SINGAPORE MANAGEMENT UNIVERSITY

Skin Cancer **Detection**

Presented by:
Tran Binh Minh, **01323752**
Shwe Tin Aung, **01439396**
Aditya Vijay, **01522128**
Liu Chih Yuan, **01500507**
Lope Immanuell IV Cudiamat Santos, **01520025**

Table of Contents

Background and Motivation

Early detection of melanoma is **crucial** for survival. However, it highly depends on dermatological expertise and access to diagnostic tools.

The **5-year** survival rate for early-stage melanoma is **close to 100%,** but this drops significantly if the cancer has spread. In the US, for 2025, it's estimated that about 8,430 people will die of melanoma.

Cancer Stat Facts: Melanoma of the Skin

Estimated New Cases in 2025	104,960
% of All New Cancer Cases	5.1%

Estimated Deaths in 2025	8,430
% of All Cancer Deaths	1.4%

5-Year Relative Survival
94.7%
2015–2021

Current Practicing Methods	Advantages	Drawbacks
Visual Inspection (Naked Eye)	Simple, widely used	Highly subjective, prone to human
Biopsy (Histopathology)	Gold standard for diagnosis	Invasive, time-consuming
Reflectance Confocal Microscopy (RCM)	High-resolution, non-invasive	Expensive and limited availability
Optical Coherence Tomography (OCT)	Provides tissue depth information	Limited resolution for pigmented

Objective of Project

Address the **challenge** of skin cancer detection by developing a deep learning system that **classifies** images as **benign** or **malignant**.

Ideal World (Perfect Dataset)

- Perfectly balanced dataset (Malignant & Benign)
- Fitzpatrick skin type
- Anatomic location
- No occlusions (hairs, veins, etc.)
- Geographically diverse

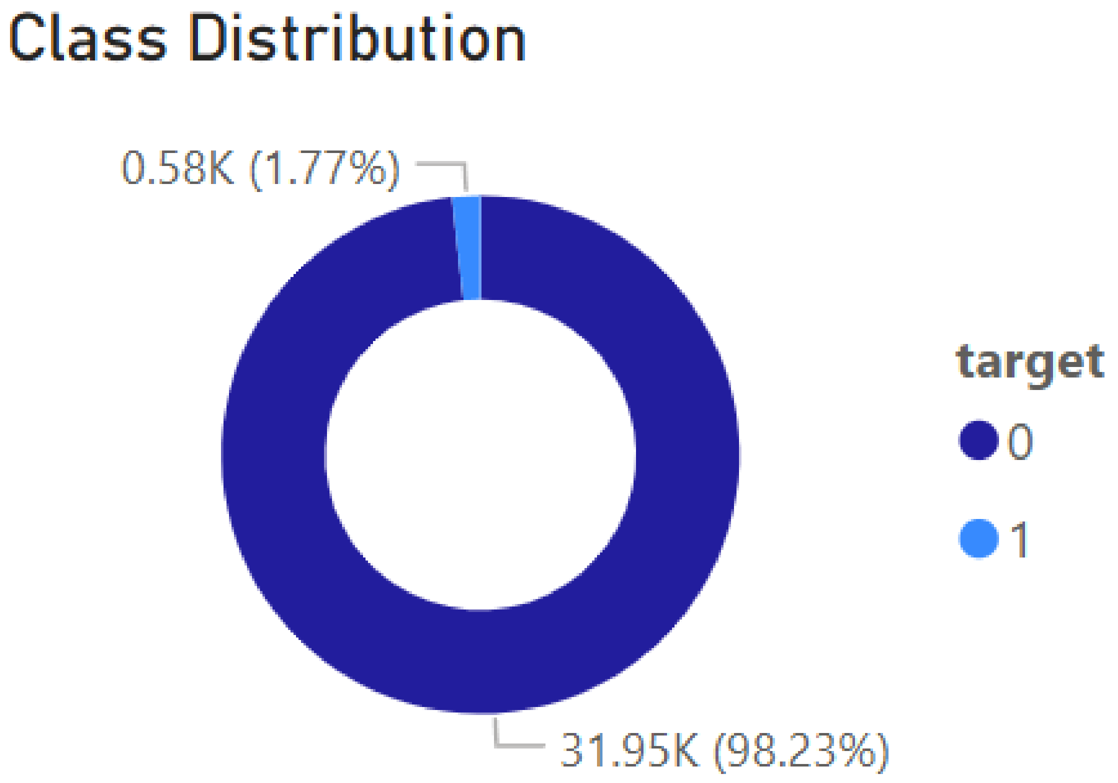
Reality(Actual Dataset)

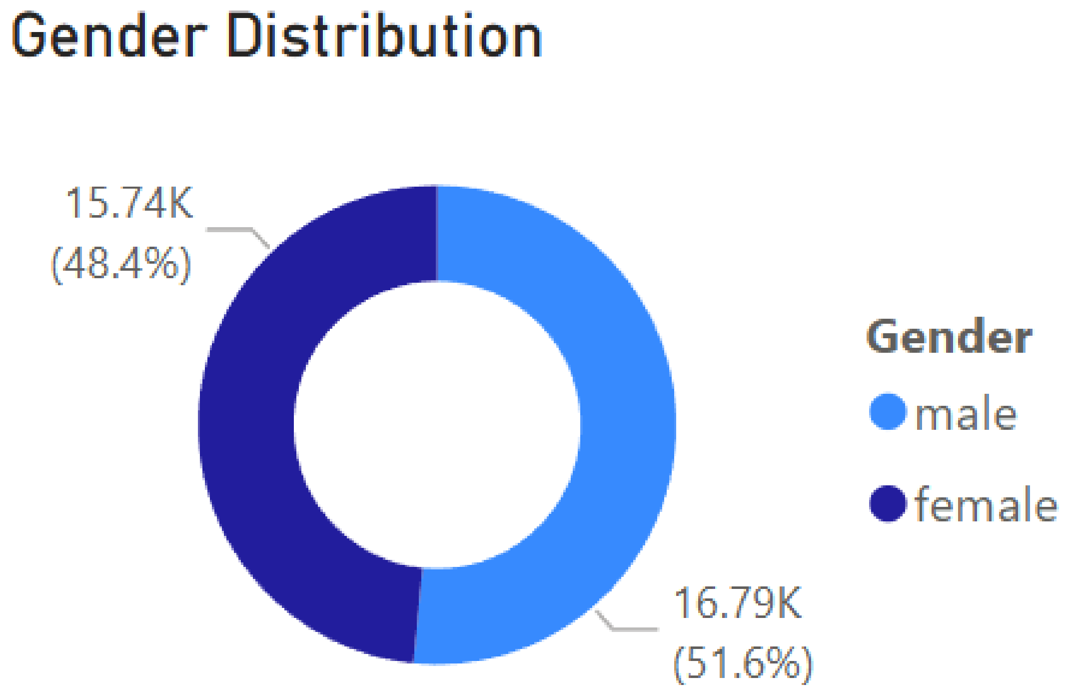
- Severe class imbalance
- Skin type metadata missing
- Skin tone bias
- Inconsistent quality images

Machine Learning Objective

- Provide consistent, high-sensitivity classification
- Prioritize model explainability to meet clinical decision-making needs
- Ensure transparent rationale for every prediction
- Foster clinician trust through clear, interpretable AI outputs

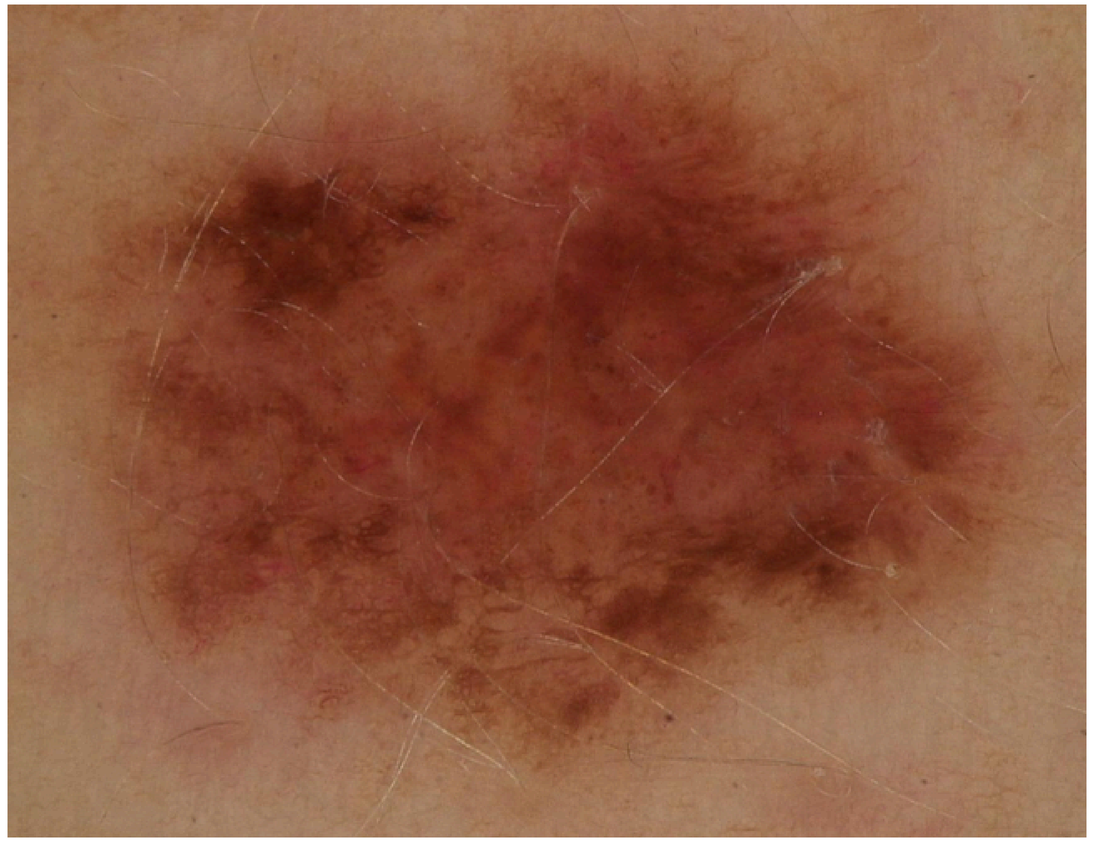
Dataset Analysis



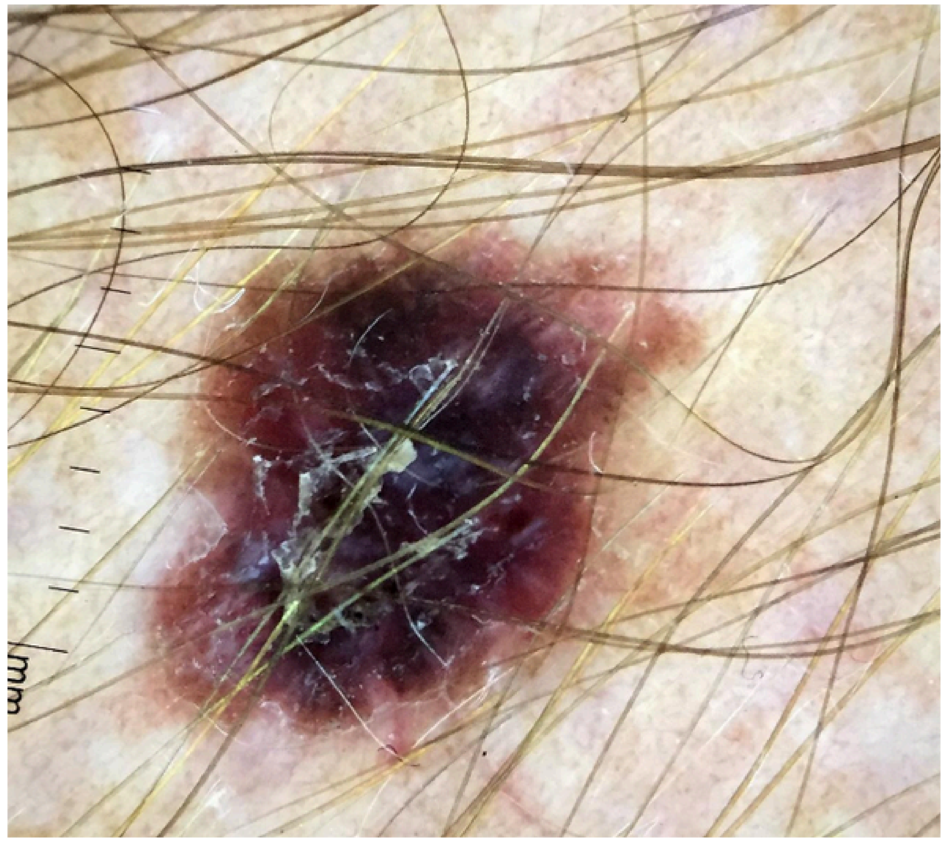


<u>Dataset Summary</u>

- ISIC skin classification dataset (2016 - 2020)
- Varying image resolutions
- Class (benign/malignant) imbalance

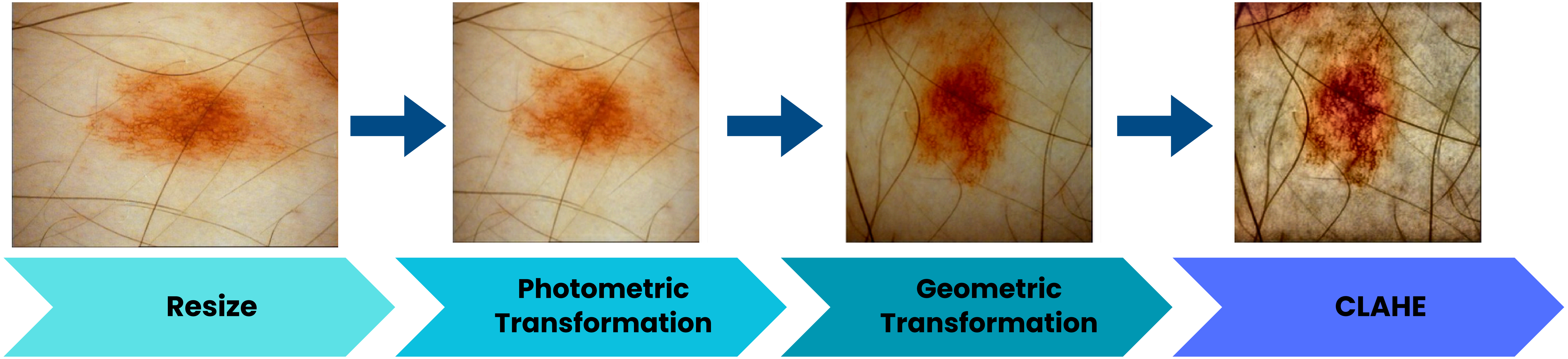


Benign



Malignant

Augmentation Methodolgy



Resizing

224 x 224 (EfficientNet-B0)

To benefit from transfer learning, best practice to use the same input size the model was trained on

Geometric Transformation

Transpose → Vertical/ Horizontal Flip

Improving special robustness and reducing reliance on fixed layouts

Photometric Transformation

RandomBrightnessContrast (adjusts skintone variations)

HieSaturationValue (tweaks color tones)

Improves robustness by simulating diverse lighting and conditions for generalization

CLAHE

Enhances local contrast (tile by tile) → reveals fine details in shadowed or bright areas

Avoids over-amplifying noise (due to clip limit)

Recap on previous researches

For skin cancer detection, while several models have been used, **EfficientNet** stands out for its superior performance in CNN model branch, and **Vision Transformer** for its innovative application

Model Name	Architecture Type	Key Feature/Rationale	Typical Performance (Accuracy/AUC range)
ResNet50	CNN	Residual connections for deep networks, mitigates vanishing gradient	High
Xception	Depth-wise Separable CNN	Efficient feature extraction via depth-wise separable convolutions	Very High
MobileNetV2	Depth-wise Separable CNN	Lightweight, efficient, suitable for mobile/edge devices	Good for mobile
VGG16	CNN	Uniform architecture with small filters, deep feature extraction	Moderate to High
EfficientNet	CNN (AutoML-scaled)	Scalable architecture for efficiency and accuracy	High to Very High
Vision Transformer (ViT)	Transformer	Global context, self-attention, processes image patches as sequences	High/Emerging

Insights:

- Research papers had proven EfficientNet is among the best CNNs model for legion detection
- Vision Transformer is also a new trending method of skin cancer detection

Approach: Focus on the evaluation of **EfficientNet** and **VisionTransformer**, and select one for **in-depth testing.**

Transformer (ViT) vs. CNN (EfficientNet)

Two training processes were designed with the same setup and dataset for model comparison:

Visual Transformer

Metric	Result
Loss	0.9981
Accuracy	0.75
Precision	0.6744
Recall	0.725
AUC	0.8342

EfficientNet

Metric	Result
Loss	1.6058
Accuracy	0.72
Precision	0.65
Recall	0.65
AUC	0.814

Insight:
This is aligned with previous researches stating ViT has better performance but performs worse for GradCAM.

> *1) Pneumonia X-ray dataset analysis*
>
> The ViT-l-16 model surpasses all CNN-based models evaluated in this study, as discussed in Section III-A. Our evaluation of fifteen Xray image samples demonstrated a high level of agreement between the ViT-l-16 model, the EfficientNet-B5 model (the best performing CNN-based model), and the radiologists' diagnoses regarding the normality or abnormality of the X-ray images. However, in terms of visualization, the radiologists expressed a clear preference for the Grad-CAM visualizations produced by the CNN-based model (EfficientNet-B5) when it came to highlighting the affected regions in the images, as shown in Table II. This performance aligns closely with the

Explainable Disease Classification: Exploring Grad-CAM Analysis of CNNs and ViTs (2025)

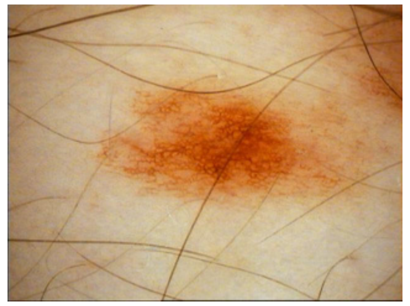
The decision is to use EfficientNet moving forward due to its high explainability, which is very important in the medical industry

EfficientNet Processing Flow

Focuses computational resources on most relevant features

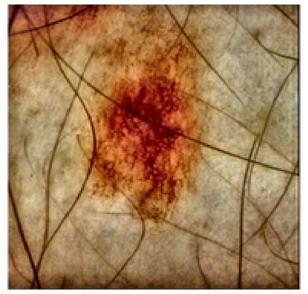
Final dense layer outputs class probabilities

Raw Image

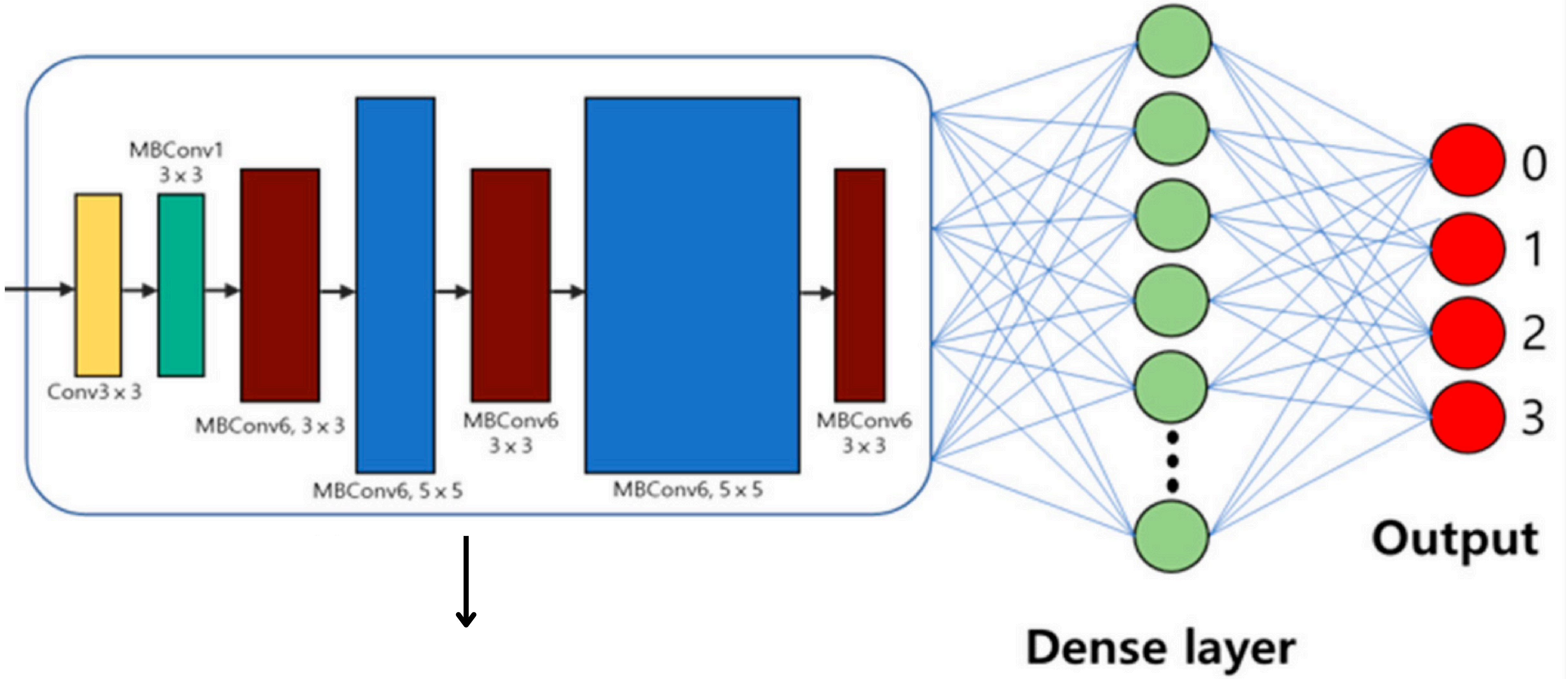


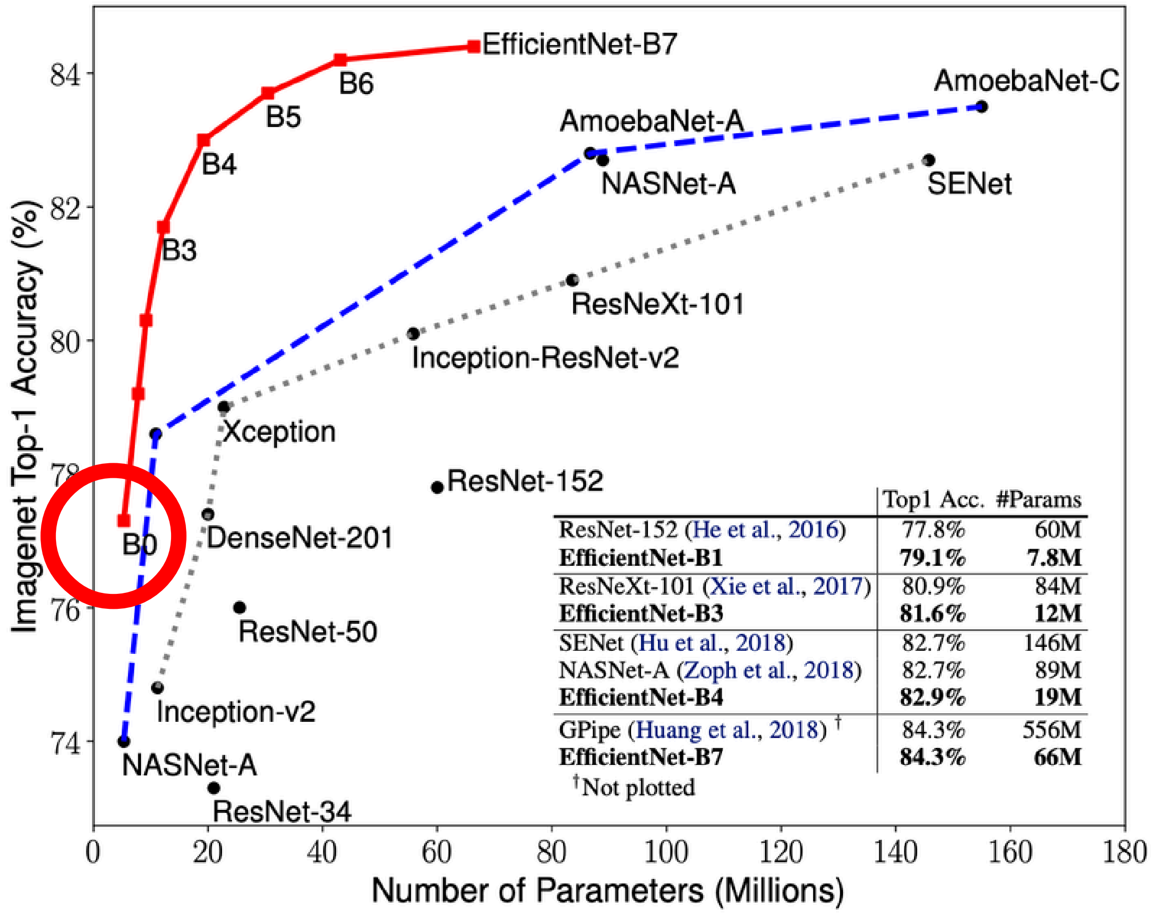
Augmentation

Input



Augmented 224 x 224





Progressive Feature Extraction

Stage 1	**Stage 2 → 7**
Basic edge and texture detection	Increasingly complex pattern recognition

Key Efficiency Gains

- 10x fewer parameters than ResNet
- 5x faster inference time
- Maintains comparable or better accuracy

Training Protocols

1. Tested both original and resized datasets with the goal of decreasing computational training time and seeing how it affects across performance metrics

Original

Final Test Results	
Accuracy	0.9825
AUC	0.8475
Recall	0

Training Time
4 hours (CPU Time)
1 hour 30 min (A100 GPU)

Resized

Final Test Results	
Accuracy	0.9825
AUC	0.8061
Recall	0

Training Time
2 hours 30 min (CPU Time)
40 min (A100 GPU)

Resized (with 2 epochs)

Final Test Results (Epoch 1)	
Accuracy	0.9825
AUC	0.7397
Recall	0

Final Test Results (Epoch 2)	
Accuracy	0.9825
AUC	0.8369
Recall	0

The AUC score increase at the 2nd epoch indicates the model may benefit from extended training periods.

Evaluation:

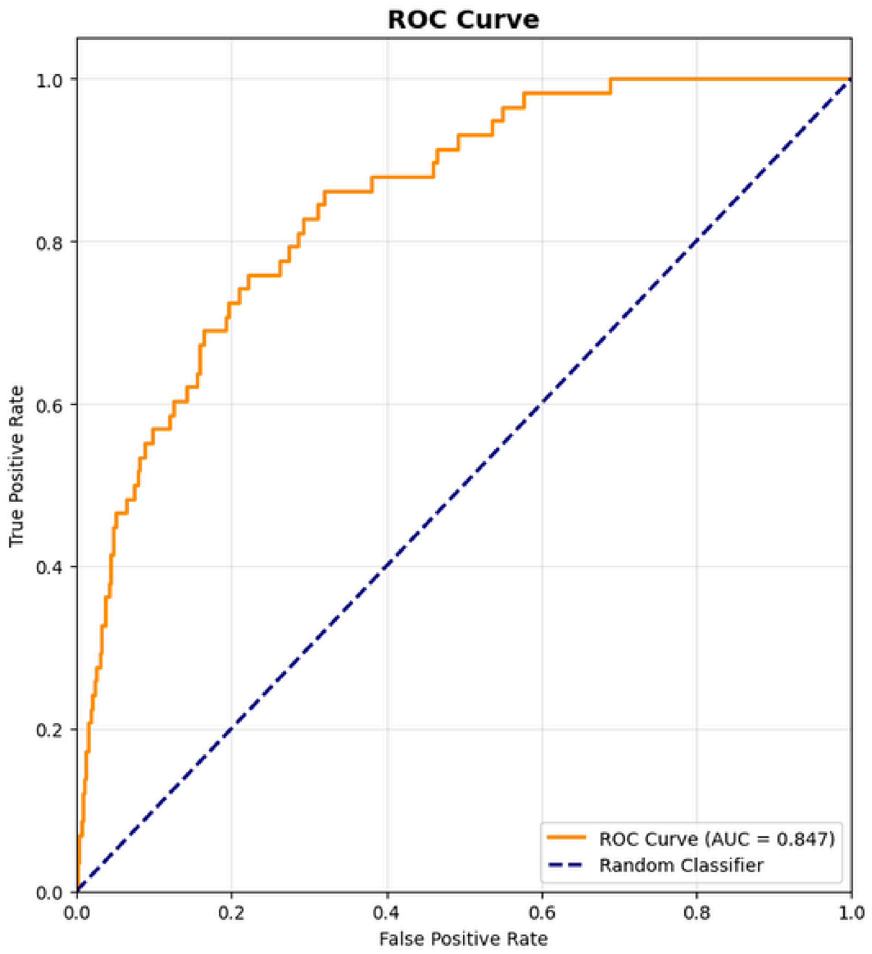
No significant difference can be used in real cases to save computational resources and for retraining scenarios

Training Protocols

Original

Confusion Matrix

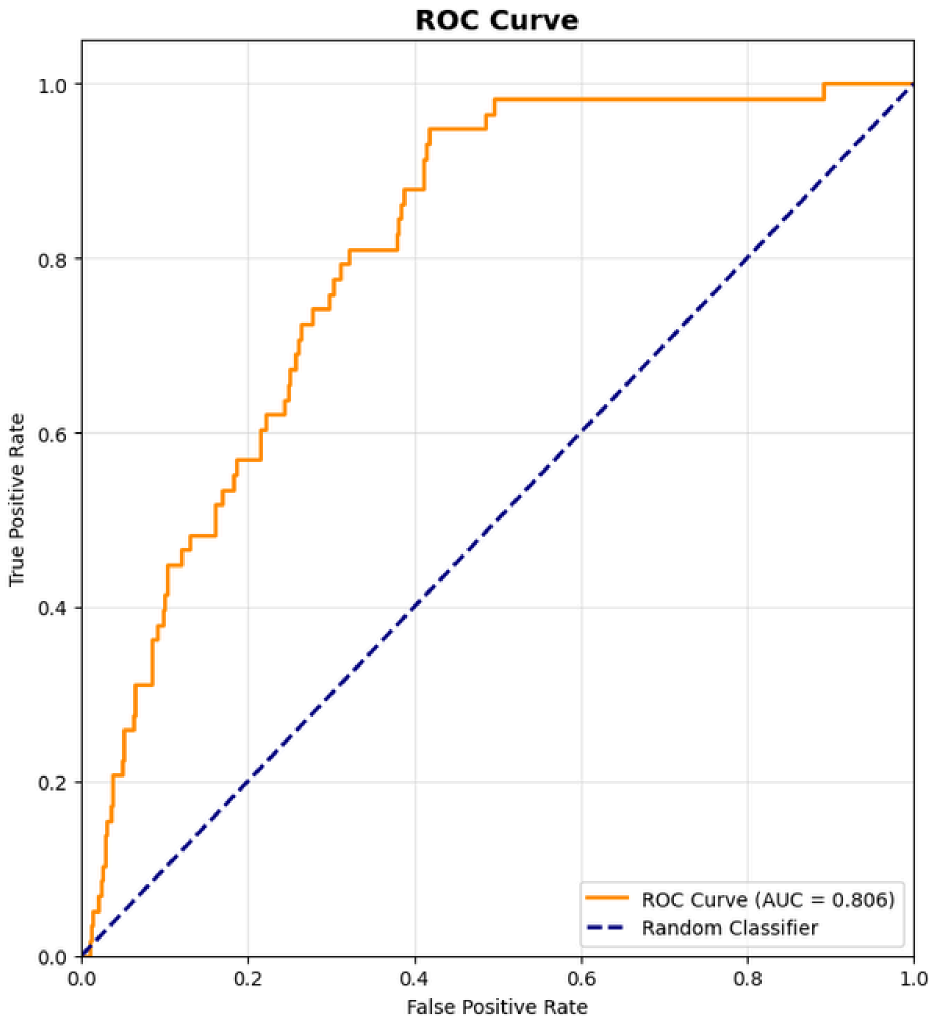
True Label \ Predicted Label	Benign	Malignant
Benign	3255 (98.2%)	**0**
Malignant	58 (1.8%)	0



Resized

Confusion Matrix

True Label \ Predicted Label	Benign	Malignant
Benign	3255 (98.2%)	**0**
Malignant	58 (1.8%)	0



Training Protocols

2. Epoch Training on three different datasets- Original, Resized, and Augmented (1000 images)

- Conduct comparative training experiments using various dataset types to assess performance differences across key metrics, prioritizing recall improvement as the primary objective.
- Identify the optimal epoch count for training when working with expanded datasets.

Original

Final Test Results (Original)	
Accuracy	0.73
AUC	0.8079
Recall	0.6

Resized

Final Test Results (Resized)	
Accuracy	0.875
AUC	0.64
Recall	0.5

Augmented

Final Test Results (Augmented)	
Accuracy	0.835
AUC	0.7211
Recall	0.6

Training Protocols

3. Tackling Class Imbalance (Oversampling)

The original dataset contained 500 malignant cases out of 33,125 images (1.51%).
Our sampling experiments proceeded as follows:
Initial small-scale test maintaining the original ratio 985-15: no results
1,000 images with exact percentage split: no results
950-50 split: no results
600-400 split: improved recall
This progression indicates that artificially increasing malignant case representation through data augmentation can potentially boost recall performance.

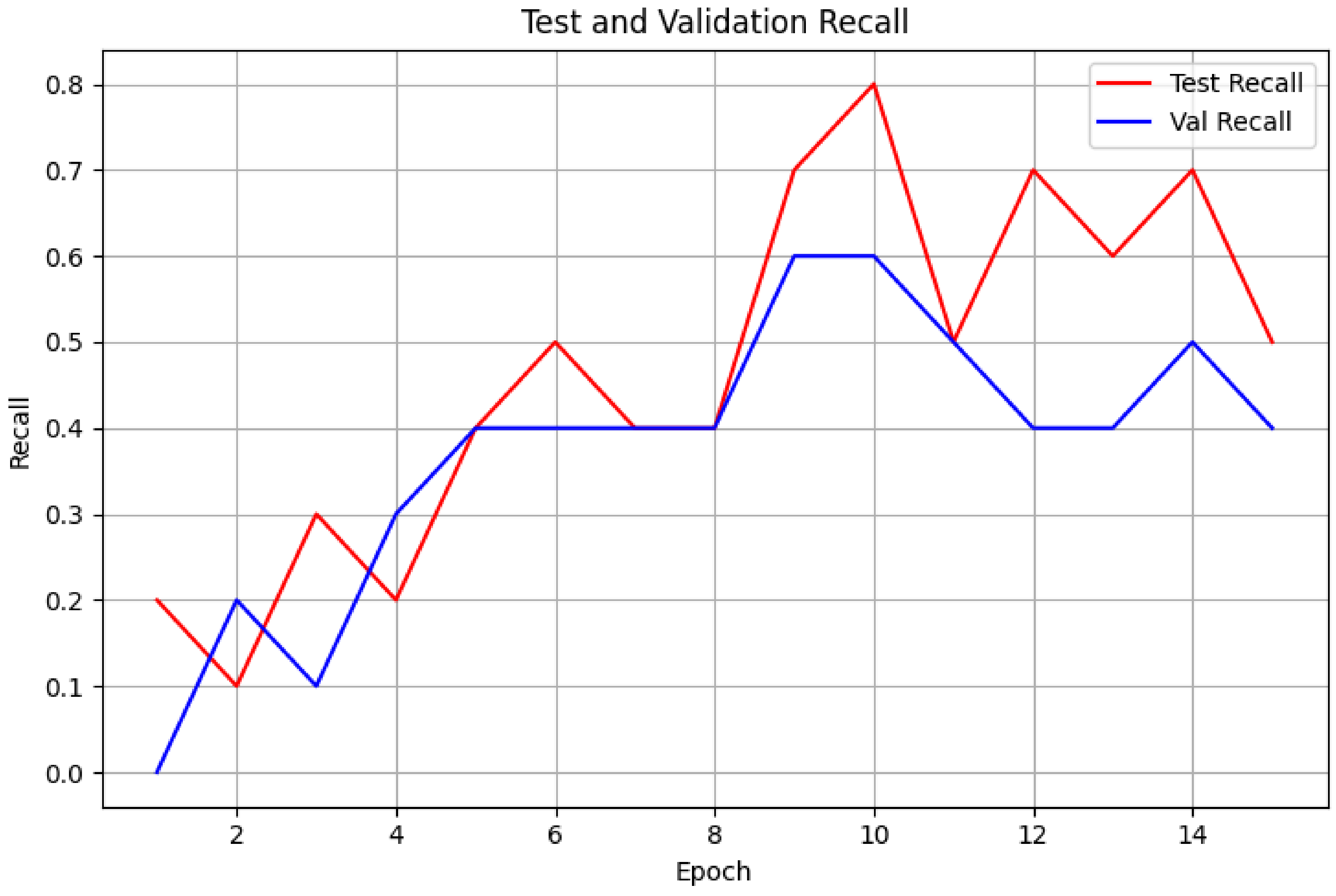
950 benign vs 50 malignant

- No significant results

600 benign vs 400 malignant

Final Test Results	
Accuracy	0.7
AUC	0.7975
Recall	0.8

600 benign vs 400 malignant

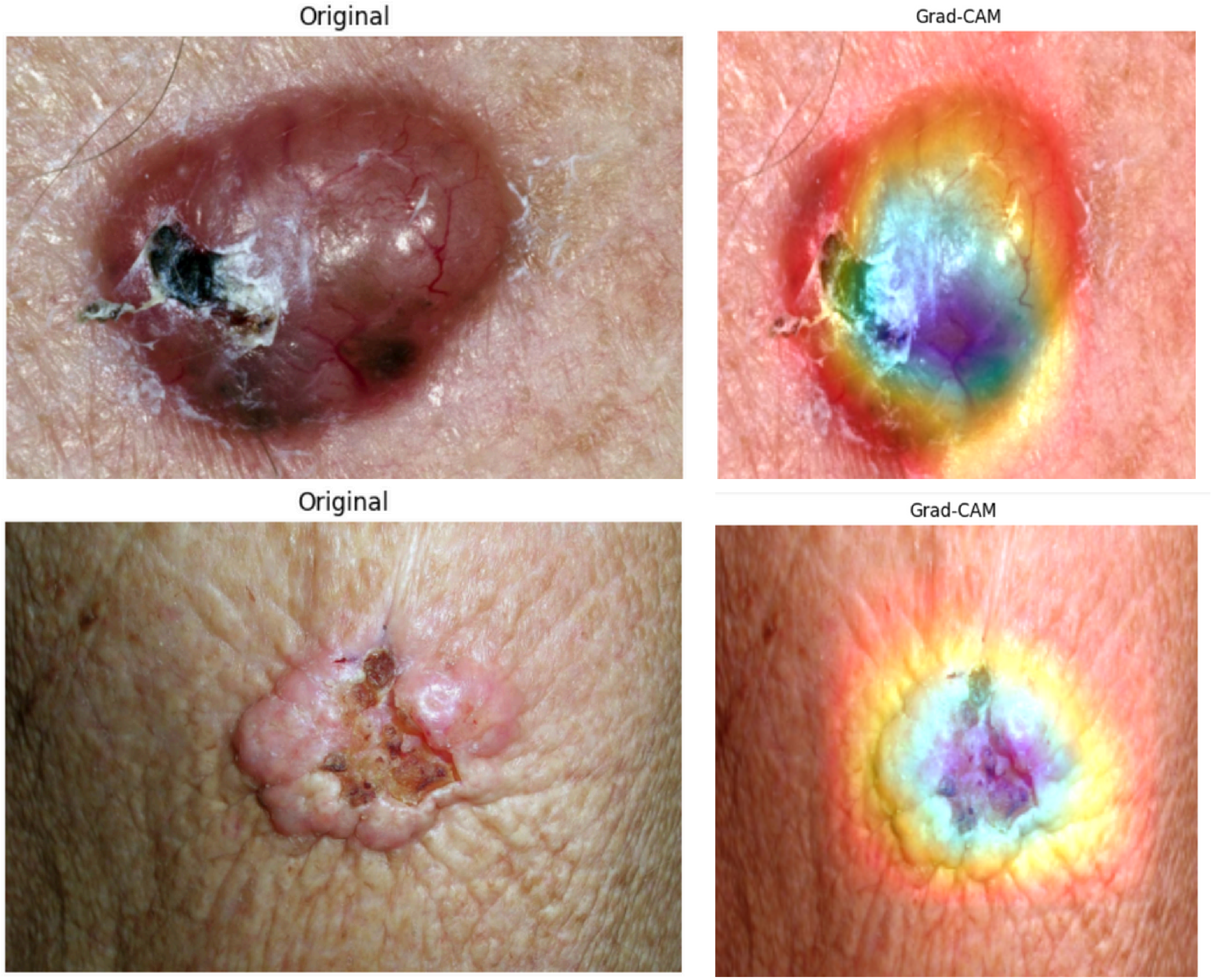


EfficientNet with Grad-CAM Explainability

EfficientNet: Prediction Process (CNN Forward Pass)
Input Image → Convolutional Layers/Blocks (EfficientNet) → Feature Maps → Pooling Layer → Fully Connected Layer → Final Prediction

Grad-CAM: Explanation Process (Reverse/Gradient Pass)
Target Class Prediction Score ← (Gradients) Feature Maps ← (Calculate) Weights → Heatmap (on original image)

Grad-CAM uses gradients flowing back to CNN's final convolutional layers to create a heatmap, visually highlighting the image regions most critical for a specific prediction



EfficientNet's generalization with different skin tone

Limitation on Generalization

- Lack of quality datasets for dark skin tones
- No widely accepted medical tools specifically designed to recolor skin data to simulate darker skin tones

Our experiment

- *Augment data to simulate darker skin tones*, enabling the model to detect tumors on dark skin
- However, *recoloring augmentation may alter tumor color*, affecting detection

Solution

- Combine *normal augmented dataset* with *dark skin augmented dataset* to improve generalization and maintain tumor color integrity

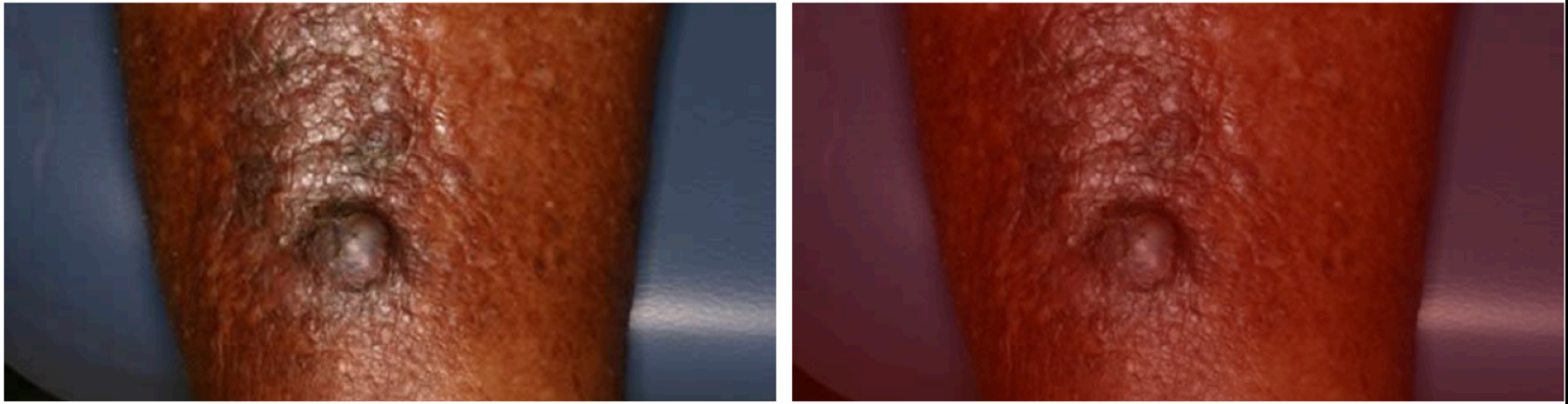
EfficientNet’s generalization with different skin tone

Model performance based on **normal augmented dataset**

```
image_path = '/content/drive/MyDrive/Colab Notebooks/EfficientNet + GradCAM/darkskin_tumor_5.png'
predict_image_with_gradcam(model_normal, image_path)
```

```
Prediction: Benign (Prob: 0.0000)
```

Original Image | Grad-CAM (Benign)



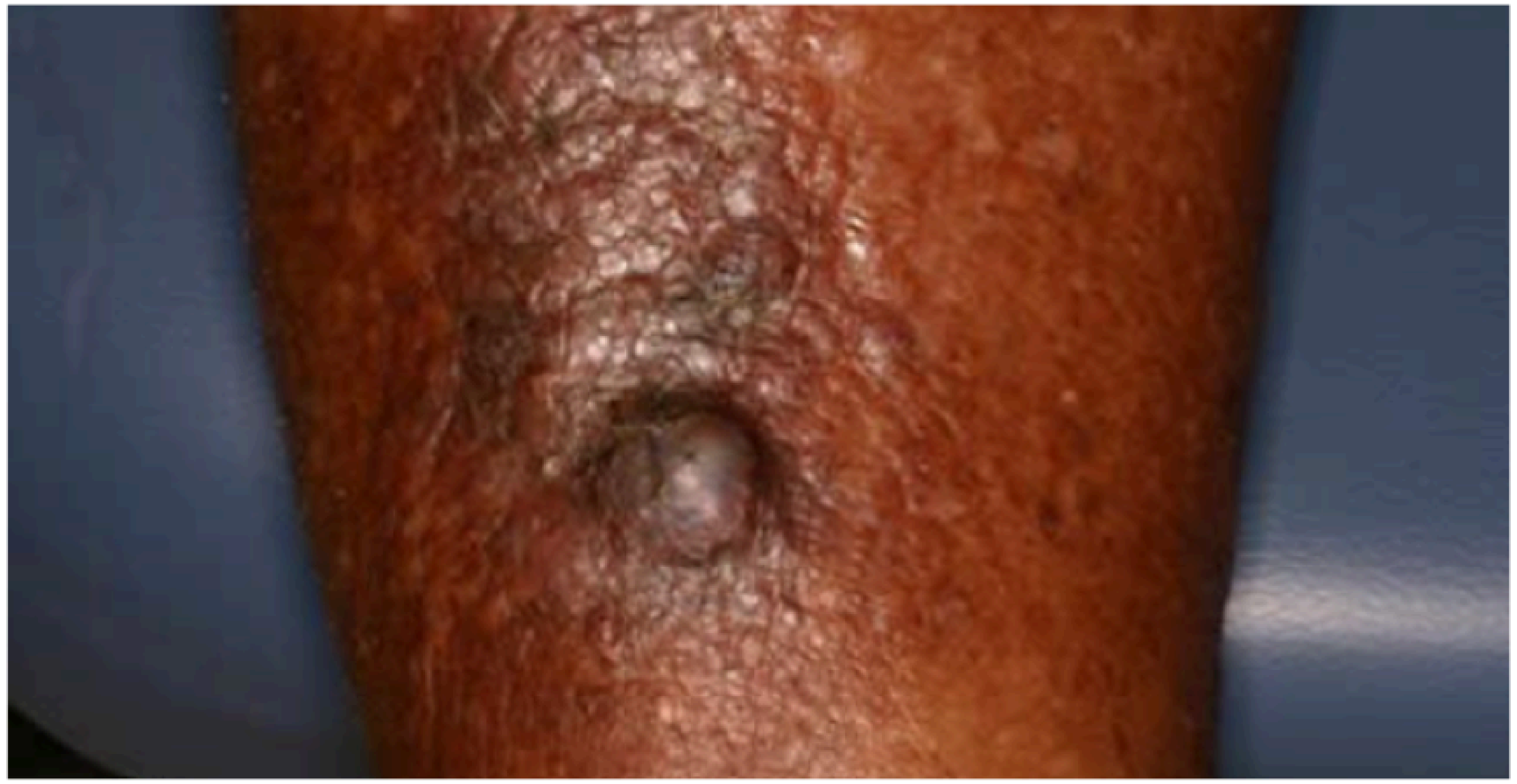
EfficientNet’s generalization with different skin tone

Model performance based purely on **dark-skin augmented dataset**

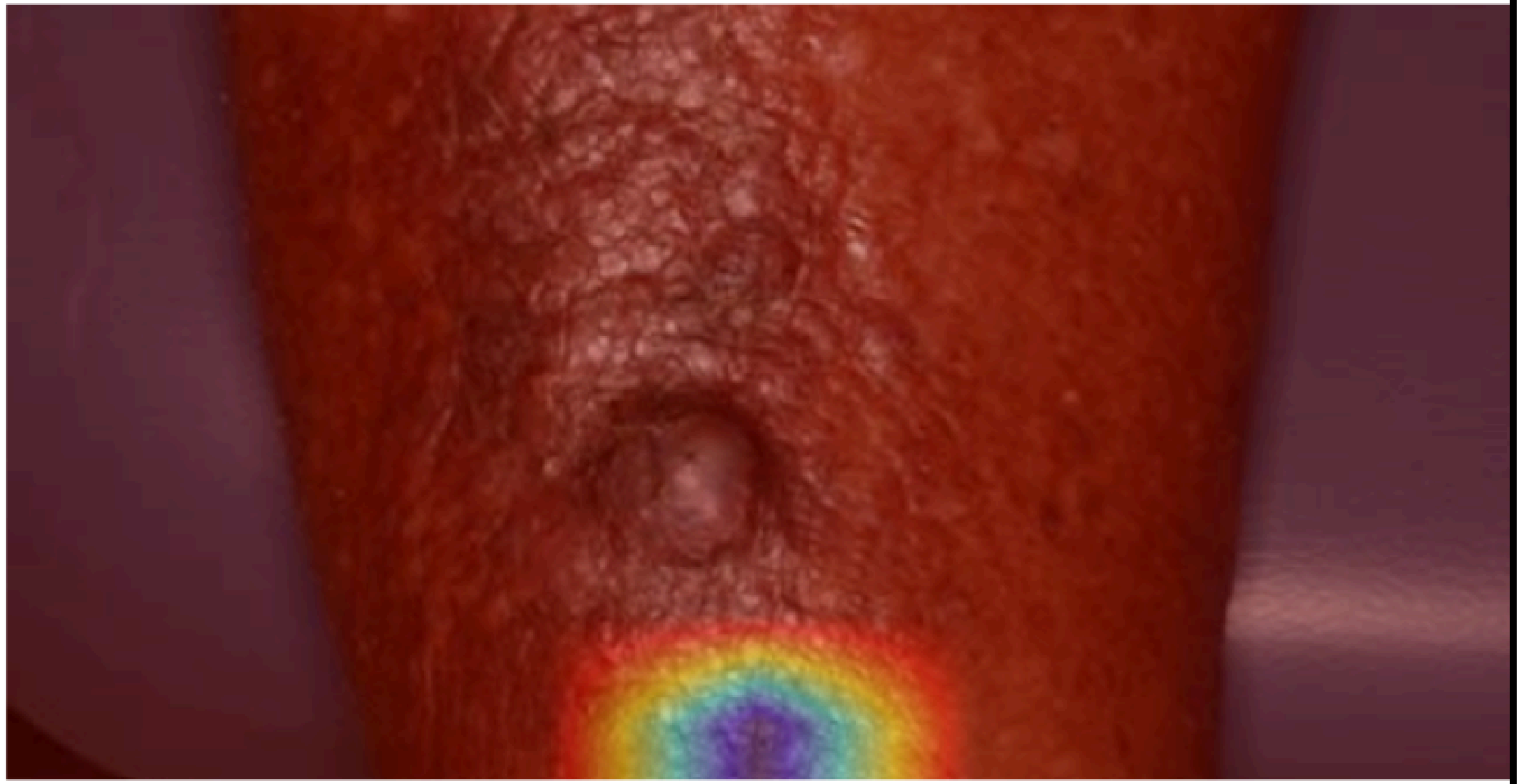
```
image_path = '/content/drive/MyDrive/Colab Notebooks/EfficientNet + GradCAM/darkskin_tumor_5.png'
predict_image_with_gradcam(model_darksin, image_path)
```

```
Prediction: Benign (Prob: 0.0330)
```

Original Image



Grad-CAM (Benign)

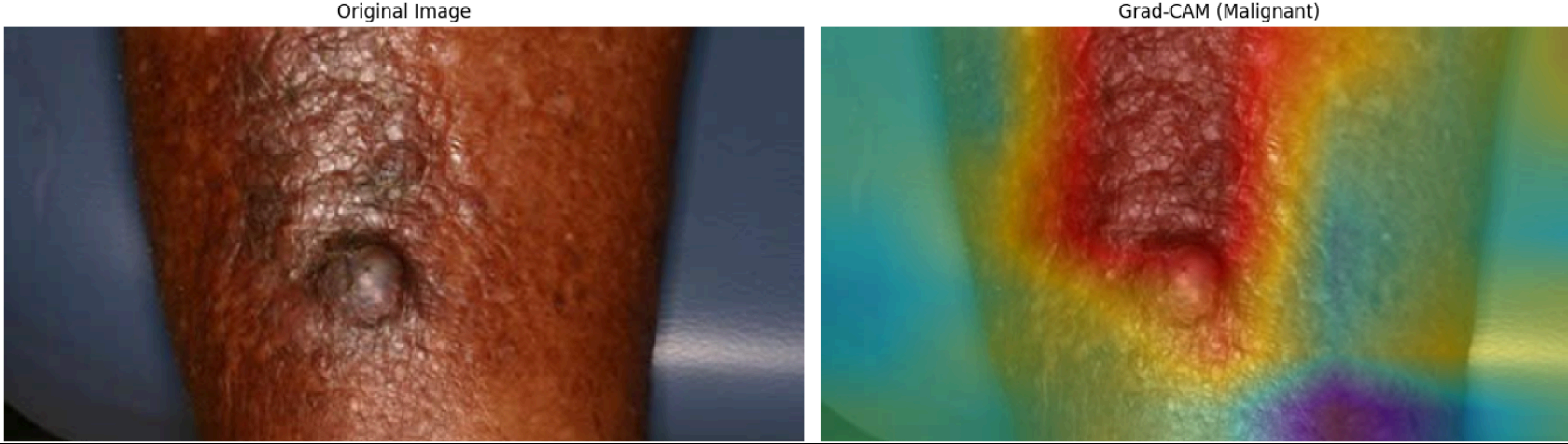


EfficientNet’s generalization with different skin tone

Combine **augmented images of dark skin tones** with **normal augmented images**
Grad-CAM shows a much better capture of the skin cancer

```
image_path = '/content/drive/MyDrive/Colab Notebooks/EfficientNet + GradCAM/darkskin_tumor_5.png'
predict_image_with_gradcam(dark_skin_normal_combined, image_path)
```

```
Prediction: Malignant (Prob: 0.8338)
```



However, this positive result is not consistently observed in all images, indicating the need for further tuning and analysis

Future Improvements

- **Ensemble: Vision Transformer & EfficientNet**
  - ViT: Higher accuracy, lower explainability
  - EfficientNet: Efficient & interpretable

- **Training with Full Augmented Dataset (33k)** to improves generalization for diverse skin tones and lesions

- **Experiment new Augmentation Techniques (to address class/skintone imbalance)**
  - Random Erasing: Adds random noise to encourage lesion identification even with occlusion
  - Mixing Different Lesions: Combines lesions to create new training examples
  - Try other technique for better skin tone transformation

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Questions?

Reference

1. **Venu Gopal et al. (2023):** "An Integrated Deep Learning Model for Skin Cancer Detection Using Hybrid Feature Fusion Technique." (Based on the arXiv preprint with similar title and year).
2. **Rahman et al. (2021):** "A Method for Detecting Skin Cancer Disease Based on Deep Learning in Dermoscopic Images."
3. **Gouda et al. (2022):** "Skin Cancer Diagnosis Utilizing Hybrid Discrete Cosine Transform and High-Performance Convolutional Neural Networks."
4. **Su et al. (2024):** "Early Stage Skin Cancer Detection Using Deep Learning: A Comprehensive Model for Improved Treatment Outcomes and Survival Rates."
5. **Oumoullyte et al. (2023):** The provided information is limited. It's possible this refers to a study using CNNs and transfer learning for skin feature extraction, but a specific publication title is not evident from the table.
6. **Nawaz et al. (2025):** "Optimizing Deep Learning for Skin Cancer Classification: A Computationally Efficient CNN with Minimal Accuracy Trade-Off." (Based on the arXiv preprint with similar title and year).
7. **Ha et al. (2020):** "Skin Cancer Disease Detection Using Transfer Learning Technique."
8. **Nawaz et al. (2023):** Explainable Disease Classification: Exploring Grad-CAM Analysis of CNNs and ViTs