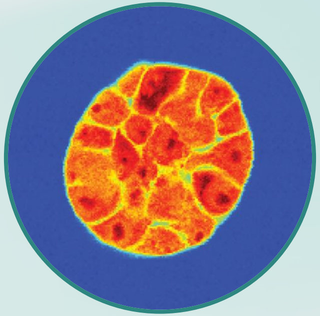
Skin Cancer Detection: An Ensemble Strategy

Ananya Kulshrestha
ananya_k@mit.edu

David Chaudhari
davidc03@mit.edu

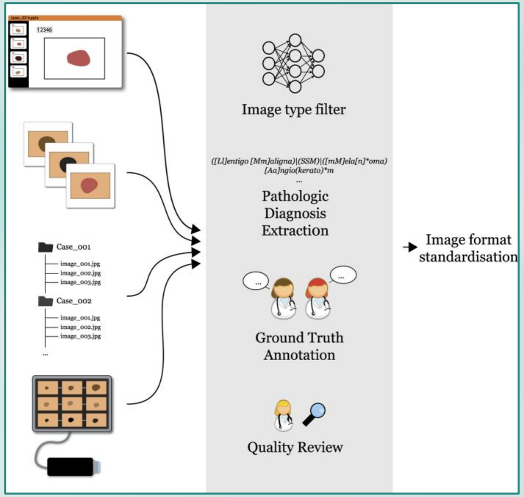
Background/Motivation

- Early detection of Skin Cancer remains a critical challenge
  - Most effective way to increase survival rate
- The project aims to enhance early skin cancer detection using Convolutional Neural Networks (CNNs)
  - We can take advantage of patterns in skin lesions to identify skin cancer more effectively
- We will utilize the ResNet, AlexNet, VGG, and DenseNet models
  - Then we will ensemble the models to get a more robust result
  - We will also use segmentation to examine changes



Relevant Works

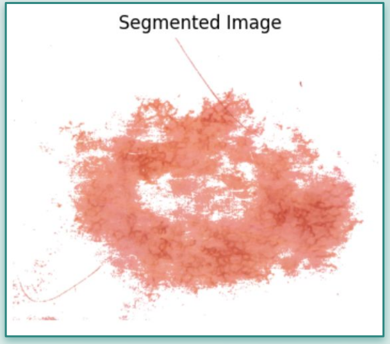
- The primary advancements in skin cancer detection have involved the use of convolutional neural networks (CNNs) [Naqvi et. al]
    - Follows the process image acquisition, preprocessing, segmentation, feature extraction, and classification.
- Segmentation via Contours
    - Have been very successful in medical imaging contexts [Chen et al.]
- AUC (Area Under Curve) [Kadampur et. al]
    - Generally used as the benchmark for model accuracy
- Popular datasets are HAM10000 and ISIC [Wu et. al, Tschandl et. al]



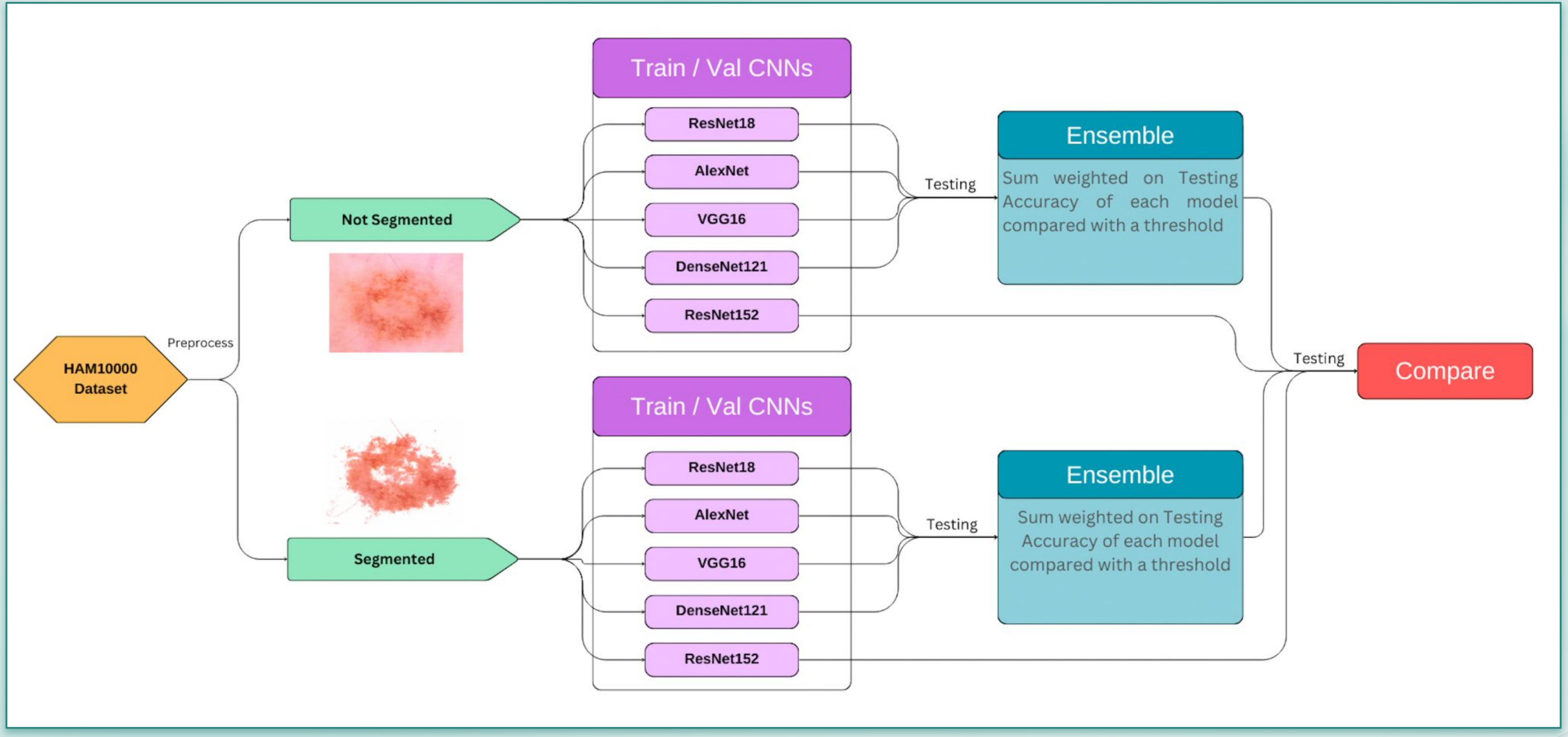
Source:https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6091241/

Segmentation

- We explored whether the background skin color affects results
  - Segmentation approach
    - We segment the lesion part of the image by way of contours
    - Our theory is that segmenting the image will increase the accuracy of the predictions of our models
  - Non-segmentation approach
    - We feed the original image from our dataset straight to our models for training

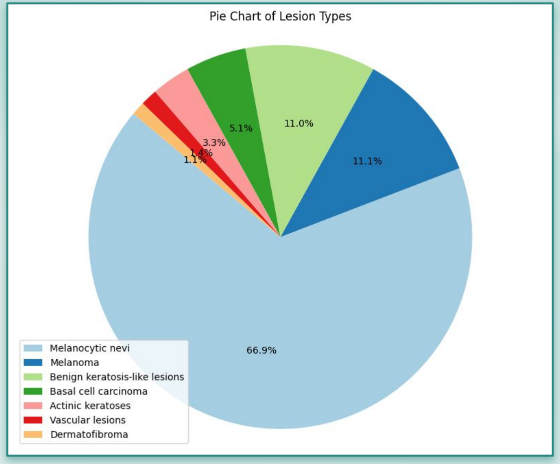


Methodology



Methodology contd.

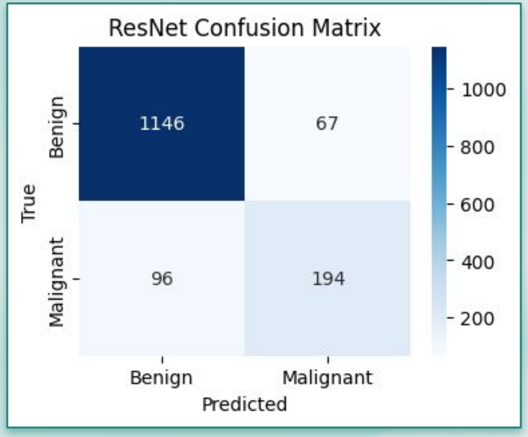
- Took HAM10000 Dataset, relabeled to have two classes: Benign, Malignant
  - Split the Data into **Train**: 70%, **Val**: 15%, **Test**: 15%
- We set our parameters to be same across different models
  - Batch size = 32
  - Learning Rate = 0.001
  - Momentum = 0.9
  - Epochs = 25
- Criterion: Cross Entropy Loss
- Optimizer: Stochastic Gradient Descent
- Ensemble Strategy
  - threshold = 0.5

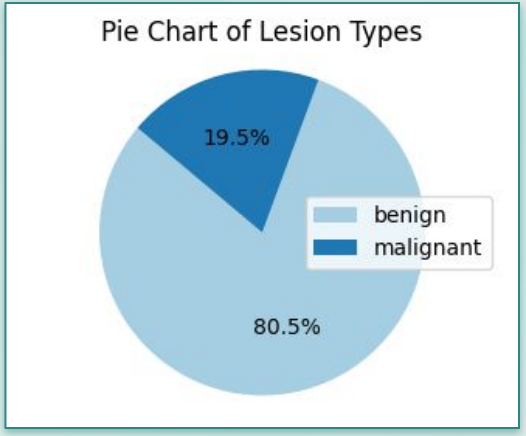


Experiments

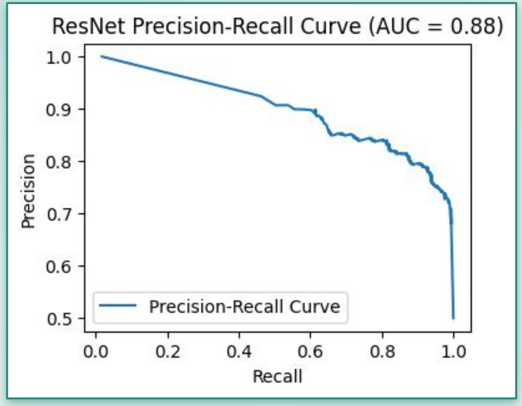
- Ran ResNet to see results
  - Testing Accuracy: 89.16
  - Benign good, Malignant terrible
  - Data imbalance

- Kept the parameters same
- Handled Data imbalance
  - Undersampled: randomly sampled from Benign images
  - Made new dataset 50-50

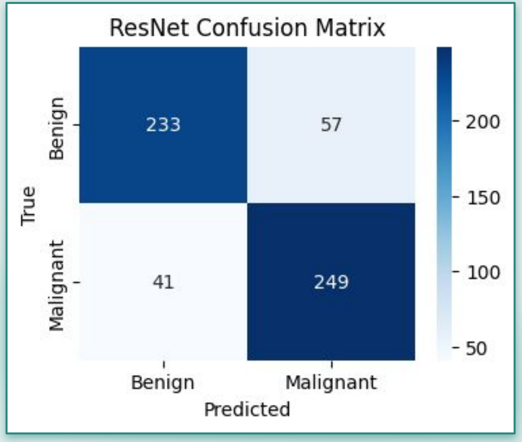




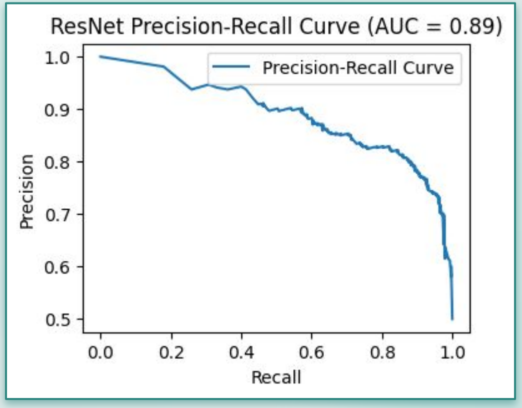
ResNet



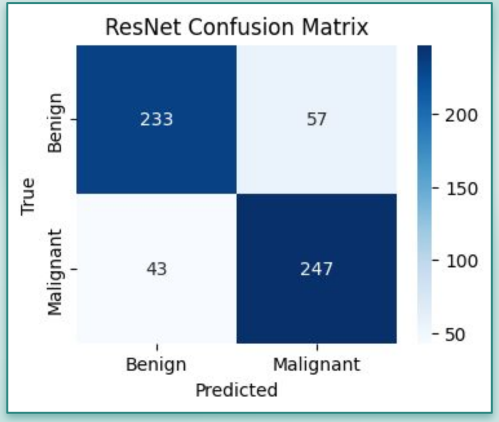
Testing Accuracy: **0.8310**



Segmented

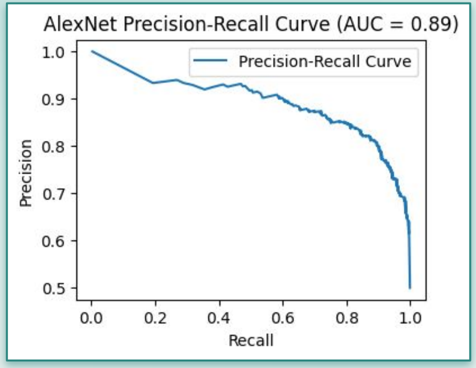


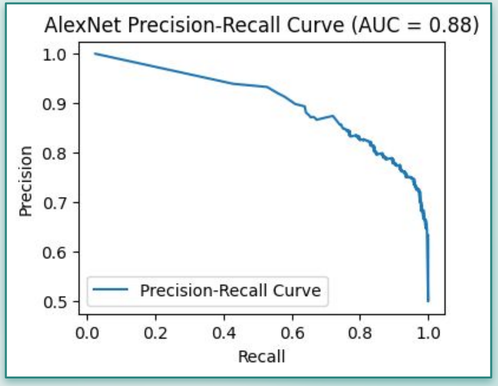
Testing Accuracy: **0.8276**



Not Segmented

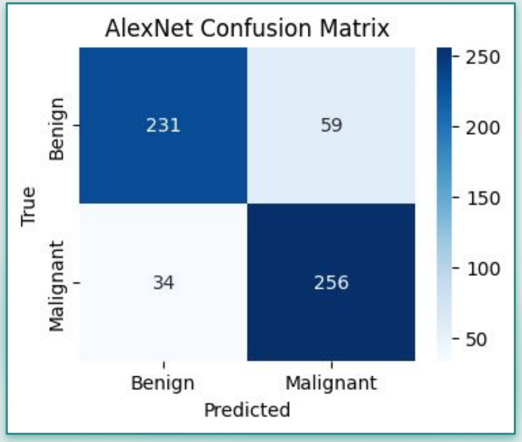
AlexNet



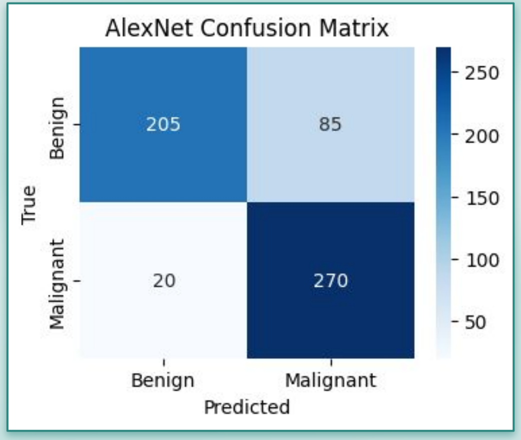


Testing Accuracy: **0.8397**

Testing Accuracy: **0.8190**

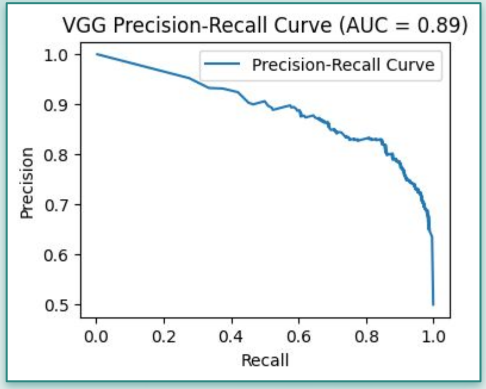


Segmented

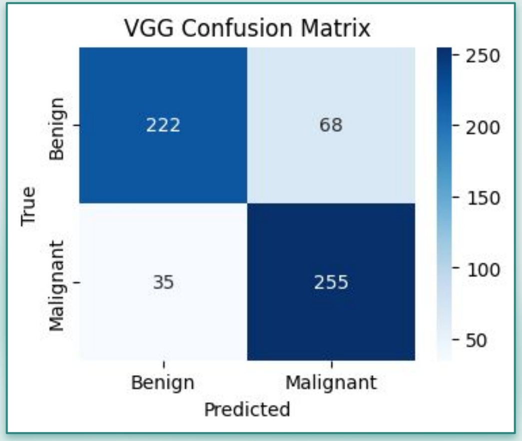


Not Segmented

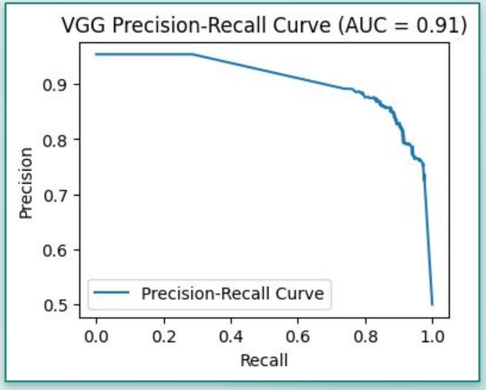
VGG



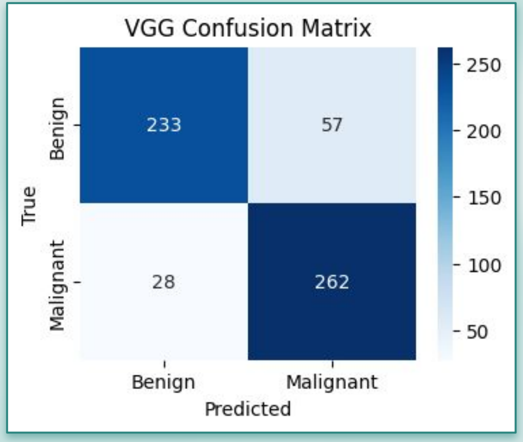
Testing Accuracy: **0.8224**



Segmented

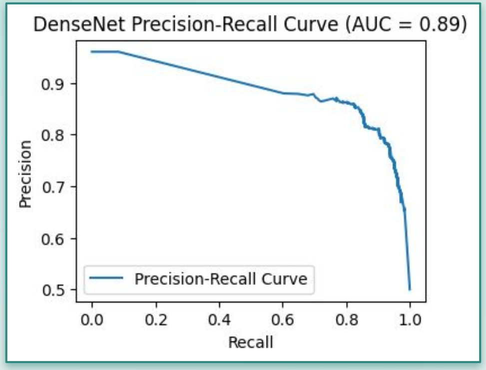


Testing Accuracy: **0.8534**

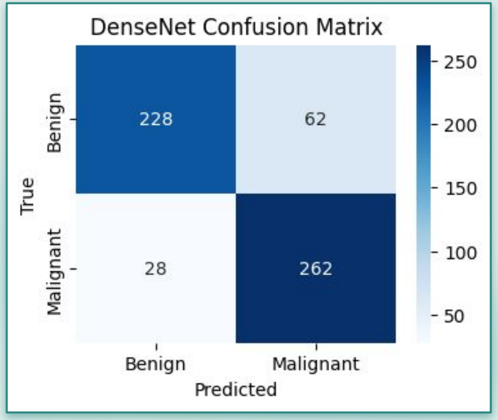


Not Segmented

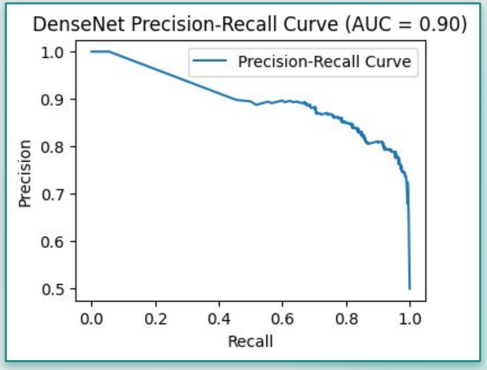
DenseNet



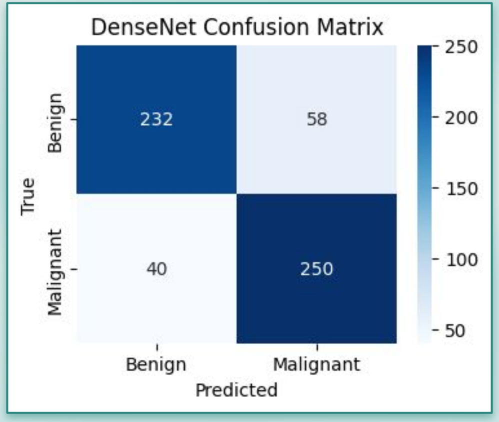
Testing Accuracy: **0.8448**



Segmented

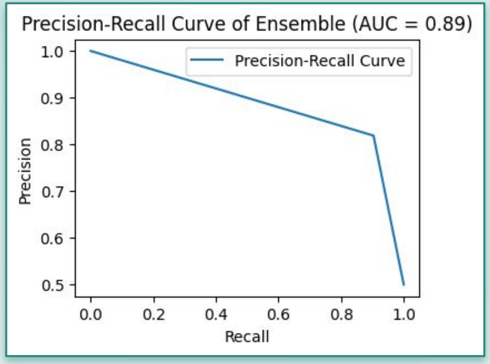


Testing Accuracy: **0.8310**

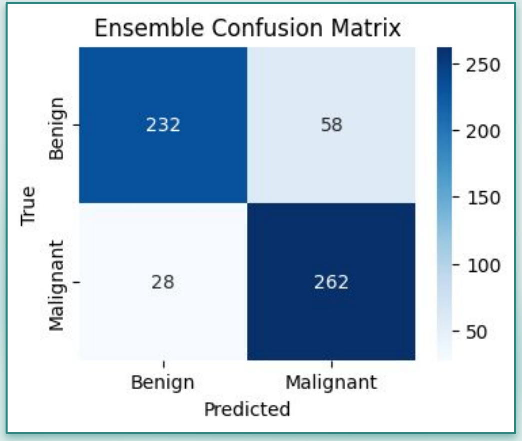


Not
Segmented

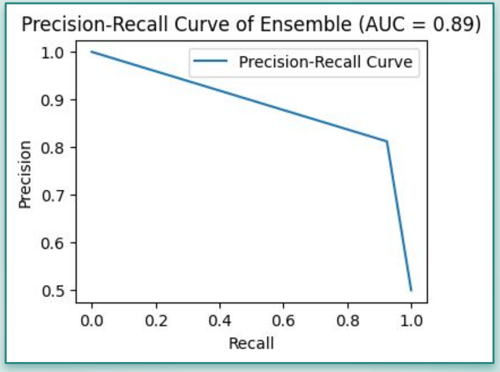
Ensemble



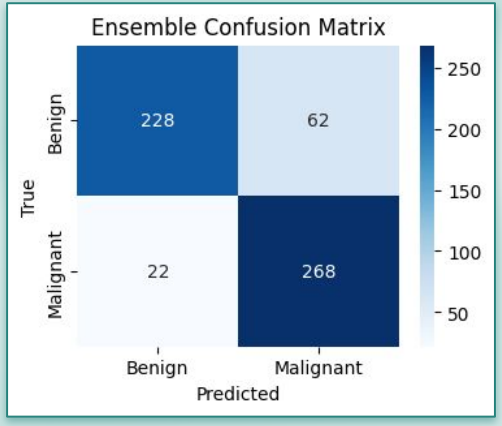
Testing Accuracy: **0.8517**



Segmented



Testing Accuracy: **0.8552**



Not Segmented

Conclusions

- Ensemble method performed better than the individual models for both segmented and non-segmented
    - However, accuracy between both is very similar
- Models generally classify malignant tumors more successfully compared to benign tumors
    - Could be because the models catch an underlying pattern within the malignant samples
- Limitations:
    - Data Imbalance: can use GAN to oversample data
    - Weaker Ensemble Method
    - Model Choices
- In the paper we will explore ResNet152 for further analysis

References

- Naqvi et. al, Skin cancer detection using deep learning—a review, 2023. https://www.mdpi.com/2075-4418/13/11/1911
- Kadampur et. al, Skin cancer detection: Applying a deep learning based model driven architecture in the cloud for classifying dermal cell images, 2020. https://www.sciencedirect.com/science/article/pii/S2352914819302047
- Tschandl et. al, The ham10000 dataset, ..., 2014. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6091241/
- Wu et. al, Skin cancer classification with deep learning: A systematic review, 2022. https://www.frontiersin.org/journals/oncology/articles/10.3389/fonc.2022.893972/fullB78

Ananya Kulshrestha
ananya_k@mit.edu

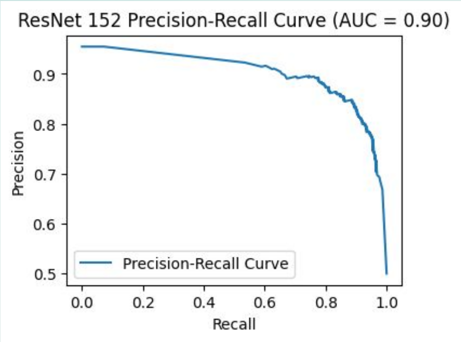
David Chaudhari
davidc03@mit.edu

Report Grading Rubric: Please refer to this link. Presentation Grading Rubric (/100)

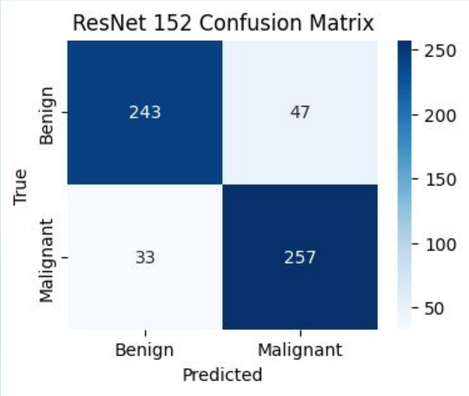
Note: Since presentations are holistic, we won't assign percentages to the individual sub-points.

- Clarity (/40)
  - Was your presentation easy to follow? Was it clear to someone who took 6.8300/1 but isn't familiar with the particulars of the project?
  - Were your slides clear, concise, and well organized? Did they get the message across?
  - Was your talk supported by meaningful visuals with a clear message? Were they easy to understand and legible?
  - Did you articulate your thoughts effectively? Did you speak clearly and at a good pace to ensure everyone can follow your talk?
  - Was the delivery of your presentation well organized? Did you cover all your slides within the time limit without rushing through them?
    - For teams of two: Was the coordination between joint speakers handled smoothly? Was the presentation divided up effectively and the presentation time well balanced between joint speakers?
- Content (/60)
  - Intro/Motivation & Related work: What are you trying to do? How is it done today, and what are the limits of current practice?
  - Methodology: What is your approach, how is it different from previous work, and why is it better?
  - Results & Discussion: How did you evaluate your approach or conduct your analysis? Which experiments did you run to support your claims? What other experiments will you run and why? What are the limitations of your approach or analysis? (As mentioned above, not having your results ready for the presentation is perfectly fine. In that case, please make sure to discuss the experiments you will run and what you think they will demonstrate.)

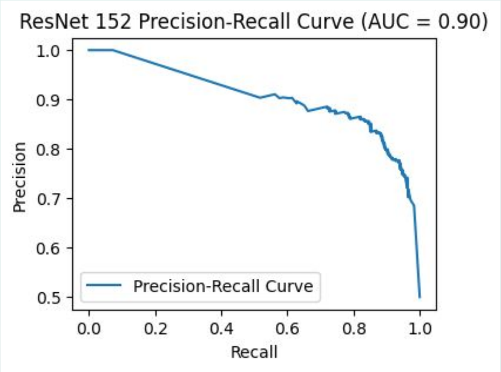
ResNet 152



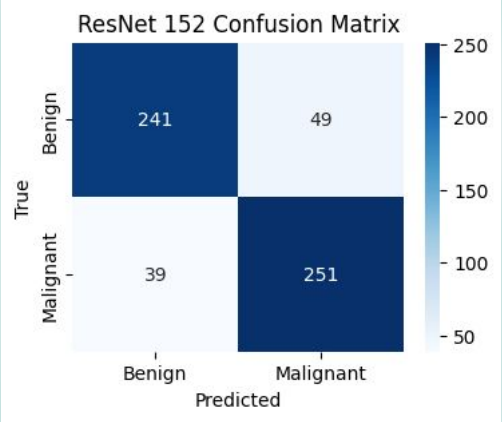
Testing Accuracy: **0.8621**



Not Segmented



Testing Accuracy: **0.8483**



Segmented