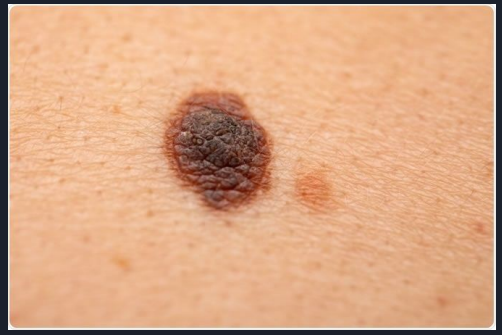
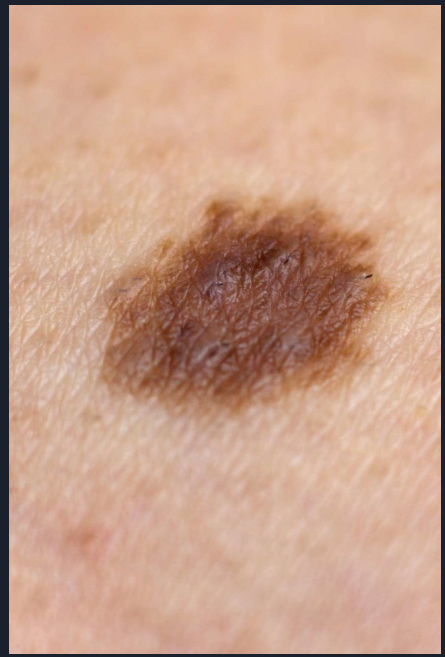
Melanoma Classification Using DenseNet

Ramiz Akhtar (rsa5wj), Rayaan Faruqi (raf9dz), and Kunaal Sarnaik (kss7yy)
October 11th, 2020

Motivation

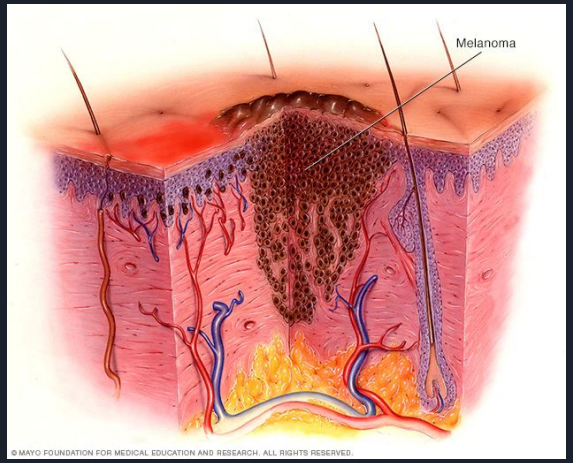


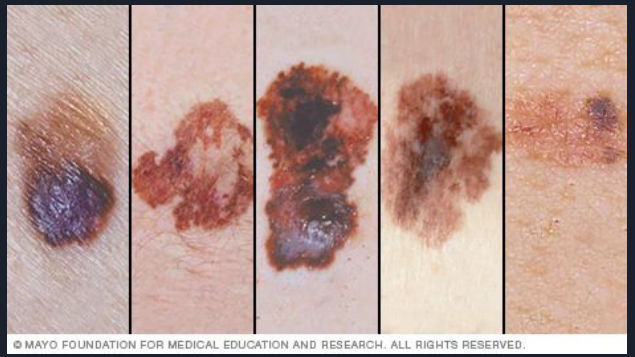
- **What is the difference between the two images on the right?**
  - Top - Melanoma
  - Bottom - Typical skin mole
  - *Extremely difficult to differentiate which one is which*
- An estimated 76,380 Americans will be diagnosed with an invasive melanoma[1]
  - While melanoma is less common than other skin cancers, it is more likely to metastasize relative to other types of skin cancers.[2]
  - *What if there was a way to classify this early on?*
    - *In the safety and comfort of one's own home (i.e., without having to obtain an expensive clinical diagnosis in person)*
- Providing such diagnostic information via a smartphone applications makes early detection both more feasible, accessible, and economical!
  - Higher likelihood of successful treatment outcome(s)
  - Ease of Use, Operability, and Accessibility
  - Raises awareness of a serious disease that affects the lives of many



Background

- What is Melanoma?
  - Per the Mayo Clinic, melanoma is the most serious type of skin cancer[3]
  - Develops in the melanocytes → The cells that gives one's skin its pigmentation
- Why is hard to diagnose?
  - Can develop anywhere on the body
  - Most melanoma cells still make melanin → usually brown or black
    - However, some melanomas do not make melanin → can appear pink, tan, or white
- What are the clinical, financial, and psychosocial implications of a late diagnosis?
  - Clinical → Harder to manage melanoma progression in later stages
  - Financial → Higher cost of emergent operations
  - Psychosocial → Withdrawal from activities, avolition, and dysphoria



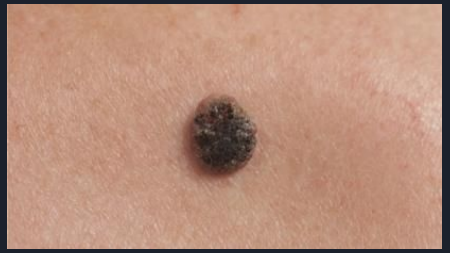


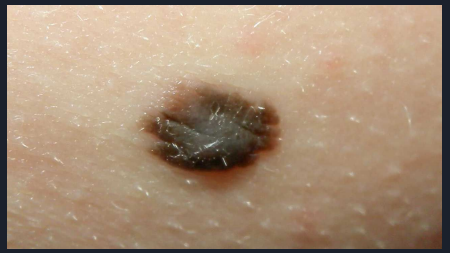
Related Work

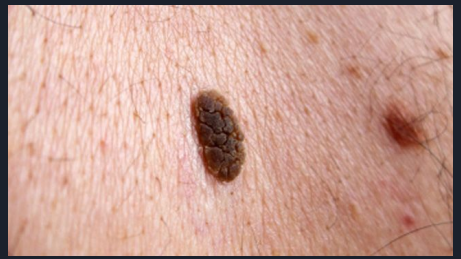
- Esteva et al. demonstrated dermatologist level classification of skin cancer with machine learning techniques, specifically a deep convolutional neural network (CNN)[4]
- In addition to disease classification via images, as mentioned above, Chan et al. lists[5]
  - Disease classification using dermatopathology images
  - Assessment of skin disease using mobile applications and personal monitoring devices
  - Facilitation of larger scale dermatological and epidemiological research
  - Precision medicine
- Substantial research has been funded to classify breast cancer tumors based on size and malignancy
  - Ray et. al. utilized a supervised machine learning approach to classify BC tumors based on both image and numeric datasets[6]
  - Dhahri et. al. optimized automated BC tumor classifiers using kNN, Random Forest, AdaBoost, etc.[7]
- Relevant melanoma classification models based on *numeric* data and *genomic* data conducted
  - Current need: *IMAGE DATA!* [8]

Claim/Target Task

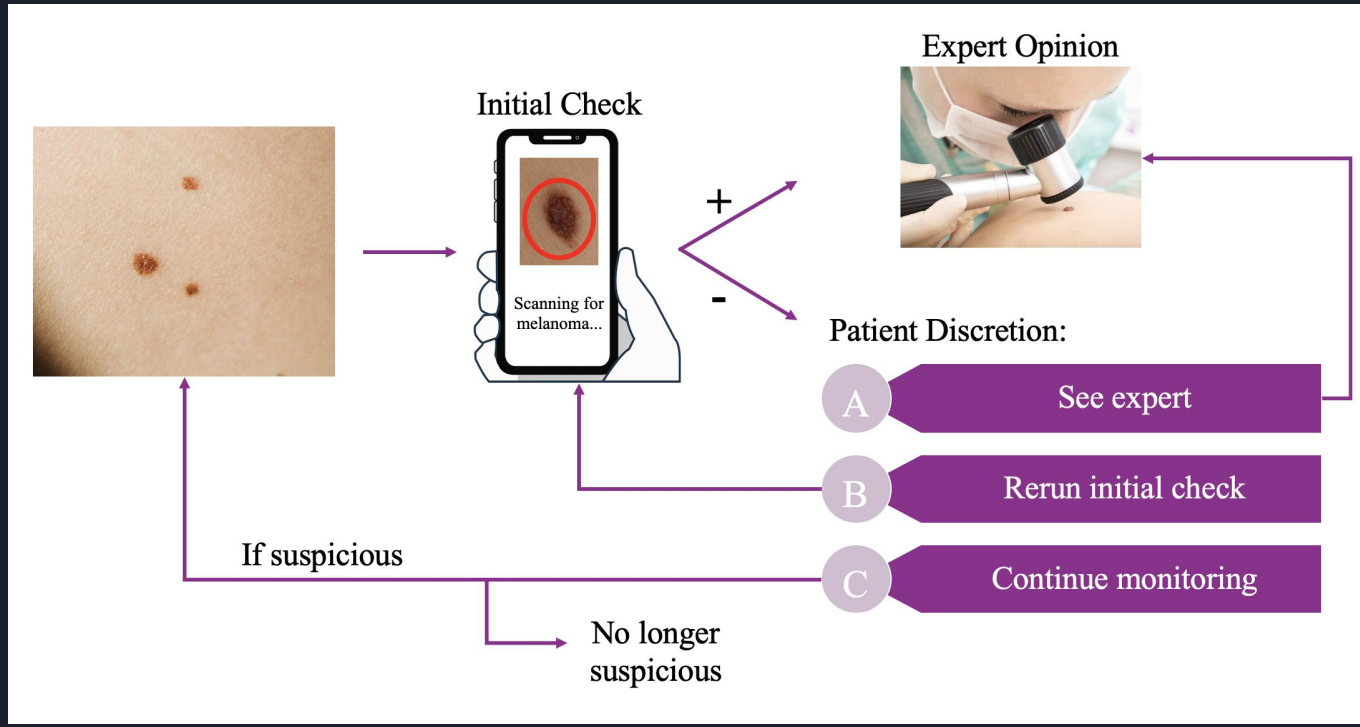
- **Using supervised multi-class classification to train and validate a model that can robustly distinguish the properties of a skin image as either:**
  - Melanoma (top)
  - Non-melanoma skin (nevus) (middle)
  - Seborrheic Keratosis (bottom)
- Dataset being utilized (2750 images of the three outcomes) : https://www.kaggle.com/wanderdust/skin-lesion-analysis-toward-melanoma-detection





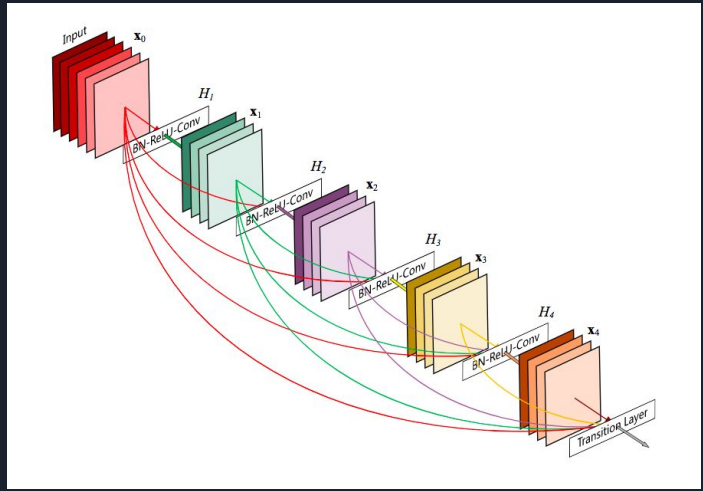


Intuitive Figure



Proposed Solution

- **A DenseNet implemented with Keras using a TensorFlow backend**
  - Consists of dense blocks with internal convolutional layers, followed by a transition layer
  - Dense Blocks: Internal convolutional layers each consecutively connected to one another
  - Convolutional Layer: Batch normalization of data → ReLU activation → ending with a 3x3 convolution. Dropout of 0.2
  - Transition Layer: Another convolutional layer, except ending with a 1x1 convolution instead of 3x3, and ending with a max pooling layer



Example architecture visualization for a single 5-layer dense block with 4 convolutional layers, each consisting of batch normalization, ReLU activation, and 3x3 convolution. Terminates with a transition layer.

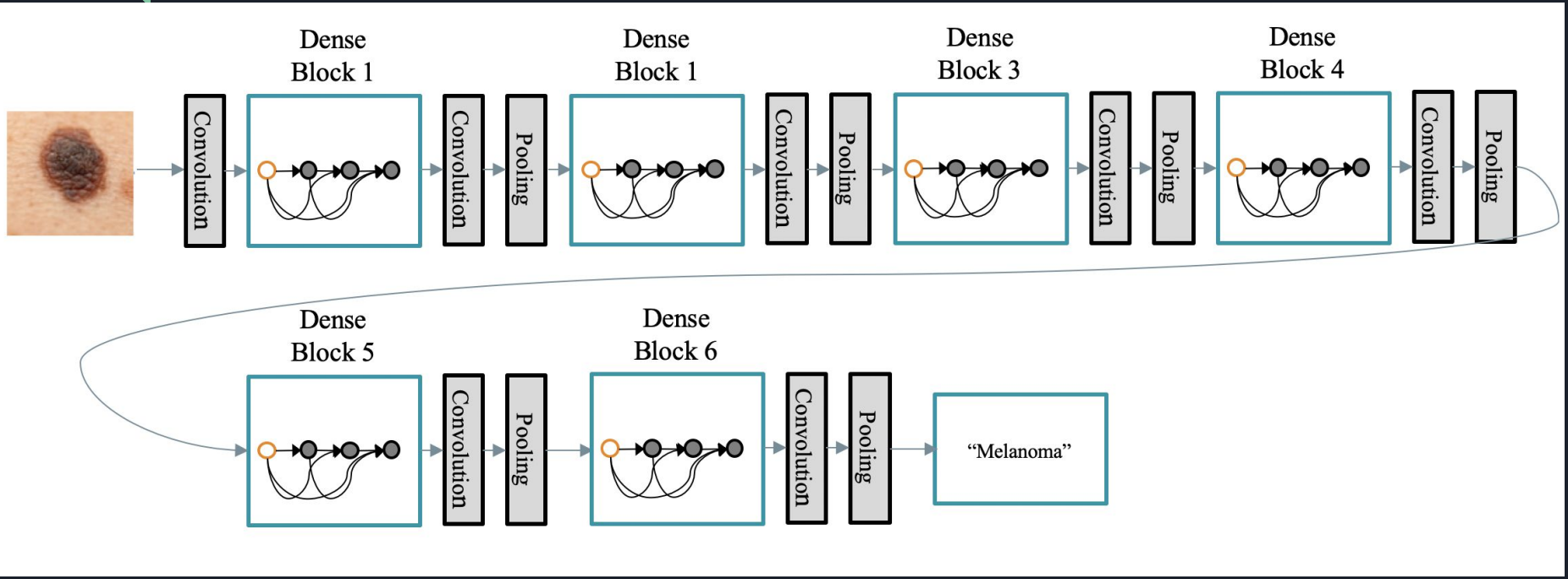
Implementation

- Preprocessing:
  - Images resized → from median size of **3008 x 2000** to **128x128**
    - Extreme downsampling
  - Image pixel normalization → floating point (0, 1)
  - Keras ImageDataGenerator
    - Automatically scales, performs vertical/horizontal flipping, rotation, modifies brightness, and randomly zooms in on images within the dataset[9]

- The Model:
  - 6 dense blocks
  - 4 convolutional layers within each dense block
  - A single transition layer after each dense block, including a 1x1 convolution bottleneck layer
  - Batch size = 12 ; Epochs = 12

- Why DenseNet?
  - Compact internal representations that allow for reduced feature redundancy
  - Each layer receives all preceding layers as input → more diversified features with richer patterns
  - Maintains lower complexity features
    - Smoother decision making boundaries → **performs well with insufficient training data**

Model Architecture

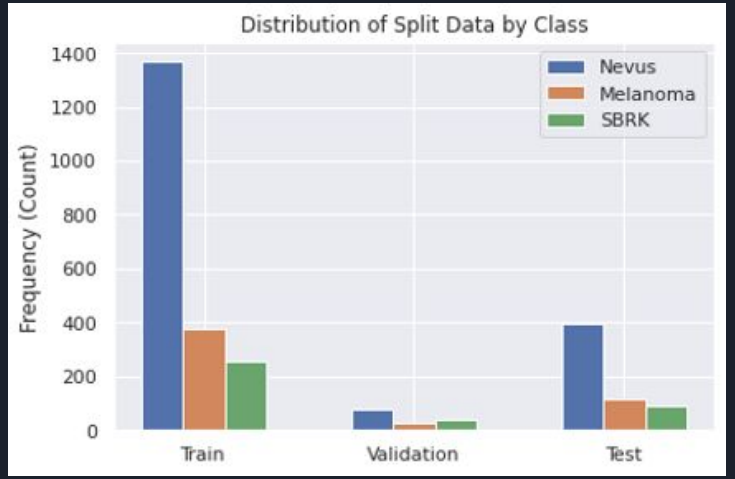


Contributions

- Demonstrates use of a DenseNet for melanoma classification, on a small and disproportionate data set, with highly downsampled images
  - Displays effectiveness of Keras automated data augmentation and DenseNet architecture in compensating for these factors
- Tuned DenseNet architecture's hyperparameters to further advance work on accurately classifying nevus skin, melanoma, and seborrheic keratosis
  - Advancing work on biomedical image classification
- Displayed effectiveness of the cutting-edge DenseNet model versus a more traditional CNN model (see **Appendix**)

Data Summary

- Dataset → 2750 images of the three outcomes
- Provided as a publicly available dermoscopy dataset by the International Skin Imaging Archive (ISIC) in 2018
  - Sourced from “leading clinical centers internationally from a variety of devices used at each center”



Distribution of Split Data by Class (Frequency)			
	Nevus	*Melanoma*	*Seborrheic Keratosis (SBRK)*
Training	1372	374	254
Validation	78	30	42
Testing	393	117	90

Experimental Results

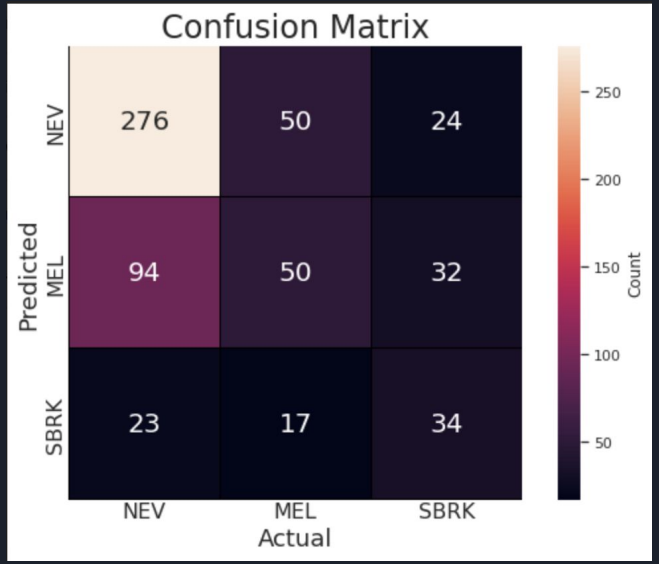
The best results were for a DenseNet architecture with a growth rate of **12**, and hyperparameters as follows:

- Optimizer: **nadam**
- Learning Rate: **1e-05**

- **Validation accuracy: 64.66%**
- **Testing Accuracy: 60%**

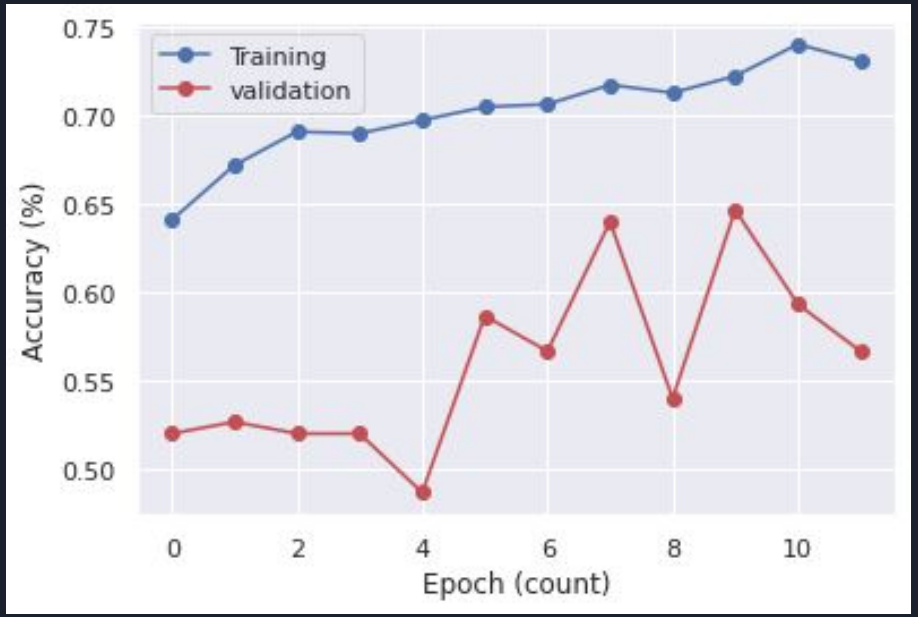
Other DenseNet or traditional CNN architectures were either too low in accuracy or not generalizable enough for practical use (see **Appendix** for results from other architectures)

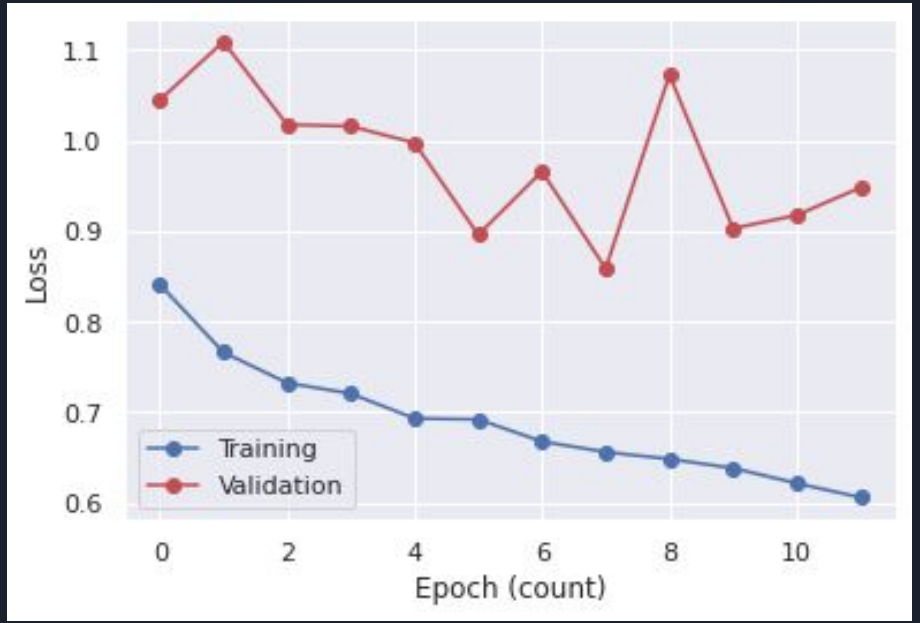
Low MEL and SBRK precision and recall/sensitivity values. Moderately high specificity values for MEL and SBRK.



	ACCURACY	PRECISION	RECALL	SPECIFICITY	F1-Score
NEV	0.68	0.79	0.7	0.64	0.74
MEL	0.68	0.28	0.43	0.74	0.34
SBRK	0.84	0.46	0.38	0.92	0.41

Experimental Results





Experimental Analysis

- The model is not fit for practical use in melanoma classification
  - Model needs additional training and refinement → Low recall, precision, and insufficiently high values for specificity in MEL and SBRK
    - Insufficiently high specificity → model frequently misclassifies image as either MEL or SBRK
    - Low Recall and precision → model frequently misclassified an image (high false negative rate)

- DenseNet model performed admirably given skewed, heavily downsampled dataset and computational power restrictions
  - Traditional CNN model heavily overfit for NEV classification due to overrepresentation of Nevus samples in the training data
  - **DenseNet architecture was somewhat able to compensate for this imbalance in training data.**

Conclusion & Future Work

- Conclusions:
  - DenseNet performed admirably despite computational and dataset limitations
    - Appears well suited for imbalanced data and restricted computational power
  - Traditional CNN severely underfit melanoma and seborrheic keratosis images (see **Appendix**), most likely due to disproportionate number of nevus samples

- Future Work:
  - A UNET model architecture, assigning labels for each pixel in an image
  - An application of DenseNet with a larger and more balanced dataset
    - More computational power (i.e. more GPU power, RAM, etc.) to reduce downsampling of image data

References

1. Glazer AM, Winkelmann RR, Farberg AS, Rigel DS. Analysis of Trends in US Melanoma Incidence and Mortality. JAMA Dermatol. 2017;153(2):225. doi:10.1001/jamadermatol.2016.4512
2. American Cancer Society. What Is Melanoma Skin Cancer? | What Is Melanoma? Published August 14, 2019. Accessed October 11, 2020. https://www.cancer.org/cancer/melanoma-skin-cancer/about/what-is-melanoma.html
3. Melanoma - Symptoms and causes. Mayo Clinic. Accessed October 11, 2020. https://www.mayoclinic.org/diseases-conditions/melanoma/symptoms-causes/syc-20374884
4. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542(7639):115-118. doi:10.1038/nature21056
5. Chan S, Reddy V, Myers B, Thibodeaux Q, Brownstone N, Liao W. Machine Learning in Dermatology: Current Applications, Opportunities, and Limitations. Dermatol Ther. 2020;10(3):365-386. doi:10.1007/s13555-020-00372-0

References

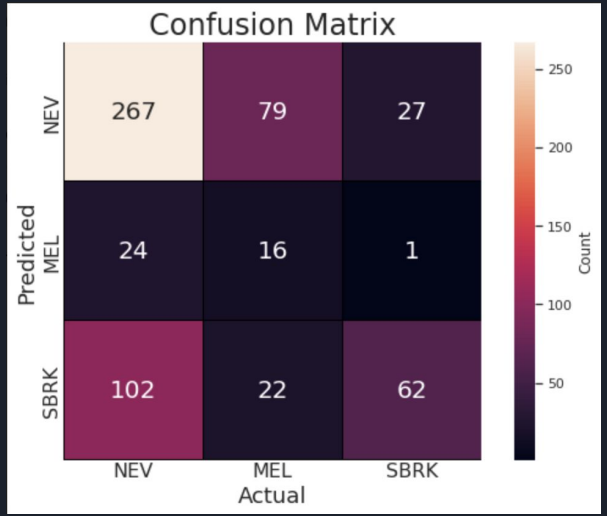
6. Ray R, Abdullah AA, Mallick DK, Ranjan Dash S. Classification of Benign and Malignant Breast Cancer using Supervised Machine Learning Algorithms Based on Image and Numeric Datasets. J Phys Conf Ser. 2019;1372:012062. doi:10.1088/1742-6596/1372/1/012062
7. Dhahri H, Al Maghayreh E, Mahmood A, Elkilani W, Faisal Nagi M. Automated Breast Cancer Diagnosis Based on Machine Learning Algorithms. Journal of Healthcare Engineering. doi:https://doi.org/10.1155/2019/4253641
8. Genomic Classification of Cutaneous Melanoma. Cell. 2015;161(7):1681-1696. doi:10.1016/j.cell.2015.05.044
9. Brownlee J. How to Configure Image Data Augmentation in Keras. Machine Learning Mastery. Published April 11, 2019. Accessed December 3, 2020. https://machinelearningmastery.com/how-to-configure-image-data-augmentation-when-training-deep-learning-neural-networks/

Appendix: Other Models

These results were for a DenseNet architecture with a growth rate of **16**, and hyperparameters as follows:

- Optimizer: **nadam**
- Learning Rate: **1e-05**

- **Validation accuracy: 66.0%**
- **Testing Accuracy: 57.5%**



	ACCURACY	PRECISION	RECALL	SPECIFICITY	F1-Score
NEV	0.61	0.72	0.68	0.49	0.7
MEL	0.79	0.39	0.14	0.95	0.2
SBRK	0.75	0.33	0.69	0.76	0.45

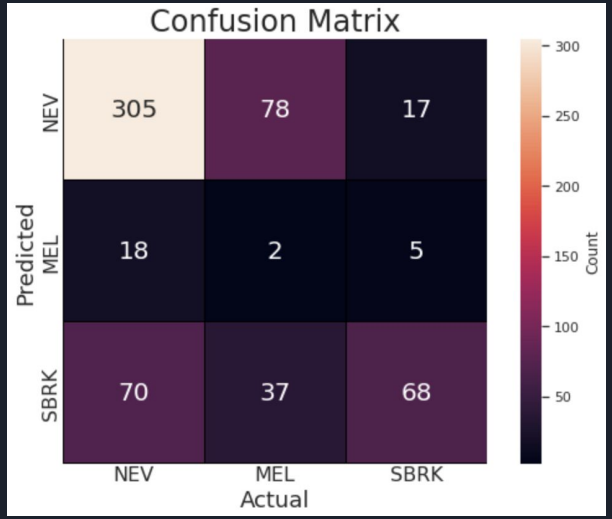
Appendix: Other Models

These results were for a DenseNet architecture with a growth rate of **6**, and hyperparameters as follows:

- Optimizer: **nadam**
- Learning Rate: **1e-04**

- **Validation accuracy: 68.0%**
- **Testing Accuracy: 62.5%**

This model was not chosen to be the best model as it performed extremely poorly in classifying melanoma.



	ACCURACY	PRECISION	RECALL	SPECIFICITY	F1-Score
NEV	0.69	0.76	0.78	0.54	0.77
MEL	0.77	0.08	0.02	0.95	0.03
SBRK	0.79	0.39	0.76	0.79	0.51

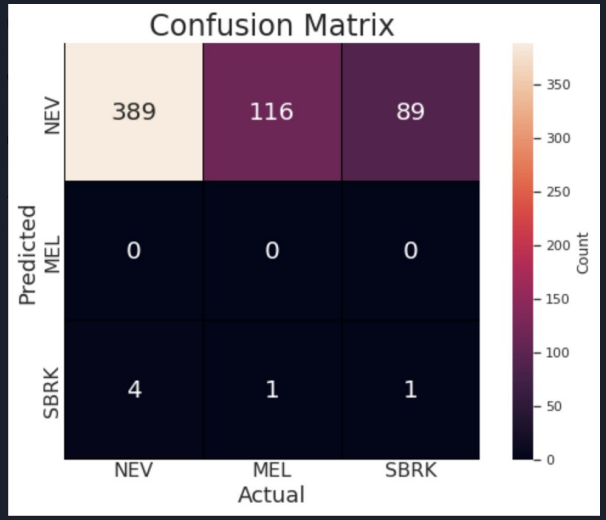
Appendix: Other Models

These results were for a traditional CNN architecture with **batch_size, epochs = 8**, and hyperparameters as follows:

- Optimizer: **adamax**
- Learning Rate: **0.0001**

- **Validation accuracy: 52.66%**
- **Testing Accuracy: 65.0%**

This model was not chosen to be the best model as it was overfit for NEV classifications .



** Statline with specificity, recall, precision, etc. not shown due to floating point division by zero error.*