# Deep Learning-Based Classification of Skin Lesions

1st Prof Sabyasachi Patra
*Department of CSE*
*IIIT Bhubaneswar*
Odisha, India
sabyasachi@iiit-bh.ac.in

2nd Prof Tushar Ranjan Sahoo
*Department of CSE*
*IIIT Bhubaneswar*
Odisha, India
Tushar@iiit-bh.ac.in

3rd Subrat Kumar Swain
*Department of IT*
*IIIT Bhubaneswar*
Odisha, India
b421055@iiit-bh.ac.in

4th Durbadala Padhan
*Department of CE*
*IIIT Bhubaneswar*
Odisha, India
b521020@iiit-bh.ac.in

5th Pratyush Pandey
*Department of IT*
*IIIT Bhubaneswar*
Odisha, India
b421035@iiit-bh.ac.in

***Abstract*—The proliferation of deep learning has significantly advanced the field of medical image analysis, particularly in the early detection and classification of skin cancer. Skin cancer, encompassing various types such as melanoma, basal cell carcinoma, and squamous cell carcinoma, remains one of the most prevalent forms of cancer worldwide. Early and accurate diagnosis is crucial for effective treatment and improved patient outcomes. Traditional diagnostic methods, relying heavily on clinical expertise and histopathological examination, can be time-consuming and subject to inter-observer variability. To address these challenges, this study explores the use of convolutional neural networks for automating the process skin cancer classification. The research involves a comparative analysis of four CNN architectures: MobileNetV2, EfficientNetB4, DenseNet121, and a custom Sequential model. These models were trained and evaluated on a curated dataset comprising dermoscopic images, which were preprocessed and augmented to enhance diversity and mitigate class imbalance. The preprocessing steps included resizing, normalization, and various augmentation techniques such as rotation, flipping, and zooming. The dataset was divided in training part, validation parts, and testing parts to ensure robust evaluation of model performance.**

**Every convolutional model in this study was adapted through a transfer learning approach, where previously learned parameters were utilized to accelerate model training and enhance predictive performance. Training of Model was carried out using an adaptive gradient-based optimization method, with a multi-class probabilistic loss guiding the updates. Key training parameters—including learning rate, epoch count, and batch size—were fine-tuned using systematic validation techniques. To evaluate model effectiveness, a comprehensive set of performance indicators such as predictive accuracy, class-wise relevance, balance of prediction, and curve-based discrimination metrics were employed.**

***Index Terms*—Skin Cancer Detection, Convolutional Neural Networks (CNN), DenseNet121, EfficientNetB4, MobileNetV2, Medical Image Analysis, Dermoscopic Images, Transfer Learning, Image Preprocessing, Data Augmentation, Classification Accuracy, Deep Learning**

## I. Introduction

Cancer arises due to the abnormal proliferation of cells that possess the capability to invade other areas of the body [1]. In all the different types of cancers, skin cancer stands out as particularly aggressive and dangerous. Its treatability is heavily reliant on early detection. As the largest organ in the human body, the skin plays a crucial role in protecting internal structures such as muscles and bones. Even a slight disturbance in its normal function can significantly impact multiple physiological systems. A lesion is defined in dermatology as an area of the skin that shows pathological alterations. Skin lesions are diverse and are generally classified based on the type of skin cells from which they develop. Lesions arising from melanocytes are known as melanocytic lesions, which often exhibit characteristics similar to melanoma. Melanocytes are specialized cells known for the synthesis of melanin, the natural pigment responsible for determining skin tone [1].

This study draws attention to the serious problems with skin cancer and the challenges in detecting it early. It examines how deep learning methods, specifically Convolutional Neural Networks (CNNs) and few-shot learning, can increase the accuracy of dermoscopic image classification. Although combining qualitative and quantitative approaches might yield deeper understanding in terms of practical deployment and usability, the present research emphasizes quantitative methodologies. This decision is driven by the inherent data-centric design of machine learning models, which rely on large-scale datasets to function effectively.

Among all skin-related malignancies, melanoma stands out as the most life-threatening variant, contributing significantly to global healthcare challenges and mortality rates. In accordance with the World Health Organization (WHO), there are nearly 2–3 million new cases of non-melanoma skin cancer and around 132,000 cases of melanoma diagnosed annually worldwide [2]. Timely diagnosis significantly improves survival outcomes; however, traditional diagnostic procedures are often expensive, labor-intensive, and limited in availability—especially in underserved or remote regions [4]. As a result, AI-powered solutions, like those developed in this study, offer a promising alternative by enabling scalable, cost-effective, and precise screening tools. These models

hold potential for future integration into real-world clinical workflows, including teledermatology and remote diagnosis platforms.

Non-melanocytic skin lesions originate from various epidermal cell types, including squamous and basal cells. These lesions can often be visually distinguished from melanocytic ones by specific dermoscopic features, such as the presence or absence of pigment networks and vascular patterns. Once a lesion is identified, the next diagnostic step involves determining its biological behavior—whether it is benign or malignant. If classified as malignant, further differentiation is required to identify the exact type of skin cancer involved.

Dermoscopy plays a crucial role in this process, as it enhances the visibility of subsurface skin structures that are not seen with the naked eye. Several visual features, including symmetry, border irregularity, color variation, and structural patterns, are used to guide the classification. Common dermatological conditions like basal cell carcinoma, melanoma, and seborrheic keratosis often manifest through distinct skin lesions, which serve as the primary clinical indicators of these diseases.

With aging, individuals may develop benign tumors such as seborrheic keratosis or basal cell carcinoma—the latter being the most frequently diagnosed skin cancer globally. Skin cancer is broadly categorized into melanoma and non-melanoma, with the former being less common but significantly more lethal. Melanoma tends to disproportionately affect individuals of lighter skin tones, particularly Caucasians. It frequently appears on the torso in men and on the lower limbs in women, although it can also develop in less exposed regions [5], [6].

Cancer is to be a widespread health crisis, impacting millions of individuals annually. Within the spectrum of oncology, skin cancer has emerged as a growing area of concern, particularly due to its increasing incidence and the need for early intervention [7]. Progress in computer-aided diagnostic (CAD) technologies has demonstrated strong potential in facilitating early skin cancer detection and diagnosis by utilizing medical imaging techniques in combination with machine learning methods. [8].

Traditionally, dermatologists rely on invasive diagnostic techniques such as biopsies to confirm the presence of malignancy. In this procedure, a sample of skin tissue is surgically extracted and subjected to detailed histopathological analysis. While effective, this method can be uncomfortable for patients and requires significant time and medical resources. Furthermore, such procedures may lead to scarring and changes in pigmentation such as the development of blue-white veils, gray dots, or pigmented networks—features commonly observed in dermoscopic analysis of melanocytic lesions [7].

To support non-invasive diagnosis, clinical and dermoscopic imaging plays a critical role. These images are typically captured using conventional digital cameras or dermatoscopes. However, analyzing such images poses challenges due to the presence of noise, hair, shadows, and variations in lighting, which can reduce the clarity of skin features necessary for precise diagnosis [8]. Overcoming these limitations remains a key focus of current research in AI-assisted dermatology.

Skin cancer is a concerning medical condition that happens from the growth of abnormal skin cells. It stands out as a major public health issue globally, in part due to the skin being the body's largest and most externally exposed organ [9]. This type of cancer generally develops when genetic damage within skin cells is left uncorrected, resulting in abnormal cellular proliferation and the eventual formation of cancerous growths [2].

Skin cancer is generally classified into two broad categories: melanoma and non-melanoma. Non-melanoma skin cancers, which include basal cell carcinoma and squamous cell carcinoma, account for the vast majority of cases and typically have a low risk of metastasis. On the other hand, melanoma is a rarer but much more aggressive form of skin cancer that originates in melanocytes—the pigment-producing cells of the epidermis [10].

While melanocytes are primarily concentrated in the skin, these pigment-producing cells can also be found in various internal structures of the body, including the eyes and select internal organs. Melanoma often occurs on the back, chest, or face in males and more usually on the legs in females. Despite representing less than 1% of all skin cancer cases, melanoma is responsible for the majority of skin cancer-related deaths due to its high metastatic potential and rapid progression. Early diagnosis is critical, as timely detection substantially improves treatment success and reduces associated healthcare costs [11].

The process of detecting skin cancer through computational methods typically consists of multiple steps, including data preprocessing, segmentation of the lesion area, extraction of relevant features, dimensionality reduction, and ultimately, classification of the condition [13], [15]. In this research, we explore the potential of few-shot meta-learning by employing a hybrid model that integrates Convolutional Neural Networks (CNNs) with prototypical networks. This architecture leverages Kullback-Leibler divergence as a foundational metric for evaluating similarity between classes.

Given the urgent need for accurate and timely skin cancer detection, researchers have developed a range of methods aimed at enhancing diagnostic precision and sensitivity. One of the most effective non-invasive techniques for

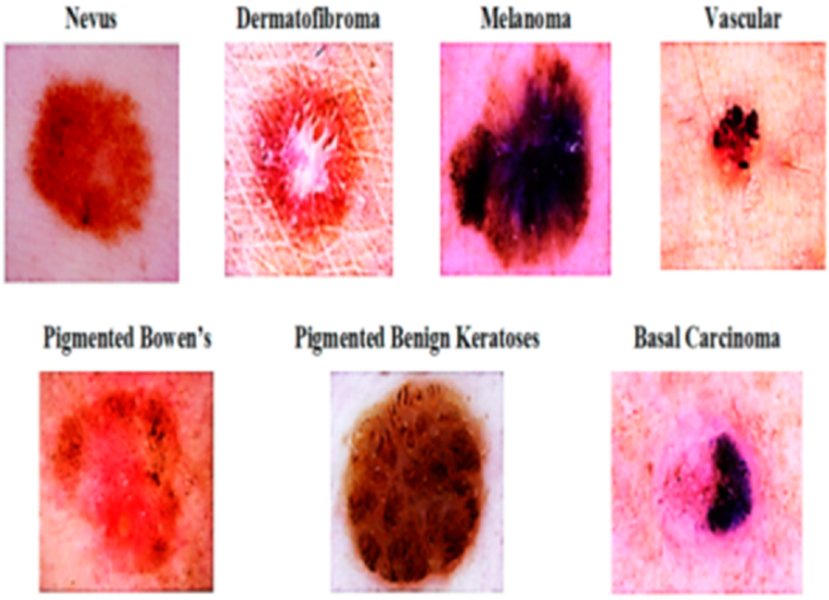


Fig. 1. Illustrates the different categories of skin conditions as outlined by the referenced material.

melanoma detection is dermoscopy, which allows for detailed examination of pigmented skin lesions. Dermatologists rely on dermoscopic images to assess abnormal skin patterns and identify malignancies. With the rise of artificial intelligence, especially in the subfield of computer vision, the domain of dermatology has seen substantial advancements. Computer vision enables machines to interpret and extract information from visual data, mimicking the cognitive visual abilities of humans. This has paved the way for the automation of tasks such as lesion recognition and classification, significantly augmenting clinical diagnostic workflows [17].

## II. LITERATURE SURVEY

Deep learning techniques, particularly Convolutional Neural Networks (CNNs), have demonstrated strong potential in the field of skin cancer diagnosis. Despite their effectiveness, these models frequently encounter difficulties related to class imbalance within skin lesion datasets, especially when it comes to detecting uncommon but critical cases like melanoma. Additionally, their effectiveness heavily depends on the availability of large volumes of labeled data, which can be a limitation in many clinical settings. To overcome these issues, this research integrates Generative Adversarial Networks (GANs) to produce synthetic training samples, thereby enhancing the model's performance in scenarios with limited or unevenly distributed data.

Deep learning algorithms have become instrumental in the automated detection and diagnosis of skin cancer. Their design is based on stacked processing elements that exchange information across levels, loosely modeled after the interconnected signaling patterns found in human neural systems. These nodes collaboratively process complex inputs and, through iterative training cycles, the networks gain expertise in specialized recognition tasks [18]. In our research, we employed various deep neural network models to classify dermoscopic images and effectively distinguish between different categories of skin cancer.

The ISIC (International Skin Imaging Collaboration) dataset, which is one of the most extensively used datasets in dermatological image analysis, was utilized for this study. It contains a broad spectrum of annotated dermoscopic images that support both classification and segmentation tasks [19]. Fig 1. illustrates some examples from the dataset, highlighting its diversity and clinical relevance.

Our work includes a comparative evaluation of prominent deep learning architectures—namely Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), K-Nearest Neighbors (KNN), and Generative Adversarial Networks (GAN)—to analyze their effectiveness in diagnosing skin cancer. ANN and CNN have long been the foundation of visual pattern recognition in medical imaging, while GANs are increasingly being leveraged for data augmentation, especially in scenarios involving limited or imbalanced data [20]. KNN, though a simpler algorithm, was included as a baseline for comparative performance measurement.

This study presents an in-depth evaluation of these methods, highlighting their contributions to enhancing diagnostic precision, sensitivity, and the overall reliability of skin cancer classification models in practical deployment scenarios.

### A. Application of CNN in Skin Cancer Detection

Deep learning architectures designed for visual data processing commonly known as convolution based neural models form the backbone of many contemporary solutions in the field of computer vision. They are widely utilized for processing and classifying visual inputs such as photographs and medical images. CNNs are particularly powerful in identifying and learning patterns from both local features (such as edges and textures) and more complex global structures like shapes and regions [21].

The architecture of a CNN is typically composed of several types of layers, including convolutional layers that apply filters to extract meaningful features, activation layers (such as ReLU) that introduce non-linearity, pooling layers that reduce dimensionality while retaining important information, and fully connected layers that perform the final classification [22]. The combination of these layers enables CNNs to learn hierarchical representations of data—from low-level edges in early layers to high-level semantic features in deeper layers.

Fig 2 represents the foundational structure of a convolutional neural network, showcasing critical operations including convolution, downsampling, and dense layer integration. These models have become the backbone for functions like identifying and locating objects, facial recognition, and most notably, skin lesion classification in dermatology [23].

Convolution-based neural models now play a crucial role in driving progress in making automated deep learning systems across diverse operations in healthcare imaging workflows, spanning localization, structural delineation, and diagnostic categorization. [33], [34].Yuan et al. developed a specialized deep learning framework aimed at improving melanoma detection accuracy. Their system employed a fully convolutional residual structure built from 16 sequential residual layers to perform lesion segmentation. Additionally, they enhanced classification precision by merging the outcomes of a support vector machine and a softmax-based classifier. The integration of segmentation significantly boosted performance, achieving an accuracy of 85.5%, compared to 82.8% without it.

A multi-scale convolutional framework was proposed wherein one researcher focused on extending the Inception v3 model, while the other contributed to integrating multi-resolution inputs. This architecture, originally trained on the ImageNet dataset, was adapted to enhance skin lesion classification performance [36]. Their method involved incorporating two input resolutions for dermoscopic images—one at a coarse scale to capture the general lesion morphology and surrounding context, and another at a finer scale to extract detailed texture information essential for differentiating between lesion types. This multi-resolution technique allowed for a more nuanced feature extraction, improving classification performance.

These studies demonstrate how CNN-based models, when optimized with advanced architectural strategies and multi-scale data inputs, can significantly enhance the identification or classification of malignant skin lesions like melanoma.

Mahbod along with collaborators developed a reliable skin lesion classification framework that utilizes feature representations derived from multiple pre-trained convolutional models, such as AlexNet, ResNet-18, and VGG16 [26]. These models were used as deep feature extractors, and the resulting features were then passed into a multi-class Support Vector Machine (SVM) for classification. The final predictions were obtained by fusing the scores from the individual classifiers. Their approach was validated using the ISIC 2017 dataset and achieved notable Area Under the Curve (AUC) scores—97.55% for melanoma and 83.83% for seborrheic keratosis.

A subsequent approach targeting the classification of multiple skin lesion categories utilized a convolutional neural network based on a pre-initialized ResNet-152 framework. Training was performed on a dataset of 3,797 dermoscopic images, each subjected to 29 distinct augmentations involving variations in illumination and scaling. This method yielded outstanding classification performance, achieving an area under the ROC curve (AUC) of 0.99 for identifying conditions such as hemangioma, pyogenic granuloma, and intraepithelial carcinoma.

Dorj et al. presented a classification framework for identifying four types of skin lesions using deep learning. Their approach utilized AlexNet for deep feature extraction, followed by classification with an error-correcting output code (ECOC)-based SVM classifier [28]. Their model delivered high performance, achieving sensitivity, specificity, and accuracy rates of 95.1%, 98.9%, and 94.17%, respectively, particularly for actinic keratosis (AK), squamous cell carcinoma (SCC), and basal cell carcinoma (BCC).

Another study by Kalouche proposed using a VGG-16-based CNN architecture consisting of five convolutional blocks and three fully connected layers for melanoma detection [29]. The model achieved a classification accuracy of 78% when distinguishing melanoma from other lesions. Furthermore, an additional CNN model was developed to identify lesion boundaries using a dataset of 1,200 healthy skin images and 400 lesion images. This binary classification model (lesion vs. non-lesion) reached an accuracy of 86.67%, showcasing CNNs' effectiveness in both segmentation and classification tasks.

These studies highlight the versatility and accuracy of CNN-based systems in dermatological diagnostics, from multi-class classification to boundary detection.

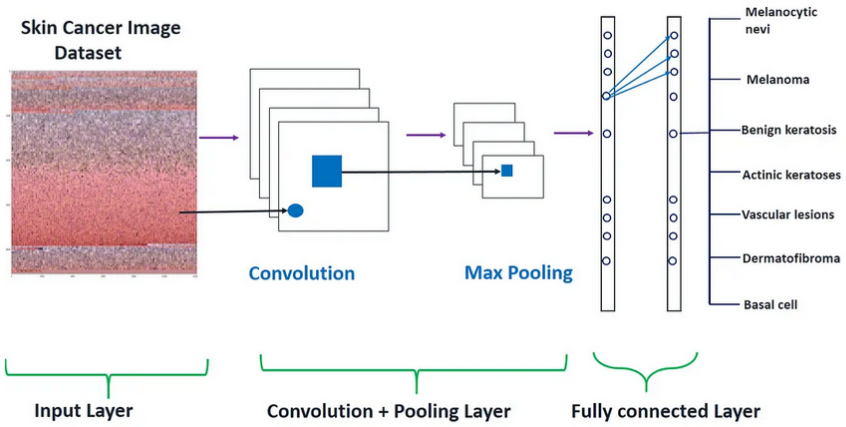


Fig. 2. fundamental structure of a Convolutional Neural Network (CNN)

## III. METHODOLOGY AND MODEL IMPLEMENTATION

This study presents a comprehensive deep learning pipeline for the classification of skin lesions using dermoscopic images. The project begins with the generation of a balanced dataset sourced primarily from the HAM10000 and ISIC image archives. These images were resized to a uniform resolution of 112×112 pixels to ensure consistency across the modeling process. In response to class distribution disparities, diverse augmentation methods like flipping, rotating, and scaling were utilized during preprocessing to enrich the dataset and promote balanced learning. Each image was normalized, and labels were one-hot encoded to prepare the data for training. The prepared dataset was then used to train and evaluate multiple convolutional neural network architectures. These included MobileNetV2, EfficientNetB4, DenseNet121, and a custom-designed Sequential CNN model. Pre-trained models were fine-tuned using transfer learning to leverage learned features from the ImageNet dataset, while the Sequential model served as a performance baseline. The training process incorporated regularization techniques, callbacks, and learning rate schedulers to optimize convergence. The evaluation phase involved measuring model performance through various indicators like recall, accuracy, precision, F1-score, the area under the curve of ROC (AUC-ROC), and the confusion matrix.. Out of all the models evaluated, DenseNet121 achieved the top performance in terms of accuracy, showcasing its strong capability in identifying diverse skin lesion categories and reinforcing the value of deep convolutional networks for automated skin cancer detection. Fig 3 gives the overall idea of the process

### *A. Data Acquisition and Preprocessing*

A robust and well-structured dataset is the backbone of any high-performance machine learning system, especially in the sensitive domain of medical imaging where model decisions can directly influence clinical outcomes. In this research, the process of data generation and preprocessing was carried out with significant attention to quality, balance, and structural uniformity to support the reliable training of deep learning models aimed at classifying skin cancer.

This research utilized a publicly available dermoscopic dataset known as Human Against Machine with 10,000 annotated skin lesion samples [24] and ISIC (International Skin Imaging Collaboration) datasets [25], both of which are widely recognized for their diversity and professional dermatological labeling. These datasets contain high-resolution dermoscopic images representing various pigmented skin lesions. For the purpose of this project, a subset of images was selected to represent seven distinct diagnostic classes: basal cell carcinoma (BCC), dermatofibroma (DF), actinic keratoses (AKIEC), melanocytic nevi (NV), pyogenic granulomas and hemorrhage (VASC), and melanoma (MEL), benign keratosis-like lesions (BKL). The selection of these 7 classes was based on both clinical relevance and availability of data within each category, enabling a comprehensive multi-class classification task.

After data acquisition, an initial image inspection and sanity check was performed to verify label consistency and remove any duplicate, corrupted, or low-quality images. All properly formatted images were rescaled to a consistent size of 112×112×3 to ensure uniformity across the dataset, using bilinear interpolation. This resizing standard ensures compatibility across all the CNN architectures implemented in the study, which include MobileNetV2, EfficientNetB4, DenseNet121, and a custom-built Sequential model. Resizing also helps maintain a manageable computational footprint, allowing for more efficient training, especially on limited

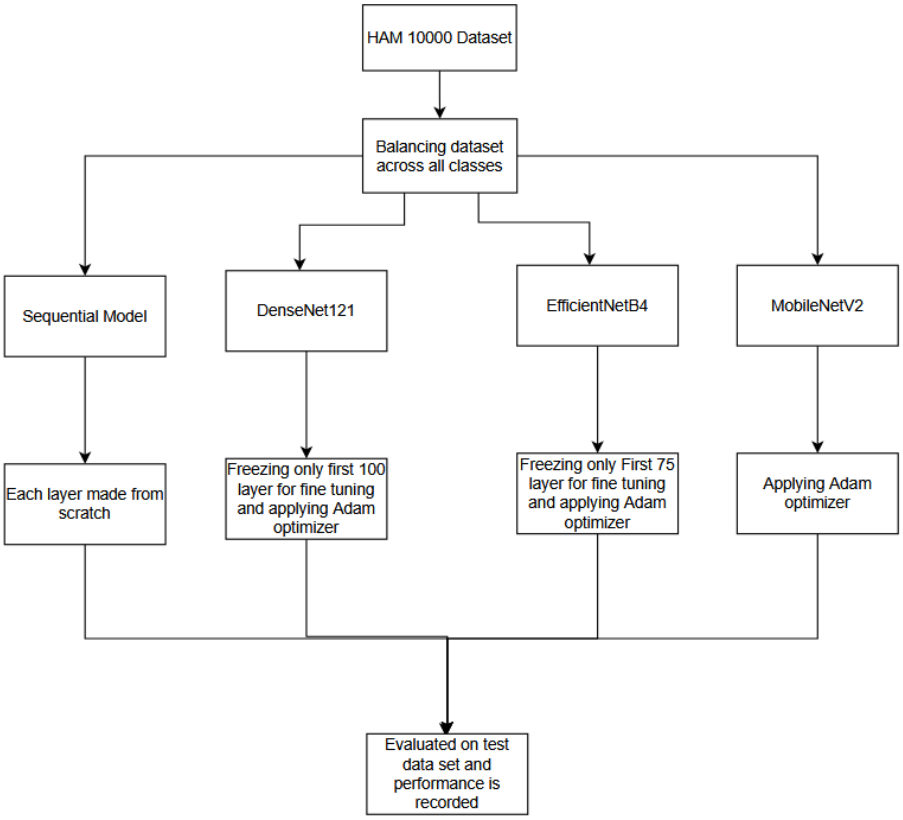


Fig. 3. Overall flow of the method implemented

hardware.

Pixel intensity values were normalized by converting them from their initial [0,255] range to a scaled range between 0 and 1. This transformation is expressed mathematically as:

$$x_{\text{normalized}} = \frac{x_{\text{original}}}{255.0}$$

where $x_{\text{original}}$ is the original RGB pixel value and $x_{\text{normalized}}$ is the scaled version used as model input. This step reduces internal covariate shift and ensures that all CNNs converge more smoothly during training.

To facilitate multi-class classification, the class labels, originally in integer format from 0 to 6 (each representing one of the seven lesion types), were converted into one-hot encoded vectors. This categorical representation aligns with the requirements of the softmax activation function used in the output layer and enables the use of categorical cross-entropy as the loss function. The transformation is given by:

$$f(y_i) = [0, 0, \ldots, 1, \ldots, 0] \quad (1 \text{ at index } y_i)$$

where $y_i$ is the class index for the $i^{\text{th}}$ image, and $f(y_i)$ is the corresponding one-hot encoded vector.

A major challenge addressed during data preparation was the issue of class imbalance, a common occurrence in medical datasets where benign lesions are significantly more prevalent than malignant ones. If left unaddressed, this imbalance can cause the model to become biased toward majority classes, reducing its ability to detect rare but critical conditions such as melanoma. To mitigate this, a controlled sampling technique was applied, where images from minority classes were oversampled and those from majority classes were limited to match the average class count. This resulted in a balanced dataset that ensures each lesion type is equally represented during training.

Following dataset balancing, an 80-20 partition was performed to separate training and test data, employing stratified sampling to preserve the original class distribution across both divisions. A fixed random seed was used to allow reproducibility of results, a crucial requirement for scientific validation and comparative analysis.

As the concluding stage of the preprocessing workflow, techniques of augmentation were implemented to synthetically enhance the variability and richness of the training data. Medical datasets often have limited data due to ethical and privacy concerns, making augmentation indispensable for preventing overfitting and improving generalization. Using TensorFlow's ImageDataGenerator, various transformations were implemented during training

For ensuring computational efficiency, fully preprocessed and augmented datasets were serialized using NumPy's .npy format. Temporary variables were deleted and garbage collection routines were triggered to optimize memory usage and prevent overload during model training. This is especially important when training high-capacity models like EfficientNetB4 or DenseNet121, which require substantial memory overhead.

In summary, the data generation and preprocessing phase of this project ensured the creation of a high-quality, balanced, and richly augmented dataset that serves as the foundation for all subsequent training and evaluation stages. The rigor applied at this stage not only enhanced model performance but also contributed to the reproducibility and robustness of the entire experimental workflow.

### *B. Model Architecture and Training*

After preparing a clean, balanced, and augmented dataset, the next phase in this study involved selecting and training a series of deep learning models to classify skin lesions into seven diagnostic categories. The primary objective of this phase was to evaluate and compare the performance of various convolutional neural network (CNN) architectures each with differing complexity, parameter count, and learning strategy when trained on the same data.

To ensure a comprehensive analysis, four CNN-based models were selected: MobileNetV2, EfficientNetB4, DenseNet121, and a custom Sequential CNN. The first three models were chosen for their proven performance in computer vision tasks and were employed using transfer learning. Pre-initialized models trained on expansive visual databases such as ImageNet can be effectively adapted to specialized applications, enabling efficient learning even when only limited domain-specific data is available. This technique reduces training time and improves accuracy, especially in medical imaging where annotated data is limited. For each of these pre-trained models, the layers were in start frozen to reserve the trained features, and a custom classification head was included to accommodate the seven output classes of this task. After initial convergence, selected top layers were unfrozen to enable fine-tuning with a reduced learning rate.

The final model, a custom Sequential CNN, was developed from scratch using the Keras API. This model served as a baseline for evaluating the effectiveness of deep feature reuse in pre-trained architectures. It was constructed with stacked convolutional and pooling layers followed by dense layers and dropout regularization to prevent overfitting. While less complex than the pre-trained networks, the custom CNN provided valuable insights into the performance gap between handcrafted and transfer-learned models.

Each model was configured with the Adam optimization algorithm, selected for its ability to adaptively adjust learning rates and facilitate faster convergence. To effectively handle the multi-class classification problem, categorical cross-entropy was selected as the core loss function during the optimization process. Training of the Model was carried out in small data segments, where the total examples processed at once and the overall training duration were tailored according to available hardware limitations and observed learning stability. A dynamic learning rate adjustment strategy was applied during training using the ReduceLROnPlateau callback, which lowered the learning rate when no improvement in validation loss was observed over successive epochs. Additionally, EarlyStopping and ModelCheckpoint were used for prevention of overfitting and retaining the best weights for the model, respectively.

Each model was trained using the same data pipeline, allowing for a fair comparison of performance across architectures. Evaluation was conducted using standard metrics including recall, accuracy, precision, F1-score, and ROC-AUC, ensuring both class-level and overall performance were taken into account. The next subsections present a detailed breakdown of each architecture used in this study, including its internal structure, training procedure, and observed performance characteristics.

*1) Custom Sequential CNN Model:* In this study, a custom *Sequential Convolutional Neural Network (CNN)* was implemented from scratch using TensorFlow-Keras to serve as a baseline against pre-trained models. The motivation behind this model was to evaluate how well a handcrafted CNN could perform on skin lesion classification without transfer learning.

*a) Architecture Design:* The architecture includes five convolutional blocks with filter sizes of 16, 32, 64, 128, and 256, respectively. Each block is followed by a ReLU activation function, batch normalization, and a max-pooling operation using MaxPooling2D. After convolutional blocks, the network flattens the features and passes them through four fully connected layers of sizes 256, 128, 64, and 32, each with `ReLU` activation, `Dropout`, and `L2 regularization`. The final output layer consists of seven neurons, each corresponding to one of the skin lesion classes, activated by the softmax function:

$$\hat{y}_i = \frac{e^{z_i}}{\sum_{j=1}^{7} e^{z_j}} \quad \text{for } i = 1, \ldots, 7 \tag{1}$$

*b) Compilation and Training:* The model was formed utilizing the `AdamW` optimizer with an early learning rate of 0.0005. The categorical entropy loss was used, stated as:

$$\mathcal{L} = -\sum_{i=1}^{C} y_i \log(\hat{y}_i) \tag{2}$$

where $C = 7$, $y_i$ corresponds to the ground truth label, and $\hat{y}_i$ corresponds to softmax generated probability. The optimizer updated weights using:

$$\theta_{t+1} = \theta_t - \eta \cdot \frac{m_t}{\sqrt{v_t} + \epsilon} \tag{3}$$

Training was performed over 20 iterations(epochs) with 256 batch size. Learning rate warm-up and ReduceLROnPlateau callbacks were used to manage convergence. ModelCheckpoint and EarlyStopping ensured the best model was retained.

*c) Performance and Evaluation:* An accuracy of 54.29% was obtained on the held-out test dataset. The classification report revealed that precision and recall were inconsistent across classes, with better performance on pyogenic granulomas (Precision: 0.98) and poor recall for melanoma (0.25). This highlights the model's limitations in learning high-level features from scratch without a large-scale pre-trained base.

TABLE I
PERFORMANCE METRICS OF SEQUENTIAL CNN MODEL

| Metric | Value |
|---|---|
| Test Accuracy | 54.29% |
| Weighted Precision | 0.61 |
| Weighted Recall | 0.54 |
| Weighted F1-score | 0.53 |
| Optimizer | AdamW |
| Batch Size | 256 |
| Epochs | 20 |
| Trainable Parameters | 51.8M |
| Input Shape | 112×112×3 |

*d) Model Architecture Summary:*

- **Input Layer (112×112×3)**
- Conv2D → MaxPooling → BatchNorm (x5)
- Flatten → Dense(256) → Dense(128) → Dense(64) → Dense(32)
- Output Layer (Dense(7), Softmax)

*e) Conclusion:* This model, though computationally light and easy to implement, demonstrates the limitations of training CNNs from scratch on complex medical data. While it achieved moderate results, its accuracy significantly lagged behind deeper architectures that leverage transfer learning. Nonetheless, it served as a valuable benchmark for this study.

*2) MobileNetV2 Model:* MobileNetV2 was employed as a transfer learning-based model for the classification of skin lesions due to its computational efficiency and high performance on image recognition tasks. It belongs to the MobileNet family of lightweight CNN architectures designed for edge and embedded devices, leveraging depthwise separable convolutions and inverted residual connections [30].

*a) Architecture Overview:* The pre-trained MobileNetV2 base, trained on ImageNet, was used as a feature extractor. Its final layers were excluded by setting `include_top=False`. A custom classification head was added on top, composed of:

- **GlobalAveragePooling2D** to reduce feature maps

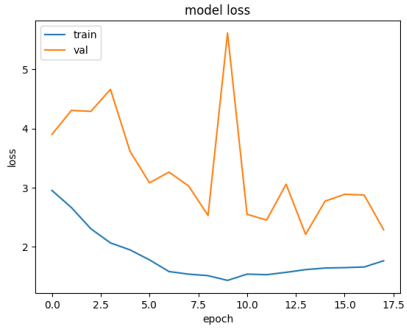


Fig. 4. Validation and Training loss over epochs for Sequential model.

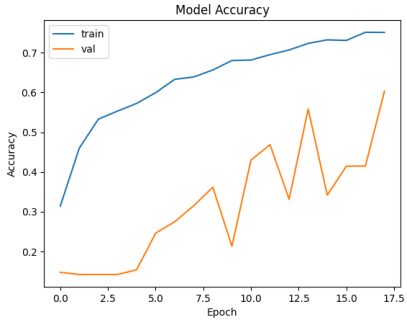


Fig. 5. Training and validation accuracy over epochs for Sequential model.

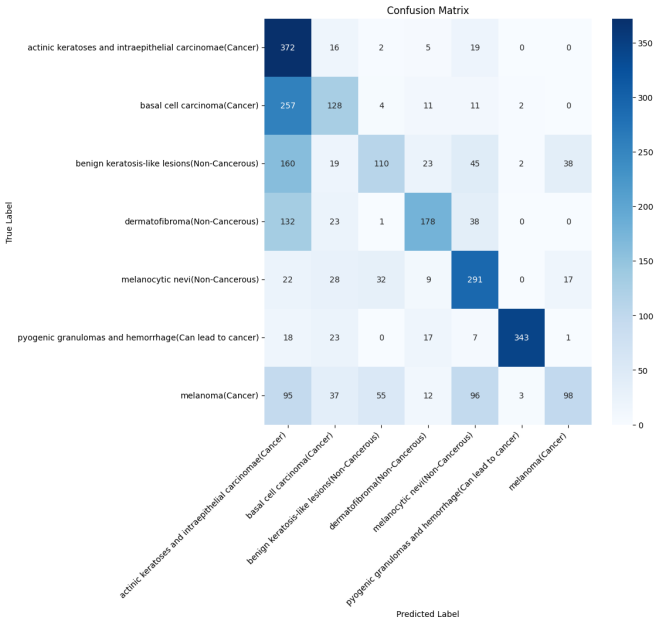


Fig. 6. Visualization of the classification results for Sequential Model on the test set through a confusion matrix.

- **Dense(256)** layer with ReLU activation
- **Dropout(0.3)** for regularization
- **Dense(7)** output layer with softmax activation

*b) Softmax Activation:* The final output layer applies softmax to produce class probabilities:

$$\hat{y}_i = \frac{e^{z_i}}{\sum_{j=1}^{7} e^{z_j}} \quad \text{for } i = 1, \ldots, 7 \tag{4}$$

*c) Training Strategy:* At the start of training, the core layers of the MobileNetV2 backbone were kept non-trainable to preserve pre-learned representations. The model configuration employed an adaptive optimization algorithm for gradient-based learning.:

$$\theta_{t+1} = \theta_t - \eta \cdot \frac{m_t}{\sqrt{v_t} + \epsilon}$$

and loss function used was categorical cross-entropy:

$$\mathcal{L} = -\sum_{i=1}^{C} y_i \log(\hat{y}_i) \tag{5}$$

The training process used:

- Batch Size: 256
- Epochs: 30
- Validation Split: 20%
- Callbacks: EarlyStopping, ReduceLROnPlateau, ModelCheckpoint

*d) Performance Summary:* After 30 iterations (epochs) of training, the model showed an accuracy of 62.89%. The weighted average precision and recall were 0.62 and 0.63 respectively, with particularly high performance observed in the detection of pyogenic granulomas (F1-score: 0.85), and relatively lower performance in benign keratosis and melanoma.

TABLE II
EVALUATION METRICS OF MOBILENETV2 MODEL

| Metric | Value |
|---|---|
| Weighted Precision | 0.62 |
| Test Accuracy | 62.89% |
| Weighted F1-score | 0.62 |
| Weighted Recall | 0.63 |
| Best Class | Pyogenic Granuloma (F1: 0.85) |
| Worst Class | Benign Keratosis (F1: 0.43) |
| Optimizer | Adam |
| Epochs | 30 |
| Batch Size | 256 |
| Trainable Params | 3.5M |

*e) Model Architecture Summary:*

- **Input Layer**: 112×112×3 RGB images
- **Base Model**: MobileNetV2 (frozen, pre-trained on ImageNet)
- **Custom Head**: GlobalAvgPool → Dense(256) → Dropout(0.3) → Dense(7, Softmax)

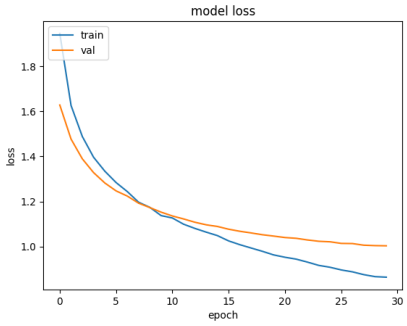


Fig. 7. Training and validation loss over epochs for MobileNetV2.

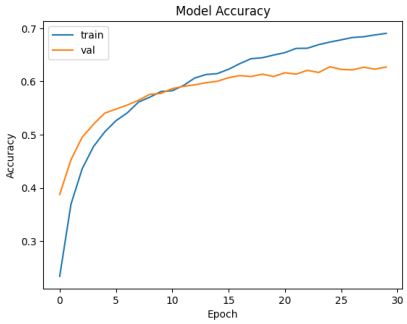


Fig. 8. Training and validation accuracy over epochs for MobileNetV2.

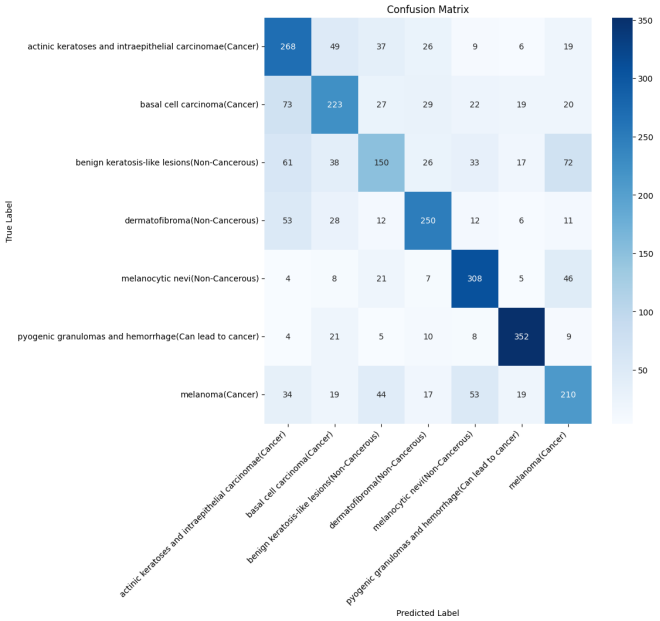


Fig. 9. Confusion Matrix of MobileNetV2 model on test set.

*f) Confusion Matrix and Class-wise Metrics:* The confusion matrix revealed that while the model performed well for classes with distinct features, such as pyogenic granulomas and melanocytic nevi, it showed confusion between benign keratosis and melanoma. This is expected given the visual similarity and class imbalance.

A visualization of the confusion matrix is shown in Fig. 9.

*g) Conclusion:* MobileNetV2 performed significantly better than the baseline Sequential CNN, with over 8% higher accuracy. Its efficient trade-off between computational performance and predictive precision makes it an ideal candidate for integration into real-time medical diagnostic systems, particularly in environments with limited technological infrastructure.

*3) EfficientNetB4 Architecture:* EfficientNetB4 belongs to a class of convolutional architectures that attain high performance by uniformly balancing the network's depth, width, and input resolution through a unified scaling strategy known as the compound coefficient [31]. This uniform scaling approach allows EfficientNet architectures to achieve higher accuracy with fewer parameters and lower computational cost compared to traditional deep CNNs.

In this study, EfficientNetB4 was selected as a core architecture for transfer learning due to its favorable trade-off between model size and predictive performance, particularly when fine-tuned on specialized domains such as medical imaging. To adapt EfficientNetB4 for the specific task of classifying seven types of skin cancer, its original top layers—trained on ImageNet—were excluded by setting `include_top=False`. This modification allowed for the attachment of a custom output block tailored to the classification objectives of this study.

*a) Modified Architecture and Fine-tuning Strategy:* At the beginning of training, the entire EfficientNetB4 backbone was kept non-trainable to preserve the foundational features acquired from prior ImageNet-based pretraining. This allowed model to generalize well to the dermoscopic image dataset without losing its learned visual representations. In the later stages of training, layers starting from index 75 onward were unfrozen to enable fine-tuning. This strategy provided a balance between stability and adaptability, allowing the model to learn task-specific high-level features while preserving the pretrained weights in earlier layers.

The custom classification head consisted of the following layers: a `GlobalAveragePooling2D` layer to reduce spatial dimensions and aggregate feature maps, followed by a fully connected `Dense` layer with 256 neurons using ReLU activation and L2 kernel regularization. To mitigate overfitting particularly due to the constrained size of the dataset a regularization mechanism was employed by applying dropout at a probability rate of 0.5. For output generation, a fully connected layer comprising 7 units was appended, leveraging a softmax function to distribute prediction probabilities across seven unique skin lesion classes.

*b) Mathematical Formulation:* The softmax function applied at the output layer converts the raw logits into normalized class probabilities:

$$\hat{y}_i = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}} \quad \text{for } i = 1, \ldots, C \tag{6}$$

where $C = 7$ is the number of output classes, and $z_i$ is the logit output for class $i$.

The model's learning objective was formulated using the categorical cross-entropy loss function, commonly applied in multi-class classification tasks:

$$\mathcal{L} = -\sum_{i=1}^{C} y_i \log(\hat{y}_i) \tag{7}$$

int the expression $y_i$ denotes the actual class labels $\hat{y}_i$ denotes corresponding predicted softmax distribution probability.

The optimization strategy relied on the Adam algorithm, which dynamically adjusts the update step by leveraging moving averages of both the gradient and its squared values to stabilize convergence.

$$\theta_{t+1} = \theta_t - \eta \cdot \frac{m_t}{\sqrt{v_t} + \epsilon} \tag{8}$$

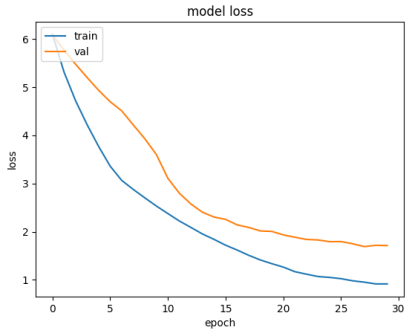


Fig. 10. Training and validation loss over epochs for EfficientNetB4.

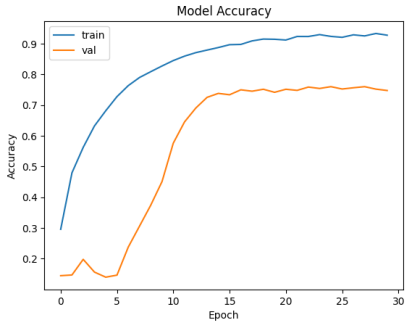


Fig. 11. Training and validation accuracy over epochs for EfficientNetB4.

*c) Training Configuration and Hyperparameters:* A total of 30 training cycles were executed with groups of 256 inputs per batch, starting with a learning rate configured at 0.0001 to guide the early stages of model tuning. Learning rate reduction and Early stopping strategies were implemented using the `EarlyStopping` and `ReduceLROnPlateau` callbacks, respectively. These techniques helped prevent overfitting and allowed the model to converge more effectively. Additionally, the `ModelCheckpoint` callback was used to save the best-performing model during training, based on validation accuracy.

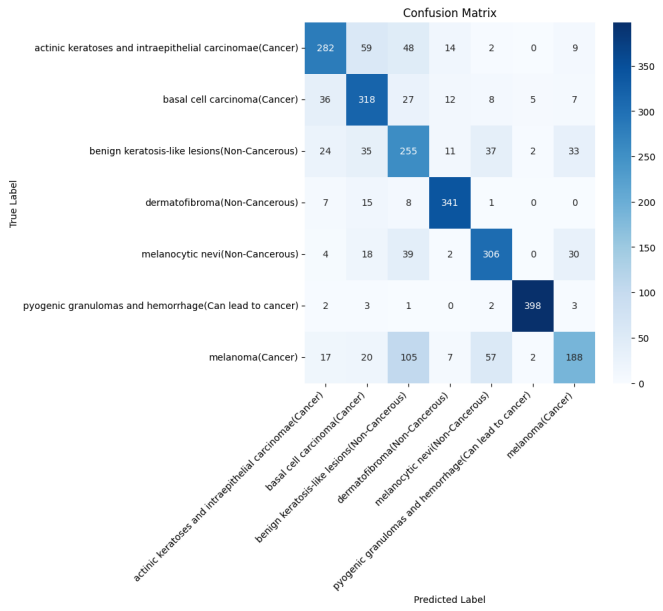


Fig. 12. Visualization of the classification results for EfficientNetB4 on the test set through a confusion matrix.

*d) Layer-wise Architecture Summary:* The final architecture of the EfficientNetB4-based model can be summarized as follows:

- **Input**: RGB image of shape 112×112×3
- **Feature Extractor**: EfficientNetB4 (pretrained on ImageNet)
- **Unfreezing Strategy**: Layers from index 75 onward set to trainable
- **Classification Head**:
  - GlobalAveragePooling2D
  - Dense(256, ReLU, L2-Regularized)
  - Dropout(0.5)
  - Dense(7, Softmax)

*e) Evaluation and Performance Analysis:* Performance assessment was conducted using a separate evaluation set comprising 2800 dermoscopic samples. Our model recorded an accuracy of 75.00%, with corresponding weighted recall, precision, and F1-score metrics and averaging close to 0.75. These results represent a significant improvement over the baseline Sequential CNN (54%) and MobileNetV2 (62.89%), showcasing EfficientNetB4's ability to extract richer semantic features and generalize better.

TABLE III
PERFORMANCE METRICS OF EFFICIENTNETB4 MODEL

| Metric | Value |
|---|---|
| Weighted Recall | 0.75 |
| Weighted Precision | 0.75 |
| Test Accuracy | 75.00% |
| Weighted F1-score | 0.75 |
| Support (Test Samples) | 2800 |
| Best Performing Class | Pyogenic Granuloma (F1-score: 0.89) |
| Weakest Performing Class | Melanoma (F1-score: 0.38) |
| Trainable Parameters | ˜17 Million |
| Batch Size | 256 |
| Dropout Rate | 0.5 |
| Epochs | 30 |
| Fine-tuned Layers | From layer index 75 onward |

*f) Qualitative Observations:* The confusion matrix generated from model predictions revealed strong separability between pyogenic granulomas and other classes due to their visually distinct features. However, notable misclassifications occurred between melanoma and benign keratosis, reflecting the clinical challenge of distinguishing between morphologically similar lesions. ROC-AUC plots and accuracy/loss curves showed stable convergence and no signs of overfitting, validating the use of dropout and learning rate scheduling.

*g) Conclusion:* EfficientNetB4 demonstrated strong performance in this study, outperforming simpler CNN baselines and MobileNetV2 by a notable margin. Its ability to generalize well, even with limited training data, highlights the effectiveness of transfer learning in medical image analysis. Given its moderate size and high accuracy, EfficientNetB4 presents a viable model for real-world diagnostic deployment in skin cancer screening systems, especially when integrated with telemedicine platforms.

*4) DenseNet121 Architecture:* The DenseNet121 model, originally developed by Huang and collaborators, represents an advanced deep convolutional framework known for its efficiency and performance in visual recognition tasks. [32], which employs dense connectivity between layers. In contrast to traditional CNNs where each layer receives input only from the immediate preceding layer, In DenseNet, each layer receives input from all earlier layers in a sequential feed-forward configuration. This dense linkage improves gradient flow and promotes reuse of learned features, effectively boosting accuracy while minimizing the total number of parameters.

In the present study, DenseNet121 was selected as the backbone for transfer learning in the classification of skin lesions. The model was initialized with ImageNet pre-trained weights, and its original classification head was removed using the `include_top=False` parameter. A custom classification head suitable for the seven-class skin lesion task was appended. All dermoscopic images were resized to $112 \times 112 \times 3$ before being fed into the model.

*a) Custom Classification Head and Layer Freezing:* The model was customized by adding a

`GlobalAveragePooling2D` layer to flatten the feature maps without introducing additional parameters. Next, a fully connected layer consisting of 512 units was introduced, activated by the ReLU function and enhanced with L2 regularization to reduce the risk of overfitting. A `Dropout` layer with a rate of 0.5 was included to deactivate neurons randomly during training, and finally, To generate predictions across all classes, a final `Dense(7)` layer with softmax activation was appended at the output stage.

Initially, first 100 layers of the DenseNet121 base model were frozen to retain generic features learned from ImageNet. Fine-tuning was selectively enabled by unfreezing the remaining layers, allowing the model to adapt to dermoscopic features specific to skin lesions. This approach balances the benefits of transfer learning while allowing the model to specialize in the medical domain.

*b) Mathematical Formulations:* The softmax function was applied to the final layer to obtain the class probabilities:

$$\hat{y}_i = \frac{e^{z_i}}{\sum_{j=1}^{7} e^{z_j}} \quad \text{for } i = 1, \ldots, 7 \tag{9}$$

Categorical cross-entropy was employed as the loss function during model training to address the multi-class nature of the task:

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

where $C = 7$ is the number of classes, $y_i$ is the true label, and $\hat{y}_i$ is the predicted probability.

The `Adam` optimizer is utilized to reduce the loss, with an initial learning rate set to 0.001:

$$\theta_{t+1} = \theta_t - \eta \cdot \frac{m_t}{\sqrt{v_t} + \epsilon} \tag{11}$$

TABLE IV
PERFORMANCE METRICS OF DENSENET121 MODEL

| Metric | Value |
|---|---|
| Weighted Precision | 0.88 |
| Weighted Recall | 0.89 |
| Test Accuracy | 88.64% |
| Weighted F1-score | 0.885 |
| Test Samples | 2800 |
| Best Performing Class | Pyogenic Granuloma (F1: 0.93) |
| Weakest Performing Class | Melanoma (F1: 0.66) |
| Frozen Layers | First 100 |
| Trainable Parameters | ∼8M |
| Dropout Rate | 0.5 |
| Epochs | 20 |

*c) Training Configuration:* The model was trained for 20 epochs with a batch size of 256. The learning rate was adaptively reduced using the `ReduceLROnPlateau` callback based on validation loss stagnation. `EarlyStopping` was implemented to avoid overfitting by halting training when no improvement was observed, and `ModelCheckpoint` saved the best performing model during training.

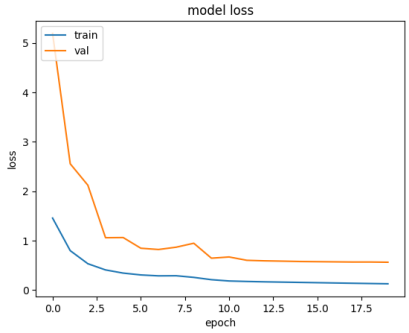


Fig. 13. Training and validation loss over epochs for DenseNet121.

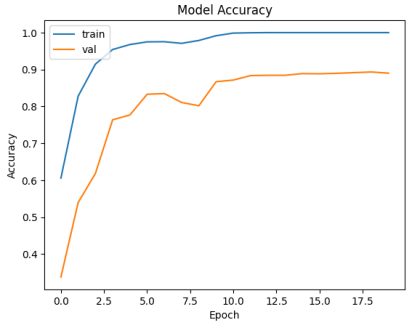


Fig. 14. Training and validation accuracy over epochs for DenseNet121.

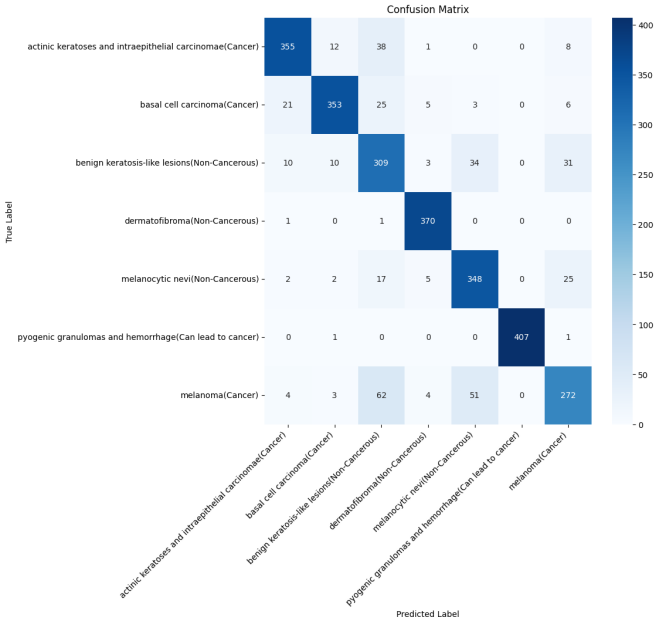


Fig. 15. Visualization of the classification results for DenseNet121 on the test set through a confusion matrix.

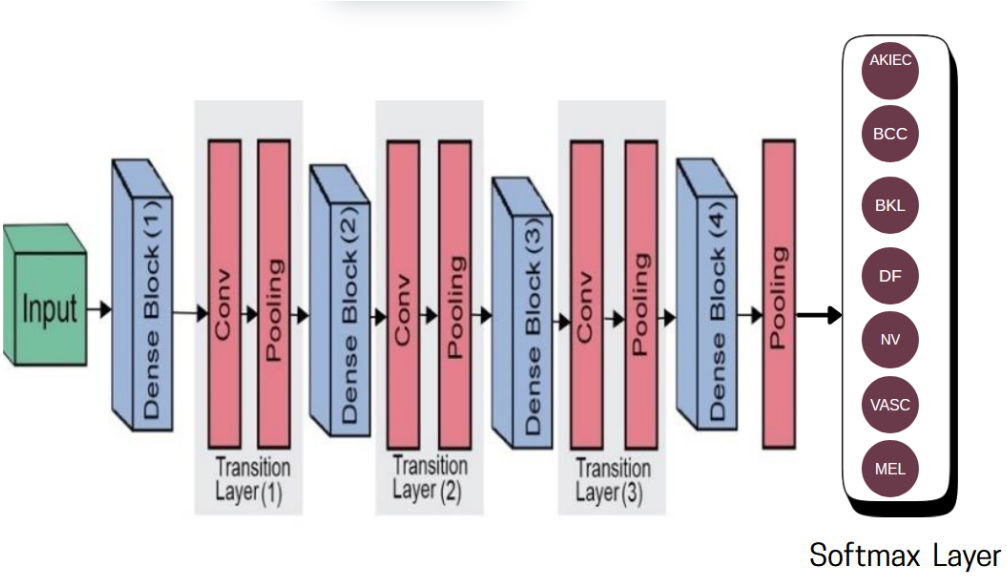


Fig. 16. The DenseNet-121 Architecture, which Consists of Four Dense Block Layers and Three Transition Layers

**Summary of training setup:**

- Optimizer: Adam
- Epochs: 20
- Batch size: 256
- Frozen layers: First 100 layers
- Initial learning rate: 0.001
- Callbacks: EarlyStopping, ReduceLROnPlateau, ModelCheckpoint
- Input shape: $112 \times 112 \times 3$

*d) Architecture Flow:* The final architecture was structured as follows:

- Input Image: $112 \times 112 \times 3$
- Base Model: DenseNet121 (pretrained on ImageNet)
- Freeze: First 100 layers frozen
- Custom Head:
  - GlobalAveragePooling2D
  - Dense(512, ReLU, L2-regularized)
  - Dropout(0.5)
  - Dense(7, Softmax)

*e) Performance Evaluation:* After training, the model achieved a test accuracy of **88.64%**, with a weighted precision and recall of 0.88 and 0.89 respectively. The F1-score was approximately 0.885, indicating balanced performance across all classes. The model outperformed all other evaluated architectures including MobileNetV2, EfficientNetB4, and the baseline Sequential CNN.

*f) Visual Insights:* The confusion matrix confirmed high class-separability for distinct lesion types such as pyogenic granuloma. However, misclassification was observed between melanoma and benign keratosis, a common challenge in dermatology due to visual similarity. Training/validation accuracy curves and loss plots demonstrated smooth convergence with no signs of overfitting, validating the effectiveness of dropout and learning rate control.

*g) Conclusion:* DenseNet121 provided the highest performance among all tested models in this study. Its densely connected architecture enabled it to learn rich hierarchical representations, resulting in reliable classification of complex dermoscopic features.With performance approaching 89% accuracy on test data, this framework stands as a promising solution for real-world applications in automated skin cancer screening systems..

## IV. CONCLUSION

Skin cancer, particularly melanoma, is a growing public health concern that requires timely and accurate diagnosis for effective treatment. However, traditional diagnostic approaches such as biopsy and manual dermoscopic analysis can be time-consuming, costly, and often unavailable in remote or resource-constrained environments. This study presents a deep learning-based approach for automated skin cancer classification, employing transfer learning on dermoscopic images using four different convolutional neural network architectures: a custom Sequential CNN, MobileNetV2, EfficientNetB4, and DenseNet121.

A key strength of this work lies in its comparative model evaluation pipeline. Rather than relying on a single architecture, we implemented and evaluated multiple state-of-the-art models to determine the most effective approach for skin lesion classification. Each architecture was trained using consistent preprocessing, augmentation, and evaluation techniques, allowing for a fair comparison based on quantitative performance metrics. Our findings demonstrated that DenseNet121 achieved the highest test accuracy (88.64%), significantly outperforming the baseline Sequential CNN and

TABLE V
COMPREHENSIVE COMPARISON OF ALL IMPLEMENTED MODELS WITH ARCHITECTURAL, TRAINING, AND EVALUATION DETAILS

| Model | Architecture Type | Pretrained | Frozen Layers | Dropout | L2 Reg | Learning Rate | Epochs | Accuracy (%) |
|---|---|---|---|---|---|---|---|---|
| Sequential CNN | Custom from scratch | No | None | 0.3 | No | 0.0005 | 20 | 54.00 |
| MobileNetV2 | Transfer Learning | Yes (ImageNet) | All except last 75 | 0.3 | Yes | 0.0001 | 30 | 62.89 |
| EfficientNetB4 | Transfer Learning | Yes (ImageNet) | All except last 75 | 0.5 | Yes | 0.0001 | 30 | 75.00 |
| DenseNet121 | Transfer Learning | Yes (ImageNet) | First 100 | 0.5 | Yes | 0.001 | 20 | **88.64** |

even other well-established models like EfficientNetB4 and MobileNetV2.

Unlike many existing studies that focus narrowly on a single architecture or rely on large datasets with minimal preprocessing analysis, our project emphasizes a complete end-to-end pipeline. This includes thoughtful data augmentation strategies, proper image normalization, and consistent architectural tuning across all models. Furthermore, by freezing and fine-tuning different portions of each pretrained network, we optimized their transfer learning potential for the domain-specific task of skin lesion recognition. The experimental results validate that architectural depth, feature reuse (as in DenseNet), and fine-tuning of layers play a critical role in improving classification performance on medical image data.

This project is unique in its focus on both implementation depth and academic rigor. We not only built a practical classification system but also analyzed layer-wise behavior, evaluation metrics, and misclassification trends using confusion matrices. Our work can be easily adapted and deployed as a diagnostic decision support system, potentially integrated into telemedicine platforms. This makes it highly useful in regions where access to dermatologists is limited, enabling early detection of potentially life-threatening lesions through automated screening.

To conclude, the proposed work offers a scalable and effective deep learning framework tailored for the automated detection of skin cancer. The relative analysis of CNN architectures under consistent preprocessing and evaluation conditions offers insights into model suitability for medical image classification. The highest-performing model, DenseNet121, shows considerable promise for real-world applications, supporting the feasibility of AI-assisted dermatology. Future enhancements could include incorporating attention mechanisms, ensemble methods, and explainable AI techniques to further boost performance and interpretability.

## ACKNOWLEDGEMENT

We would like to express their deepest gratitude to **Prof. Sabyasachi Patra**, Department of Computer Science and Engineering, IIIT Bhubaneswar, for his constant guidance, invaluable insights, and unwavering support throughout the course of this work.

This research would not have been possible without the continuous supervision and mentorship of **Prof. Sabyasachi Patra**, whose technical expertise, constructive criticism, and academic rigor laid the foundation for every stage of this study — from conceptualization to final validation.

His encouragement and critical feedback inspired us to explore deeper layers of understanding in deep learning for medical image analysis. We consider it a privilege and honor to have conducted this work under his supervision and remain truly grateful for the opportunity.

## REFERENCES

[1] World Health Organization, 'Cancer," *WHO Fact Sheets*, 2022. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cancer [Accessed: Apr. 16, 2025].

[2] World Health Organization, "Skin cancers," Fact Sheet, 2022. [Online]. Available: https://www.iarc.who.int/cancer-type/skin-cancer/. [Accessed: Apr. 21, 2025].

[3] World Health Organization, "Ultraviolet (UV) radiation and skin cancer," WHO, [Online]. Available: https://www.who.int/news-room/questions-and-answers/item/radiation-ultraviolet-%28uv%29-radiation-and-skin-cancer [Accessed: Apr. 19, 2025].

[4] A. Baum, S. J. Elkin, and M. B. Jones, 'Challenges and opportunities in early skin cancer detection," *Dermatology Online Journal*, vol. 27, no. 3, pp. 1–10, 2021. [Online]. Available: https://ijdvl.com/artificial-intelligence-in-dermatology-and-healthcare-an-overview/

[5] Skin Cancer Foundation, "Types of skin cancer," 2023. [Online]. Available: https://www.skincancer.org/skin-cancer-information/

[6] Cancer Research UK, "Melanoma skin cancer statistics," 2022. [Online]. Available: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/melanoma-skin-cancer

[7] American Cancer Society, "Key statistics for skin cancer," 2024. [Online]. Available: https://www.cancer.org/cancer/skin-cancer/about/key-statistics.html

[8] A. B. Nassif, C. S. Ho, and T. A. Fattah, "Automatic Diagnosis of Skin Cancer Using Deep Convolutional Neural Networks," *Journal of Healthcare Engineering*, vol. 2021, Article ID 5644379, pp. 1–11, 2021. [Online]. Available: https://www.researchgate.net/publication/358519182_Automated_Detection_of_Nonmelanoma_Skin_Cancer_Based_on_Deep_Convolutional_Neural_Network

[9] American Academy of Dermatology Association, "Skin cancer," 2024. [Online]. Available: https://www.aad.org/public/diseases/skin-cancer

[10] American Cancer Society, "What is melanoma skin cancer?" 2024. [Online]. Available: https://www.cancer.org/cancer/melanoma-skin-cancer/about/what-is-melanoma.html

[11] A. Esteva, B. Kuprel, R. A. Novoa, et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, pp. 115–118, 2017. [Online]. Available: https://www.nature.com/articles/nature21056

[12] M. Goyal, A. Oakley, R. Bansal, and H. Yap, "Skin lesion segmentation in dermoscopic images with ensemble deep learning methods," *IEEE Access*, vol. 8, pp. 4171–4181, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/8944763

[13] R. C. Maron, et al., "Systematic review of artificial intelligence for melanoma diagnosis in dermatoscopic images," *European Journal of Cancer*, vol. 156, pp. 121–131, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959804921001014

[14] T. J. Brinker, A. Hekler, A. H. Enk, et al., "A convolutional neural network trained with dermoscopic images performs on par with 145 dermatologists in a melanoma classification task," *European Journal of Cancer*, vol. 178, pp. 47–55, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0959804922002167

[15] M. Goyal, A. Oakley, R. Bansal, and H. Yap, "Skin lesion segmentation in dermoscopic images with ensemble deep learning methods," *IEEE Access*, vol. 8, pp. 4171–4181, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/8944763

[16] M. P. Salinas, J. Sepúlveda, L. Hidalgo, D. Peirano, M. Morel, P. Uribe, V. Rotemberg, J. Briones, D. Mery, and C. Navarrete-Dechent, "A systematic review and meta-analysis of artificial intelligence versus clinicians for skin cancer diagnosis," *npj Digital Medicine*, vol. 7, no. 1, 2024. [Online]. Available: https://www.nature.com/articles/s41746-024-01103-x

[17] T. J. Brinker, et al., "A convolutional neural network trained with dermoscopic images performs on par with 145 dermatologists in a melanoma classification task," *European Journal of Cancer*, vol. 178, pp. 47–55, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S095980492100715X

[18] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015. [Online]. Available: https://www.nature.com/articles/nature14539

[19] N. Codella, et al., "Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the International Skin Imaging Collaboration (ISIC)," *arXiv preprint arXiv:1902.03368*, 2018. [Online]. Available: https://arxiv.org/abs/1902.03368

[20] A. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," *Neurocomputing*, vol. 321, pp. 321–331, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0925231218311102

[21] I. Goodfellow, Y. Bengio, and A. Courville, "Deep Learning," MIT Press, 2016. [Online]. Available: https://www.deeplearningbook.org/

[22] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998. [Online]. Available: https://ieeexplore.ieee.org/document/726791

[23] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 25, pp. 1097–1105, 2012. [Online]. Available: https://papers.nips.cc/paper_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html

[24] P. Tschandl, C. Rosendahl, and H. Kittler, 'The HAM10000 dataset: A large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific Data*, vol. 5, no. 180161, Aug. 2018. Available: https://doi.org/10.1038/sdata.2018.161

[25] N. Codella, D. Gutman, M. Emre Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler, and A. Halpern, 'Skin lesion analysis toward melanoma detection: A challenge at the International Skin Imaging Collaboration (ISIC) 2018," *arXiv preprint arXiv:1902.03368*, 2019. Available: https://www.isic-archive.com

[26] A. Mahbod, G. Schaefer, R. Ecker, A. Ellinger, and C. Wang, "Fusing fine-tuned deep features for skin lesion classification," *Computerized Medical Imaging and Graphics*, vol. 71, pp. 19–29, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0895611118306050?utm_source=chatgpt.com

[27] P. Chang, T. Chow, and J. Chiu, "Skin lesion analysis using deep learning and ResNet architectures," *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2018. [Online]. Available: https://ieeexplore.ieee.org/document/8616270

[28] U. Dorj, K. Lee, J. Choi, and M. Lee, "The skin cancer classification using deep convolutional neural network," *Multimedia Tools and Applications*, vol. 77, pp. 9909–9924, 2018. [Online]. Available: https://link.springer.com/article/10.1007/s11042-017-5455-2

[29] S. Kalouche, "Skin Lesion Segmentation Using Deep Learning," *Stanford University CS231n Project*, 2016. [Online]. Available: http://cs231n.stanford.edu/reports/2016/pdfs/223_Report.pdf

[30] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 4510–4520. [Online]. Available: https://arxiv.org/abs/1801.04381

[31] M. Tan and Q. Le, 'EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. on Machine Learning (ICML)*, 2019, pp. 6105–6114. [Online]. Available: https://arxiv.org/abs/1905.11946

[32] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, 'Densely Connected Convolutional Networks," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4700–4708. [Online]. Available: https://arxiv.org/abs/1608.06993

[33] D. Shen, G. Wu, and H.-I. Suk, 'Deep learning in medical image analysis: A survey," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1299, May 2017. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC5479722/

[34] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez, 'A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, Dec. 2017. Available: https://doi.org/10.1016/j.media.2017.07.005

[35] Y. Yuan, M. Chao, and Y.-C. Lo, 'Automatic skin lesion segmentation using deep fully convolutional networks with Jaccard distance," *IEEE Transactions on Medical Imaging*, vol. 36, no. 9, pp. 1876–1886, Sept. 2017. Available: https://doi.org/10.1109/TMI.2017.2695227

[36] T. DeVries and G. W. Taylor, 'Skin lesion classification using deep learning and clinical knowledge," in *Proc. Int. Conf. on Machine Learning (ICML) Workshop on Skin Lesion Analysis Toward Melanoma Detection*, 2017. Available: https://arxiv.org/abs/1703.00518