LEVERAGING DEEP LEARNING FOR MELANOMA SKIN CANCER CLASSIFICATION

Sadia Khan, Om V. Patel, Sanket Patel, Ramanpreet Kaur

University of Calgary

ABSTRACT

Melanoma is a serious skin cancer originating in melanocytes [1]. Though it represents only 1% of skin cancer cases, it causes the majority of related deaths, especially in women under 30 [2]. Early, accurate detection is critical for effective treatment. In this study, we developed deep learning models to classify dermatoscopic images as malignant or benign. Using a dataset of 10,000 labeled images, we trained ResNet50V2, EfficientNetB2/B3/B4, and a custom convolutional neural network (CNN). ResNet50V2 and the CNN showed strong performance, highlighting the potential of AI-assisted tools for early melanoma detection.

Index Terms— melanoma, skin cancer, deep learning, ResNet, EfficientNet, convolutional neural network, image classification

1. INTRODUCTION

Melanoma is a type of cancer that arises in melanocytes, the skin cells responsible for producing pigment. Although it represents just 1% of all skin cancer cases, it is the leading cause of skin cancer-related deaths due to its aggressive nature and potential to metastasize rapidly [1, 2]. The disease is particularly prevalent among young individuals under 30, especially women, and its incidence has increased significantly over the last three decades due to rising ultraviolet (UV) exposure [2].

Melanoma can often be visually distinguished from benign moles using the "ABCDE" criteria: **A**symmetry, irregular **B**order, multiple **C**olours, **D**iameter greater than 6 mm, and **E**volution in appearance over time [1]. While visual examination by a dermatologist is the first step in diagnosis, confirmation of whether the lesion is malignant (cancerous) or benign (non-cancerous) requires a skin biopsy [3]. If the lesion is malignant, it is classified as melanoma; however, if it is benign, it represents a different type of skin tumour or atypical mole [4].

Given the visual nature of melanoma diagnosis and the critical role early detection plays in improving patient outcomes, there is growing interest in leveraging artificial intelligence (AI) and deep learning techniques to aid in melanoma detection. This project explores the application of convolutional neural networks (CNNs) for classifying dermatoscopic images of skin lesions as malignant (melanoma) or benign.

2. RELATED WORK

Recent advancements in medical imaging and deep learning have significantly contributed to the early detection of melanoma, a deadly form of skin cancer [1, 4]. Traditional diagnostic methods, though reliable, can be enhanced by leveraging artificial intelligence to improve accuracy and efficiency. Studies such as Vayadande et al. [5] and Dubey et al. [6] have demonstrated the effectiveness of machine learning algorithms and convolutional neural networks (CNNs) in classifying skin lesions. Samantaray et al. [7] further explored ensemble learning methods to boost CNN performance in skin cancer detection. Image augmentation techniques, such as those outlined by Bhandari [8], are essential for improving model generalization, especially in limited data scenarios. Publicly available datasets like the Melanoma Skin Cancer Dataset on Kaggle [9] have facilitated extensive experimentation and model training.

3. MATERIALS AND METHODS

3.1. Dataset

This project uses the Melanoma Skin Cancer Dataset of 10,000 images, published by Hasnain Javed on Kaggle [9]. The dataset consists of high-resolution dermoscopic images labeled into two primary categories: benign (non-cancerous lesions and malignant (cancerous lesions, including melanoma). This dataset is structured in folders, with each class represented by its respective subdirectory. It serves as a reliable resource for training deep learning models aimed at binary image classification tasks, especially in the field of medical imaging. This dataset was well-suited for our objective of training deep learning models to classify skin lesions as either malignant or benign in the context of melanoma detection.

3.2. Preprocessing

To prepare the data for model training, we first extracted and preprocessed the images by resizing them to a uniform

size of 224x224 pixels, which aligns with the input requirements of most pre-trained models, such as ResNet and EfficientNet. The images were also normalized by scaling pixel values to the range [0, 1], ensuring consistency in input across the models. To reduce the risk of overfitting, which can happen because some images are very similar or because of class imbalance, we used data augmentation with TensorFlow's ImageDataGenerator [10]. This involved applying random changes to the images, like rotating, flipping, zooming, and adjusting brightness, to make the dataset more varied and help our models perform better on new, unseen data.

3.3. Model Architecture

This project compares ResNet50V2, EfficientNetB2/B3/B4, and a custom CNN to identify the most effective model for skin cancer classification.

3.3.1. RestNet50V2
The ResNet50V2 model is known for its ability to capture features in deeper layers, making it effective for image classification tasks [11]. Using it with transfer learning allows us to leverage its powerful pre-trained features. Fine-tuning will be performed on the last few layers to adapt the model to the specific task of distinguishing between benign and malignant skin lesions.

3.3.2. EfficientNet (B2, B3, or B4)
These models are part of the EfficientNet family, optimized for both performance and computational efficiency [12]. The EfficientNetB2 is the smallest and fastest and is ideal for low-latency tasks. EfficientNetB3 offers a good balance of accuracy and speed, while EfficientNetB4 is the most accurate but demands more computational resources. By using a frozen base and fine-tuning the top layers, these models can be adapted for skin cancer classification.

3.3.3. CNN
A simpler Convolutional Neural Network (CNN) is also tested for comparison. This model uses basic layers like Conv2D and MaxPooling2D and is augmented with dropout and batch normalization to improve performance and reduce overfitting [13].

3.4. Training and Evaluation

To ensure stable and effective training, several common techniques were applied. The AdamW optimizer (learning rate 1e-4) enabled efficient learning, while binary cross-entropy suited the binary classification task. To improve generalization and prevent overfitting, batch normalization, dropout, and global average pooling were used. Training was further optimized with early stopping and ReduceLROnPlateau, which halted training or adjusted the learning rate when improvements plateaued.

The model's performance was assessed using multiple evaluation metrics to ensure its effectiveness in classifying melanoma skin cancer:

- Model Accuracy & Loss Plots: The training and validation accuracy/loss were monitored across epochs to analyze learning progression. A steady increase in training accuracy and a corresponding decrease in loss indicate effective learning.
- Confusion Matrix: Provided a breakdown of correctly and incorrectly classified cases, helping assess model reliability.
- Classification Report: Included precision, recall, and F1-score to evaluate the balance between false positives and false negatives.
- ROC Curve & AUC Score: Illustrated the model's ability to differentiate between benign and malignant cases, with a higher AUC value indicating strong classification performance [14].

4. Results and Discussion

4.1. ResNet50V2 Model

The ResNet50V2 model performed well in melanoma classification, achieving a final validation accuracy of 91.50%. Training and validation accuracy increased steadily, with validation closely tracking training trends. Loss also decreased consistently, though occasional spikes in validation loss suggest mild overfitting.

The classification report confirms strong performance, with overall accuracy at 92%. Precision was 0.88 for benign and 0.96 for malignant cases, indicating high confidence in predictions. Malignant recall was 0.87, meaning 13% of melanoma cases were missed. F1-scores were strong for both classes, reflecting balanced performance. Overall, further tuning could help reduce overfitting and improve consistency.

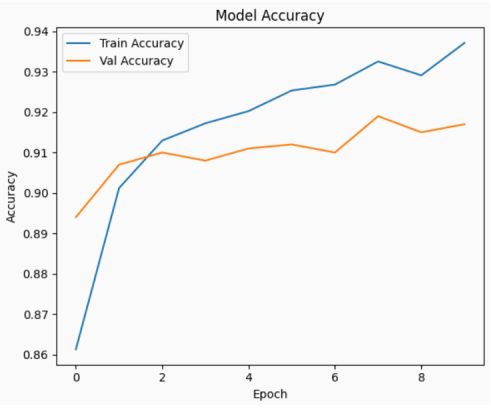

Fig. 1. ResNet50V2 Model Accuracy

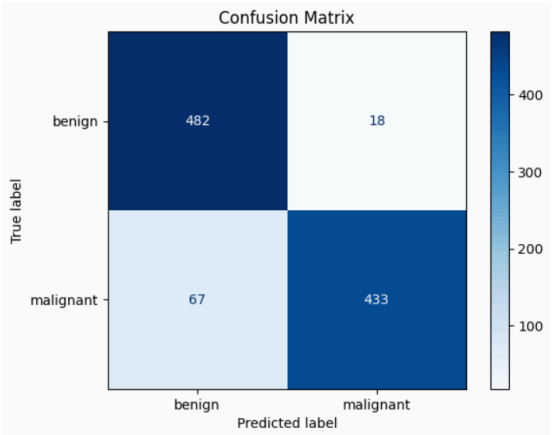

Fig. 2. Confusion matrix for ResNet50V2 model

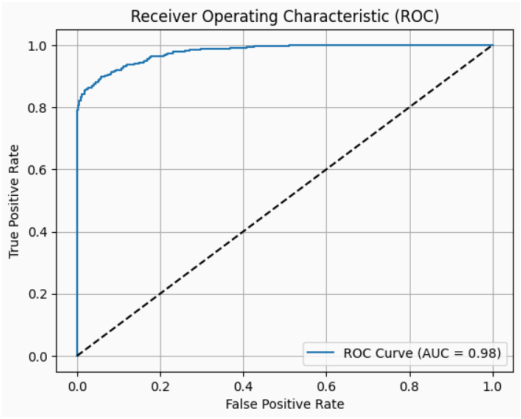

Fig. 3. ROC curve for the ResNet50V2 model

4.2. EfficientNetB2 Model

The EfficientNetB2 model showed suboptimal performance in melanoma classification. Training accuracy remained low (~52%), and validation accuracy was unstable, peaking at 66% but dropping sharply. While training loss decreased, the nearly constant validation loss suggests underfitting or limited learning.

The model misclassified 220 malignant cases as benign—an alarming rate of false negatives in a medical context. Its ROC AUC was only 0.67, well below the 0.85 benchmark for reliable classifiers.

The classification report reflected these issues: 71% precision and 56% recall for malignant cases, with overall accuracy at 67% and final validation accuracy at just 57.3%. These metrics indicate the model is not suitable for real-world deployment without significant improvements.

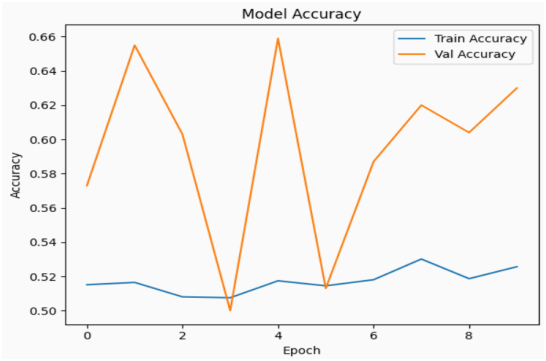

Fig. 4. EfficientNetB2 Model Accuracy

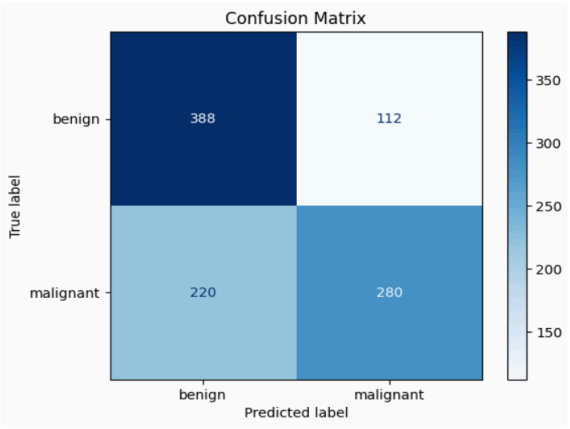

Fig. 5. Confusion matrix for EfficientNetB2 model

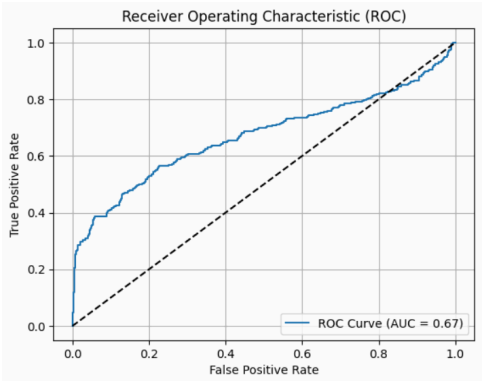

Fig. 6. ROC curve for the EfficientNetB2 Model

4.3. EfficientNetB3 Model

The EfficientNetB3 model for melanoma classification demonstrates moderate performance with signs of overfitting. The training accuracy improves gradually, but validation accuracy fluctuates, stabilizing at 71.5%. The loss curves further indicate that while training loss decreases, validation loss remains relatively unchanged.

The confusion matrix shows that the model correctly classifies 434 malignant cases but misclassifies 66 as benign. However, it struggles more with benign cases, misclassifying 171 as malignant. The ROC curve (AUC = 0.82) indicates a reasonable ability to differentiate between classes, though not optimal. The classification report reflects an overall accuracy of 76%, with a higher recall for malignant cases (0.87) than benign (0.66), suggesting a bias toward detecting malignancies.

While the model effectively identifies malignant cases, its high false positive rate and unstable validation accuracy indicate the need for further optimization, such as data augmentation and regularization, to improve generalization.

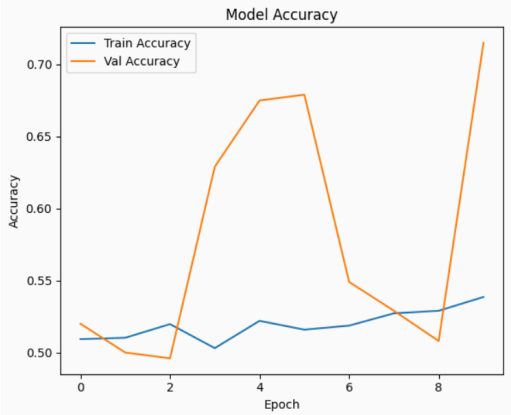

Fig. 7. EfficientNetB3 Model Accuracy

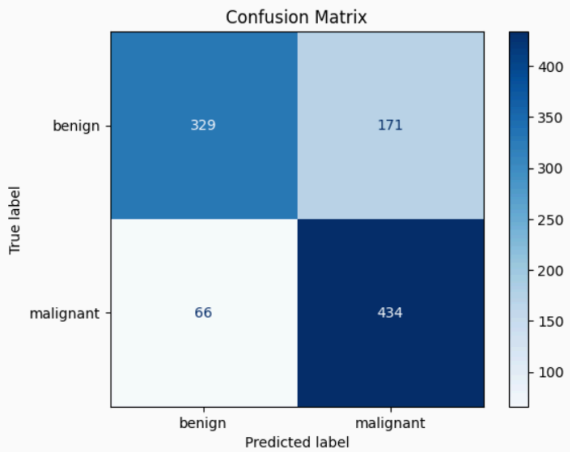

Fig. 8. Confusion matrix for EfficientNetB3 model

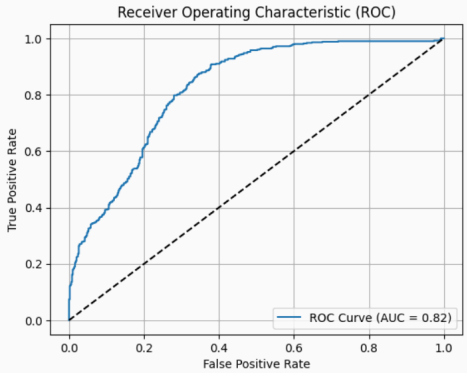

Fig. 9. ROC Curve for EfficientNetB3 model

4.4. EfficientNetB4 Model

EfficientNetB4 showed moderate performance, with 67.70% peak validation accuracy and an AUC of 0.75. It misclassified 248 malignant cases, highlighting low recall (0.50) despite high precision (0.82). Overall accuracy was 70%, with F1-scores of 0.69. The model shows potential but needs improvement in sensitivity to be clinically reliable.

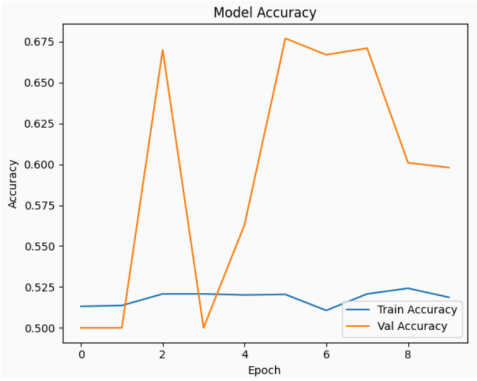

Fig. 10. EfficientNetB4 Model Accuracy

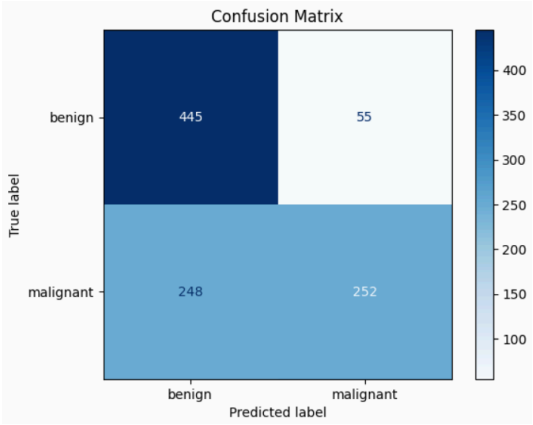

Fig. 11. Confusion matrix for EfficientNetB4 model

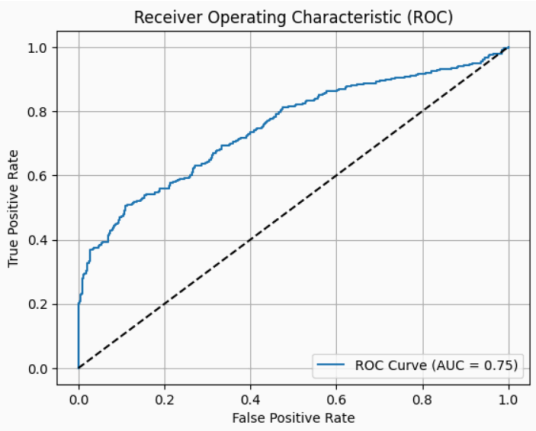

Fig. 12. ROC Curve for EfficientNetB4 model

4.5. CNN Model

The custom CNN model achieved strong performance, with a final validation accuracy of 90.00%, indicating effective differentiation between benign and malignant lesions. Accuracy and loss plots show consistent training improvement, though minor validation fluctuations suggest slight overfitting or dataset variability.

The confusion matrix shows 457 benign and 443 malignant cases were correctly classified, with 43 benign and 57 malignant misclassified, reflecting a balanced performance aligned with the dataset.

The ROC curve yielded an AUC of 0.96, confirming excellent discriminative ability. The classification report showed precision, recall, and F1-score of 0.90 for both classes, further supporting the model's robustness and balance.

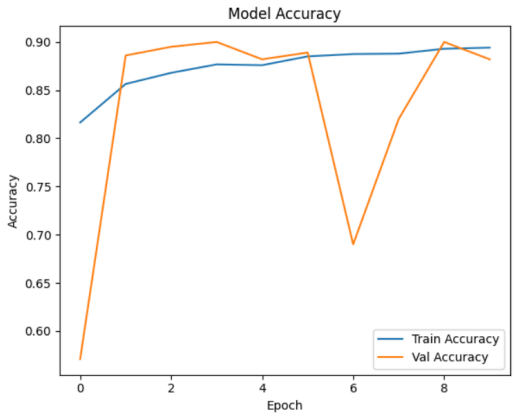

Fig. 13. Custom CNN Model Accuracy

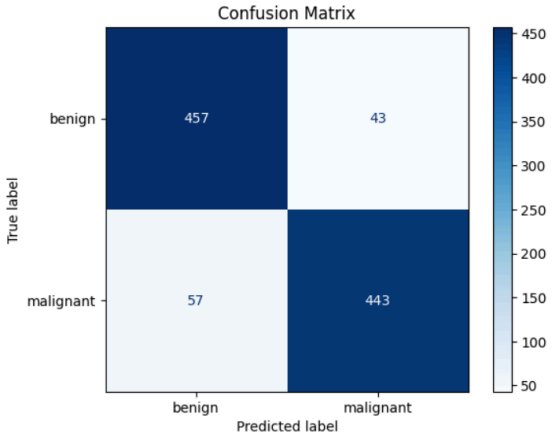

Fig. 14. Confusion matrix for CNN model

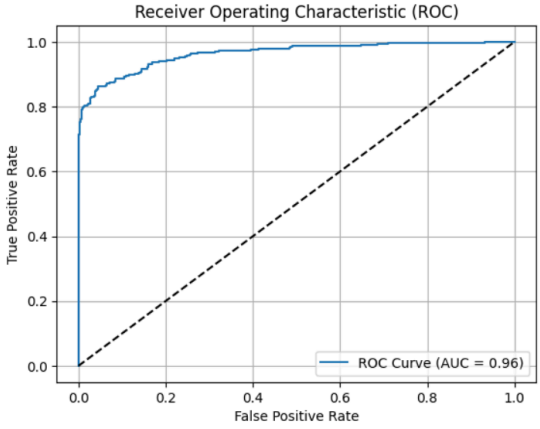

Fig. 15. ROC Curve for CNN model

5. Conclusion

This project demonstrated the potential of deep learning models in classifying melanoma skin cancer from dermatoscopic images. Among the tested models, ResNet50V2 and the custom CNN achieved the highest performance, with validation accuracies around 90% and excellent AUC scores, making them suitable for real-world applications. In contrast, EfficientNet variants underperformed, particularly in sensitivity to malignant cases. These findings suggest that carefully selected and tuned CNN architectures can effectively support early melanoma detection and assist clinicians in improving diagnostic accuracy.

6. Repository Link

https://github.com/OMPATEL20/SkinCancer-Classification-Group15-ENEL-645

7. References

[1] J. Y. Wang, E. B. Wang, and S. M. Swetter, "What is melanoma?," JAMA, vol. 329, no. 11, p. 948, Mar. 2023. doi:10.1001/jama.2022.24888

[2] "Melanoma: Symptoms, staging & treatment," Cleveland Clinic, https://my.clevelandclinic.org/health/diseases/14391-melanoma (accessed Mar. 30, 2025).

[3] "Diagnosis of melanoma skin cancer," Canadian Cancer Society, https://cancer.ca/en/cancer-information/cancer-types/melanoma-skin/diagnosis (accessed Mar. 30, 2025).

[4] "Moles and benign skin growths," UC Health, https://www.uchealth.com/en/conditions/moles-and-benign-skin-growths (accessed Mar. 30, 2025).

[5] K. Vayadande, O. Mirajkar, S. Mhetre, H. Mehta, S. Markande and S. Mohol, "The Future of Skin Cancer Detection: Machine Learning and Convolutional Neural Networks," *2024 5th International Conference on Data Intelligence and Cognitive Informatics (ICDICI),* Tirunelveli, India, 2024, pp. 1282-1290.

[6] P. Dubey, P. Bande, et. al, "AI-Powered Dermatology: Revolutionizing Skin Cancer Detection Through Machine Learning Algorithms," *2024 International Conference on Cybernation and Computation (CYBERCOM),* Dehradun, India, 2024, pp. 455-460.

[7] R. Samantaray, et. al, "Skin Cancer Detection Using CNN and Ensemble Learning," *2024 IEEE 1st International Conference on Advances in Signal Processing, Power, Communication, and Computing (ASPCC),* India, 2024, pp. 247-252.

[8] A. Bhandari, "Image augmentation on the fly using keras imagedatagenerator!," *Analytics Vidhya*, https://www.analyticsvidhya.com/blog/2020/08/image-augmentation-on-the-fly-using-keras-imagedatagenerator (accessed Apr. 1, 2025).

[9] M. H. Javid, "Melanoma skin cancer dataset of 10000 images," Kaggle, https://www.kaggle.com/datasets/hasnainjaved/melanoma-skin-cancer-dataset-of-10000-images (accessed Mar. 30, 2025).

[10] "Image Data Generator", https://www.tensorflow.org/api_docs/python/tf/keras/preprocessing/image/ImageDataGenerator (accessed March 31, 2025).

[11] "RestNet50" ,https://www.tensorflow.org/api_docs/python/tf/keras/applications/ResNet50V2 (accessed Mar. 25, 2025).

[12] "Understanding EfficientNet- The most powerful CNN architecture"https://arjun-sarkar786.medium.com/understanding-efficientnet-the-most-powerful-cnn-architecture-eaeb40386fad (accessed Mar. 25, 2025).

[13] "Convolutional Neural Networks", https://www.ibm.com/think/topics/convolutional-neural-networks (accessed Mar. 25, 2025).

[14] "Machine Learning - AUC - ROC Curve," W3 Schools, https://www.w3schools.com/python/python_ml_auc_roc.asp (accessed Apr. 1, 2025).