Skin Cancer Detection (Melanoma Classification)

Authors:

Vinyas Naidu Karri, karri.vi@northeastern.edu

Sri Sai Tarun Vemu, vemu.s@northeastern.edu

Ramappa, Sathvik, ramappa.s@northeastern.edu

1.Abstract

This project investigates the application of deep learning in skin cancer detection, specifically melanoma classification using dermoscopic images. Leveraging the ISIC Archive dataset, the study developed Convolutional Neural Network (CNN) models and utilized transfer learning with architectures such as EfficientNet and DenseNet. Data preprocessing and augmentation were employed to address the challenges posed by a highly imbalanced dataset and enhance model generalization. The models were evaluated using metrics like accuracy, precision, and recall. Results demonstrate that DenseNet121 achieved superior performance in distinguishing malignant from benign lesions, highlighting the potential of AI-driven diagnostic tools to support early and accurate melanoma detection. The study underscores the importance of addressing class imbalance and proposes future enhancements including synthetic data generation, cost-sensitive learning, and explainability frameworks for clinical adoption.

2.Introduction

2.1 Background

Skin cancer is one of the most prevalent forms of cancer worldwide, with melanoma being its most dangerous variant. Early detection significantly improves treatment outcomes, but traditional diagnostic methods require expert dermatological analysis, which is time-consuming and subjective. Advancements in digital imaging and deep learning have opened new avenues for automated and accurate diagnosis of skin cancer using dermoscopic images.

2.2 Motivation

The complexity and high variability of skin lesions present challenges in achieving reliable diagnoses. Many healthcare datasets suffer from class imbalance, further complicating the performance of machine learning models. A deep learning-based approach, utilizing CNNs, can not only address these challenges but also assist dermatologists by providing reliable preliminary classifications, potentially reducing diagnostic time and errors.

2.3 Objectives

The main objective of this project is to develop a deep learning model capable of classifying dermoscopic images of skin lesions into benign or malignant categories. Specifically, the goals are:

1. To preprocess and explore the ISIC Archive dataset for melanoma detection.
2. To design and train a CNN-based model for binary classification of skin lesions.
3. To evaluate model performance using appropriate metrics, including sensitivity and specificity.
4. To explore transfer learning techniques using pre-trained architectures like EfficientNet and DenseNet for enhanced generalization.
5. To provide insights into the practical application of deep learning in clinical dermatology.

2.4 Our Approach

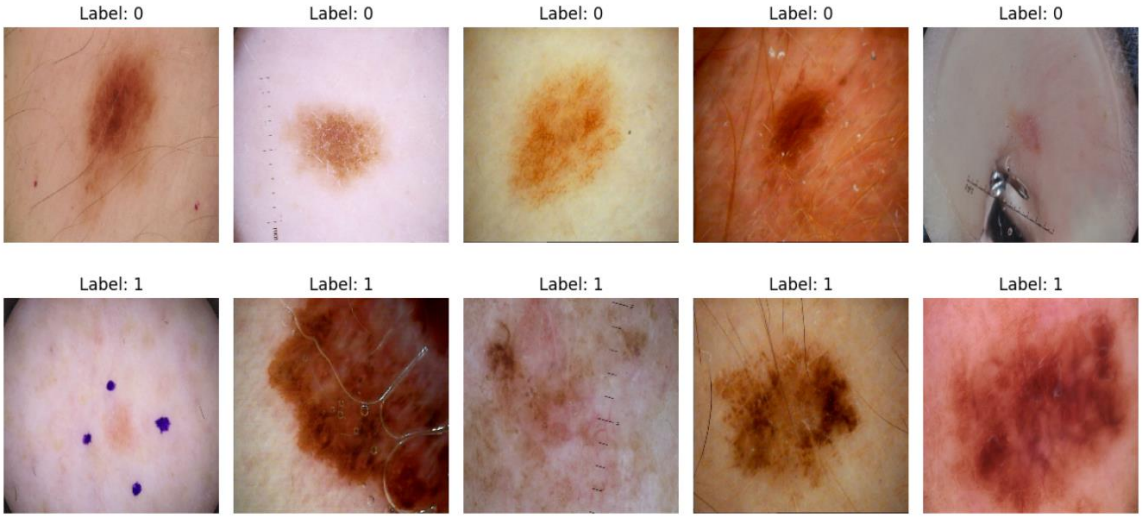
The project focused on developing a robust CNN-based model to classify dermoscopic images as benign or malignant, using the ISIC Archive dataset. Preprocessing included resizing, normalization, and data augmentation (e.g., flipping, rotation, zooming) to address class imbalance and improve generalization. A baseline CNN model was developed, followed by transfer learning with pre-trained architectures like EfficientNet and DenseNet to enhance accuracy and reduce training time. The models were trained using the Adam optimizer and evaluated with metrics such as accuracy, precision, recall, sensitivity, and specificity. Regularization techniques, including dropout and early stopping, mitigated overfitting.

Performance trends were visualized using accuracy-loss plots and confusion matrices, providing insights into the model's effectiveness and areas for improvement.

2.5 Dataset for Experiments and Evaluation

The ISIC Archive dataset served as the primary data source for this project, providing 33,126 dermoscopic images of unique benign and malignant skin lesions from over 2,000 patients. Each image in the dataset is linked to a unique patient identifier, ensuring traceability and dataset integrity. Malignant diagnoses were verified via histopathology, while benign cases were confirmed through expert agreement, longitudinal follow-up, or histopathology.

The dataset, curated by the International Skin Imaging Collaboration (ISIC), is a comprehensive resource for dermatological research. It includes images sourced from renowned institutions such as the Hospital Clínic de Barcelona, Medical University of Vienna, Memorial Sloan Kettering Cancer Center, Melanoma Institute Australia, University of Queensland, and the University of Athens Medical School. This diverse collection ensures a wide representation of lesion types, patient demographics, and anatomical locations, making it a robust foundation for developing and evaluating melanoma classification models.



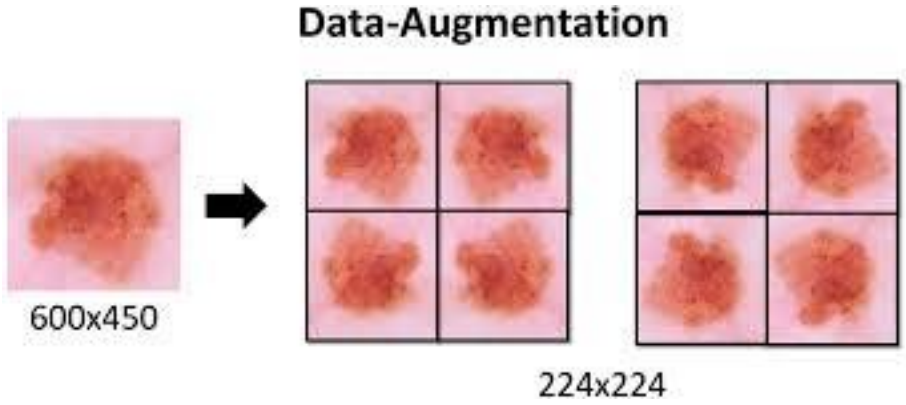
3. Methodology:

3.1 Model Development

The approach centered on designing a robust Convolutional Neural Network (CNN)-based deep learning model to perform binary classification of dermoscopic images into benign and malignant categories. The ISIC Archive dataset, consisting of thousands of labeled skin lesion images, was utilized as the primary data source for training and evaluation.

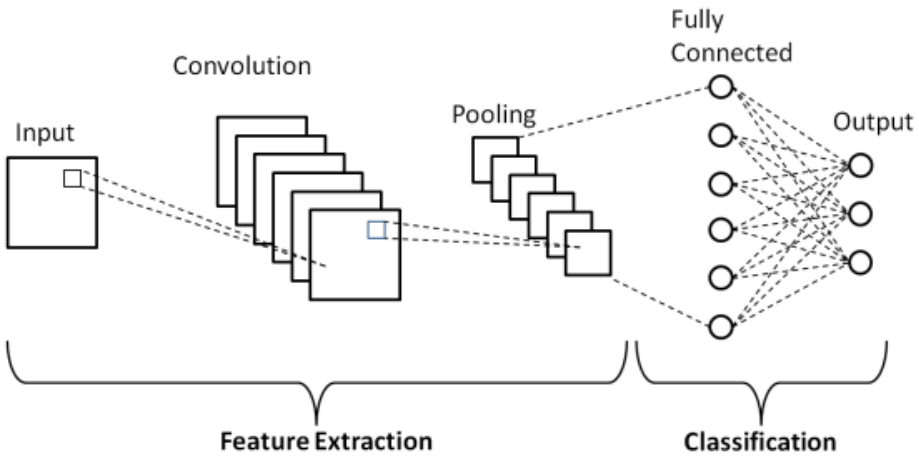
3.2 Data Preprocessing and Augmentation

Preprocessing steps were implemented to ensure data consistency and optimize model performance. These steps included resizing images to a standard input size, normalization to scale pixel values, and applying data augmentation techniques. Augmentation strategies such as flipping, rotation, and zooming were employed to address class imbalance and increase the diversity of the training set. These techniques enhanced the model's ability to generalize across variations in skin lesions, including differences in size, orientation, and color.



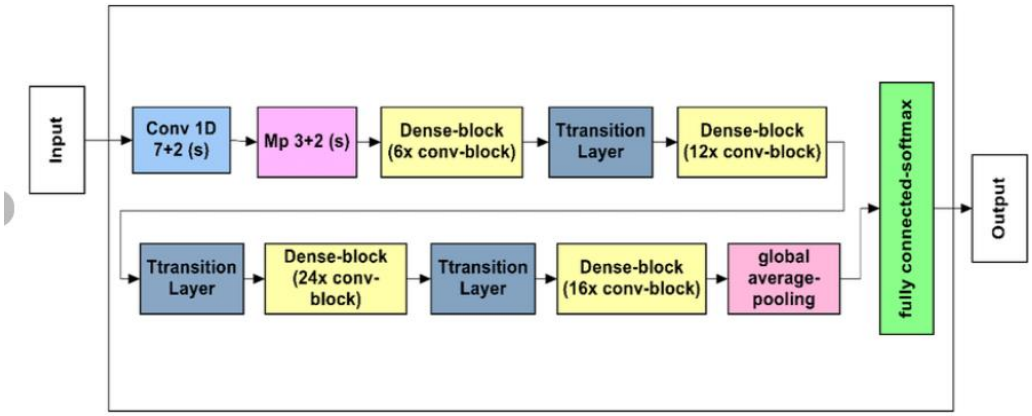
3.3 Baseline Model

An initial CNN-based model was built from scratch, incorporating convolutional layers, max-pooling, and fully connected layers. This model served as a baseline to benchmark the performance before implementing more advanced methods.

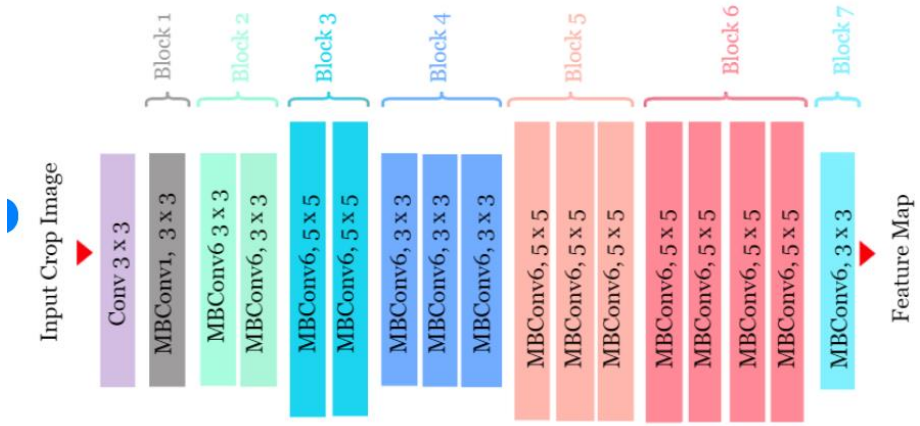


3.4 Transfer Learning

To improve accuracy and generalization, transfer learning was employed by fine-tuning pre-trained models such as EfficientNet and DenseNet. These architectures, which had already been trained on large datasets, provided a strong foundation for melanoma classification. Fine-tuning involved updating the weights of selected layers while preserving the learned features from the original dataset. This approach significantly reduced training time while boosting performance.



DenseNet121 architecture.



4: EfficientnetB0 Model Architecture

3.5 Training and Optimization

The models were trained using a stratified split of the dataset to ensure balanced representation of classes in the training, validation, and test sets. The Adam optimizer and sparse categorical cross-entropy loss function were used to train the models over a specified number of epochs. Various hyperparameters, such as learning rate and batch size, were fine-tuned to achieve optimal performance. Regularization techniques, including dropout and early stopping, were also applied to mitigate overfitting.

3.6 Evaluation Metrics

The models were evaluated using a comprehensive set of metrics, including:

- Accuracy: Overall correctness of the model in classifying lesions.
- Precision and Recall: Specific focus on the model's performance for malignant cases.

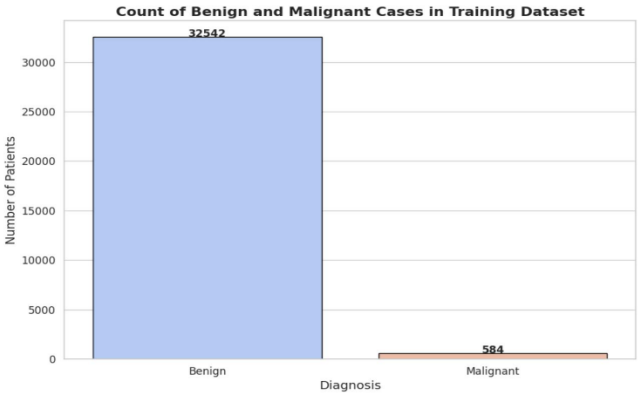
3.7 Visualization and Insights

Training and validation performance were visualized through plots of accuracy and loss over epochs. These visualizations provided insights into the model’s learning behavior,

highlighting areas for further improvement. Additional visualizations, such as confusion matrices, were used to analyze the model's performance in detail.

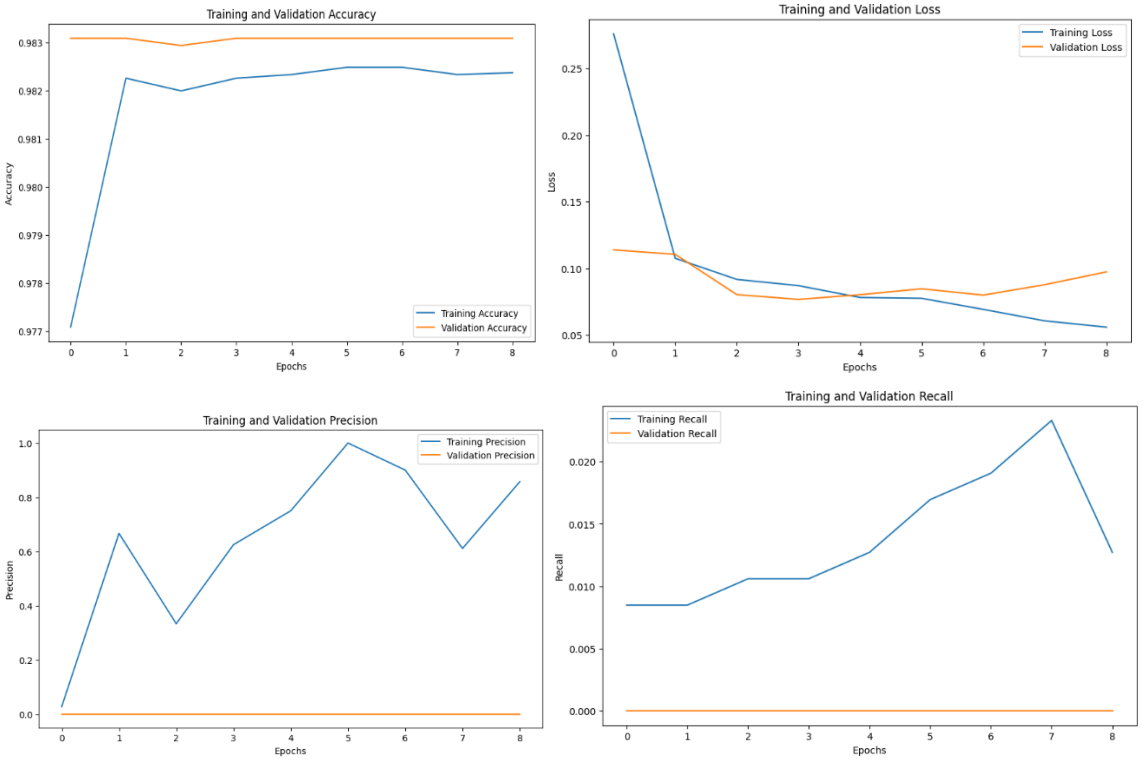
5.Results:

5.1 Imbalance Data



5.2 Training and Validation Performance of models

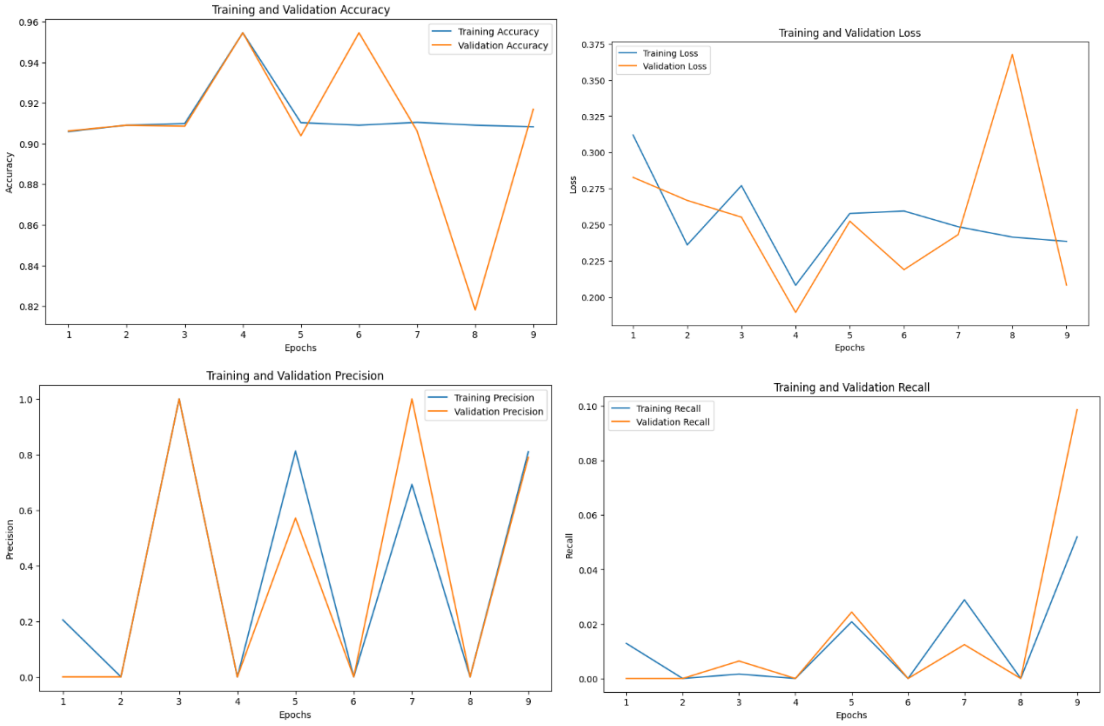
Baseline CNN Model



The baseline CNN provided a benchmark for evaluation.

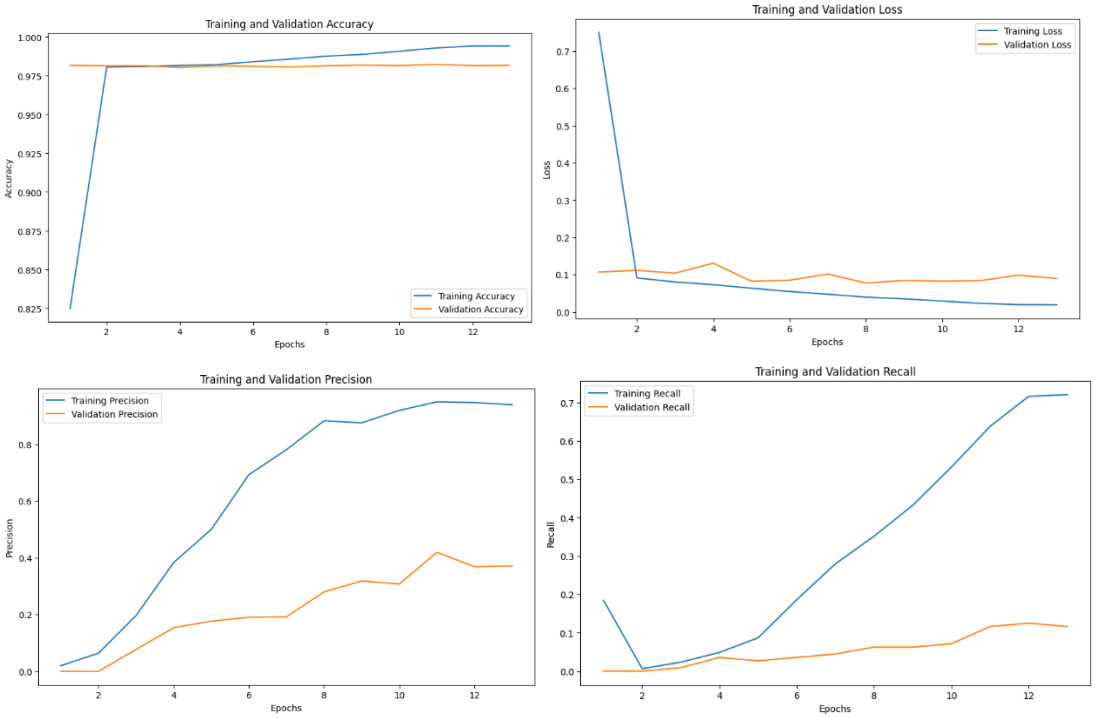
- Accuracy & Loss: The model achieved reasonable accuracy but struggled slightly with generalization, as seen in validation accuracy and loss trends.
- Precision & Recall: Performance was limited, particularly for the minority malignant class, reflecting challenges with dataset imbalance.

CNN after Augmentation



- **Improved Accuracy & Loss**: Data augmentation enhanced generalization, leading to higher validation accuracy and reduced validation loss, minimizing overfitting.
- **Better Precision & Recall**: Augmentation significantly improved sensitivity to malignant lesions, addressing class imbalance effectively.

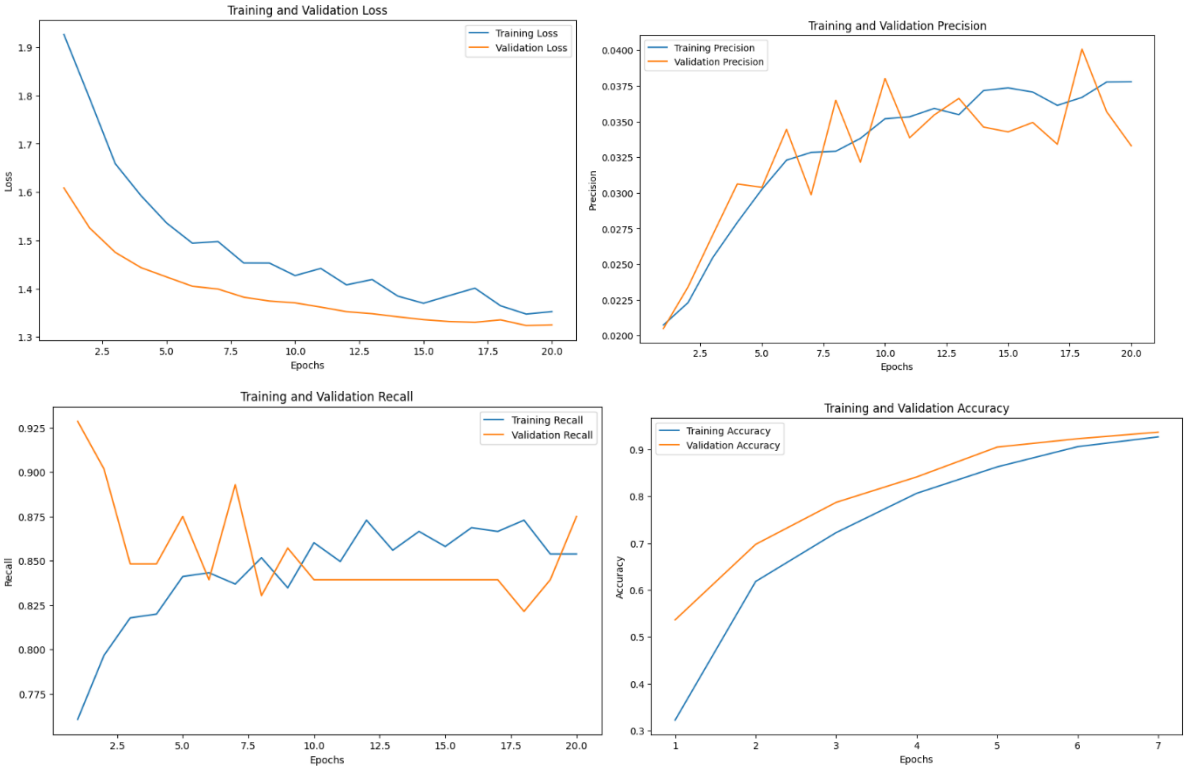
EfficientNetB0 (Finetuned)



EfficientNetB0, fine-tuned for the task, significantly improved performance.

- Accuracy & Loss: Near-perfect training accuracy with high validation accuracy and low validation loss.
- Precision & Recall: Effectively addressed class imbalance, achieving high sensitivity for malignant cases.

DenseNet121(Finetuned)



DenseNet121 outperformed other models.

- Accuracy & Loss: Slightly better validation accuracy than EfficientNetB0, with stable and consistent loss.
- Precision & Recall: Achieved robust detection of malignant lesions, balancing sensitivity and specificity.

6.Discussion:

6.1 Interpretations

The results demonstrated the effectiveness of deep learning models in addressing the challenges of melanoma classification. Transfer learning models like EfficientNetB0 and DenseNet121 significantly outperformed the baseline CNN, showcasing their ability to extract complex features and generalize well to unseen data. Despite the imbalance in the dataset, these models achieved high precision and recall for malignant lesions, a critical factor in clinical settings.

6.2 Challenges

The primary challenge was the highly imbalanced dataset, with benign lesions significantly outnumbering malignant cases. This imbalance led to a natural bias in predictions toward the majority class, complicating the detection of malignant cases. Standard training approaches often failed to adequately prioritize the minority class, impacting recall and overall model reliability. Data augmentation and loss weighting partially mitigated these issues, but achieving a perfect balance remains challenging.

6.3 Implications

The study highlights the necessity of addressing class imbalance in medical datasets to ensure reliable and sensitive detection of critical conditions. The results emphasize the importance of using advanced models like DenseNet121, which effectively handle imbalanced data while maintaining robust generalization. This work underscores the potential of AI-assisted diagnostic systems to support dermatologists in early melanoma detection, thereby improving patient outcomes. However, further exploration into techniques like synthetic data generation and cost-sensitive learning could enhance performance on imbalanced datasets.

7.Conclusion

7.1 Summary of Findings

This study demonstrated the application of deep learning models, particularly transfer learning techniques, for melanoma classification in a highly imbalanced dataset. Baseline CNN provided a solid starting point, but transfer learning models like EfficientNetB0 and DenseNet121 significantly outperformed it in terms of accuracy, precision, and recall. DenseNet121 emerged as the best-performing model, effectively addressing the dataset imbalance while maintaining robust generalization. The results highlight the potential of AI-driven diagnostic systems in enhancing early melanoma detection.

7.2 Future Work Directions

Future work will address dataset imbalance through techniques like oversampling, synthetic data generation (e.g., GANs), and cost-sensitive learning to enhance minority class detection. Ensemble methods and advanced architectures such as Vision Transformers or hybrid models with attention mechanisms will be explored to improve performance. Expanding datasets with diverse demographics and lesion types will boost generalization, while explainability frameworks like Grad-CAM or SHAP will ensure transparency and trust in clinical adoption. Additionally, efforts will focus on optimizing computational efficiency for real-time deployment in resource-constrained and underserved settings.

8.References

- ISIC Archive. (2020). *ISIC 2020 Challenge: Skin Lesion Analysis Towards Melanoma Detection*. Retrieved from https://challenge2020.isic-archive.com/
- SFU CSPMP. (2020). *An Introduction to Convolutional Neural Network (CNN)*. Medium. Retrieved from https://medium.com/sfu-cspmp/an-introduction-to-convolutional-neural-network-cnn-207cdb53db97
- ResearchGate. (2023). *DenseNet121 Architecture*. Retrieved from https://www.researchgate.net/figure/DenseNet121-architecture_fig1_363850803
- PeerJ. (2020). *Data Augmentation Techniques for Medical Imaging*. PeerJ Computer Science, 371. Retrieved from https://peerj.com/articles/cs-371/
- ResearchGate. (2023). *EfficientNetB0 Model Architecture*. Retrieved from https://www.researchgate.net/figure/EfficientnetB0-Model-Architecture_fig2_378395574