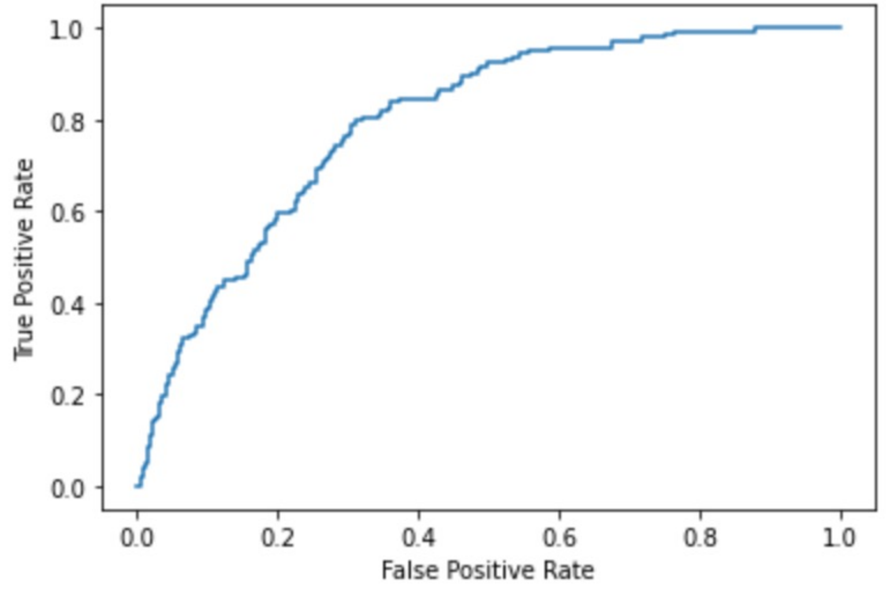
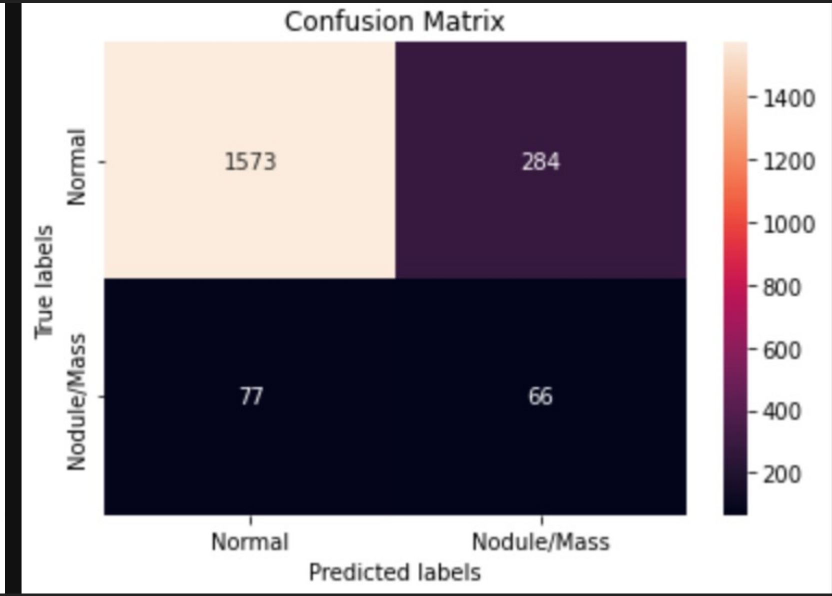
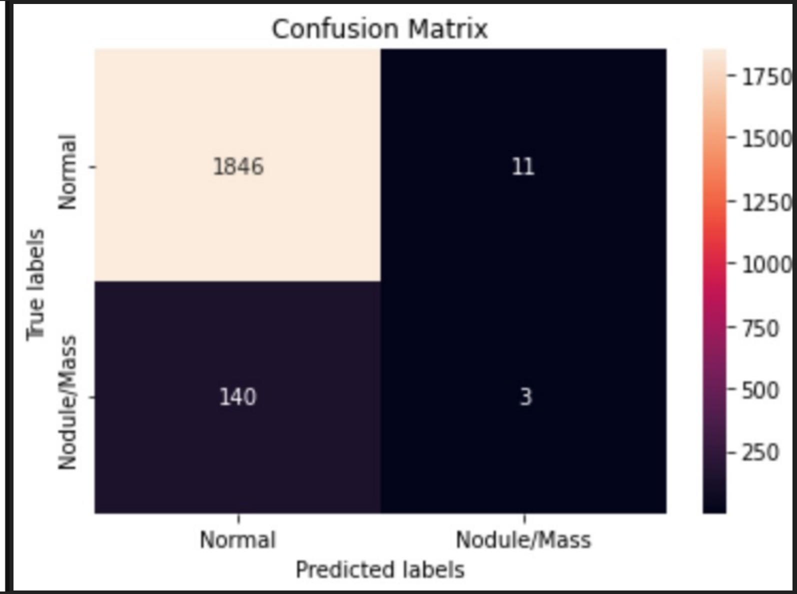
Behold AI Task







Initial Model & Analysis

Recall = 0.46

Precision = 0.19

F1 score = 0.27

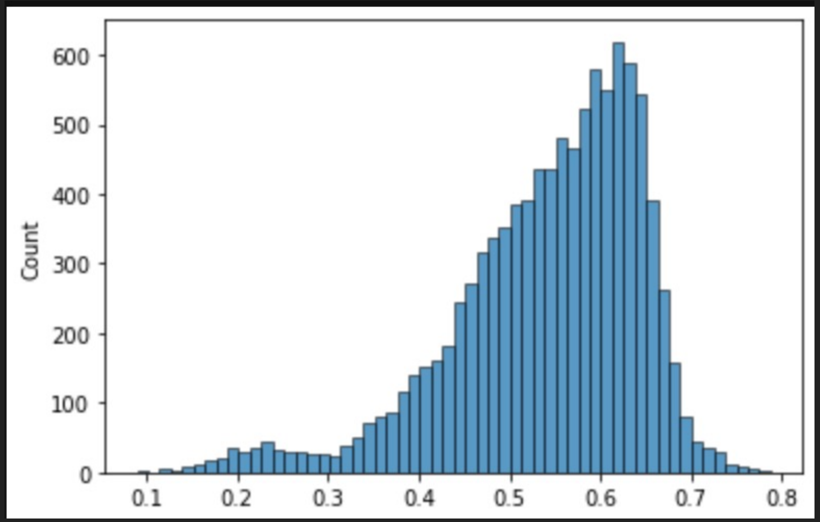
AUC = 0.795

Confusion matrix at 0.1 & 0.5 threshold

Data Augmentation

- Brightness variation:
  - Analysed mean brightness variance in training dataset
  - Used standard deviation (std) of brightness
  - Used torchvision class to randomly vary brightness within range
  [max(0, 1 – mean brightness std), 1 + brightness std]

```
transforms.ColorJitter(brightness=img_brt_std),
```



- Horizontal Flipping
  - Random horizontal flipping (p=0.5)

Loss Function & Optimiser

- ML Problem: Imbalanced dataset.
- Business problem: False Positive preferrable to False Negative

- Reweighted Binary Cross Entropy to optimize for recall (of nodule/mass class)
  - Positive weight = ratio of positive examples to negative examples

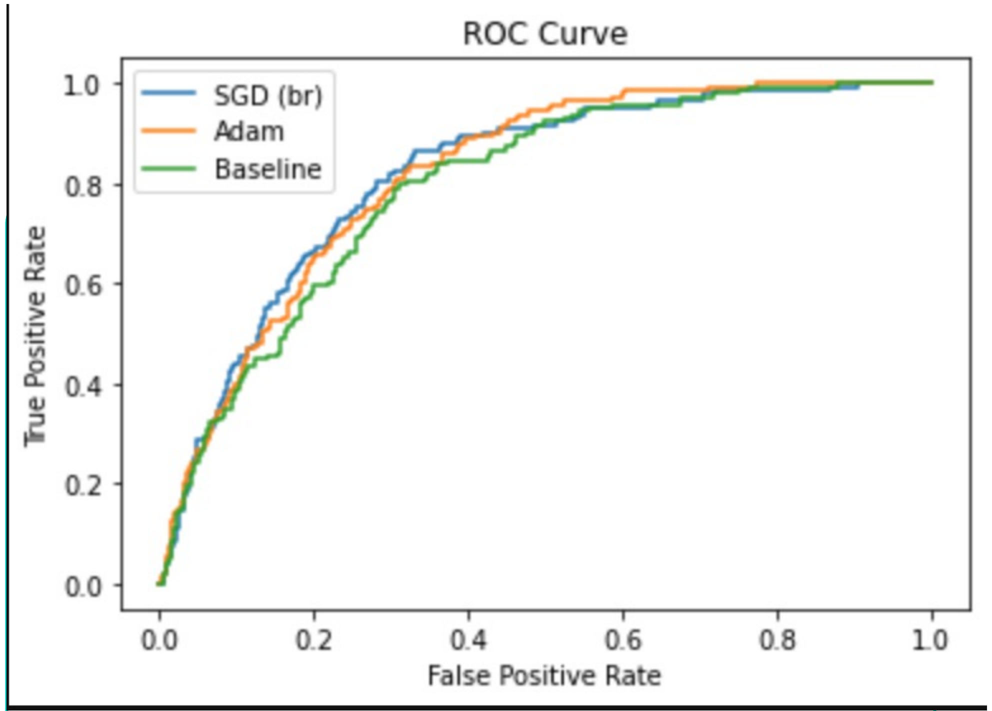
```
self.BCELoss = nn.BCEWithLogitsLoss(pos_weight=self.pos_weight)
```

- Experimented with SGD w/ momentum vs Adam

Results

Adam vs SGD

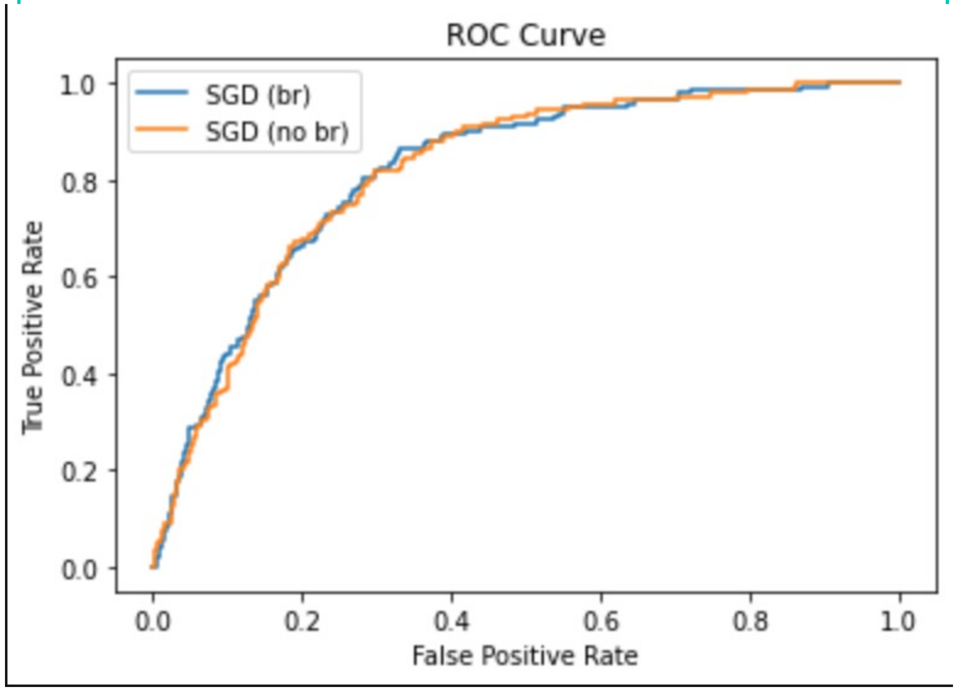
- Trained both models for 20 epochs.
- Adam AUC (brightness) = AUC 0.818
- SGD (no brightness) AUC = 0.815
- SGD (brightness) = 0.821



Brightness Variation

- Brightness Variation of 1 std slightly improves model performance

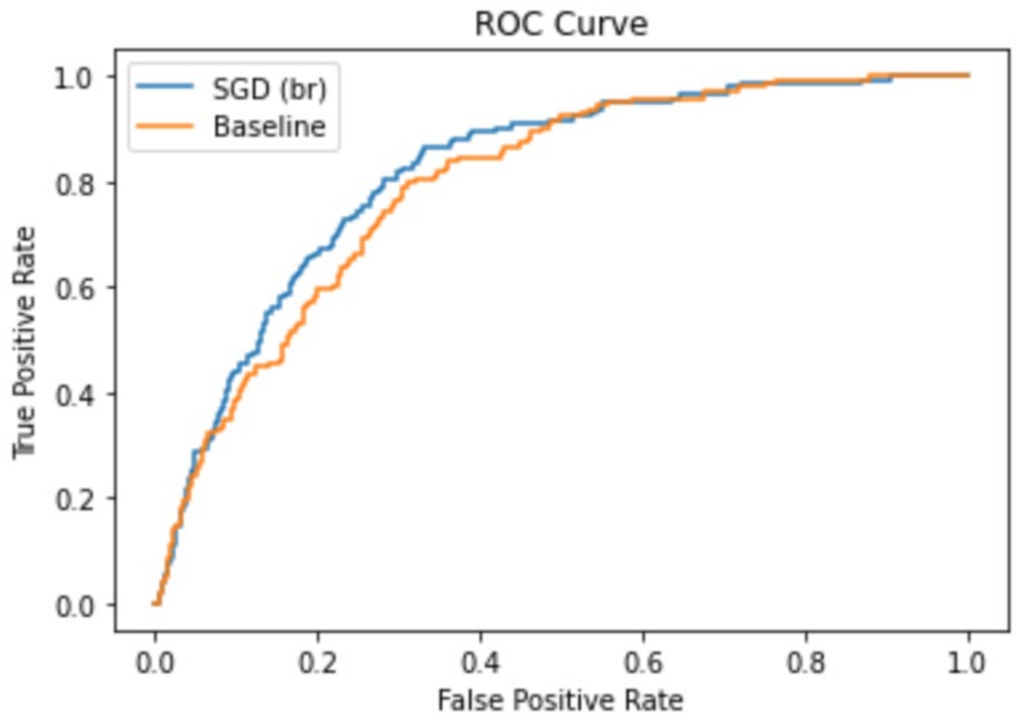
- Doesn’t have significant impact

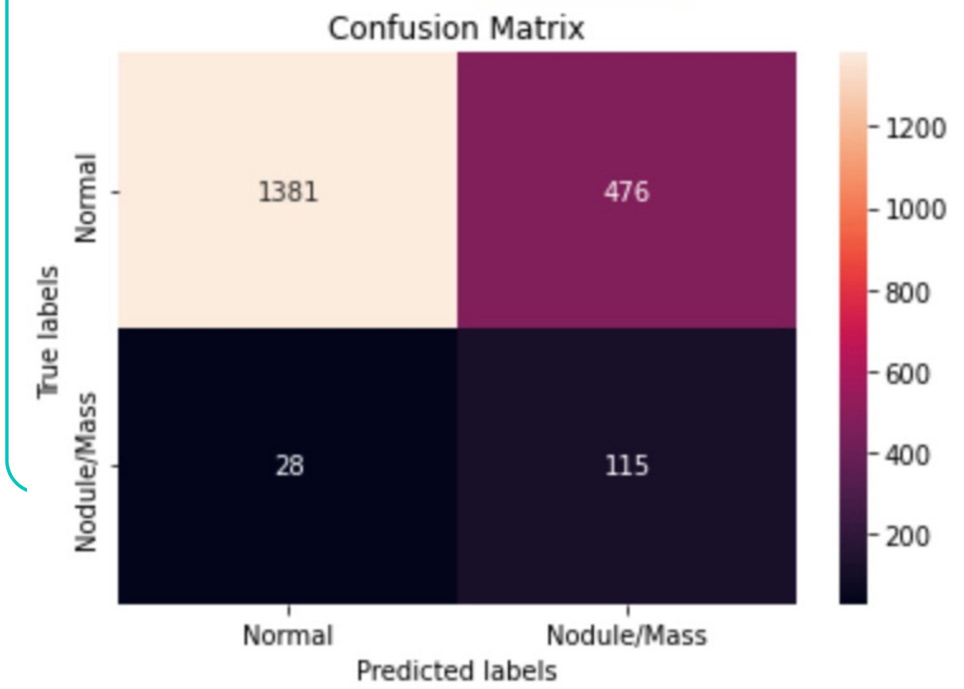


Results

Best model

- Chose SGD w/ Brightness
- Because of the weighting to positive examples, the ROC curve has a higher gradient initially.
- Similar AUC overall to Adam.
- But better recall at lower thresholds. Given consideration of FP cost vs FN cost, I believe this makes it preferrable.
- Threshold tuned to 0.16.
  - Nodule/Mass Recall = 0.80
  - Nodule/Mass Precision = 0. 2
  - Nodule Mass F Score = 0.61 (beta = 3)
  - Maximisied F-score for beta = 3





What to label?

- Prioritise labelling of images closest to decision boundary (i.e. that the model is most uncertain about).
- From the chosen model, the decision boundary/threshold was empirically chosen as 0.16.
- Choose the 300 assets closest to the decision boundary.