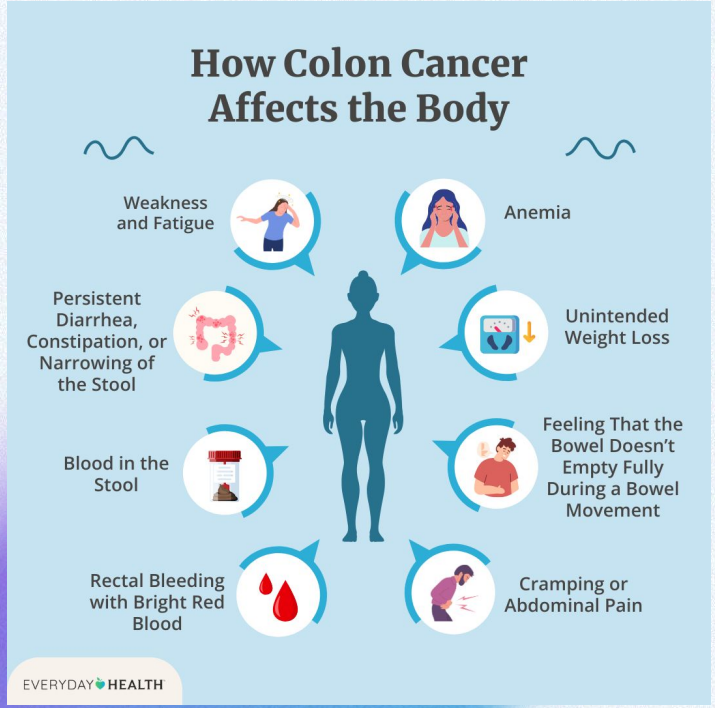
Precision Medicine in Colorectal Cancer

Raise your hand…
if you have heard of Colorectal Cancer

What is Colorectal Cancer and Why Does it Matter?

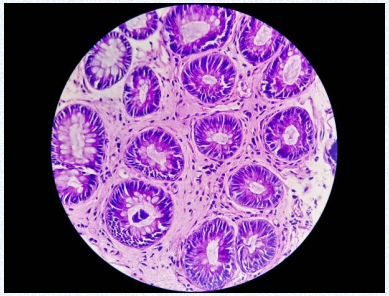
1. Starts in the large intestine
2. Spreads (metastases) until it becomes impossible to treat
3. Colonoscopies used to diagnose
    a. Allows doctors to look inside the colon through a flexible tube and camera
4. Treated with chemotherapy, Radiation, Surgery
5. On the rise due to aging, unhealthy diet → obesity
6. Hypothesis - machine learning could achieve clinical accuracy in classifying between tissue type;

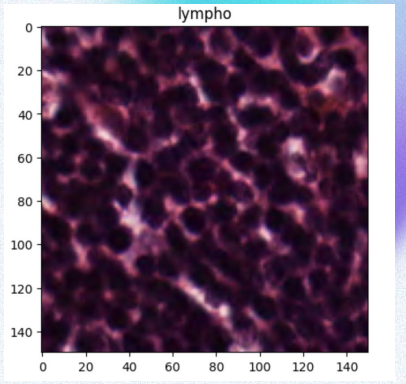


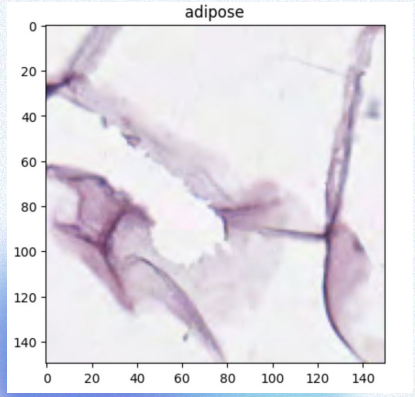
Why use Machine Learning?

Diagnosis is inefficient:

1. Symptoms are not apparent until cancer has spread
2. Under the microscope, it can be very time consuming to see if it is cancerous at all (cancer can manifest in different ways)
3. Pathologists can interpret tissue slides in variable ways, leading to different diagnoses



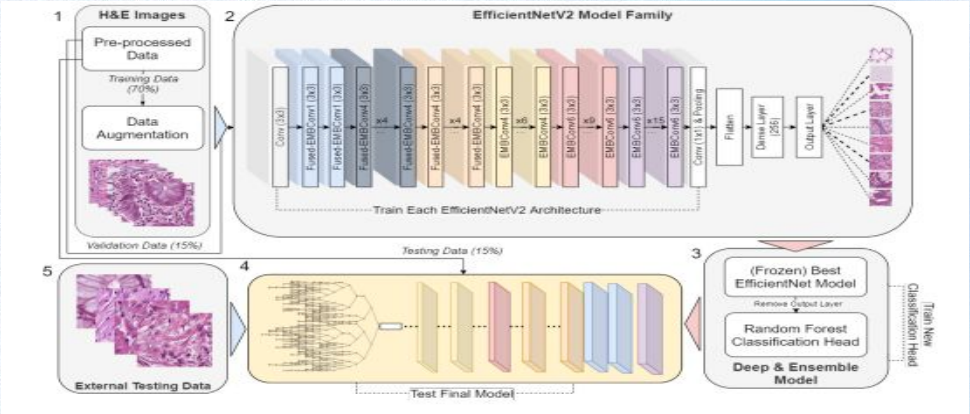




A few images from our dataset

ML in Colorectal Cancer Histology Imaging

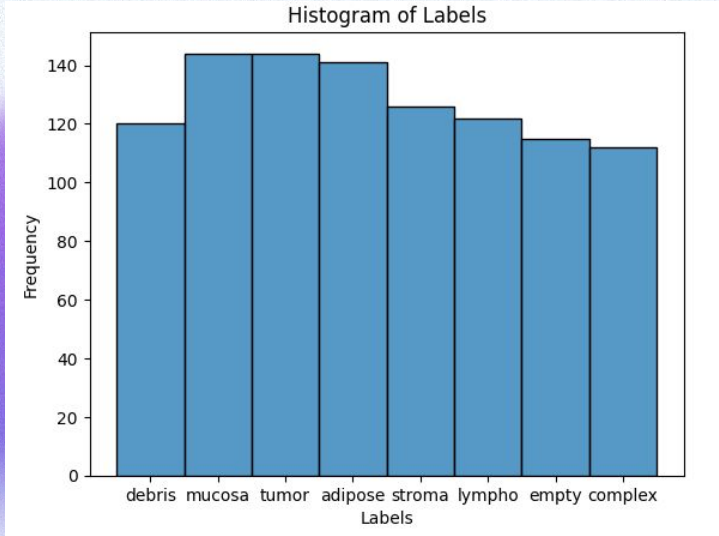
1. We found multiple examples
2. Prezja, 2023
3. Limitations:
    a. Training of data requires a lot of annotated data
    b. Label noise of specific dataset
4. Strengths:
    a. Relatively accurate program
    b. Multiple layers in model
    c. Augmentation

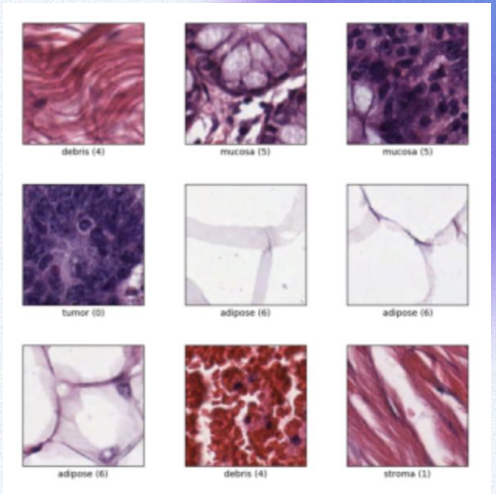


Model	ImageNet Top-1 Accuracy	ImageNet Top-5 Accuracy	Base Parameters
EfficientNetV2B0	78.7%	94.3%	7.2M
EfficientNetV2B1	79.8%	95.0%	8.2M
EfficientNetV2B2	80.5%	95.1%	10.2M
EfficientNetV2B3	82.0%	95.8%	14.5M
EfficientNetV2S	83.9%	96.7%	21.6M
EfficientNetV2M	85.3%	97.4%	54.4M

What is Our Dataset and Labels?

- TensorFlow colorectal_histology dataset used to train model






```
plt.imshow(images[i]) # Use matplotlib imshow() function to visualize an image.
plt.title(labels[i]) # Set the label of the image as the title
plt.show() # Show your plot.
```

- The labels label each image

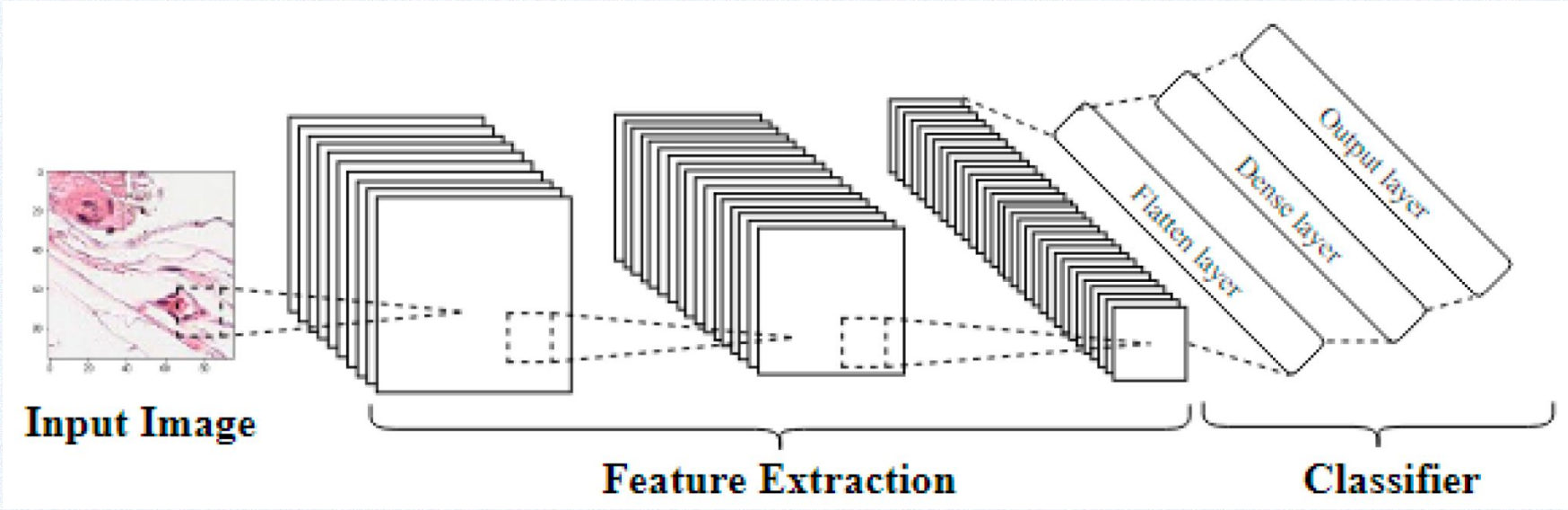
Number of imgs Dimensions of image RGB (3 values)

```
Dimensions of dataset: (1024, 150, 150, 3)
```

Model Architecture

Sequential Model!

- Adding layers

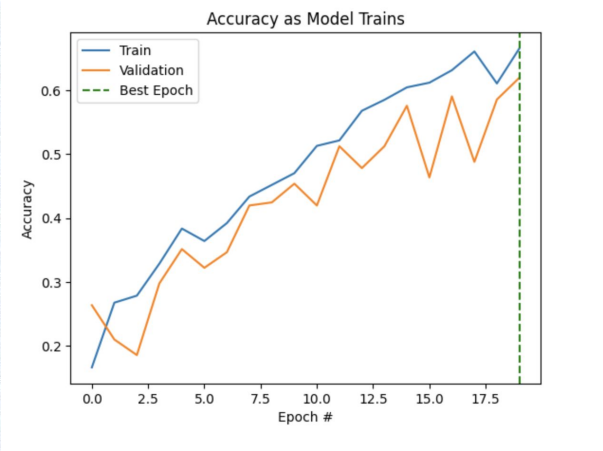


Results

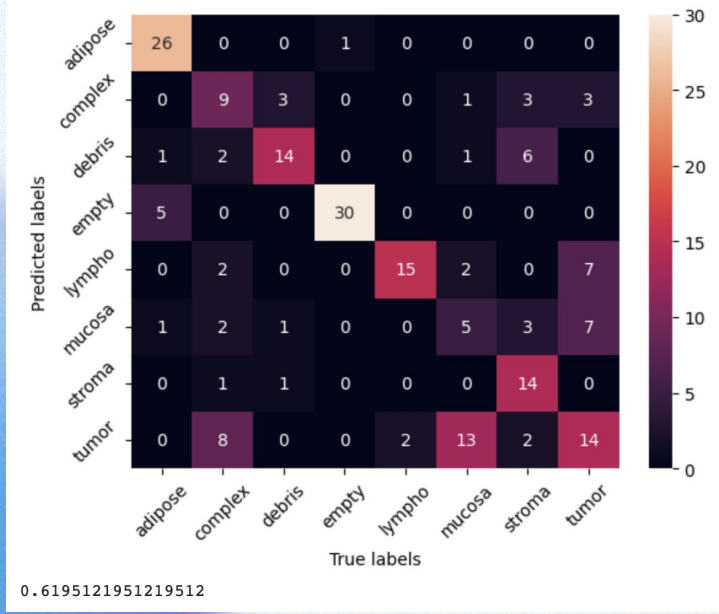
Loss: 1.18

Accuracy: 0.619

Training-Test Curve



Confusion Matrix

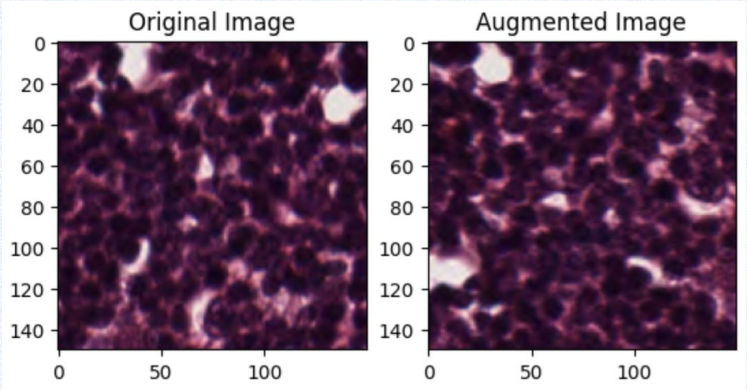


Augmented Learning

1. Augmented learning specifically applied to this project means editing our images to create “more data”
    a. For example, we can rotate the images, flip the images, etc
    b. In the image below, there is a colorectal cancer tissue and it was rotated 90 degrees

In order to augment the image:

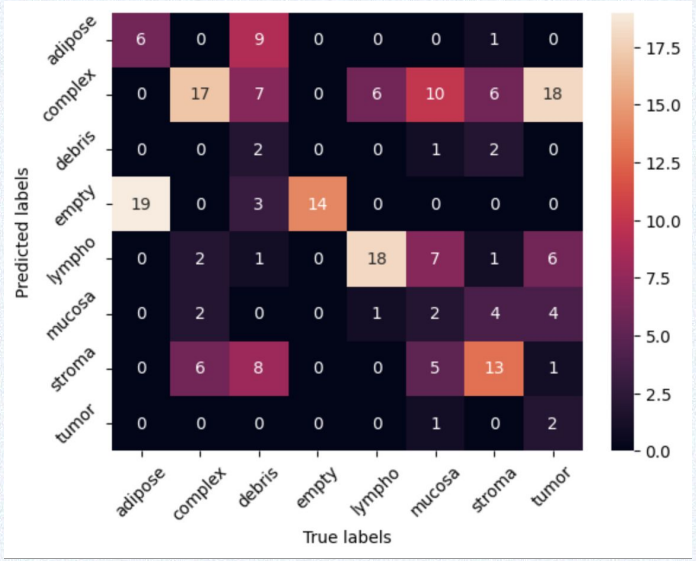
1. Create function that takes in the original image
2. Within the function, we can either rotate or flip the images
3. Return that the augmented image out of the function



Augmented Learning Results

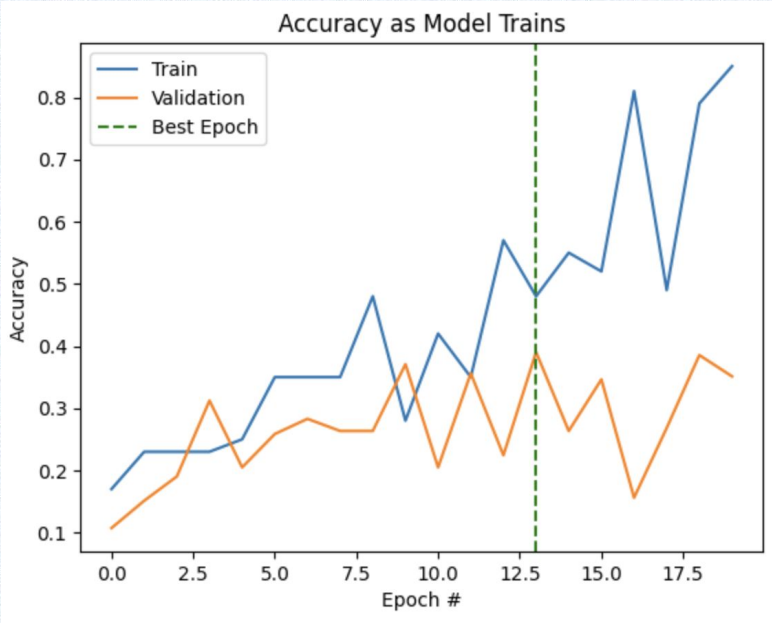
1. Using our original model, we can input our new training data with all of the augmented images, we use the same testing data and the results are as follows:

Confusion Matrix



Accuracy: 0.3609756

Training-Test Curve

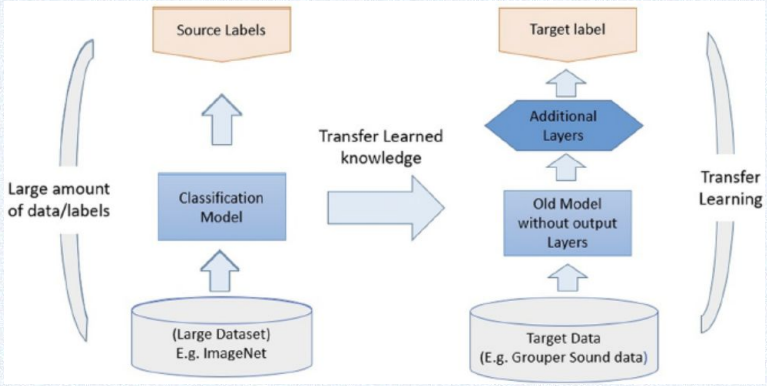


Transfer Learning

1. Transfer learning is using an already established model to run your data through
2. The image below shows what a transfer model does

In order to use the model:

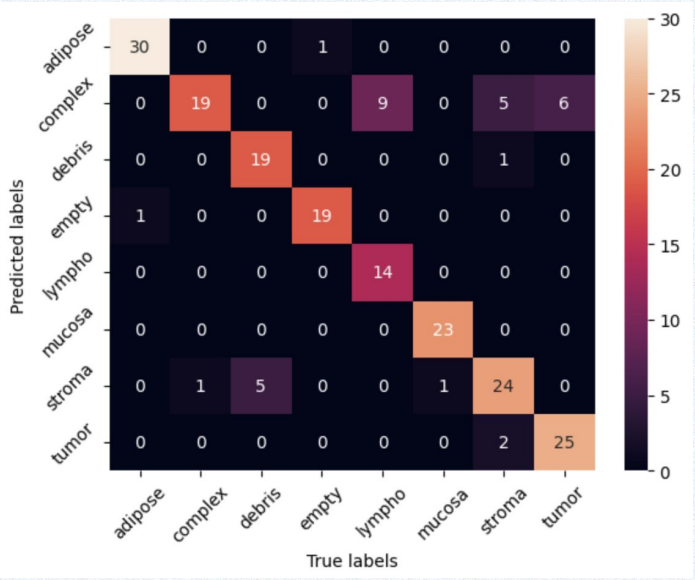
1. The predictions output layer to 8 instead of 1000 + softmax + dense layer
2. Freeze all the weights except the output layer
3. 150 x 150 pixels to 224 x 224 pixels



Transfer Learning Results

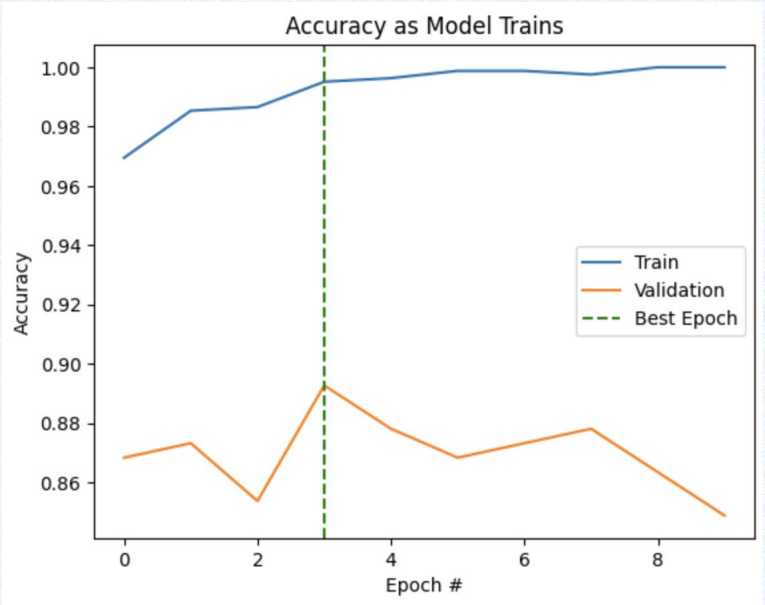
1. Using our new transfer model, we can input the resized images while using the same testing data
2. Below shows the results of the model:

Confusion Matrix



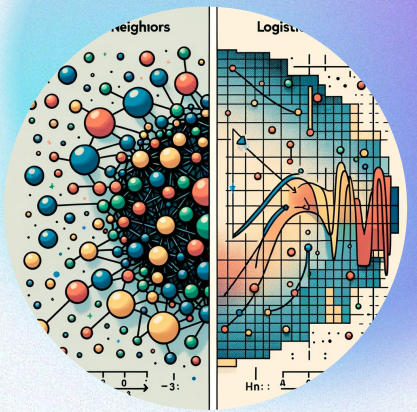
Accuracy: 0.843902439

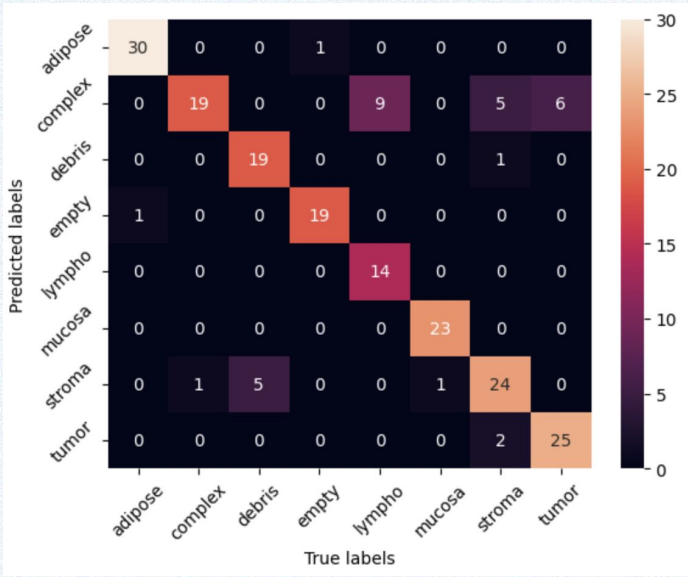
Training-Test Curve



Strengths

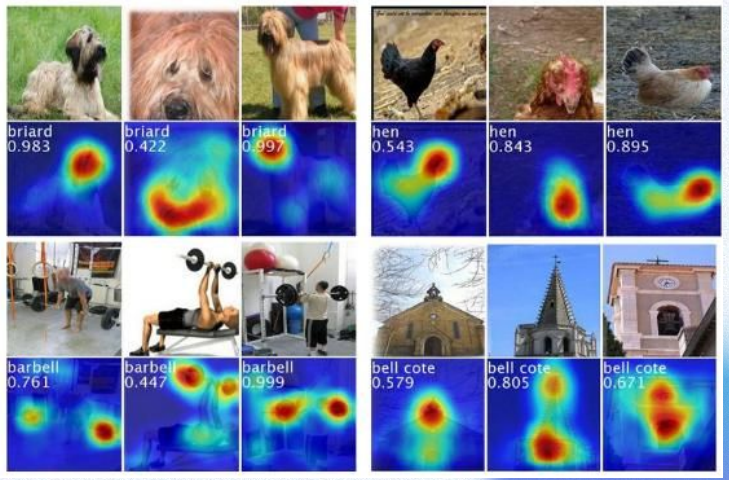
1. Predicting over eight categories of colorectal tissue
2. Accuracy is continuously improving
    a. Learn from a wide variety of cases
3. Integration
    a. Can be used with other models to predict disease progression, type of treatment response, and patient outcomes
4. Detection & Diagnosis
    a. Analyzing tissue much quicker than normal ways

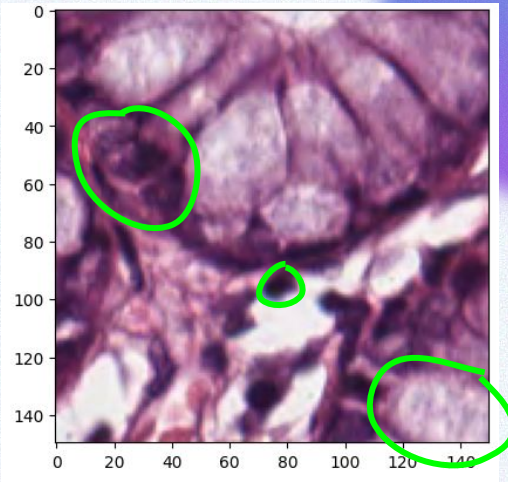




Limitations

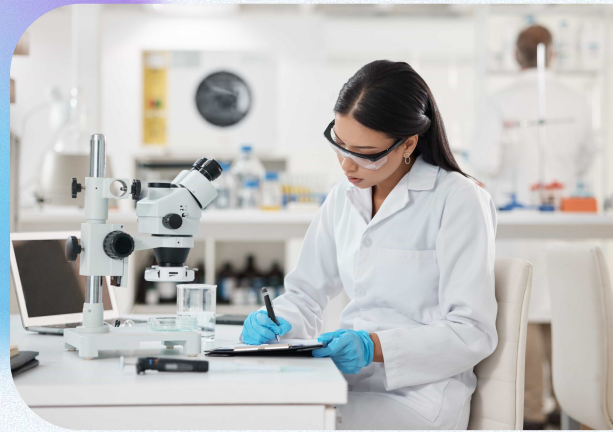
1. Bias
   a. Limited data set from a single hospital; not representative of all
2. Interpretability
   a. What is the driving force behind the model's predictions?
      i. Saliency Map
3. Trust and transparency
   a. Difficult for consumers to adopt AI recommendations
4. Integration into healthcare
   a. Changes to healthcare infrastructure might be difficult

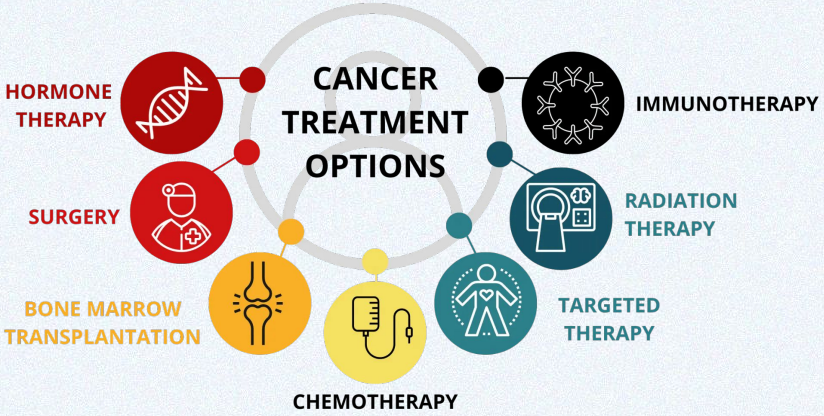




Implications

1. Field of Pathology & Pathologists
    a. Changing responsibilities
        i. Second opinion
    b. Decrease or increase in the job market
2. Future Use
    a. Disease severity
    b. Treatment
    c. Outcomes
3. Possible expansion into other types of cancer





Thank You!