions with tight uncertainty bands, demonstrating that the trained weights translate effectively from batch inference to user-facing conditions.

Chapter 6: Conclusion and Future Works

6.1 Conclusion

This study presents a comprehensive investigation into multiclass skin disease classification using advanced deep learning techniques and pre-trained convolutional neural networks. Various models, including DenseNet161, ResNet101, EfficientNetB0 and EfficientNetB3 were employed, achieving classification accuracies of 88.12%, 67.27%, 87.52% and 86.13%, respectively. The most promising results were obtained using ConvNeXt-Tiny, which reached a peak accuracy of 95.83%, outperforming all other architectures. Recognizing the challenges posed by class imbalance in the dataset, the study incorporated weighted random sampling, extensive data augmentation strategies, and focal loss to enhance model robustness and generalization. These techniques significantly contributed to mitigating bias toward majority classes and improving classification performance across all skin disease categories. Furthermore, a Gradio-based interactive interface was developed to test the model on unseen data, allowing for real-time evaluation and demonstration of its classification capability. The results underscore the effectiveness of the ConvNeXt-Tiny architecture, combined with targeted training enhancements and an interactive deployment setup, in achieving high accuracy and practical usability for real-world skin disease diagnosis applications.

6.2 Future works

To further improve the clinical reliability and fairness of the skin lesion classification model, the following future directions are proposed:

Bias Mitigation through Demographic Stratification:
Conduct a detailed bias and fairness analysis by evaluating model performance across demographic variables such as age, sex, and skin tone—if such metadata becomes available. This will help identify and address disparities in prediction accuracy, ensuring equitable performance across diverse patient populations.

Segmentation-Aided Classification:

Integrate lesion segmentation into the classification pipeline using models like U-Net or BiSeNet. Segmenting the lesion area before classification will reduce background noise and help the model focus on the lesion itself, potentially improving accuracy, especially for clinically complex classes such as melanoma.

Improved Clinical Applicability:

These enhancements aim to make the model more robust, fair, and accurate, ultimately moving it closer to being a reliable diagnostic support tool suitable for real-world dermatological applications.

References

[1] N. Hameed, A. M. Shabut, M. K. Ghosh, and M. A. Hossain, "Multi-class multi-level classification algorithm for skin lesions classification using machine learning techniques," *Expert Syst Appl*, vol. 141, p. 112961, Mar. 2020, doi: 10.1016/j.eswa.2019.112961.

[2] R. J. Hay *et al.*, "The Global Burden of Skin Disease in 2010: An Analysis of the Prevalence and Impact of Skin Conditions," *Journal of Investigative Dermatology*, vol. 134, no. 6, pp. 1527–1534, Jun. 2014, doi: 10.1038/jid.2013.446.

[3] D. Seth, K. Cheldize, D. Brown, and E. E. Freeman, "Global Burden of Skin Disease: Inequities and Innovations," *Curr Dermatol Rep*, vol. 6, no. 3, pp. 204–210, Sep. 2017, doi: 10.1007/s13671-017-0192-7.

[4] C. Kim, M. Jang, Y. Han, Y. Hong, and W. Lee, "Skin Lesion Classification Using Hybrid Convolutional Neural Network with Edge, Color, and Texture Information," *Applied Sciences*, vol. 13, no. 9, p. 5497, Apr. 2023, doi: 10.3390/app13095497.

[5] S. S. Mohammed and J. M. Al-Tuwaijari, "Skin Disease Classification System Based on Machine Learning Technique: A Survey," *IOP Conf Ser Mater Sci Eng*, vol. 1076, no. 1, p. 012045, Feb. 2021, doi: 10.1088/1757-899X/1076/1/012045.

[6] M. Q. Khan *et al.*, "Classification of Melanoma and Nevus in Digital Images for Diagnosis of Skin Cancer," *IEEE Access*, vol. 7, pp. 90132–90144, 2019, doi: 10.1109/ACCESS.2019.2926837.

[7] T. Wollina, "Recent advances in managing and understanding seborrheic keratosis," *F1000Res*, vol. 8, p. 1520, Aug. 2019, doi: 10.12688/f1000research.18983.1.

[8] K. Tanese, "Diagnosis and Management of Basal Cell Carcinoma," *Curr Treat Options Oncol*, vol. 20, no. 2, p. 13, Feb. 2019, doi: 10.1007/s11864-019-0610-0.

[9] D. Dianzani *et al.*, "Current therapies for actinic keratosis," *Int J Dermatol*, vol. 59, no. 6, pp. 677–684, Jun. 2020, doi: 10.1111/ijd.14767.

[10] E. R. Allugunti, “A machine learning model for skin disease classification using convolution neural network,” *International Journal of Computing, Programming and Database Management*, vol. 3, no. 1, pp. 141–147, Jan. 2022, doi: 10.33545/27076636.2022.v3.i1b.53.

[11] B. Ahmad, M. Usama, C.-M. Huang, K. Hwang, M. S. Hossain, and G. Muhammad, “Discriminative Feature Learning for Skin Disease Classification Using Deep Convolutional Neural Network,” *IEEE Access*, vol. 8, pp. 39025–39033, 2020, doi: 10.1109/ACCESS.2020.2975198.

[12] P. N. Srinivasu, J. G. SivaSai, M. F. Ijaz, A. K. Bhoi, W. Kim, and J. J. Kang, “Classification of Skin Disease Using Deep Learning Neural Networks with MobileNet V2 and LSTM,” *Sensors*, vol. 21, no. 8, p. 2852, Apr. 2021, doi: 10.3390/s21082852.

[13] Z. Wu *et al.*, “Studies on Different CNN Algorithms for Face Skin Disease Classification Based on Clinical Images,” *IEEE Access*, vol. 7, pp. 66505–66511, 2019, doi: 10.1109/ACCESS.2019.2918221.

[14] S. A. AlDera and M. T. Ben Othman, “A Model for Classification and Diagnosis of Skin Disease using Machine Learning and Image Processing Techniques,” *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 5, 2022, doi: 10.14569/IJACSA.2022.0130531.

[15] T. M. Alam *et al.*, “An Efficient Deep Learning-Based Skin Cancer Classifier for an Imbalanced Dataset,” *Diagnostics*, vol. 12, no. 9, p. 2115, Aug. 2022, doi: 10.3390/diagnostics12092115.

[16] M. Maduranga and D. Nandasena, “Mobile-Based Skin Disease Diagnosis System Using Convolutional Neural Networks (CNN),” *International Journal of Image, Graphics and Signal Processing*, vol. 14, no. 3, pp. 47–57, Jun. 2022, doi: 10.5815/ijigsp.2022.03.05.

[17] A. Adebiyi *et al.*, “Accurate Skin Lesion Classification Using Multimodal Learning on the HAM10000 and ISIC 2017 Datasets,” May 31, 2024. doi: 10.1101/2024.05.30.24308213.

[18] M. Fraiwan and E. Faouri, “On the Automatic Detection and Classification of Skin Cancer Using Deep Transfer Learning,” *Sensors*, vol. 22, no. 13, p. 4963, Jun. 2022, doi: 10.3390/s22134963.

[19] K. Ali, Z. A. Shaikh, A. A. Khan, and A. A. Laghari, "Multiclass skin cancer classification using EfficientNets – a first step towards preventing skin cancer," *Neuroscience Informatics*, vol. 2, no. 4, p. 100034, Dec. 2022, doi: 10.1016/j.neuri.2021.100034.

[20] S. Inthiyaz *et al.*, "Skin disease detection using deep learning," *Advances in Engineering Software*, vol. 175, p. 103361, Jan. 2023, doi: 10.1016/j.advengsoft.2022.103361.

[21] S. Jinnai, N. Yamazaki, Y. Hirano, Y. Sugawara, Y. Ohe, and R. Hamamoto, "The Development of a Skin Cancer Classification System for Pigmented Skin Lesions Using Deep Learning," *Biomolecules*, vol. 10, no. 8, p. 1123, Jul. 2020, doi: 10.3390/biom10081123.

[22] M. A. Kassem, K. M. Hosny, and M. M. Fouad, "Skin Lesions Classification Into Eight Classes for ISIC 2019 Using Deep Convolutional Neural Network and Transfer Learning," *IEEE Access*, vol. 8, pp. 114822–114832, 2020, doi: 10.1109/ACCESS.2020.3003890.

[23] A. Mahbod, G. Schaefer, C. Wang, R. Ecker, and I. Ellinge, "Skin Lesion Classification Using Hybrid Deep Neural Networks," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, May 2019, pp. 1229–1233. doi: 10.1109/ICASSP.2019.8683352.

[24] E. Maher and A. Anokhin, "Bacterial Skin and Soft Tissue Infections in Older Adults," *Clin Geriatr Med*, vol. 40, no. 1, pp. 117–130, Feb. 2024, doi: 10.1016/j.cger.2023.09.006.

[25] S. K. Singh, V. Abolghasemi, and M. H. Anisi, "Fuzzy Logic with Deep Learning for Detection of Skin Cancer," *Applied Sciences*, vol. 13, no. 15, p. 8927, Aug. 2023, doi: 10.3390/app13158927.

[26] M. S. Akter, H. Shahriar, S. Sneha, and A. Cuzzocrea, "Multi-class Skin Cancer Classification Architecture Based on Deep Convolutional Neural Network," in *2022 IEEE International Conference on Big Data (Big Data)*, IEEE, Dec. 2022, pp. 5404–5413. doi: 10.1109/BigData55660.2022.10020302.

[27] A. Karambele, M. Tamhane, P. Jayanty, R. Patil, and K. Shirsat, “eDermaCare: Enhancing Skin Disease Classification with ResNet50,” in *2024 11th International Conference on Computing for Sustainable Global Development (INDIACom)*, IEEE, Feb. 2024, pp. 1008–1014. doi: 10.23919/INDIACom61295.2024.10499009.

[28] L. Faris, R. Qayyum, C. Pepe, M. F. Ali, and S. Maria Zanoli, “Multi-Class Skin Disease Detection Using Deep Learning Hybrid Method,” in *2025 26th International Carpathian Control Conference (ICCC)*, IEEE, May 2025, pp. 1–5.doi: 10.1109/ICCC65605.2025.11022791.

[29] P. Tschandl, C. Rosendahl, and H. Kittler, “The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions,” *Sci Data*, vol. 5, no. 1, p. 180161, Aug. 2018, doi: 10.1038/sdata.2018.161.

[30] IBM, “Convolutional neural networks (CNNs),” IBM Think, [Online]. Available: https://www.ibm.com/think/topics/convolutional-neural-networks.

[31] D. Kim, “Activation functions,” Wikidocs, [Online]. Available: https://wikidocs.net/255187.

[32] M. Yani, B. Irawan, and C. Setiningsih, “Application of Transfer Learning Using Convolutional Neural Network Method for Early Detection of Terry’s Nail,” Journal of Physics: Conference Series, vol. 1201, no. 1, 2019, Art. no. 012052, doi: 10.1088/1742- 6596/1201/1/012052.

[33] Shaees, Shamoil & Naeem, Muhammad Rashid & Naeem, Hamad & Syed, Hamza & Arslan, Muhammad & Aldabbas, Hamza. (2020). Facial Emotion Recognition Using Transfer Learning. Doi: 10.1109/ICCIT-144147971.2020.9213757.

[34] A. Dey, "The Architecture of VGGNet: Breaking Down VGG16," Plain English, Jun. 9, 2024. [Online]. Available: https://ai.plainenglish.io/the-architecture-of-vggnet-breaking-down-vgg16-4f9fe45327d1

[35] J. Lowe, "What is EfficientNet?" Roboflow Blog, Mar. 15, 2023. [Online]. Available: https://blog.roboflow.com/what-is-efficientnet/

[36] A. Debray, "Mastering ResNet: Deep Learning for Image Recognition," Ikomia Blog, Feb. 27, 2024. [Online]. Available: https://www.ikomia.ai/blog/mastering-resnet-deep-learning-image-recognition.

[37] A. Erdogan, "ConvNeXt: Next Generation of Convolutional Networks," Medium, May 6, 2024. [Online].
Available: https://medium.com/@atakanerdogan305/convnext-next-generation-of-convolutional-networks-325607a08c46