# Melanoma and Skin Cancer Detection

Mar 26, 2023 by Akhil Vintas and Jeffrey Yang

Skin cancer is one of the most prevalent cancers in the US and often misdiagnosed or underdiagnosed globally. Benign and malignant lesions often appear visually similar and are only distinguishable after intensive tests. However, deep learning models have proven to be extremely accurate in the classification of melanoma and other skin cancers. Recently, convolutional neural network (CNN) models such as Resnet-50 have achieved over 85% classification accuracy on the ISIC binary melanoma classification datasets and ensemble classifiers have reached over 95% accuracy. In this project, we will explore the relevant high-performing CNN models and their efficacy when utilized for skin cancer classification. We will run various experiments on these models to explore performance-related differences and potential issues with current datasets available. We find from our experiments that pre-trained CNNs are already quite robust in their ability to categorize seven different skin lesion categories from the ISIC 2018 dataset. However, the specific errors of the networks depend significantly on the data, demonstrating potential input-related biases and issues with current data available.

# Introduction

Ranked number one as the most diagnosed cancer in the US and ranked number 17 globally, skin cancer is a prominent medical issue that continues to challenge medical practitioners in the diagnosis process. From the delayed timeline of diagnosis due to skin grafts and the similarity of benign and malignant lesions and tumors for pathologists to distinguish, the current process severely overdiagnoses skin cancers to avoid potential missed cases.

In recent years, doctors have begun to use AI algorithms to augment their decision making. Although medical faith in AI has yet to reach a threshold to significantly alter current decision pipelines, early studies show marginal improvement by decreasing the number of missed cases. Many of the strongest computer vision models have been spun out from the annual International Skin Imaging Collaboration (ISIC) challenge. Their dataset of over 10,000 benign and malignant skin lesions is an extremely popular dataset to train models on, and great strides have been made in recent years to provide the medical world with potential tools to use. In this project, we take inspiration from two specific projects related to the ISIC challenge. A strong submission to their annual challenge from Milton compared AlexNet, ResNet, and VGG networks and used an ensemble of these classifiers to achieve high marks [1], and a 2022 paper using the ISIC 2020 dataset showed that transfer learning can be efficiently applied to achieve an accuracy of over 95% [4]. Using these two ideas, we implement these three base models and the pretrained models on the ImageNet dataset to explore differences in the models and facets of the 2018 ISIC dataset.

In our project, we set up the data and collab file in this following link.

We have the youtube video here: link.

## Introduction to CV Models used in thie Project:

AlexNet

AlexNet is a convolutional neural network with 8 layers. This makes it the least complex model we will utilize, which is noticeable in the relatively low test accuracy. However, this model trains very fast, and is remarkably accurate for its size. AlexNet is the first CNN to win ImageNet. In context, AlexNet can be considered the CNN structure that first popularized CNN's, which in turn revolutionized the computer vision industry.

Model Diagram:
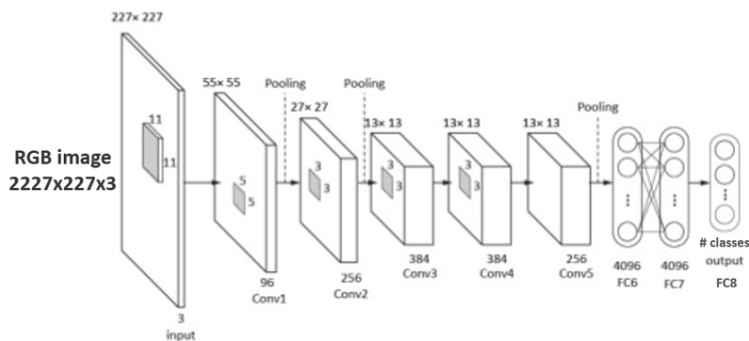


"*Fig 1. AlexNet: CNN architecture figure* [4]
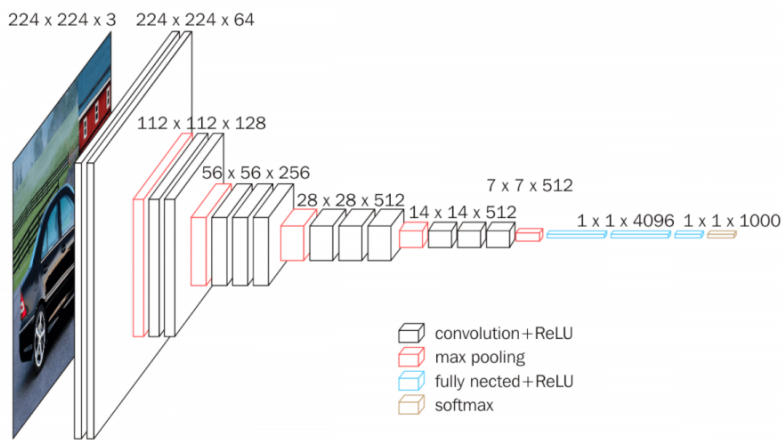
Model Code:

                    AlexNet_Code

"*Fig 2. AlexNet Code: CNN architecture code* [4]

AlexNet consists of 5 convolutional layers followed by 3 linear perceptron layers. The non-linear activation layer used is ReLU, and there are maxpool layers to reduce dimensionality and feature complexity.

VGG

VGG is a deep learning CNN model that consists of 16 (VGG-16) or 19 (VGG-19) convolutional layers. VGG-16 achieves over 74% top-1 accuracy on ImageNet, making it a premier classification model, but not the industry standard by any means. It is significantly larger than both AlexNet and ResNet.

Model Diagram:

"*Fig 3. VGG: CNN architecture figure* [5]

Model Code:

```python
class VGG(nn.Module):

    def __init__(self, features, num_classes=1000, init_weights=True):
        super(VGG, self).__init__()
        self.features = features
        self.classifier = nn.Sequential(
            nn.Linear(512 * 7 * 7, 4096),
            nn.ReLU(True),
            nn.Dropout(),
            nn.Linear(4096, 4096),
            nn.ReLU(True),
            nn.Dropout(),
            nn.Linear(4096, num_classes),
        )
        if init_weights:
            self._initialize_weights()

    def forward(self, x):
        x = self.features(x)
        x = x.view(x.size(0), -1)
        x = self.classifier(x)
        return x

    def _initialize_weights(self):
        for m in self.modules():
            if isinstance(m, nn.Conv2d):
                n = m.kernel_size[0] * m.kernel_size[1] * m.out_channels
                m.weight.data.normal_(0, math.sqrt(2. / n))
                if m.bias is not None:
                    m.bias.data.zero_()
            elif isinstance(m, nn.BatchNorm2d):
                m.weight.data.fill_(1)
                m.bias.data.zero_()
            elif isinstance(m, nn.Linear):
                m.weight.data.normal_(0, 0.01)
                m.bias.data.zero_()
```

"*Fig 4. VGG: CNN architecture code* [7]

## ResNet

ResNet is a convolutional neural network with multiple layers. It is one of the best models in current use, with optimized versions boasting over 84% top-1 accuracy on ImageNet. Previous complicated models such as VGG often ran into what is known as the vanishing gradient problem, which occurs when one variable in a long chain of multiplied chain derivatives is a small value, eventually resulting in a gradient that is near zero. To sidestep this issue, Resnet introduces the idea of residual blocks.

Block Diagram:
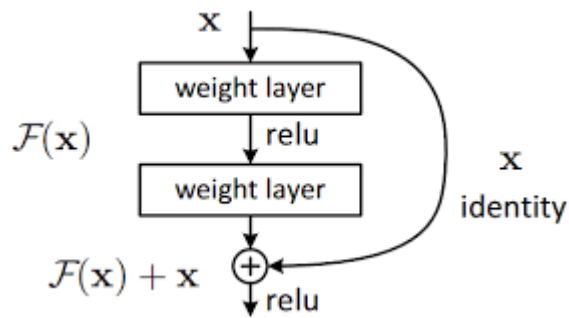


"*Fig 5. Resnet: Resnet block architecture figure* [6]

Model Code:

```python
class BasicBlock(nn.Module):
    expansion = 1

    def __init__(self, in_planes, planes, stride=1):
        super(BasicBlock, self).__init__()
        self.conv1 = nn.Conv2d(in_planes, planes, kernel_size=3,
stride=stride, padding=1, bias=False)
        self.bn1 = nn.BatchNorm2d(planes)
        self.conv2 = nn.Conv2d(planes, planes, kernel_size=3, stride=1,
padding=1, bias=False)
        self.bn2 = nn.BatchNorm2d(planes)

        self.shortcut = nn.Sequential()
        if stride != 1 or in_planes != self.expansion*planes:
            self.shortcut = nn.Sequential(
                nn.Conv2d(in_planes, self.expansion*planes, kernel_size=1,
stride=stride, bias=False),
                nn.BatchNorm2d(self.expansion*planes)
            )

    def forward(self, x):
        out = F.relu(self.bn1(self.conv1(x)))
        out = self.bn2(self.conv2(out))
        out += self.shortcut(x)
        out = F.relu(out)
        return out
```

"*Fig 6. Resnet: Resnet code* [6]

The idea behind this residual block is the shortcut, which forwards the output value of a layer x to that of the next layer directly, which allows the gradient to flow through, eliminating the vanishing gradient problem.

The structure of a ResNet simply consists of numerous convolutional layers, residual blocks, max-pooling, and ReLU. Models range from ResNet-18 to ResNet-152, and they differ in number of layers and other complexity differences. ResNet-50 is the model that we will use in this project. It's architecture is as follows below:

Model Diagram:
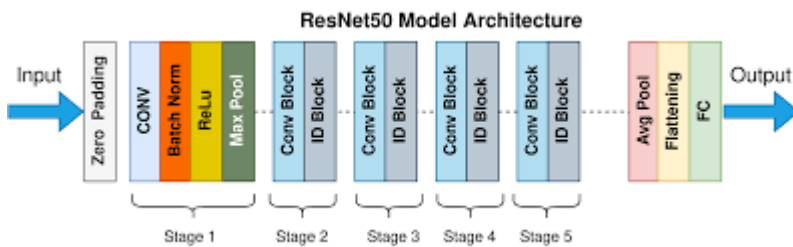
"*Fig 6a. Resnet: Resnet model architecture* [6]

# Methods

For our model comparisons, we are using the ISIC 2018 image database. This database consists of 10015 training images, which are all labeled with ground truths in:

Melanoma (MEL), Melanocytic nevus (NV), Basal cell carcinoma (BCC), Actinic keratosis / Bowen's disease (intraepithelial carcinoma) (AKIEC), Benign keratosis (solar lentigo/seborrheic keratosis/lichen planus-like keratosis) (BKL), Dermatofibroma (DF), Vascular lesion (VASC)

We will be using the three training models defined above (AlexNet, VGG-16, and Resnet-50) to classify the training images into these 7 categories. We will observe both accuracy and training time on all three of these models. Then, we will experiment on various properties of these three training models.

## Baseline Control Model:

Train baseline models and see what their relative loss, runtime, and accuracies are on 20 epochs.

## Experiment 1

Vary the amount of training data that is given to each model, and see how each model reacts. We measure the accuracy and loss of each model when given the following number of training samples trained on 5 epochs.

## Experiment 2

Import versions of the three models that have been pre-trained on millions of images from ImageNet. Then, modify these models in one of two schemes:

Linear: For this scheme, the model will be frozen, then a linear classifier is trained which takes the features before the fully-connected layer. Then, a new fully-connected layer is written, which takes the in-features and outputs scores of size num_classes. Finetune: Same as Linear, except that features do not need to be frozen and the model can finetune on the pretrained model.

Performances will be compared between all 3 pretrained models in both Linear and Finetune schemes. All models will be trained on 10015 images, for 5 epochs.

## Experiment 3

So far our experiments have focused largely on measuring the efficacy of the training models, and the emphasis has not been skin-cancer specific. Here, we look at the accuracy by skin category, to see if certain labels are worse performing than others.

## Experiment 4

Apply model visualization techniques to compare feature extraction for 7 categories from three deep learning models.

# Results

## Baseline models
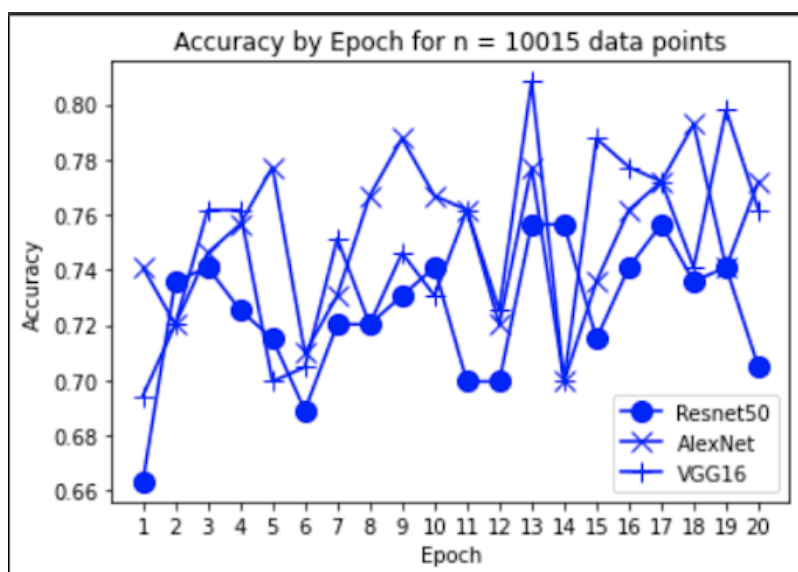

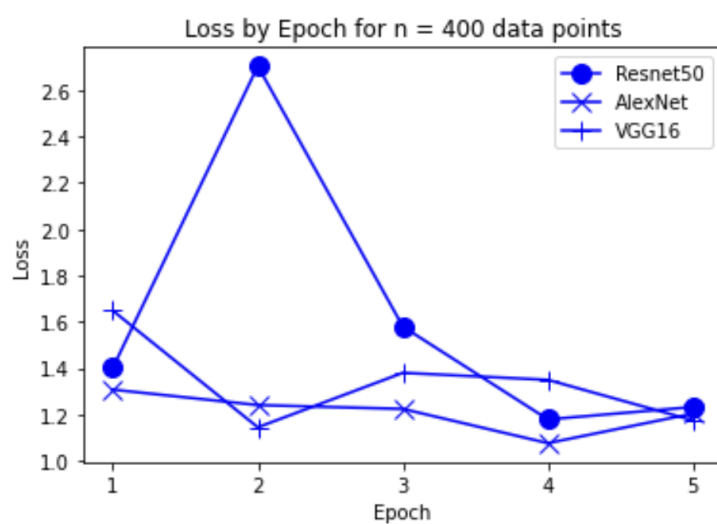
"*Fig 7. Accuracy of baseline models*
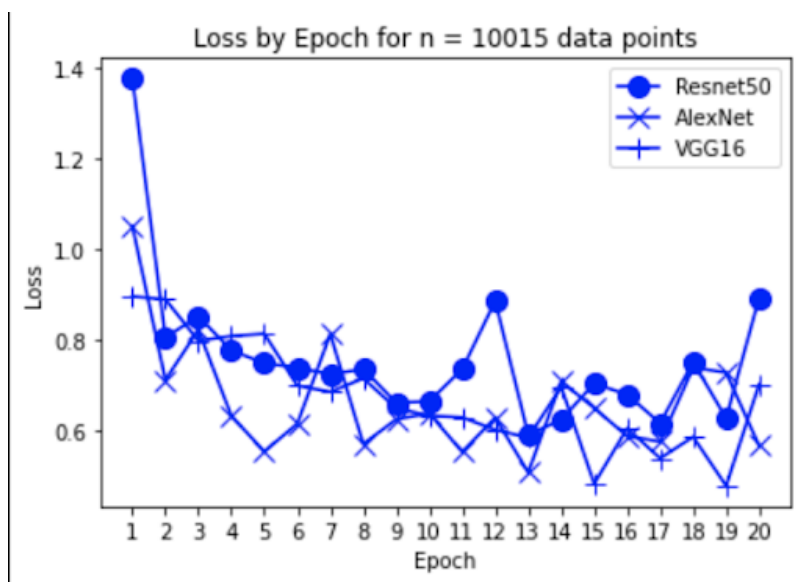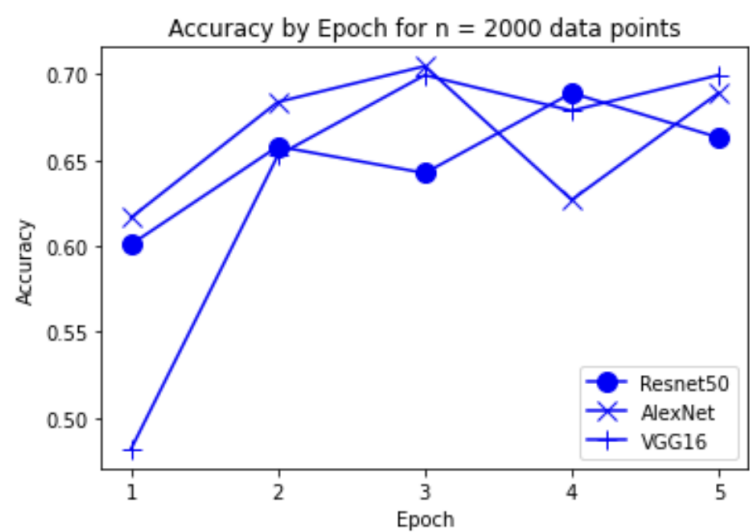
"*Fig 7a. Loss of baseline models*
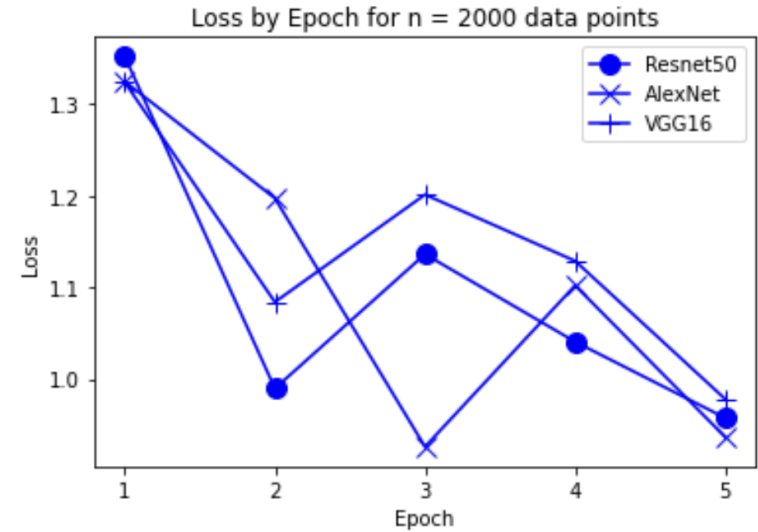
# Experiment 1
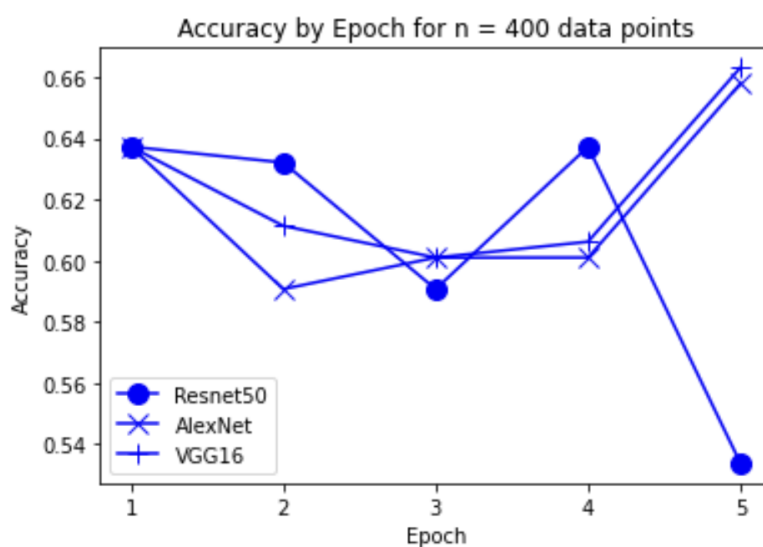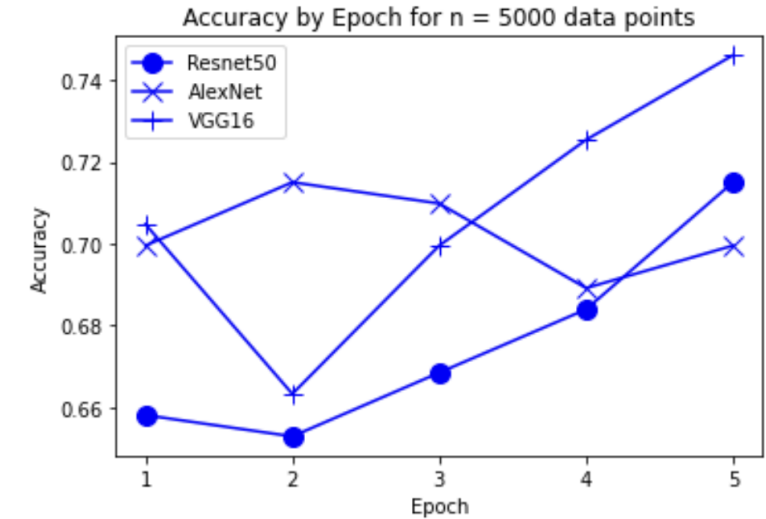


"*Fig 8. Accuracy of models on 400 inputs*
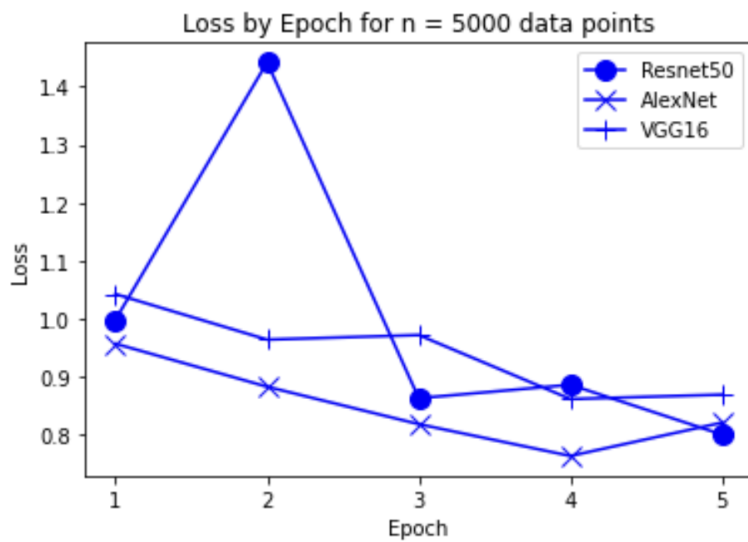
*"Fig 9. Loss of models on 400 inputs*



*"Fig 10. Accuracy of models on 2000 inputs*
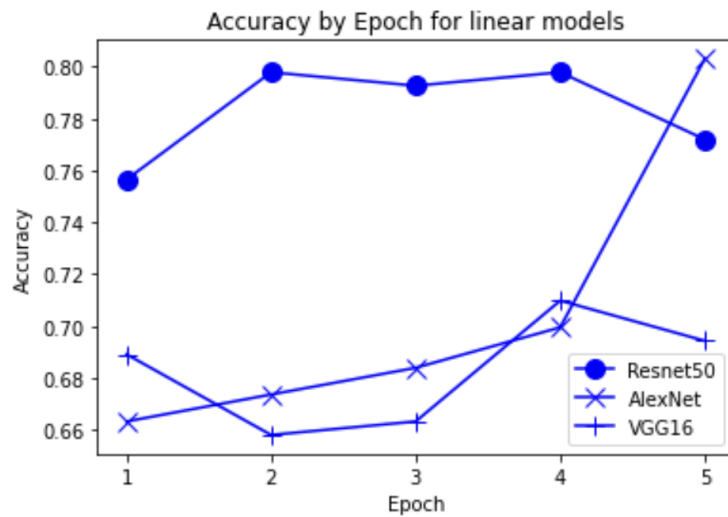


*"Fig 11. Loss of models on 2000 inputs*



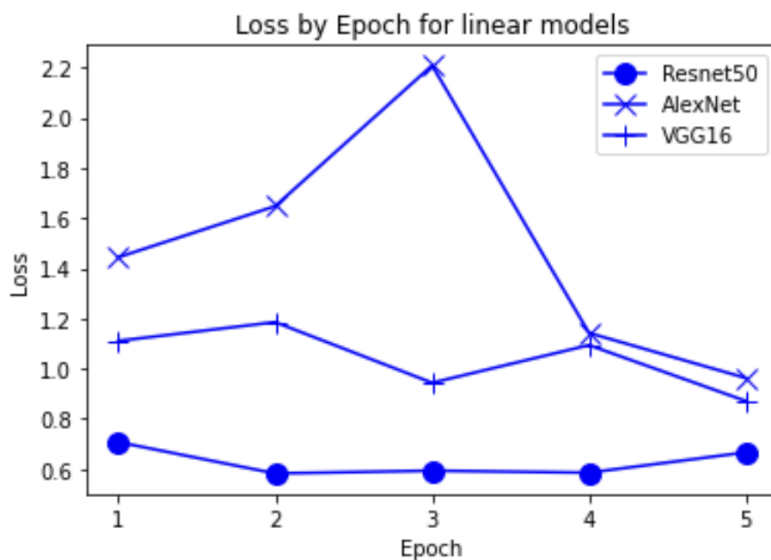*"Fig 12. Accuracy of models on 5000 inputs*

"*Fig 13. Loss of models on 5000 inputs*
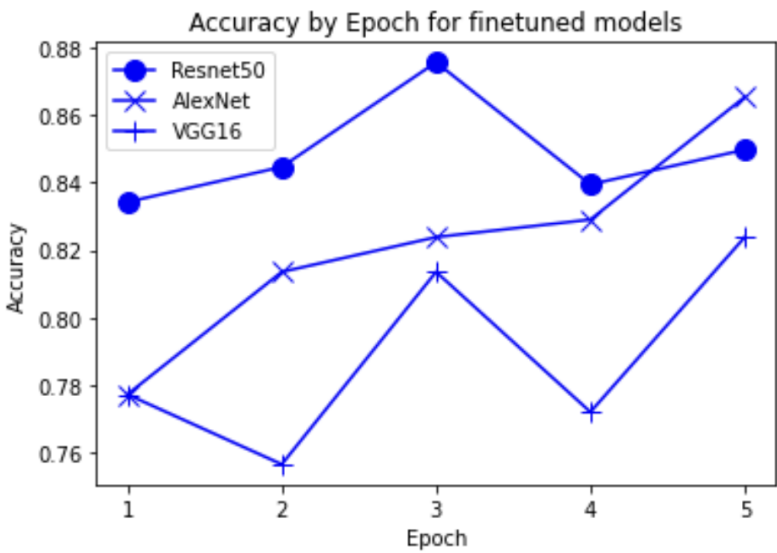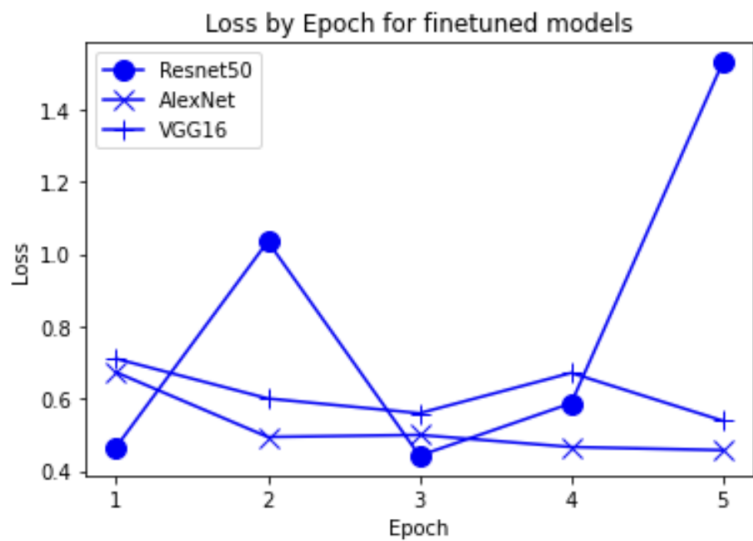
# Experiment 2



"*Fig 14. Accuracy of linear pretrained models*

"*Fig 15. Loss of linear pretrained models*



"*Fig 16. Accuracy of finetune pretrained models*
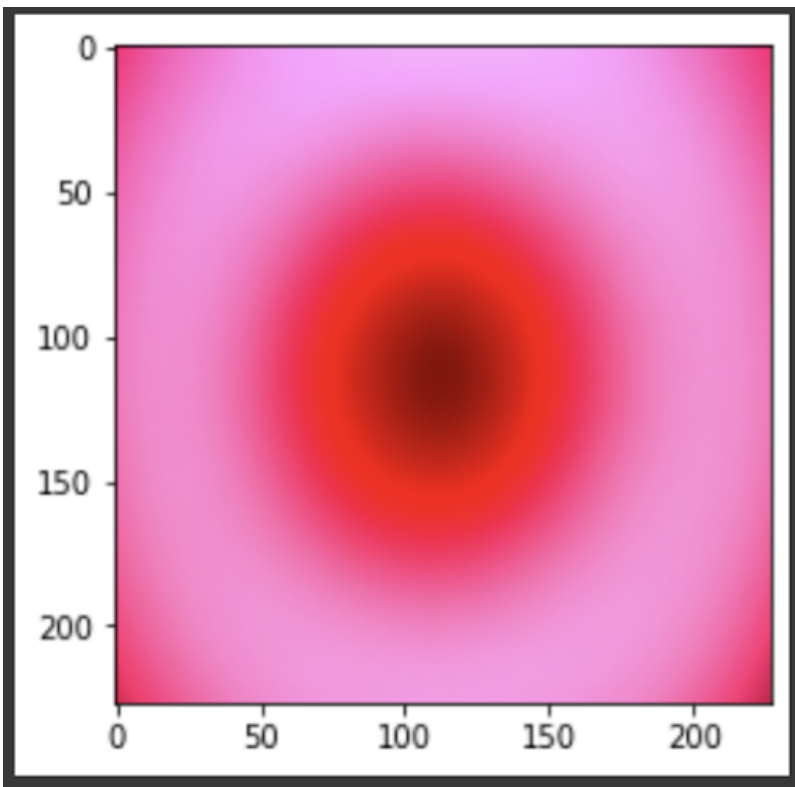


"*Fig 17. Loss of finetune pretrained models*
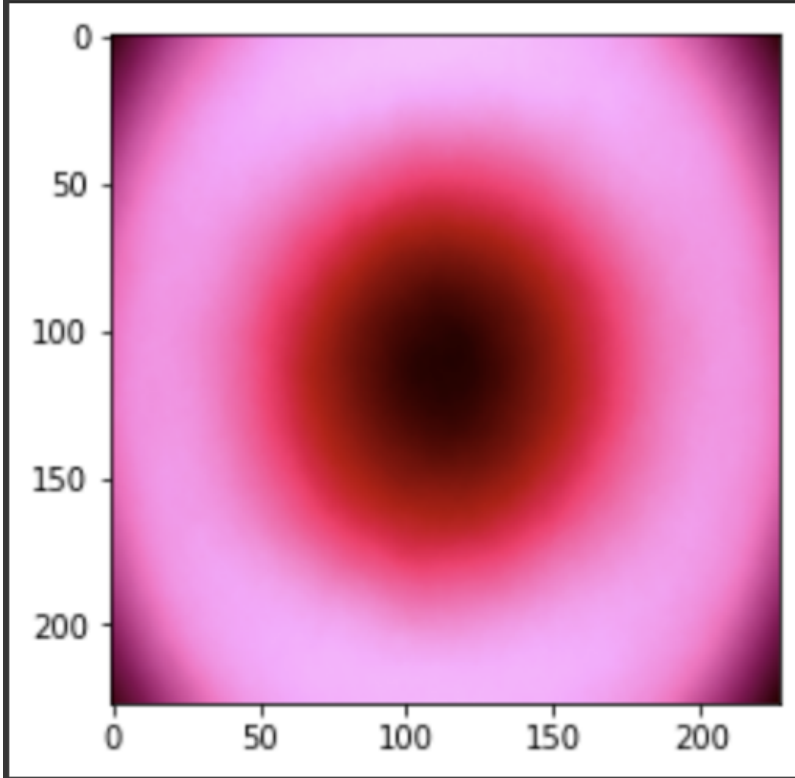
# Experiment 3

Resnet
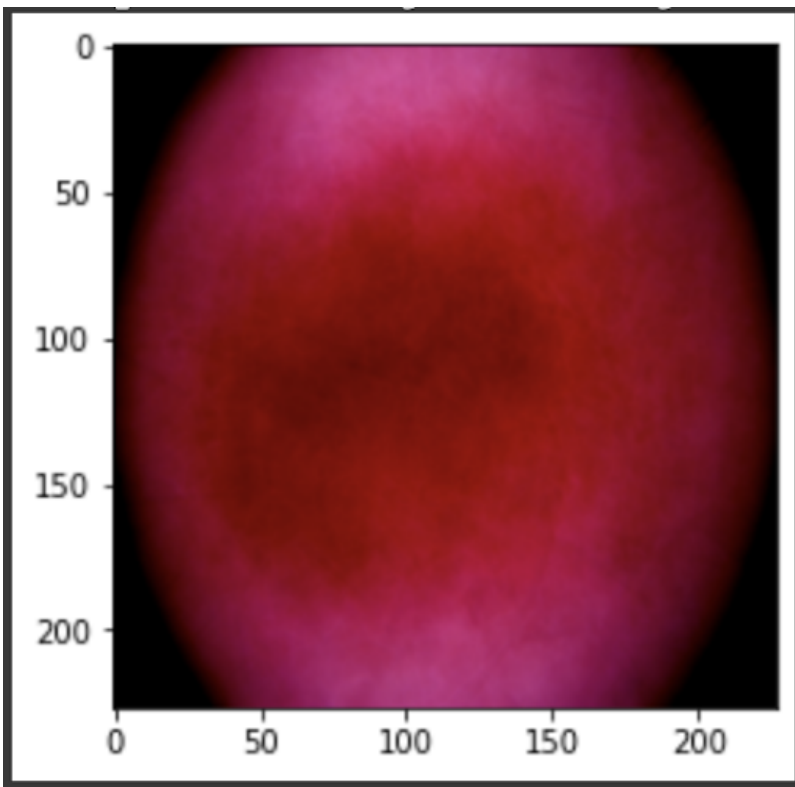
"*Fig 18. Error totals of finetune pretrained models*
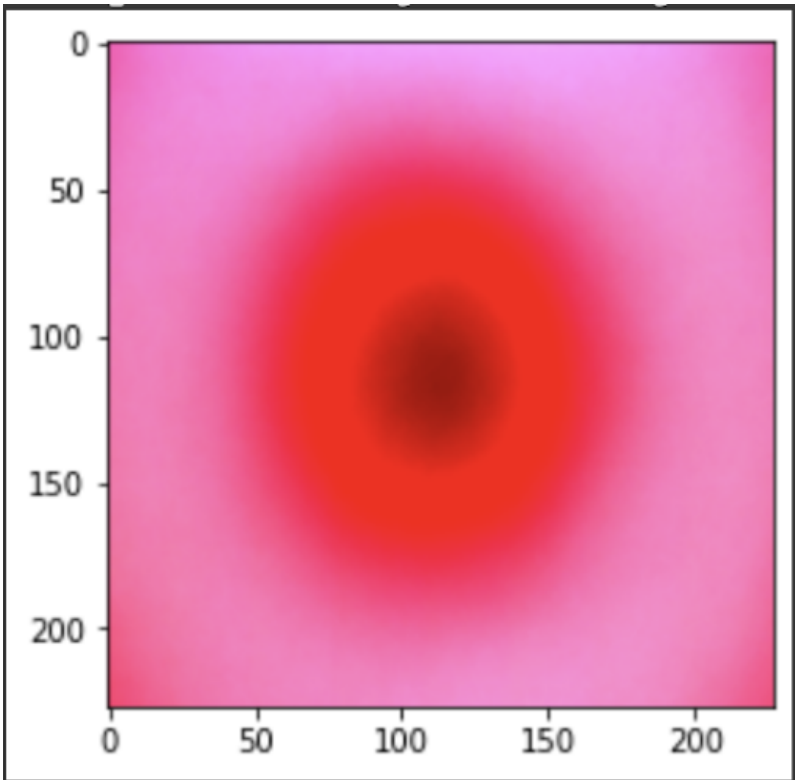
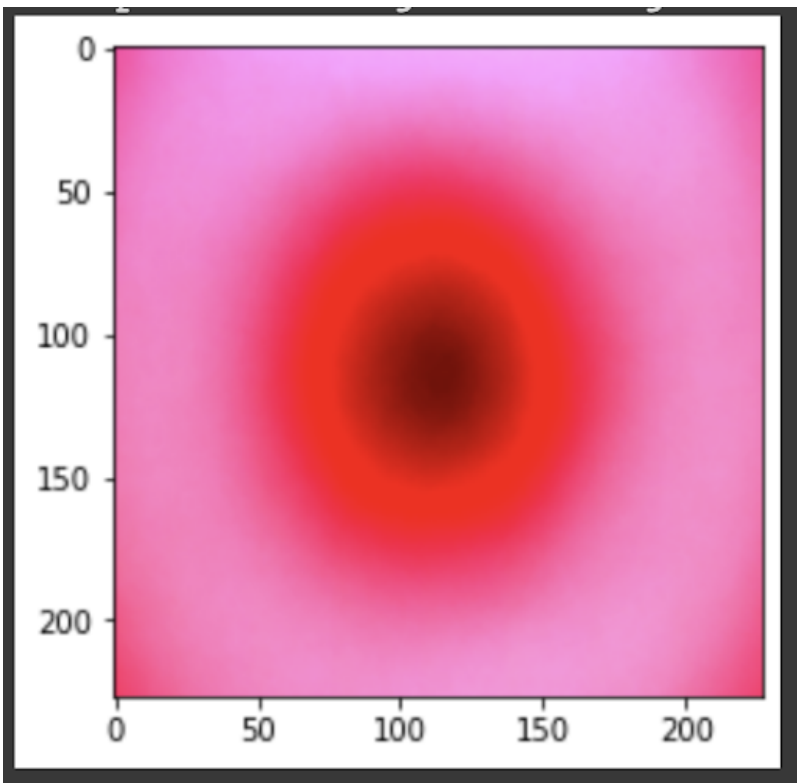# Experiment 4

"*Fig 19. Mean of all images*



"*Fig 20. Mean of melanoma images*

"*Fig 21. Mean of prediction for melanoma*



"*Fig 22. Mean of MV images*

"*Fig 23. Mean of prediction for MV*

# Discussion

The results show an interesting trend for the robustness of the three models. While ResNet50 does have the most robust accuracy by most measures, AlexNet is not far behind despite the relative simplicity. Specifically, when finetuning AlexNet, the network approaches the performance of ResNet50, and performs quite well considering the relative number of layers between ResNet and AlexNet. This suggests that, perhaps, the deeper layers in ResNet50 are not providing much value in extracting more information from the general features in the initial layers. This task, however, continues to provide the same challenge for all three models as from experiment 4, when we collapse the errors for all the models, the most commonly misidentified lesions are Melanocytic nevi (NV) and Benign lesions of the keratosis (BKL), which are benign skin lesions visually similar to Melanoma. All three models have similar error rates for these two categories, suggesting that like human performance, the most common misdiagnosis is of these visually benign and malignant lesions. Currently, these lesions are often diagnosed as cancer, and the only way to verify this diagnosis is to track the progression of the lesion rather than using the initial lesions. Furthermore, from experiment 4, it is clear that more data leads to more accurate representations. However, it seems that the model is extracting a feature of MV that may lead to misclassification, as the darker center closely mirrors that of melanoma. The similarity of these two classes likely leads to many shared

features, as shown in the mean images of prediction. More images for all classes are linked in our presentation slides and in our video.

Potential future research can image the progression of these three categories — Melanoma, Melanocytic Nevi, and Benign Keratosis Lesions, to provide time-stepped images of these lesions to train a model on. This dataset could potentially provide models more features to extract and generalize from malignant lesions that previously were unavailable. Furthermore, the rather low error count for Dermatofibroma (DF) and Vascular lesion (VL) could be attributed to the low counts of data in the training and validation dataset (around 1 - 2%). This could be attributed to both the similarity with melanoma as the classification of these lesions often falls under Melanoma. However, these rarer lesions can also be both benign and malignant, and the behavior of these lesions is not yet understood compared to other forms of lesions.

For future research, it is imminent that more data is collected for specific lesions that may not be as common. These less common lesions should be tracked by time to collect more development-related information to classify early-stage lesions as benign or malignant.

# Reference

[1] Milton, M. "Automated Skin Lesion Classification Using Ensemble of Deep Neural Networks in ISIC 2018: Skin Lesion Analysis Towards Melanoma Detection Challenge." *arXiv*. 2019.

[2] Pham, TC., Luong, CM., Hoang, VD. et al. "AI outperformed every dermatologist in dermoscopic melanoma diagnosis, using an optimized deep-CNN architecture with custom mini-batch logic and loss function." *Sci Rep 11, 17485 (2021).* https://doi.org/10.1038/s41598-021-96707-8

[3] Gorriz, M., Carlier A., Faure, E., Giro-i-Nieto, X. "Cost-Effective Active Learning for Melanoma Segmentation" *arXiv*. 2017.

[4] Rashid J, Ishfaq M, Ali G, Saeed MR, Hussain M, Alkhalifah T, Alturise F, Samand N. Skin Cancer Disease Detection Using Transfer Learning Technique. Applied Sciences. 2022; 12(11):5714. https://doi.org/10.3390/app12115714

[5] https://medium.com/mlearning-ai/an-overview-of-vgg16-and-nin-models-96e4bf398484

[6] https://towardsdatascience.com/residual-blocks-buildingda-blocks-of-resnet-fd90ca15d6ec

[7] https://github.com/ashushekar/VGG16

[8] https://www.wcrf.org/cancer-trends/skin-cancer-statistics/

[9] https://www.aad.org/public/diseases/skin-cancer/find/know-how

[10] https://www.aad.org/public/diseases/skin-cancer/types/common

[11] https://www.washington.edu/news/2022/05/03/many-pathologists-agree-overdiagnosis-of-skin-cancer-happens-but-dont-change-diagnosis-behavior/

[13] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7519424/

[14] https://www.sciencedirect.com/science/article/pii/S1361841521003509

[15] https://www.mdpi.com/2072-4292/9/8/848

---