Credit: allaboutvision.com

Eye disease multi-classification using retinal images

By:
Maybelle Auw
22-Oct-2022

Problem statement

Health Promotion Board is searching for innovative, cutting-edge technological **solutions to facilitate mass eye screening** of common eye diseases in the general adult population.

They hope to **automate interpretation of retina screening images**, shorten the time taken to flag high-risk persons for further health assessment.

This would **allow early intervention,** lower risk of disease progression and **lower healthcare cost burden**. By minimising number of people in the population with severe eye disease, we minimise the use of more costly therapies.

Possible stakeholders: public health authorities, eye clinics, optical shops

My proposed solution

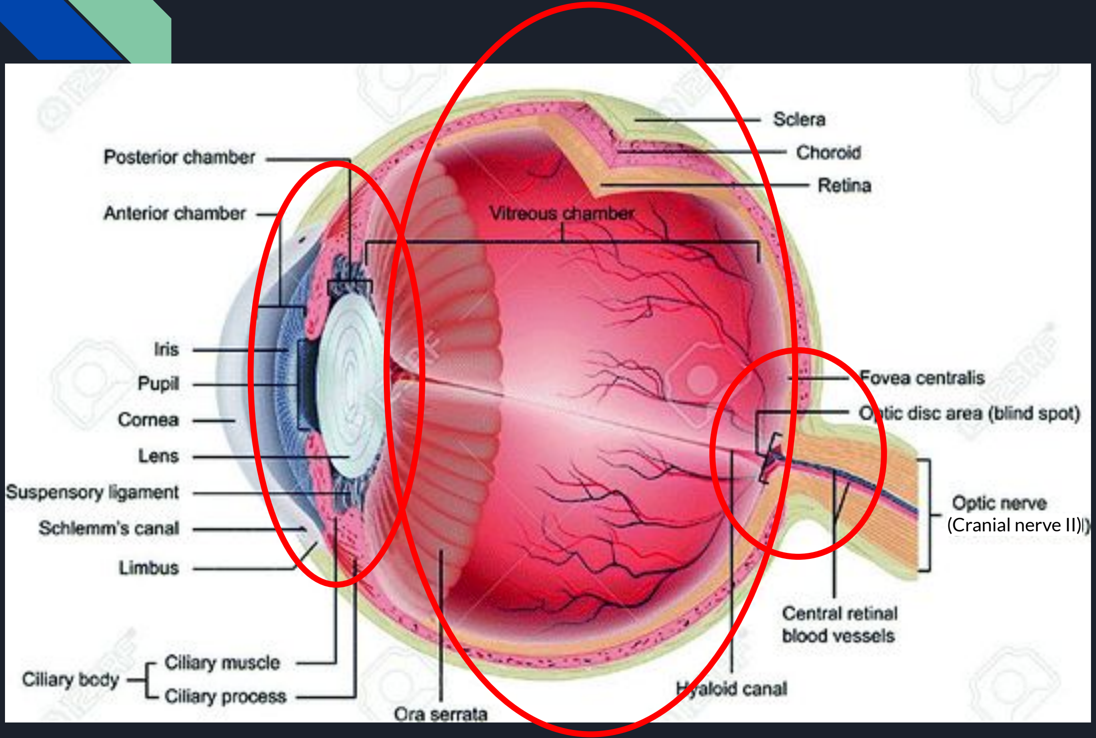
Using 4,217 retinal images collected from various sources, I constructed a machine learning algorithm to help classify a RGB retina image to any of 3 classes of common adult eye diseases and a **normal** control group:

- Cataract
- Diabetic retinopathy
- Glaucoma
- normal

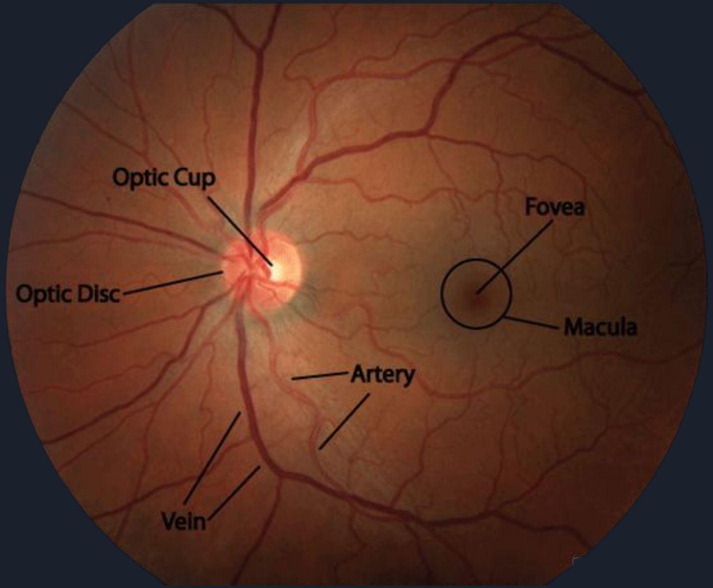
What do I want the model to look for in the images?

Terminology

Basic eye anatomy

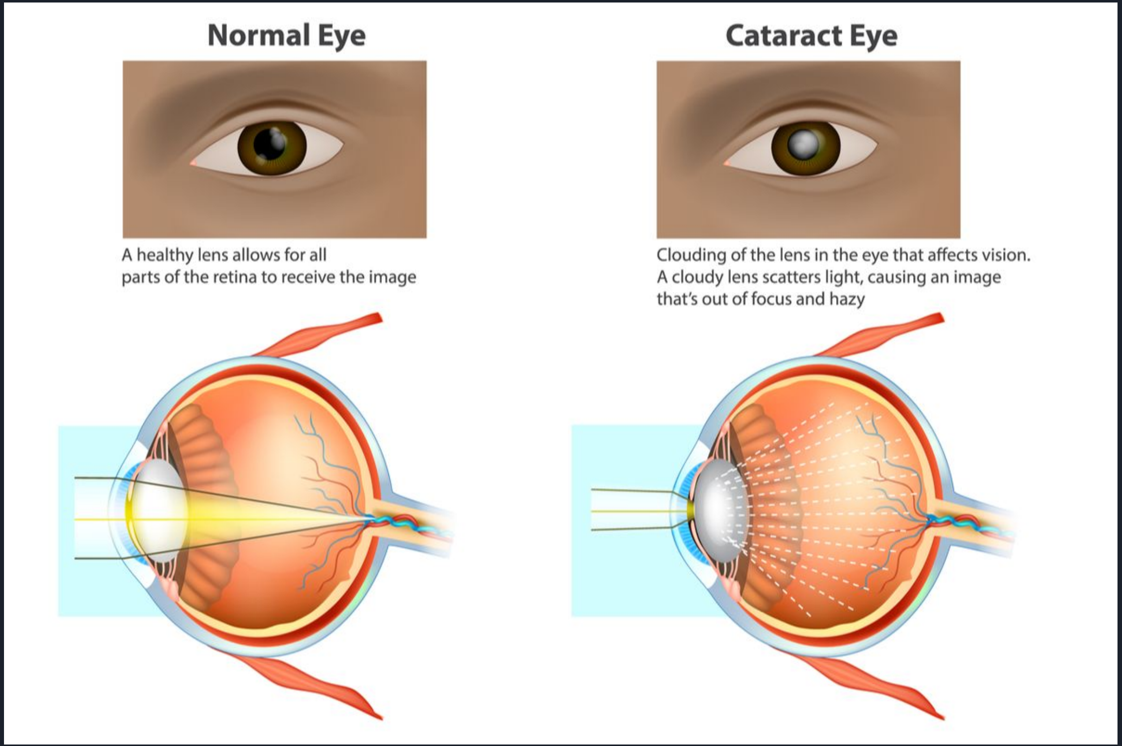


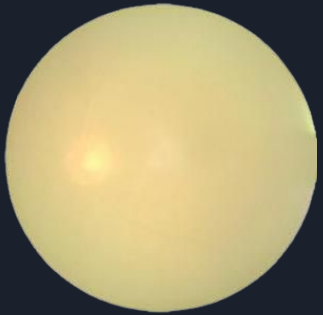
Credit: Springer



Credit: stanfordmedicine25.stanford.edu

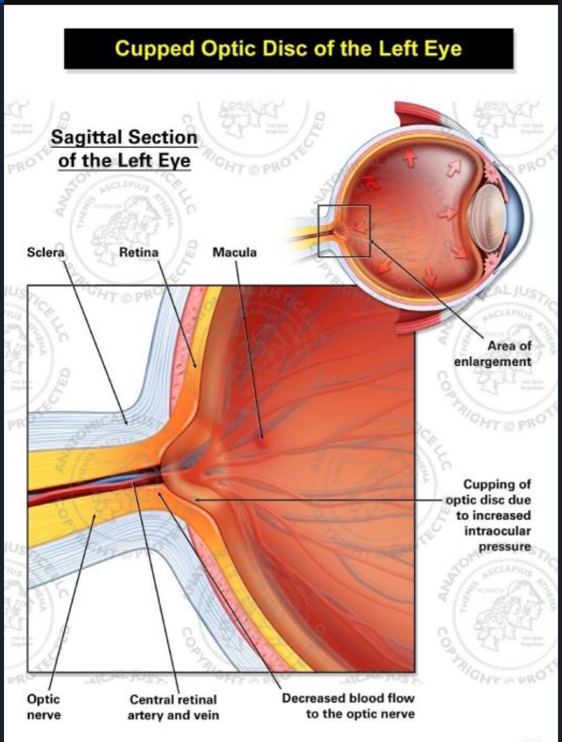
Cataract





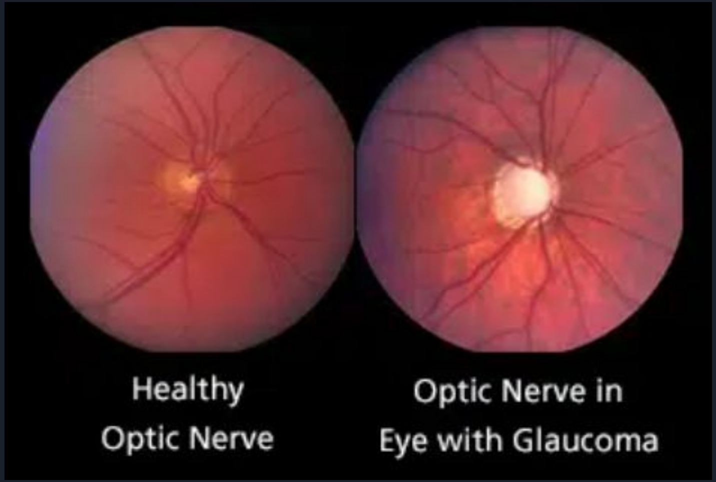
Credit: nvisioncenters.com

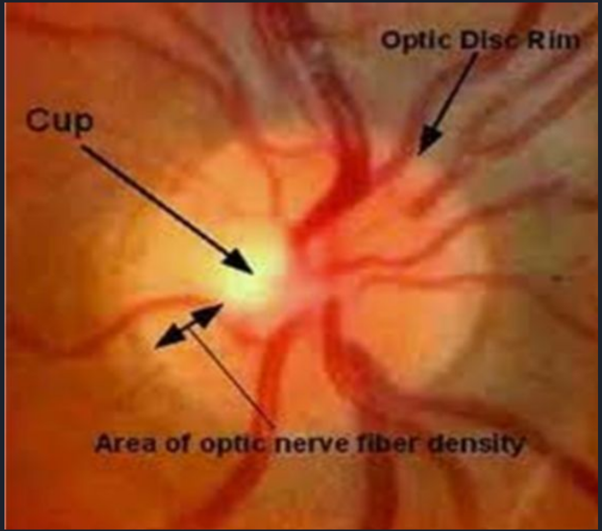
Glaucoma



Credit: anatomicaljustice.com

Credit: glaucoma.org

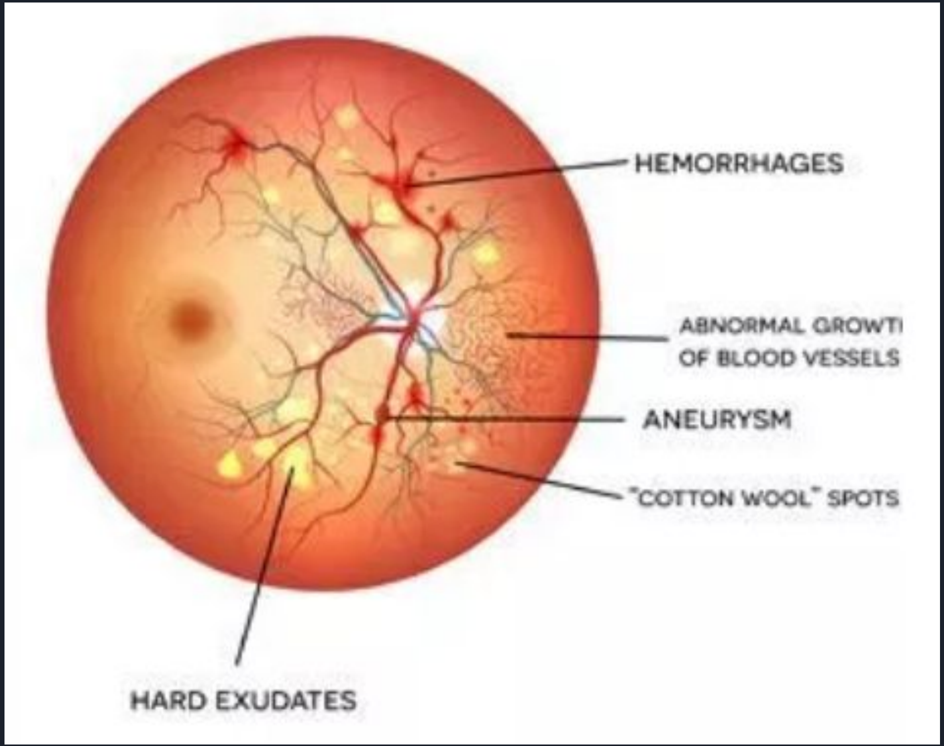




- cup-to-disc ratio > 0.5 = abnormal
- physiologic limit of 0.5

Credit: intechopen.com

Diabetic retinopathy



Back to Project

Image dataset

- 4,217 images
  - Cataract: 1038
  - Diabetic retinopathy: 1098
  - Glaucoma: 1007
  - Normal: 1074
- Attributes:
  - Variety of sizes
  - RGB
  - *.png, *.jpg, *.bmp
- Various sources

Image hashing

- 4215 unique images

Train / Validation / Hold-out:

- 0.5 / 0.25 / 0.25

Preprocessing

Resize images for specific CNN requirements

- EfficientNet (160, 160)
- InceptionResNetV2 (299, 299)

Re-scale image pixels from (0, 255) to between 0 and 1

Run base models

Metrics

- Categorical accuracy
- Precision
- Recall
- AUC
- F1 score (custom)

Optimizer = Adamax

Loss = Categorical crossentropy

Callbacks:

- EarlyStopping (patience=10)

Models:

- EfficientNetV2S
- InceptionResNetV2

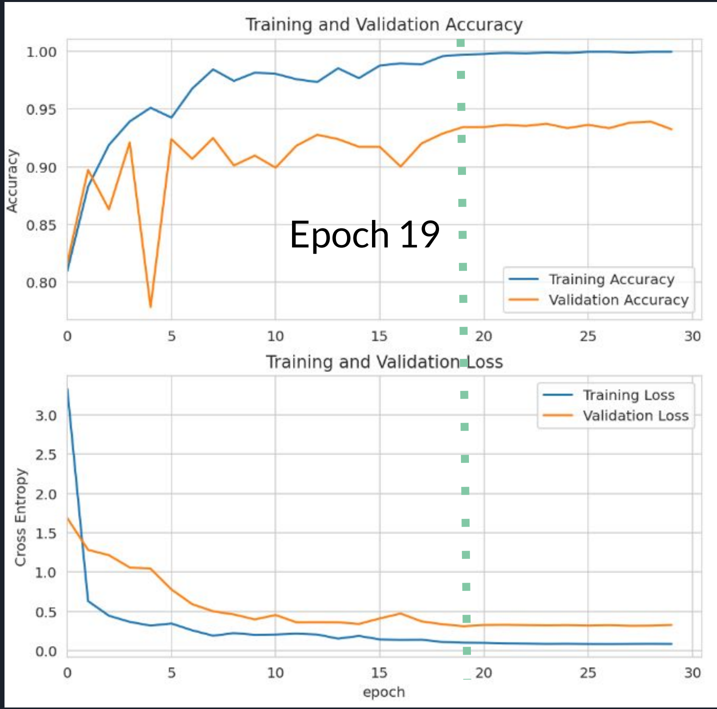
Weights

- ImageNet as reference
- Trainable = True

Added top layers:

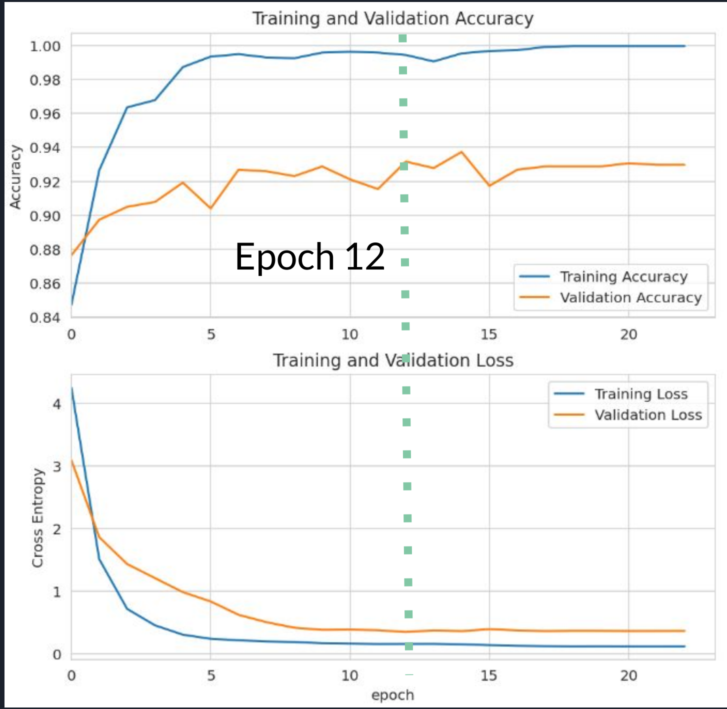
1. BatchNormalization
2. Dense (regularization)
3. Dropout (0.4)
4. Dense (4, softmax)

Base EfficientNetV2S epoch run



Best val loss: 0.311
Val accuracy: 93.44%
Accuracy generalisation: 6.29%

Base InceptionResNetV2 epoch run



Best val loss: 0.347
Val accuracy: 93.16%
Accuracy generalisation: 6.31%

Run models with augmentation layers

Added

- RandomFlip(‘horizontal’)
- RandomRotation(0.1)
- RandomContrast(0.1)

Weights

- ImageNet as reference
- Trainable = True

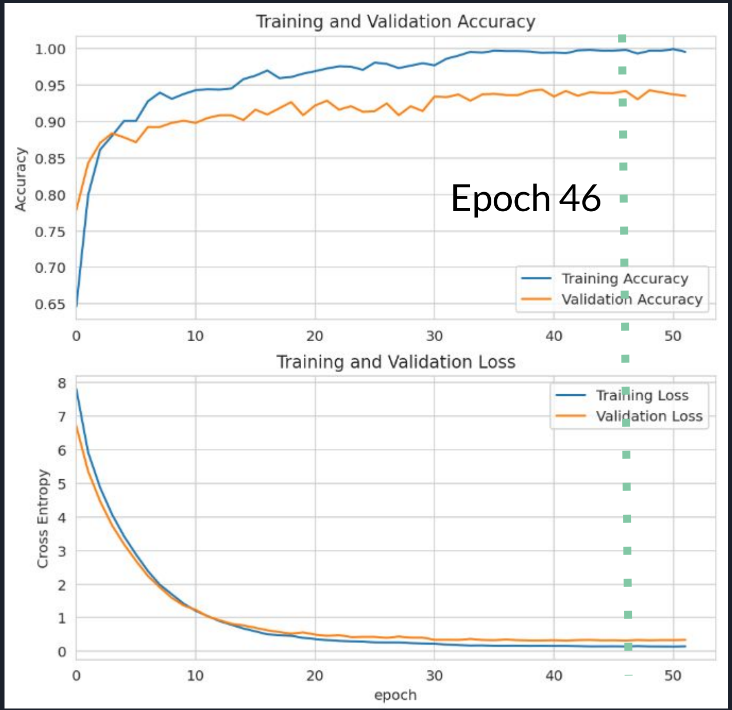
Same metrics

- Categorical accuracy
- Precision
- Recall
- AUC
- F1 score (custom)

Optimizer = Adamax

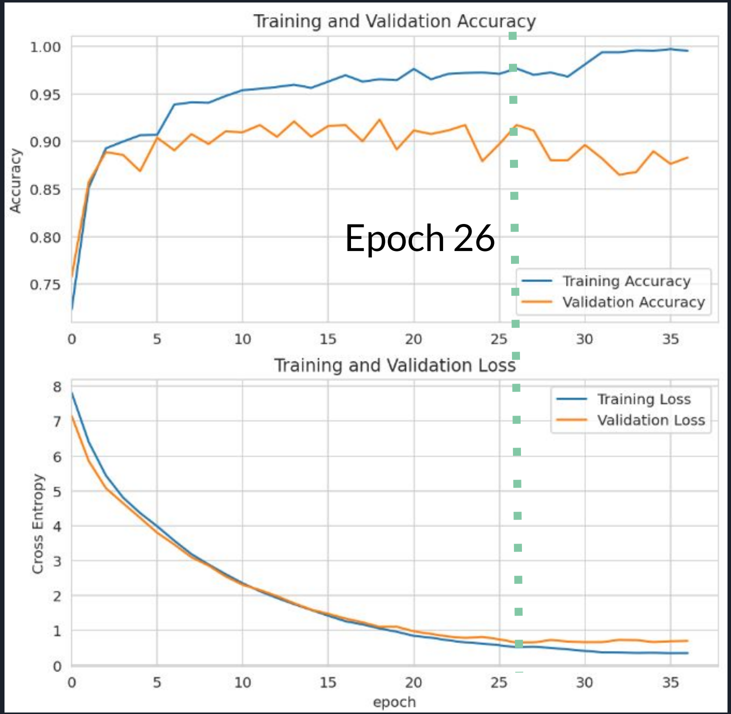
Loss = Categorical crossentropy

EfficientNetV2S + augment layers



Best val loss: 0.312
Val accuracy: 94.2%
Accuracy generalisation: 5.66%

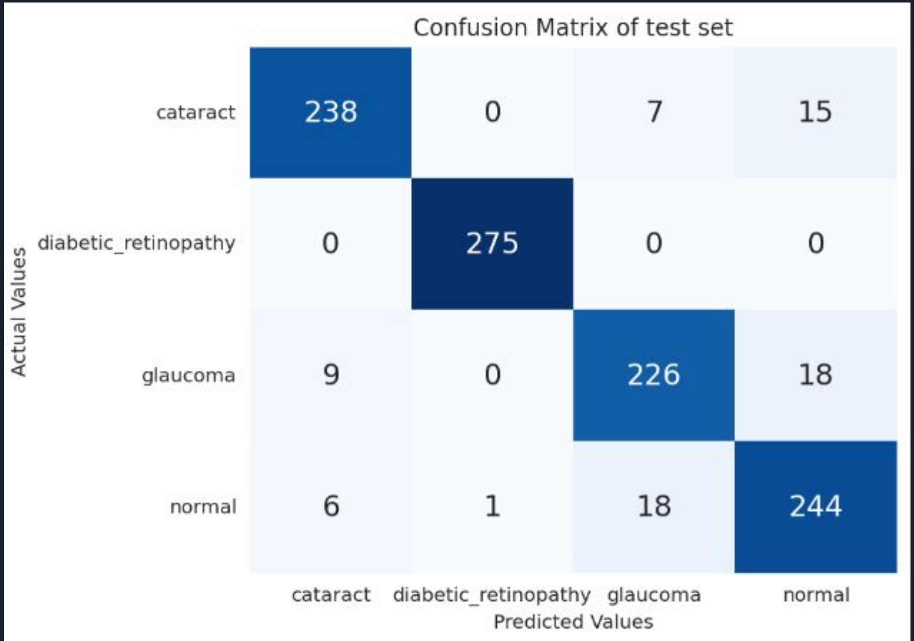
InceptionResNetV2 + augment layers



Best val loss: 0.650
Val accuracy: 91.73%
Accuracy generalisation: 6.09%

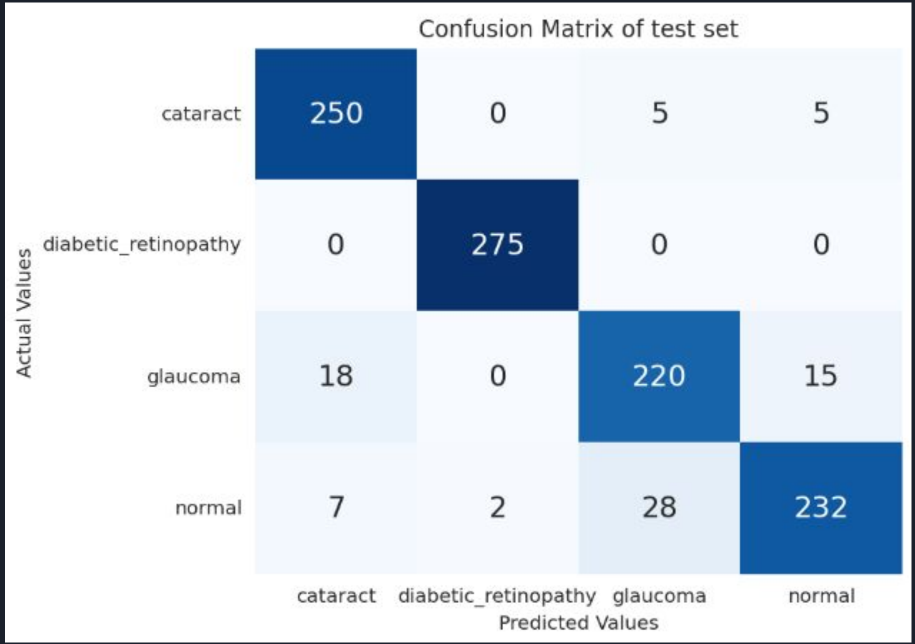
Confusion matrices

Base EfficientNetV2S



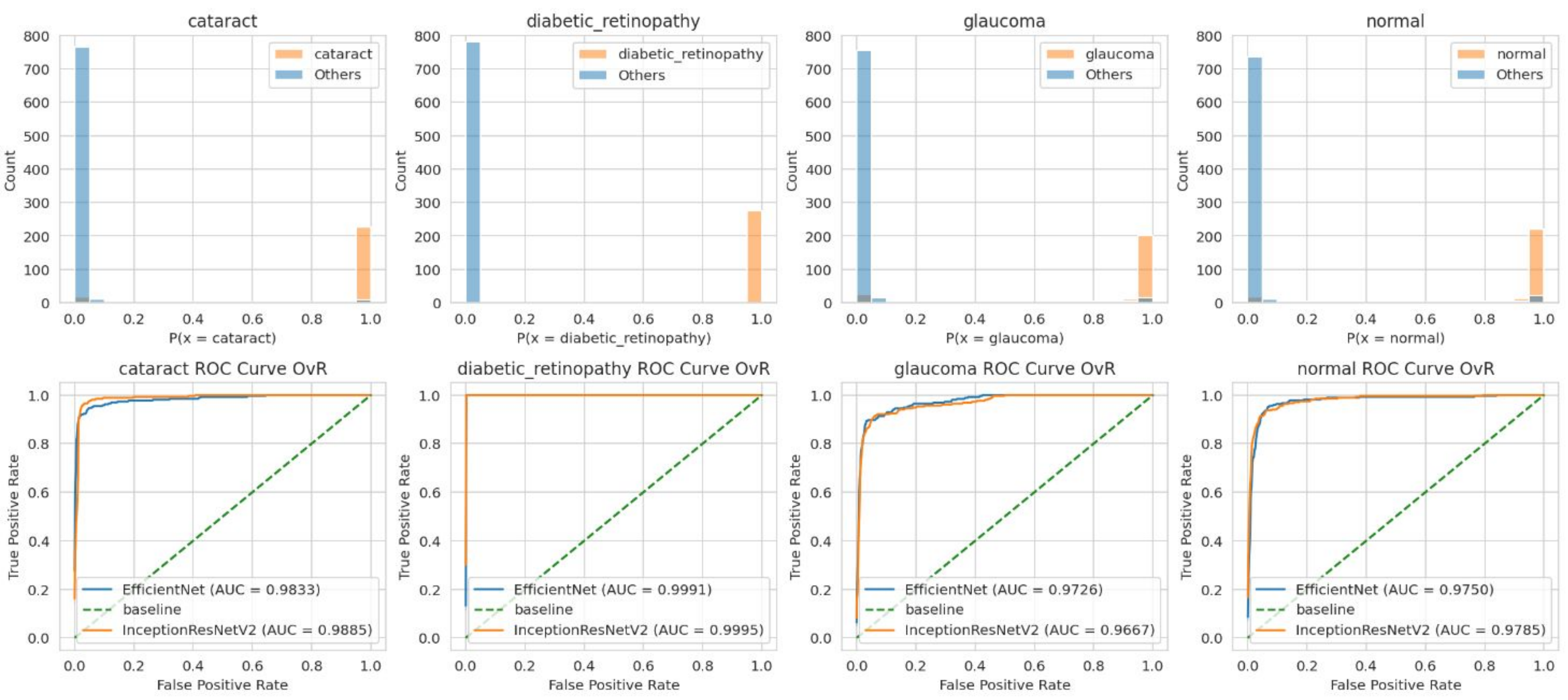
Accuracy: 93%

Base InceptionResNetV2



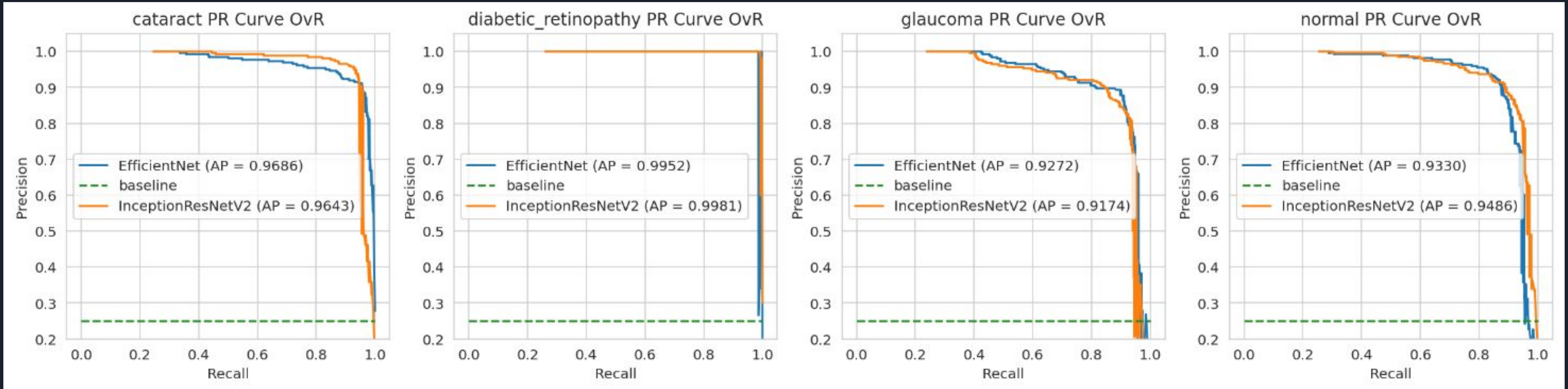
Accuracy: 92.43%

ROC curves



EfficientNet slightly higher false positive rate for Cataract

PR curves

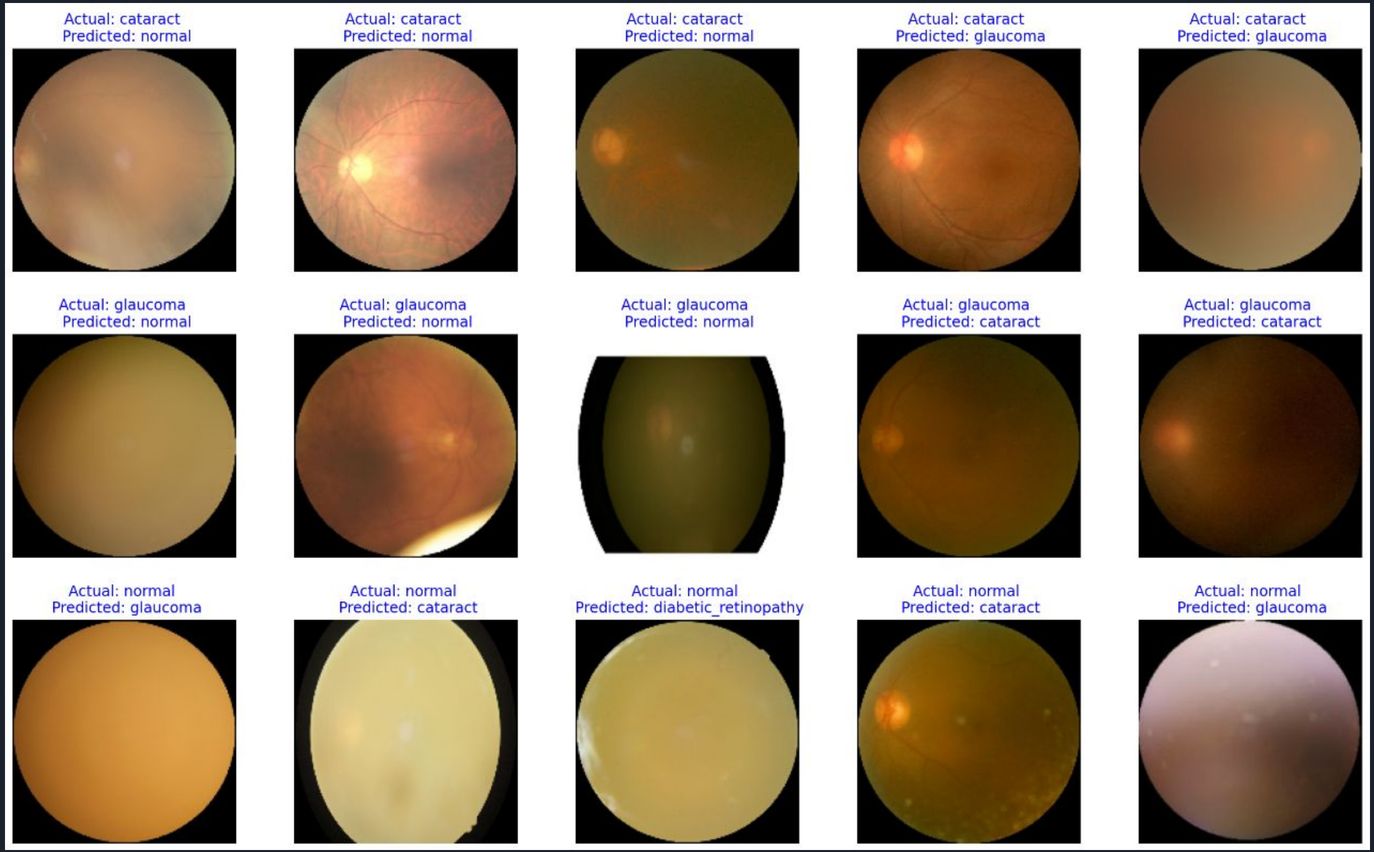


- Aim to minimise false negatives (not miss presence of disease)
- Aim for maximising recall for disease.
- EfficientNet has better recall at precision lower than .9

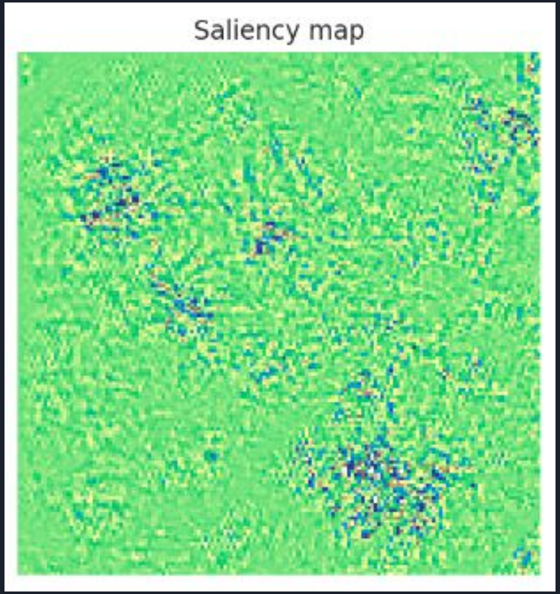
Misclassified images

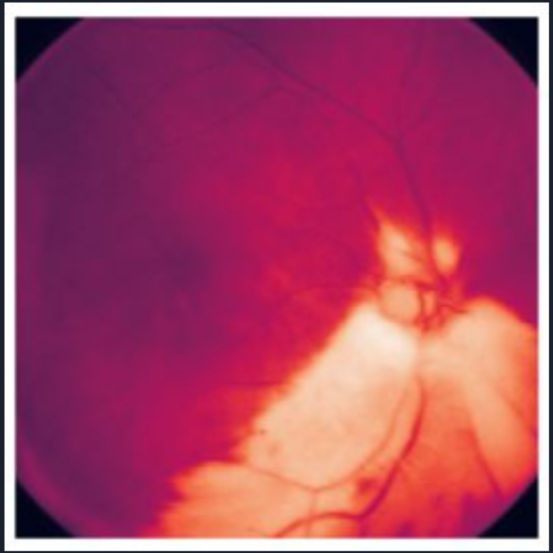
Reasons:

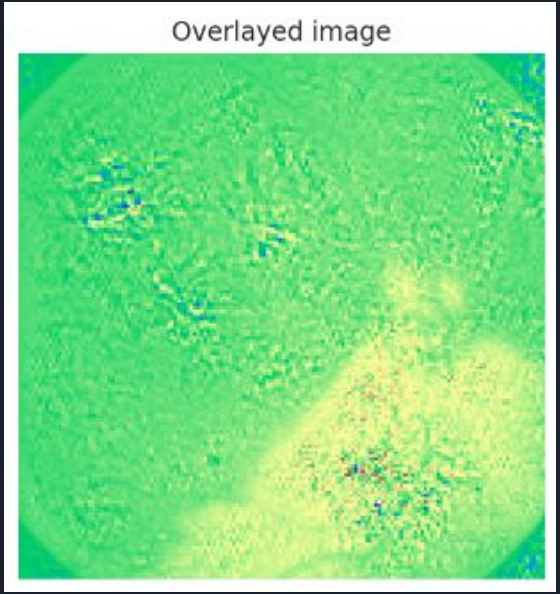
- Mislabelled
- Ambiguous
- Poor quality
- Low differentiation



Saliency map - Diabetic retinopathy example





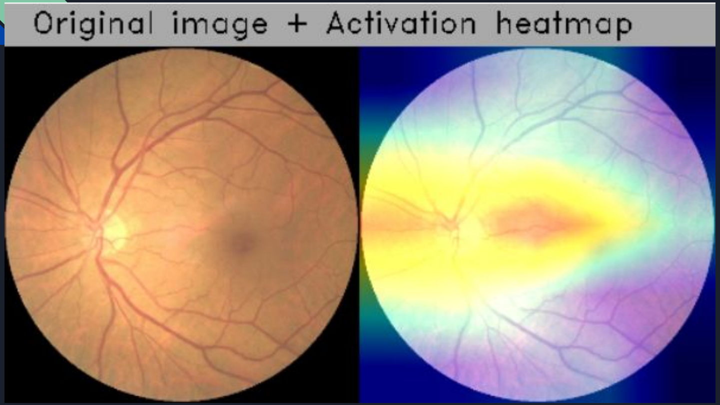


Model doesn’t really know where to focus

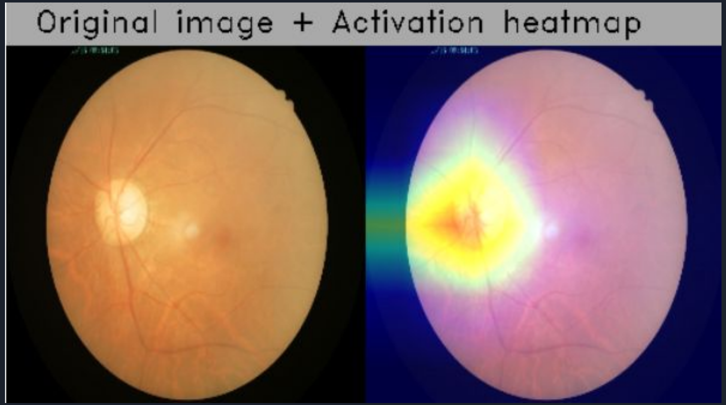
Activation heatmaps

Model **not** activated by appropriate pixels for DRE

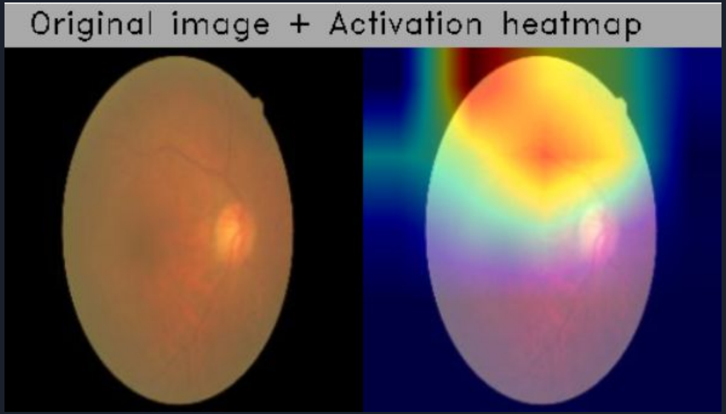
Normal



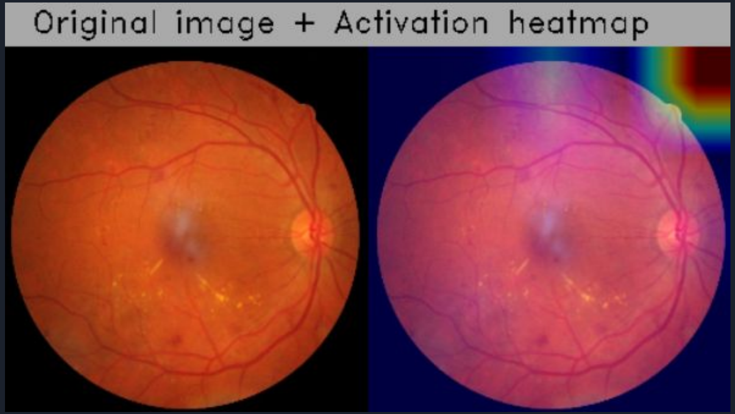
Glaucoma



Cataract



Diabetic retinopathy



Limitations

- Model is restricted to identifying 1 disease class per image. Model is unable to identify multiple diseases if multiple abnormalities are present in retinal image
- Images with Low quality, Low differentiating features from normal may be wrongly classified
- Diabetic retinopathy abnormalities are not giving the strongest signals to the model, meaning images outside of dataset with diabetic retinopathy may be wrongly classified

Conclusions

- EfficientNetV2S base model without augmentation layers is chosen for deployment with 93% accuracy.
- This model is best used with other screening modalities to increase precision and sensitivity of diagnosis e.g.
  - tonometer (for anterior chamber pressure) for glaucoma
  - snellen chart (visual acuity) for overall vision ability
  - auto perimeter visual field analyser (glaucoma)
  - Optical coherence tomography (OCT) - (study retina layers, and subretinal deposits)

Future steps

- Try input images with specific colour channels (e.g. green channel only)
- Remove background, replace with transparency (alpha channel)
- Exclude ambiguous, low differentiation, possibly wrongly labelled images from dataset
- Train a specific model to differentiate between diabetic retinopathy and normal - maybe improve the right pixels activation
- Try Meta Pseudo Label model (semi-supervised learning method)

Q & A