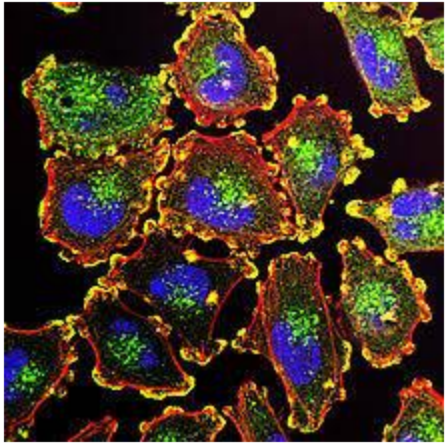
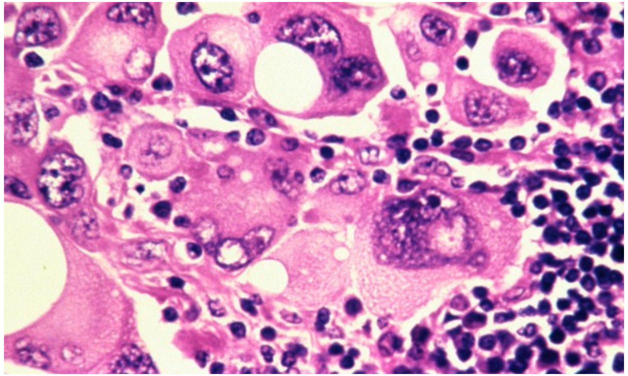
Classifying Melanoma Tumor Images

Gwyneth Volkmann

Melanoma Introduction

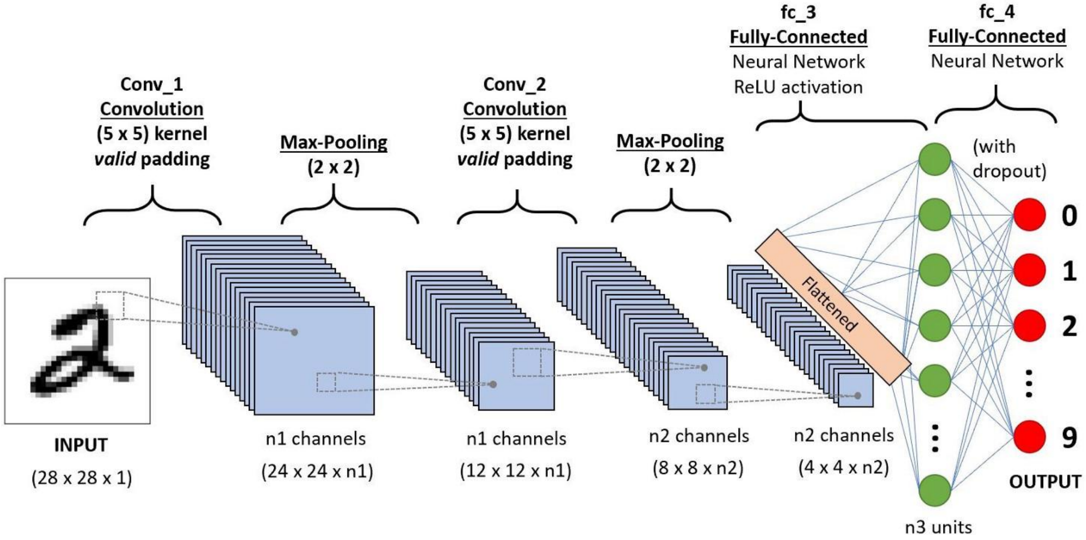
- 106,000 new melanomas will be diagnosed and around 7,180 Americans will die in 2021
- The risk of melanoma increases over time but it is one of the most common cancers diagnosed in young adults, particularly young women
- Some of the most common risk factors are sun exposure, family history, fair skin, and the presence of moles
- Survival rates are very high if the cancer is diagnosed at an early stage, but drop rapidly once melanoma has spread from the initial site





Taken from the American Cancer Society website: cancer.org

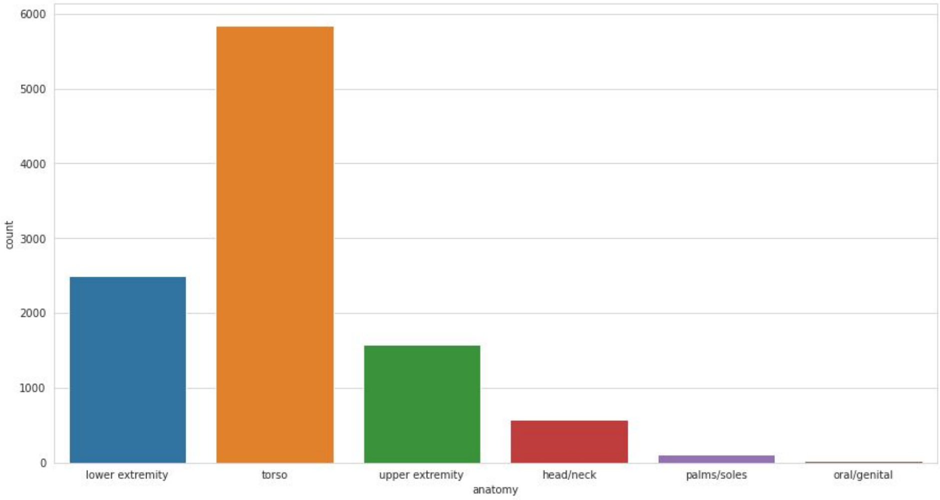
Machine Learning and Medical Imaging

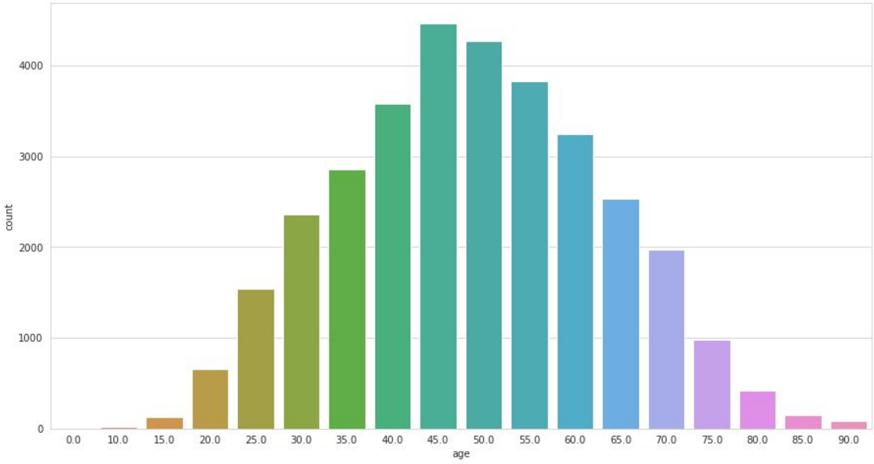


- Convolutional Neural Networks can take an input image, assign weights and biases to various features and and be able to differentiate the classifications of these images
- This offers the promise of accelerating disease identification in the future
- Machine learning has become an important tool in radiology already

Data Cleaning and EDA

- There were missing values for both the testing and training datasets. These were imputed using various methods covered in the documentation.
- Training dataset: 33,126 rows   Testing dataset: 10,982 rows

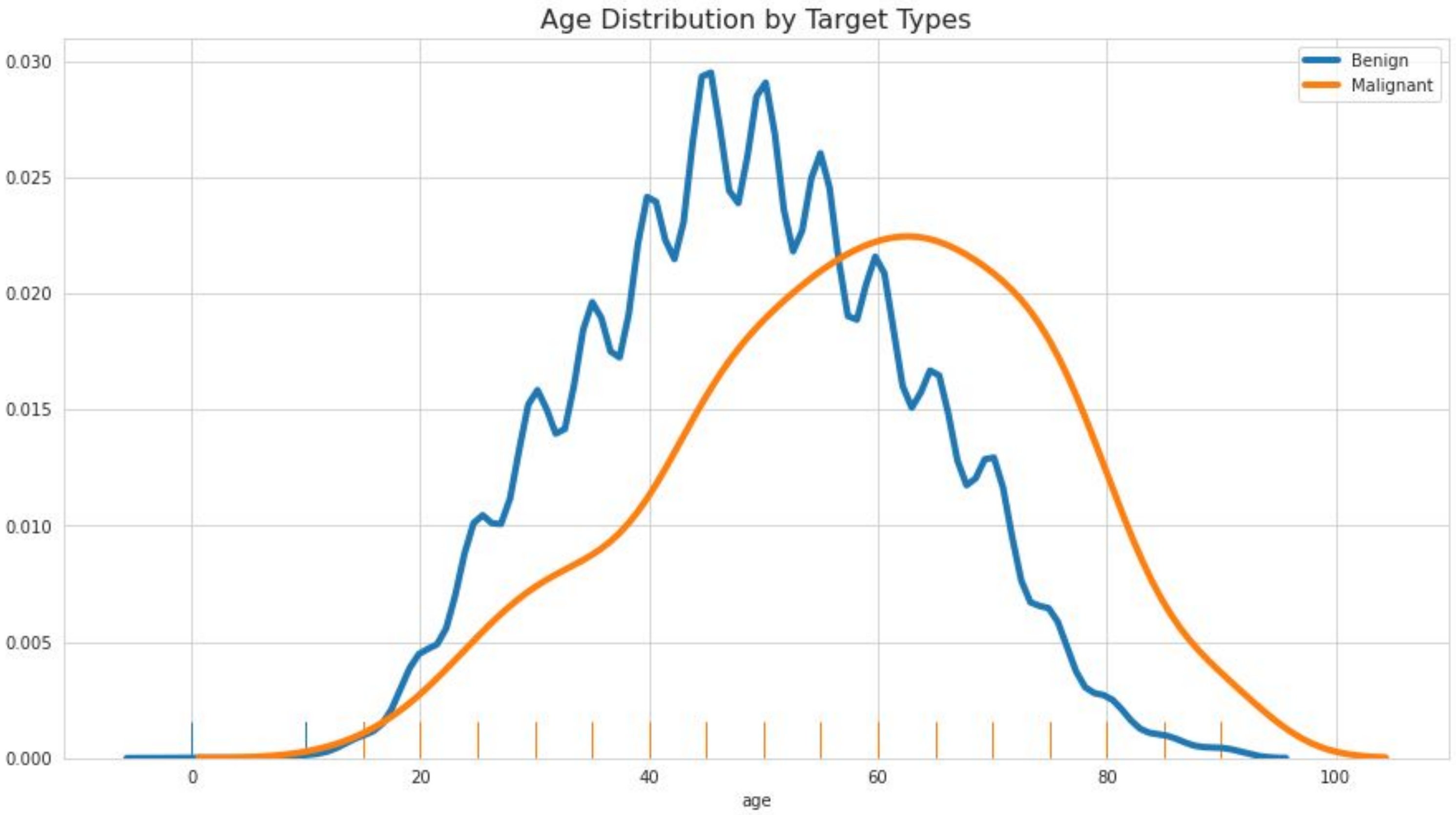




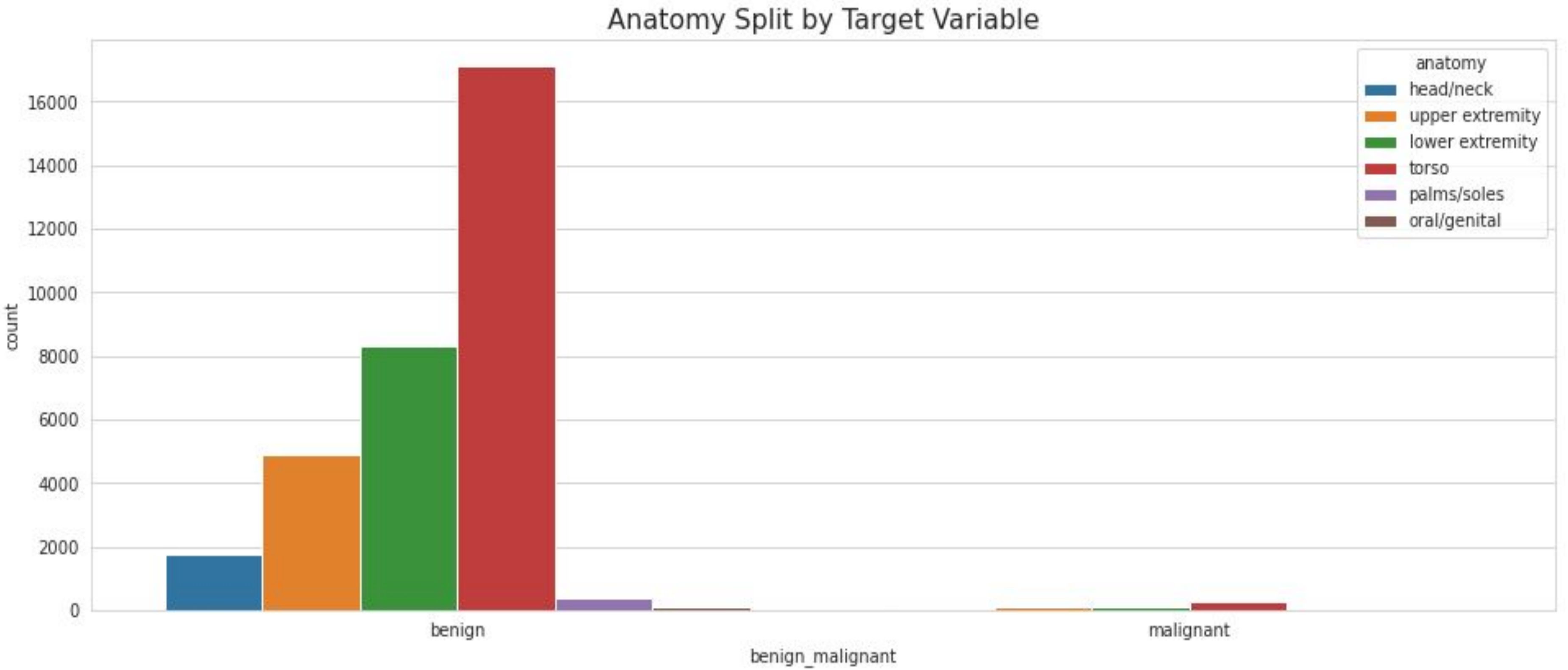
The most common tumor location is the torso.

The distribution of ages is normal. I imputed missing ages using the median since the sample size is also very large.

Age Distribution by Target Type

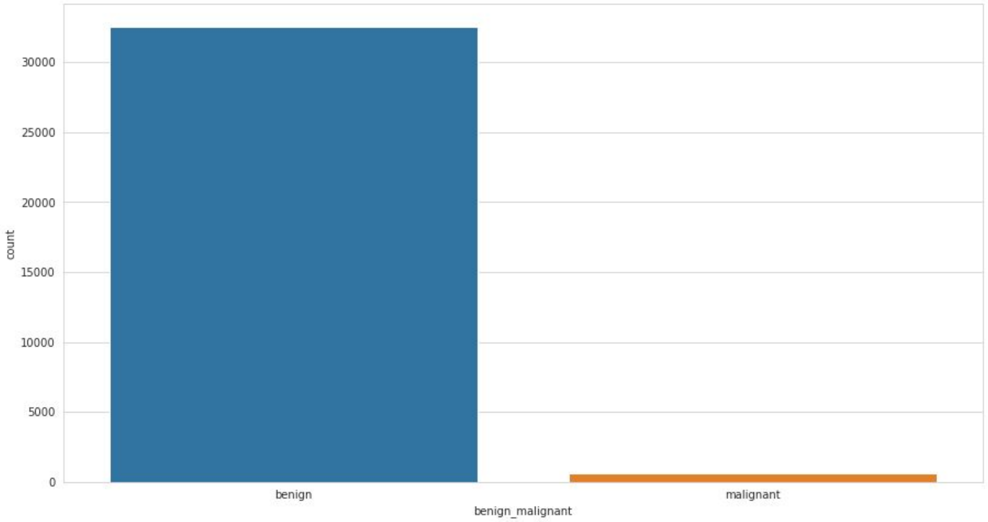


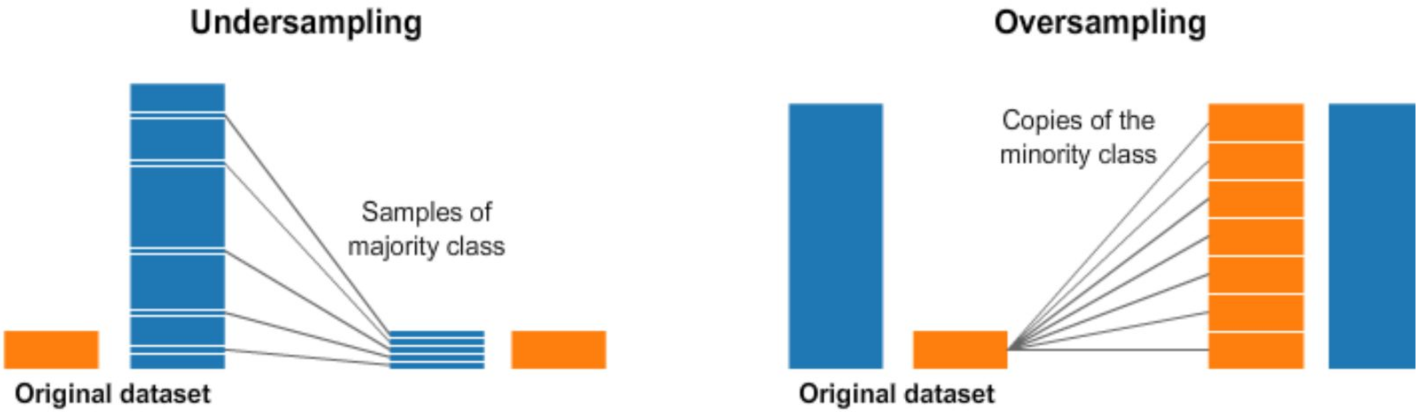
Anatomy by Target



EDA

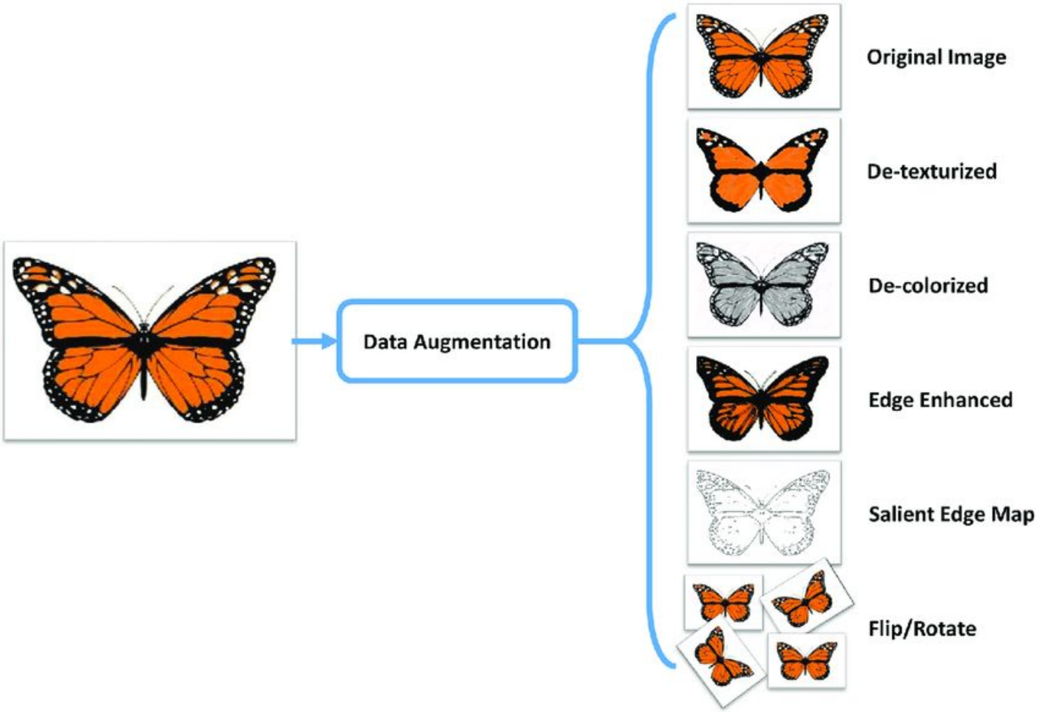
Data classes are very unbalanced. There are a few ways to deal with this problem. For this project we will be oversampling and generating images for the malignant class of images.



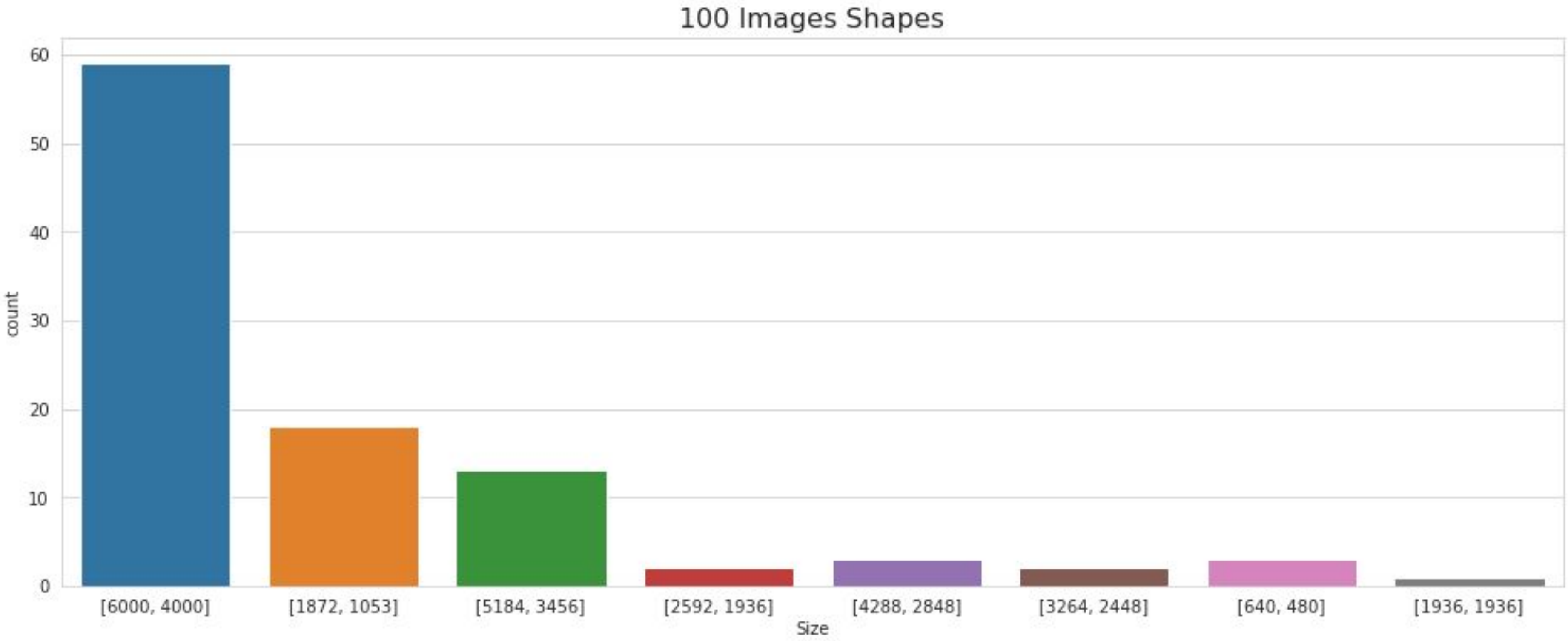


Data Augmentation

- We can perform random vertical/horizontal flips, change color channels, add Gaussian blur, change pictures to black and white
- This is one way of creating enough images for the model to fit the malignant class more accurately even though the class is so much smaller than the benign class.



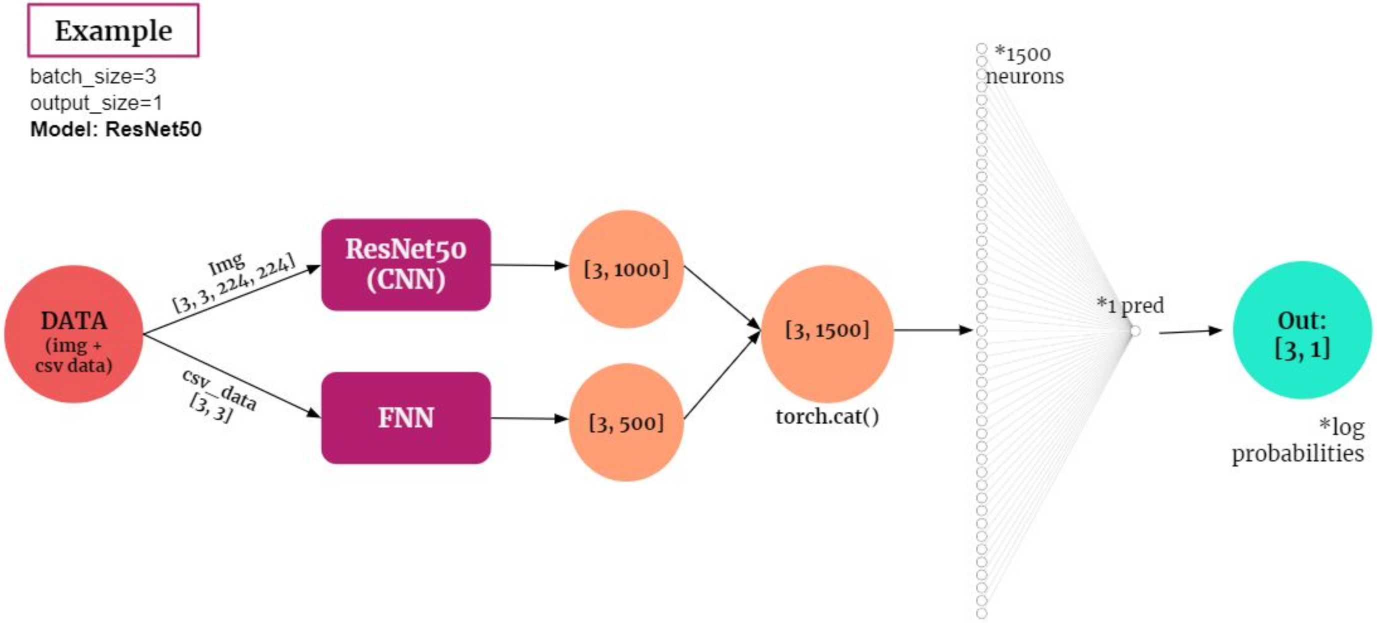
EDA



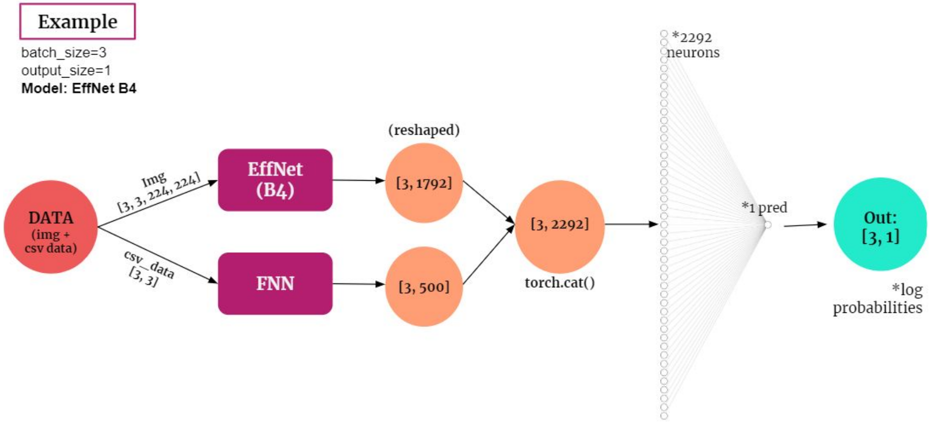
There are a variety of shapes so these will need to be transformed before training the model.

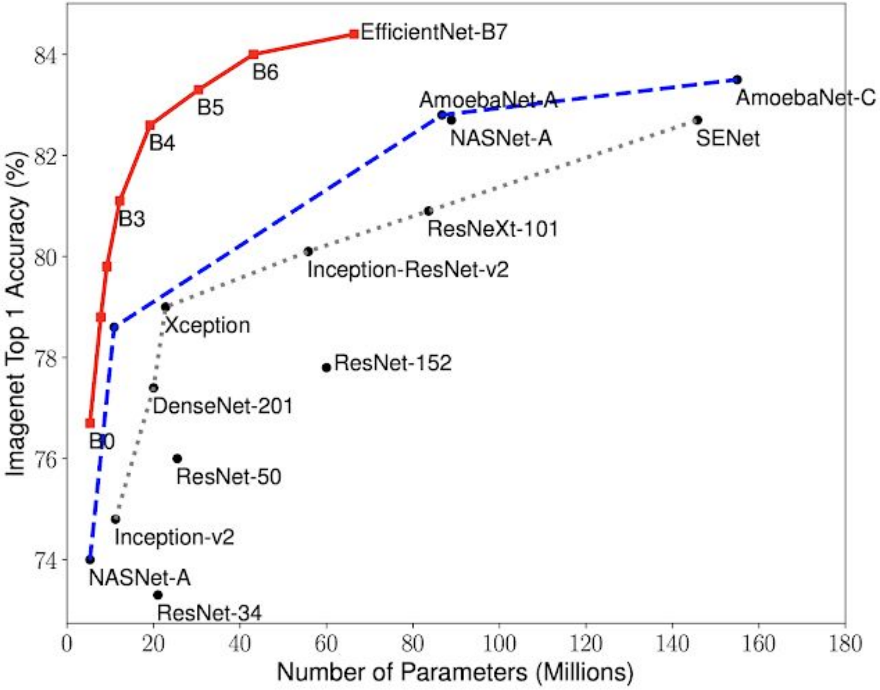
Modeling

- ResNet50
  - Convolutional network
  - 50 layers deep
  - Pre-trained



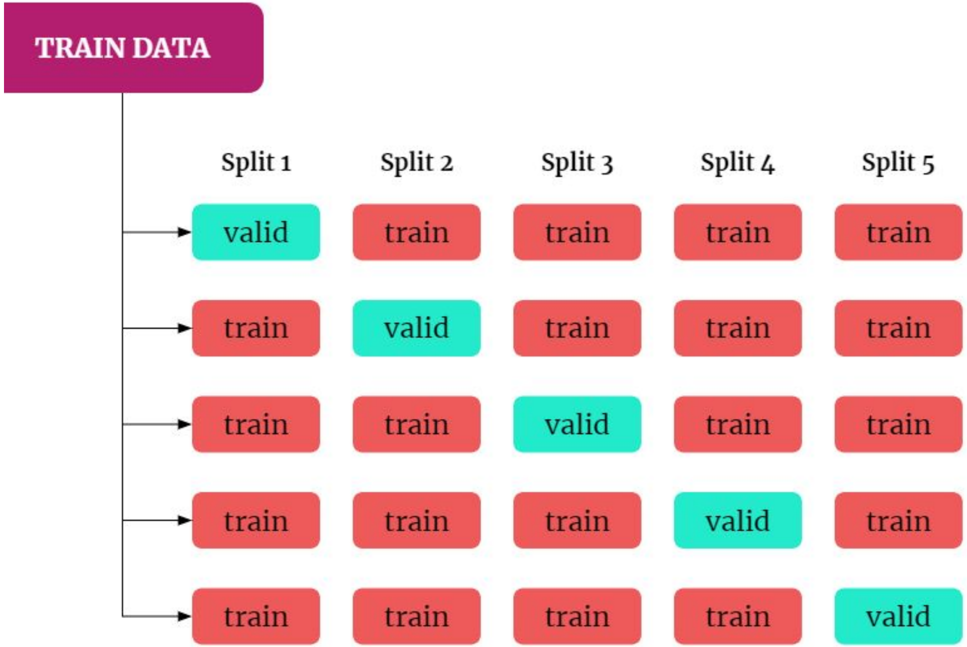
- EfficientNet
    - Convolutional Neural Network
    - Uses uniform compound scaling method for all dimensions





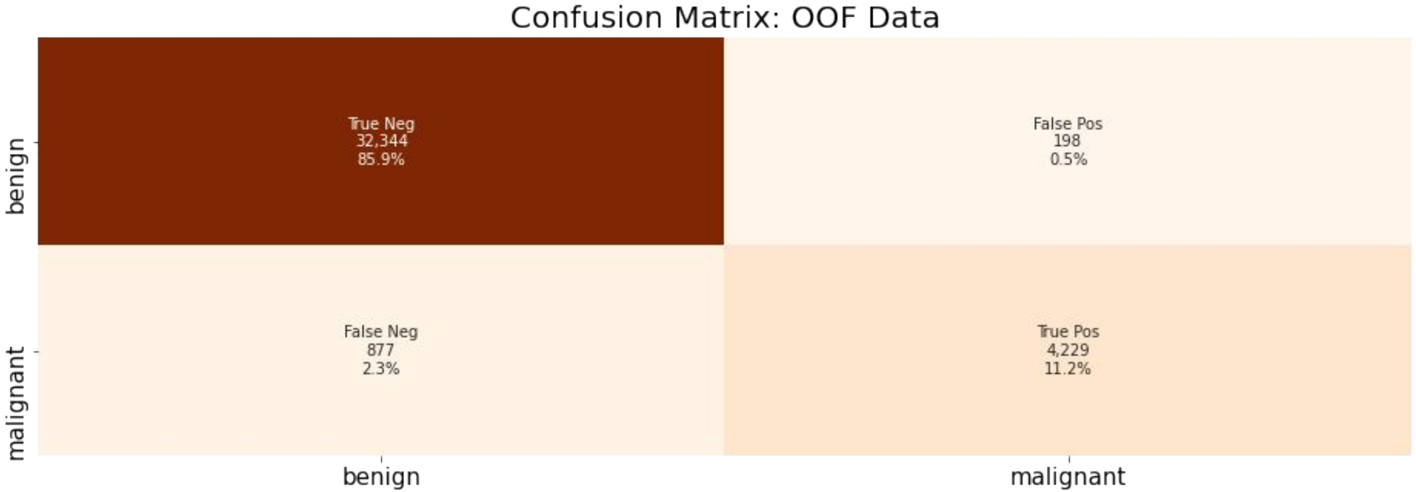
Training

- K-Fold Cross Validation
  - Resampling procedure used to evaluate machine learning models
  - Less biased than simple train/test split
- Training Loop
  - Epochs =15
  - Patience =3
  - TTA
    - Will perform random augmentations to the images in the data set
  - Learning rate = 0.05



Model Evaluation

- Confusion Matrix
  - True negative value is good, true positive value is very low
  - This model is good at identifying benign images, but did not correctly identify the malignant melanoma images.
- OOF ROC: 0.976



Classification Report

- The classification values look good for this model
- The recall for the malignant category is the lowest value
    - Recall is the number of relevant items that have been retrieved
    - Precision is the number of selected items that are relevant
        - Precision values for both classes are high

	precision	recall	f1-score	support
Benign	0.97	0.99	0.98	32542
Malignant	0.96	0.83	0.89	5106
accuracy			0.97	37648
macro avg	0.96	0.91	0.94	37648
weighted avg	0.97	0.97	0.97	37648

Conclusions

- The model is skilled at classifying benign images but will need further tuning in order to classify a high percentage of malignant images correctly
- A larger sample of actual malignant images might be a helpful tool to train the model further in the future.