Classification of Benign and Malignant Melanoma Skin Cancer using Deep Neural Network

Minh Phuong

Abstract

Melanoma skin cancer is one of the less common cancerous diseases, however, if not detected early, could spread and become highly dangerous. In this project, we aim to use Deep Learning framework to investigate and classify images of melanoma skin cancer cells. Using a fully connected layer as the baseline, we compare different architecture of convolutional network and pre-trained network on their performances of benign and cancerous cell classification. Overall, the pretrained models seem to have a better performance.

1 Credits

The dataset was adapted from Kaggle: https://www.kaggle.com/datasets/hasnainjaved/melanoma-skin-cancer-dataset-of-10000-images

This dataset contains 10000 images of benign and malignant melanoma skin cancer photographs.

2 Introduction

Melanoma is a special type of skin cancer. It develops in the cell that produce melanin which creates the color of our skin. One of the main causes for melanoma is UV exposure. Risk of getting melanoma seems to increase in middle-aged people. If detected early, it could be treated well. (C. Halpern et al., 2021).

However, it is a challenge to detect melanoma just by normal eyes. It is often developed around a mole on one's skin and is better observed over time. If the mole develops irregularities such as strange border, invasive color or diameter growth, it is highly melanoma. Risk factors include fair skin, history of sunburn or excessive UV exposure. Populations residing closer to the equator or higher elevation (hence closer to the sun) are also susceptible. (AIM at Melanoma, 2021)

3 Approach

Using the provided dataset, we modeled this project as a binary classification problem. The model should be able to distinguish between benign and malignant melanoma skin cancer with high accuracy. We utilized GPUs on Google Colab for the project.

We first attempted to analyze the image characteristics of the benign and malignant melanoma moles using different techniques. One of them is to create an "average image" of the two classes. There are 9605 files for training and 1000 for testing. The train-validation split is 80-20. This allows for plenty of data for cross validation and testing. Upon closer inspection, the data is also very balanced between the classes.

We then used Keras to construct different network architecture for the classification task. The first model is a fully connected network with drop out. The second model is a convolutional neural network (CNN) model with drop out and batch normalization. Then, we also experimented with pretrained State-of-the-art models such as VGG16 and ResNet50.

4 Data Analysis

We first looked through the images manually to get a sense of what benign and malignant melanoma appeared to be. After several rounds of inspections, it seemed that the malignant moles are darker in color, have raised surface as well as a more "blooming" pattern.

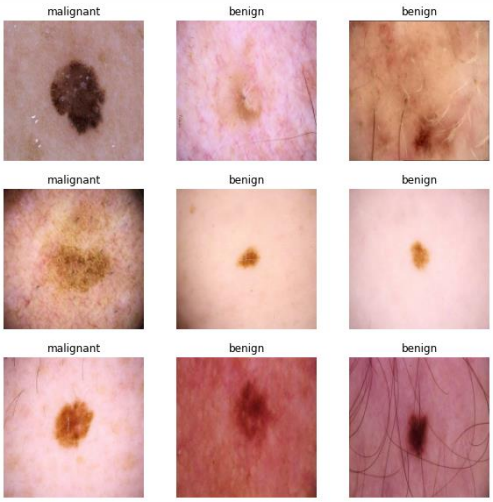

Figure 1: Inspection of malignant and benign melanoma

For a closer look, average image of the two classes could be useful (Byeon, 2021). We created an aggregate image of the melanomas by converting them into grayscale and then Numpy matrices. The matrices are then divided to get the "average image" for malignant and benign moles.

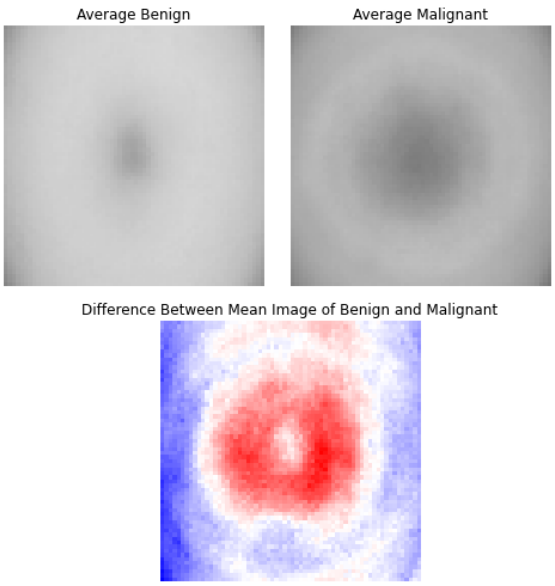

Figure 2: Average images of malignant and benign melanoma

This helped to strengthen our hypothesis that the blooming pattern and larger area of coverage could be the tell-tale signs of malignant melanoma.

5 Experiments

5.1 Data Augmentation

From our data analysis, we can see that moles could be off centered, zoomed in or out. Thus, we created different versions of the training image through Keras built-in data augmentation functions. We utilized Random Flip (both vertical and horizontal), Rotation, Translation and Zoom. The images are also normalized before training.

5.2 Baseline fully connected layers

We built a simple fully connected layers for the baseline of our project. For the fully connected layer, we experimented with 2, 3, 4 layers respectively, with the number of hidden units divided by half as the layer increased. After that, we introduced drop out layers with different ratio from 0.1 to 0.5. We experimented with 20 epochs and keep the highest accuracy.

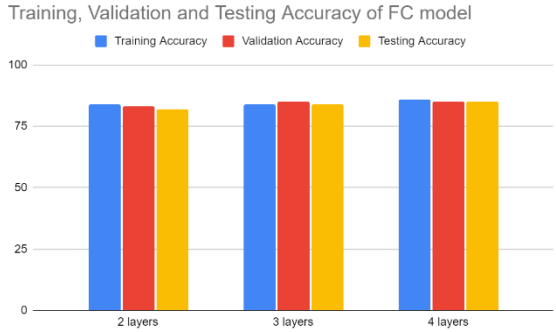

Figure 3: Accuracy of fully connected model

The highest testing accuracy is 87% while the training accuracy is 85%. Dropout rate of 0.2 was the best. However, the loss curve was not stable and overfitting was observed at times. 4 layers gave the best result.

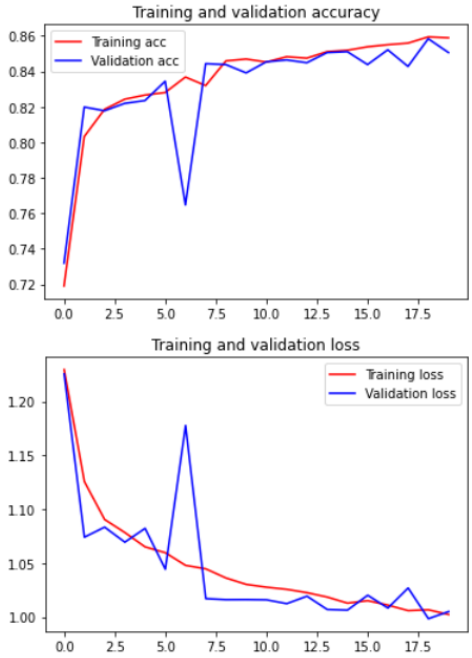

Figure 4: Loss curves of fully connected layers

5.3 Convolutional Network Layers

Subsequently, we experimented with convolutional network architecture. With the convolutional network architecture, we slowly increase each convolutional block while experiment with different regularization parameters. We also introduced batch normalization at each iteration. We observed that without dropout layers, the network performed better than with dropout. Additionally, global average pooling made the model worse off, hence, flattening layer was used instead.

We also experimented with 2, 3 and 4 blocks of convolutional layers, with increasing number of hidden units. 4 blocks was the best performing model

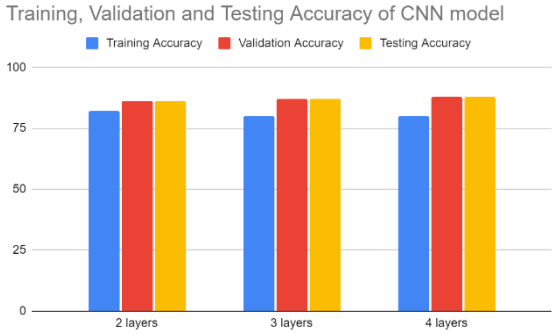

Figure 5: Accuracy of convolutional network model

The highest testing accuracy is 88% while the training accuracy is 80% and validation accuracy is 88%. We can see that the model is not overfitting and the loss curves are more stabilized.

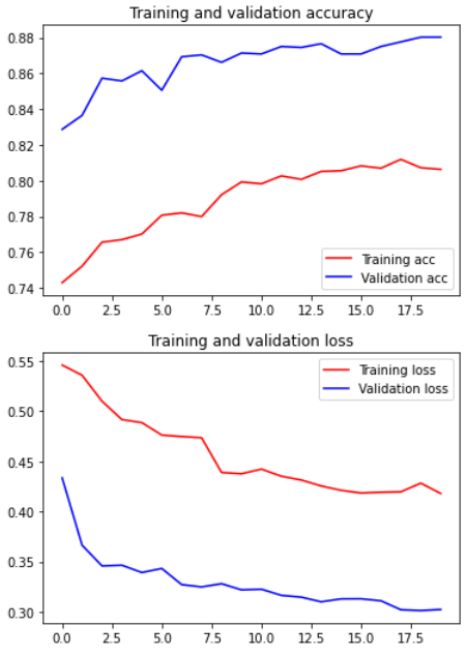

Figure 6: Loss curves of convolutional layers

5.4 Transfer Learning with VGG16

Subsequently, we experimented with different pretrained state-of-the-art models. For the first pretrained model, we are utilizing VGG16.

For VGG16, we first experimented with freezing everything but the last layer. Then we implemented our own fully connected block, referenced from the fully connected architecture. Subsequently, we also fine tune more of the pre-trained networks.

We also experimented with 2, 3 and 4 fully connected layers at the end. In our experiments, we noticed that increasing dropout rate to 0.5 was helping the model performing better. Increasing number of fully connected layers at the end improved the result overall.

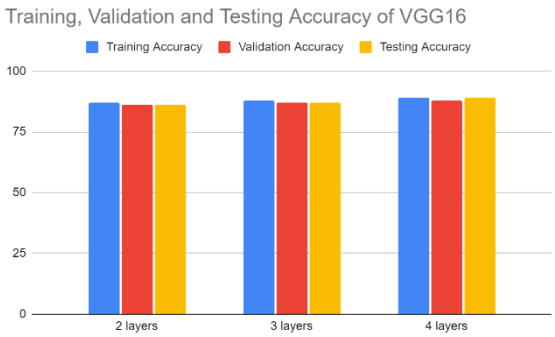

Figure 7: Accuracy of VGG16

With transfer learning on VGG16, the test accuracy is 89%, an improvement from our CNN Model. The training accuracy is 88% while validation accuracy is 89%. However, the loss curve was once again not stable. Some low dips were visible, and overfitting could be a potential issue.

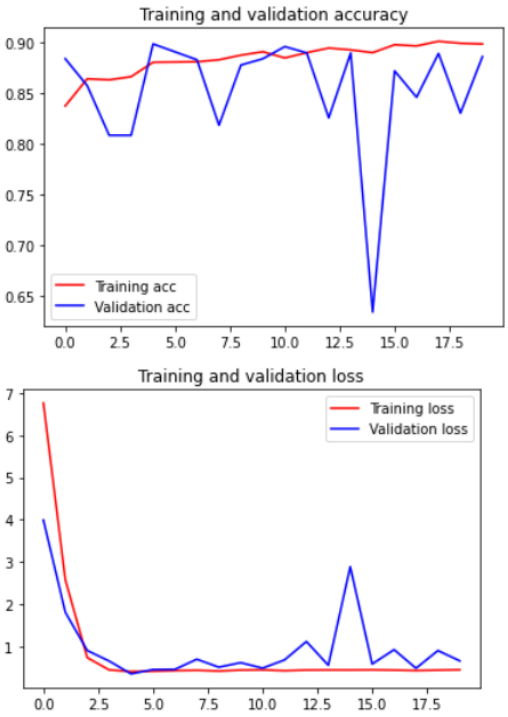

Figure 8: Loss curves of VGG16

5.5 Transfer Learning with ResNet50

Finally, we experimented with transfer learning on ResNet50. We also froze everything but the last layer. Implemented our own fully connected block, referenced from the fully connected architecture in 2,3 and 4 layers respectively.

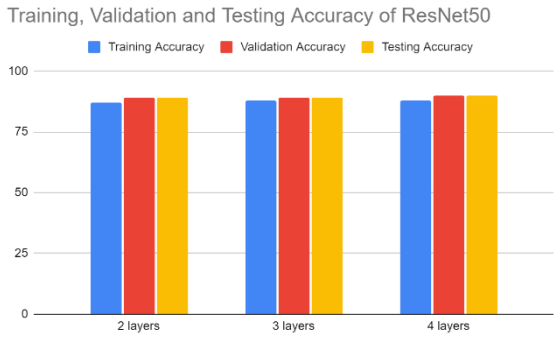

Figure 9: Accuracy of ResNet50

With transfer learning on ResNet50, the test accuracy is 90%, an improvement from VGG16. The loss curves were much more stable and overfitting was not an issue. Thus, this model combined the best performance from all the previous ones. Similar to before, more fully connected layers at the end improved the result overall.

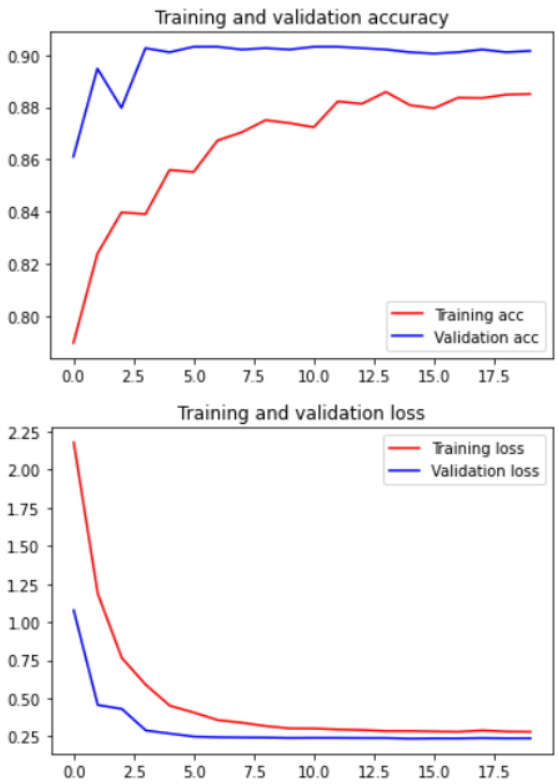

Figure 10: Loss curves of ResNet50

6 Error Analysis

By manually looking at the results, we could see that some misclassification happened when malignant cells are classified as benign. This could be due to their various shapes that are not just in the “blooming” forms. Some images are different

from the rest as well, with a strong vignette outline. Thus, future works could explore on how to make the images more uniform through data manipulation. Oversampling of the malignant class could also be another alternative.

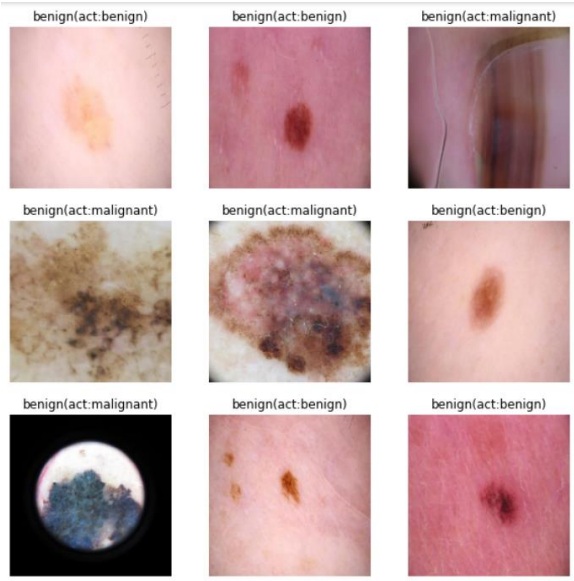

Figure 11: Sample errors made by best performing model (ResNet50)

7 Conclusion

While in practice, diagnosing melanoma skin cancer requires multi-faceted examination, not just from images but also from physical inspection, a preliminary methodology for classifying benign and malignant moles could be highly useful. Using a fine-tuned pre-trained deep learning model, we could potentially create a premise for mobile or web application to help patients predict their risk for melanoma condition early and thus, effective prevention.

In this experiment, we have seen that Pre-Trained Deep Learning frameworks achieved superior results compared to normal handcrafted fully connected or convolutional networks. Moreover, we can see that finetuning pretrained models using our custom convolutional and fully connected layers create the best result.

	Training Accuracy	**Testing Accuracy**
Fully Connected	86 %	85 %
Convolutional	80 %	88 %
VGG16	89%	89%
ResNet50	**88%**	**90%**

Figure 12: Best accuracy of all models

For many of our models, the testing accuracy is higher than the training accuracy, thus we managed to combat overfitting. The pretrained model converges faster, the ResNet50 model has the best performance and loss curves. Depending on our needs, we can still use our convolutional network with significant accuracy. However, it's worth exploring other state-of-the-art models.

Possible next steps can include training on more epochs, looking at more regularization method, experiment with different optimizers and kernel regularizers.

The model can certainly still improve, and it's best to predict on other datasets beside the one provided. Future work could expand on data manipulation and feature extraction before training for more efficient computation.

Acknowledgments

We have referenced the approach of different authors on image analysis and model training for similar medical images dataset. See References.

References

AIM at Melanoma. (2021, January 5). *Melanoma Risk Factors*. AIM at Melanoma Foundation. Retrieved May 1, 2022, from https://www.aimatmelanoma.org/melanoma-101/understanding-melanoma/melanoma-risk-factors/

Byeon, E. (2021, December 15). *Exploratory Data Analysis Ideas for Image Classification*. Medium. Retrieved May 1, 2022, from https://towardsdatascience.com/exploratory-data-analysis-ideas-for-image-classification-d3fc6bbfb2d2

C. Halpern, A., A. Marghoob, A., & Reiter, O. (2021, June 1). *Melanoma risk factors.* SkinCancer.Org. Retrieved May 1, 2022, from https://www.skincancer.org/skin-cancer-information/melanoma/melanoma-causes-and-risk-factors/