Classifying Skin Lesions

Brian Xu, Caroline Xu, Evan Zhang

12/16/2025

Team 15

Motivation

- 104,960 estimated new Melanoma skin cancer cases in 2025
- The skin is the largest organ of the human body, as it shoulders the responsibility of protecting all major systems
- Achieving automatic skin cancer classification is difficult because the majority of skin disease images used for training are imbalanced and in short supply; meanwhile, the model's cross-domain adaptability and robustness are also critical challenges.

Background

- Skin lesions are unusual lumps of skin (not always cancerous)
  - Classifying skin lesions at an early stage could aid clinical decision-making by providing an accurate disease diagnosis, potentially increasing the chances of cure before cancer spreads.

Using ML for medical imaging is not new

- MRIs - Alzheimer's
- Lung Scans - Pneumonia

Using ML algorithms in healthcare diagnoses is useful as a:

- More objective way of making conclusions about data
- More efficient way of having someone individually look through a set of scans

Related Work

https://doi.org/10.48550/arXiv.2409.03794

Evaluated two CNN architectures using Keras on the Kaggle HAM10000 dataset. Model 1 used MobileNet1 while model 2 was a custom CNN. Both models were evaluated twice, once attempting to distinguish between each lesion type, and once only to distinguish between malignant and benign lesions. Evaluated both accuracy, fairness, and feature recognition from both models.

https://doi.org/10.48550/arXiv.2305.11125

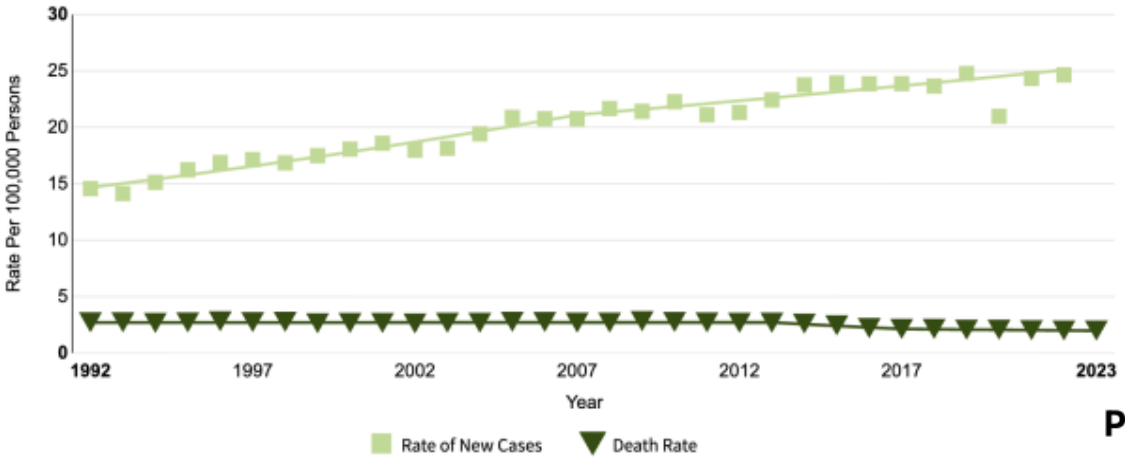
Benchmarked three architectures on the same HAM10000 dataset. Evaluated VGG16, ResNet50, and DenseNet121 with data augmentation before training. Generated confusion matrices for each model to examine which classes were most similar to the models, as well as precision and recall.

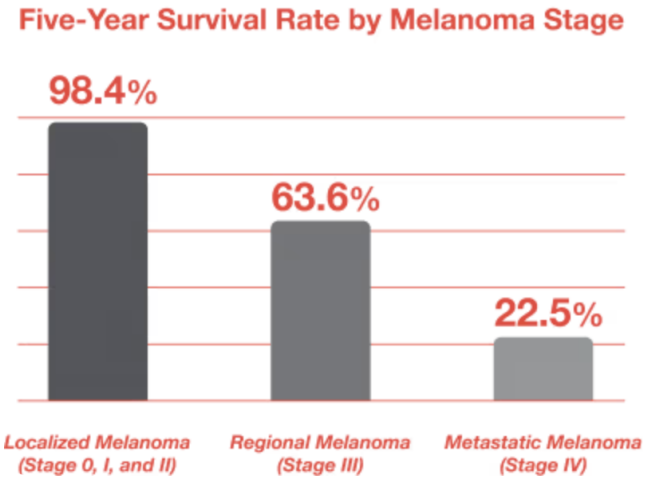
Claim / Target Task

This project proposes an **EfficientNet** to detect skin cancer using the provided Kaggle dataset, thereby achieving a better predictions compared to ResNet and other previously created models.

Given the results of ResNet in the related work on the HAM10000 dataset, we want to hit a target above ResNet's prediction rate with fewer trainable parameters.

An Intuitive Figure Showing WHY Claim





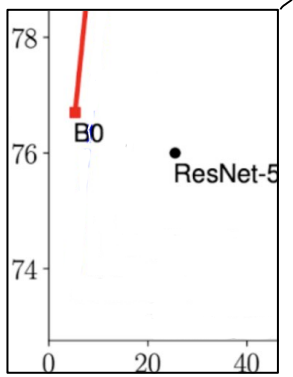
Performance Metrics

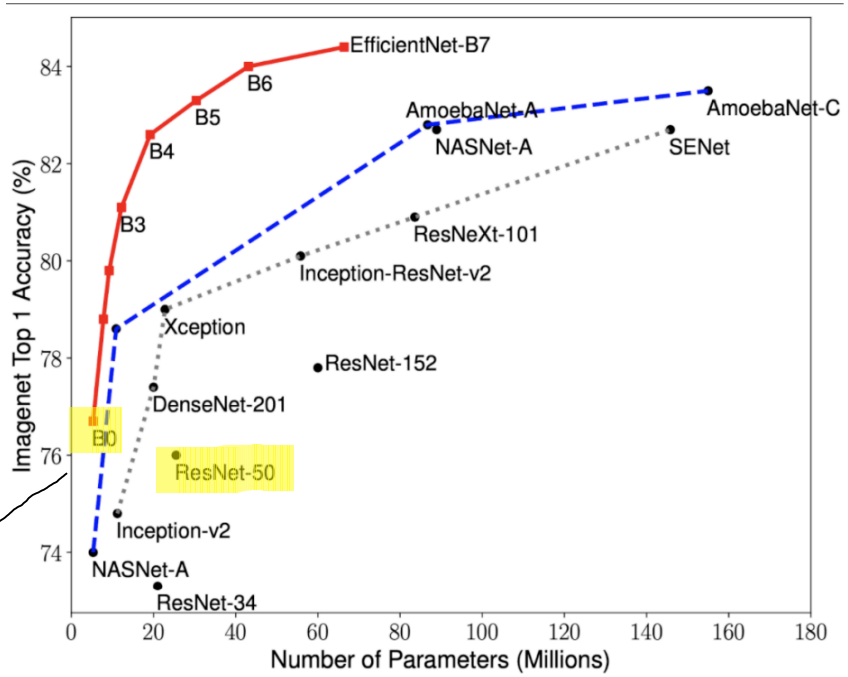
Metric	EfficientNet (B0-B7)	ResNet (50-152)
Accuracy (ImageNet)	77.1% (B0) to 84.4% (B7)	76.2% (ResNet-50) to 78.3% (ResNet-152).
Inference Speed	Faster due to fewer FLOPs for comparable accuracy.	Slower for deeper variants (ResNet-101/152).
Model Size	Smaller models (EfficientNet-B0: ~5MB).	Larger models (ResNet-50: ~98MB).
FLOPs	Significantly lower (~0.4B for B0, ~37B for B7).	Higher (~4B for ResNet-50, ~11B for ResNet-152).

Evaluated on CIFAR-10

Proposed Solution

EfficientNet is a model that has shown to be more accurate with cheaper architecture compare to models like ResNet. We propose to train an EfficientNet model on the HAM10000 dataset to benchmark its performance against the two previous studies mentioned. Since both studies showed both accuracy and recall, we have two metrics to compare to.





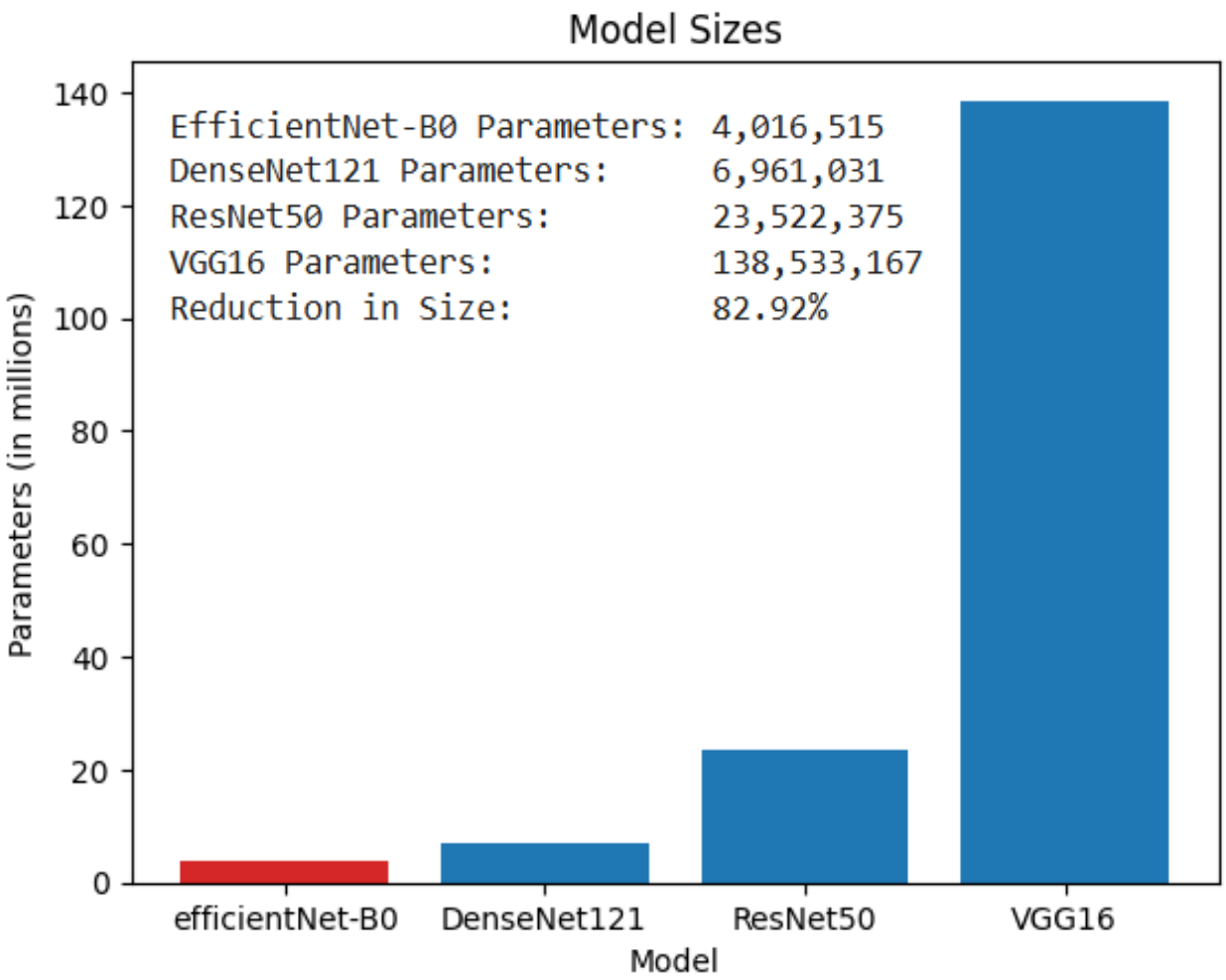
Implementation

Our model:
- EfficientNet
- 4 million params

Comparing to best model in previous study:
- ResNet
- 25 million params

Total size reduction:
- 82.92%

Model size different from normal due to in-out layers

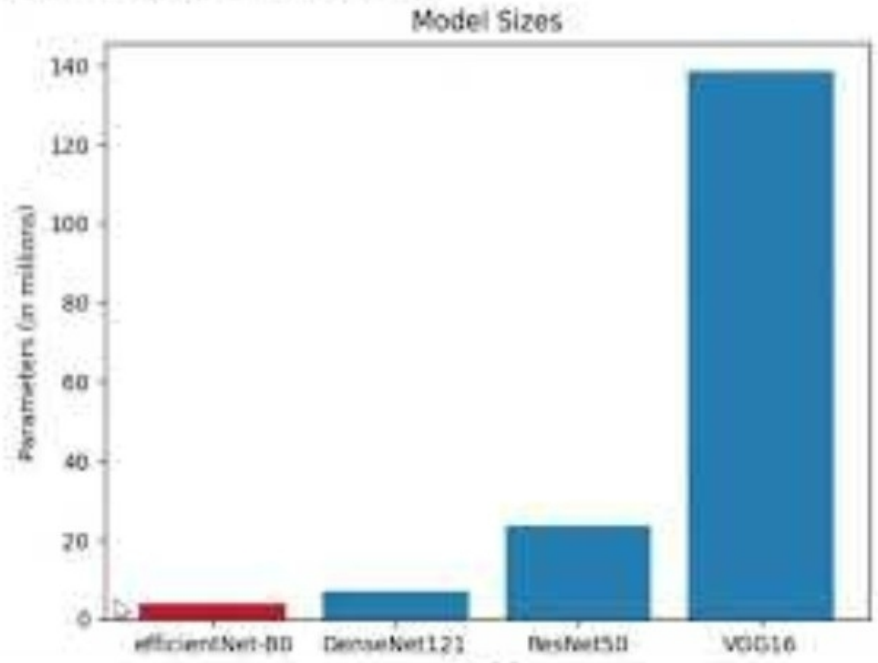


Implementation

- Followed paper as closely as possible to minimize differences just to model behavior
  - Data splitting
    - No k-fold, 80% training 20% validation, no test set, validation set used as test set
  - Batch size 64, 100 epochs
    - Due to Colab time limit only 34 epochs could be run before timing out even with GPU cuda acceleration
  - Data augmentation followed the previous paper as closely as possible
    - Resize to (460,460)
    - Composition of random augmentations
      - Random flip
      - Random rotation
      - Random color jitter
    - Resize to (224,224) <- this is native input size of efficientnet
    - Normalize

Implementation (Code Demo)

```
EfficientNet-B0 Parameters:  4,016,515
DenseNet121 Parameters:      6,961,031
ResNet50 Parameters:         23,522,375
VGG16 Parameters:            138,511,167
Reduction in Size:           82.92%
--------------------------------------------------
Using EfficientNet-B0 for training...
```

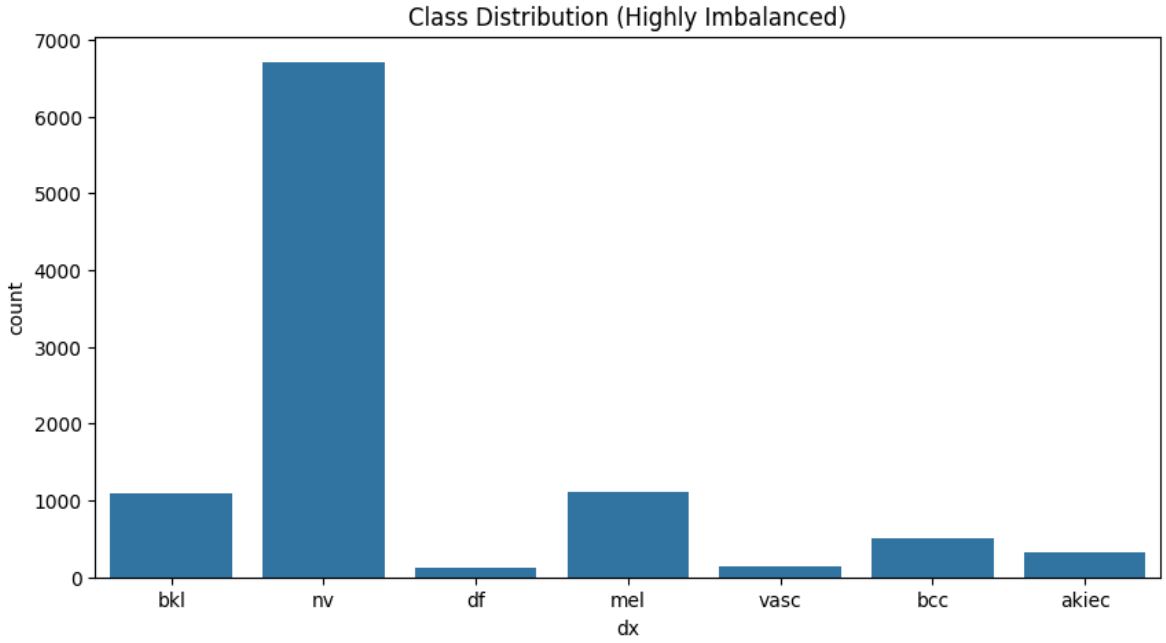


https://youtu.be/dWKxdcSMWO4

Data Summary

10,015 total images -> 80% training, 20% validation

	lesion_id	image_id	dx	dx_type	age	sex	localization	path	label
0	HAM_0000118	ISIC_0027419	bkl	histo	80.0	male	scalp	/content/drive/My Drive/Colab Notebooks/kaggle...	2
1	HAM_0000118	ISIC_0025030	bkl	histo	80.0	male	scalp	/content/drive/My Drive/Colab Notebooks/kaggle...	2
2	HAM_0002730	ISIC_0026769	bkl	histo	80.0	male	scalp	/content/drive/My Drive/Colab Notebooks/kaggle...	2



Experimental Results

Our model

EfficientNet-B0			
	Precision	Recall	F1-Score
akiec	0.85	0.69	0.76
bcc	0.78	0.83	0.80
bkl	0.81	0.77	0.79
df	0.79	0.65	0.71
mel	0.67	0.84	0.75
nv	0.95	0.92	0.94
vasc	0.71	0.89	0.79
Accuracy			0.88

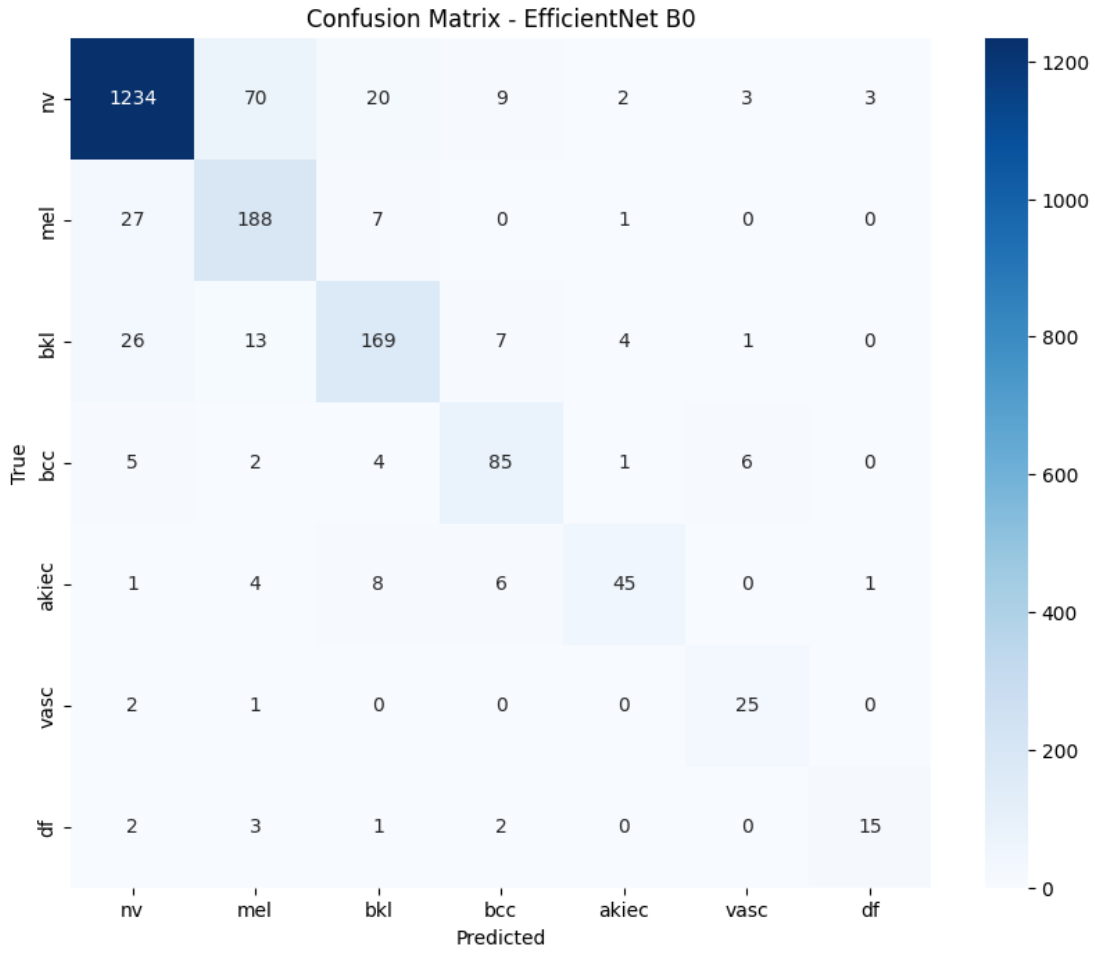
Comparison to best model

ResNet50			
	Precision	Recall	F1-Score
akiec	0.93	0.67	0.78
bcc	0.93	0.91	0.92
bkl	0.86	0.81	0.83
df	0.90	0.82	0.86
mel	0.81	0.74	0.78
nv	0.94	0.98	0.96
vasc	0.93	1.00	0.96
Accuracy			0.92

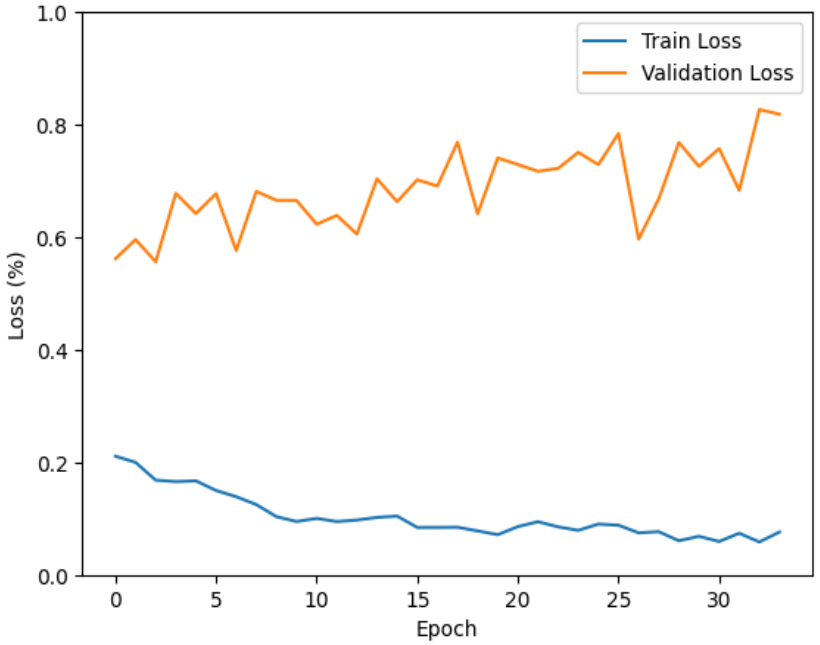
Experimental Results

Classification Report:

	precision	recall	f1-score
nv	0.95	0.92	0.94
mel	0.67	0.84	0.75
bkl	0.81	0.77	0.79
bcc	0.78	0.83	0.80
akiec	0.85	0.69	0.76
vasc	0.71	0.89	0.79
df	0.79	0.65	0.71
accuracy			0.88
macro avg	0.79	0.80	0.79
weighted avg	0.89	0.88	0.88



Experimental Results



Training time: ~3 hours (Interrupted)
Best Epoch:
Epoch [24/10] Train Loss: 0.0791 | Train Acc: 94.35% | Val Loss: 0.7507 | Val Acc: 87.92%

Experimental Analysis

- Accuracy of 0.88 vs 0.92, Macro F1 of 0.79 and F1 of 0.88
  - Benign nevi (nv) class has excellent performance which is expected as it has the most data. This shows EfficientNet is comparable in performance to ResNet despite much smaller size.
  - With Melanoma, EfficientNet has higher positive classifications of malignant cases and false positives, which may be a beneficial trade-off.
- Strong class imbalance causes minority class prediction accuracy to be slightly reduced.

Conclusion and Future Work

- 82.92% reduction in model size with final accuracy of 88%.
  - Lower inference times and memory footprint.
  - Deployment onto smaller devices such as phones.
  - Batch processing of large image volumes.
- Further fine-tuning, class-balanced loss functions, or ensemble approaches can be deployed.
  - Ensemble of multiple EfficientNet models
  - Combining with MobileNet model or Grad-CAM

References

[1]
T. Jain, “Evaluating Machine Learning-based Skin Cancer Diagnosis,” *arXiv.org*, 2024. https://arxiv.org/abs/2409.03794 (accessed Dec. 16, 2025).

[2]
D. Alonso and Y. Li, “Skin Lesion Diagnosis Using Convolutional Neural Networks,” *arXiv.org*, 2023. https://arxiv.org/abs/2305.11125 (accessed Dec. 16, 2025).

[3]
G. Varoquaux and V. Cheplygina, “Machine learning for medical imaging: methodological failures and recommendations for the future.” npj Digit. Med., 2021. https://doi.org/10.1038/s41746-022-00592-y (accessed Dec. 16, 2025).

[4]
Wu, Y., Chen, B., Zeng, A., Pan, D., Wang, R., & Zhao, S. (2022). Skin Cancer Classification With Deep Learning: A Systematic Review. Frontiers in oncology, 12, 893972. https://doi.org/10.3389/fonc.2022.893972

Work Split

- Brian
  - Related Work
  - Proposed Solution
  - Implementation
  - Code Demo
  - Github PR
- Evan
  - Target Task
  - Why Claim
  - Experimental Analysis
  - Conclusion and Future Work
- Caroline
  - Motivation
  - Background
  - Data Summary
  - Experimental Results