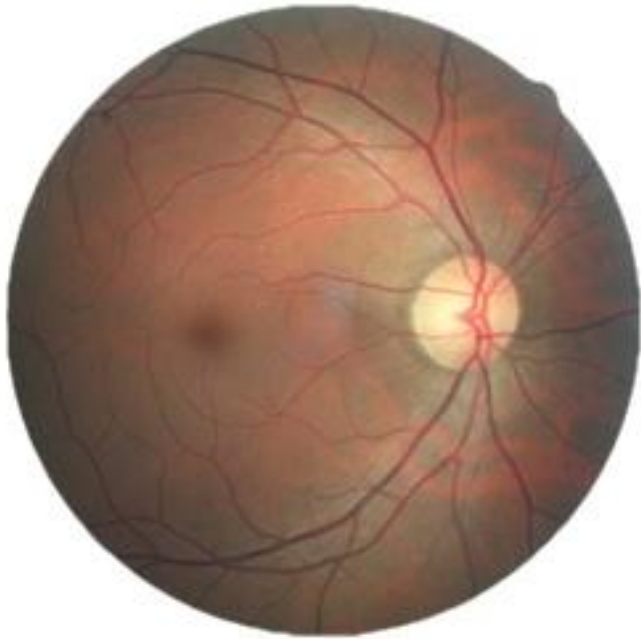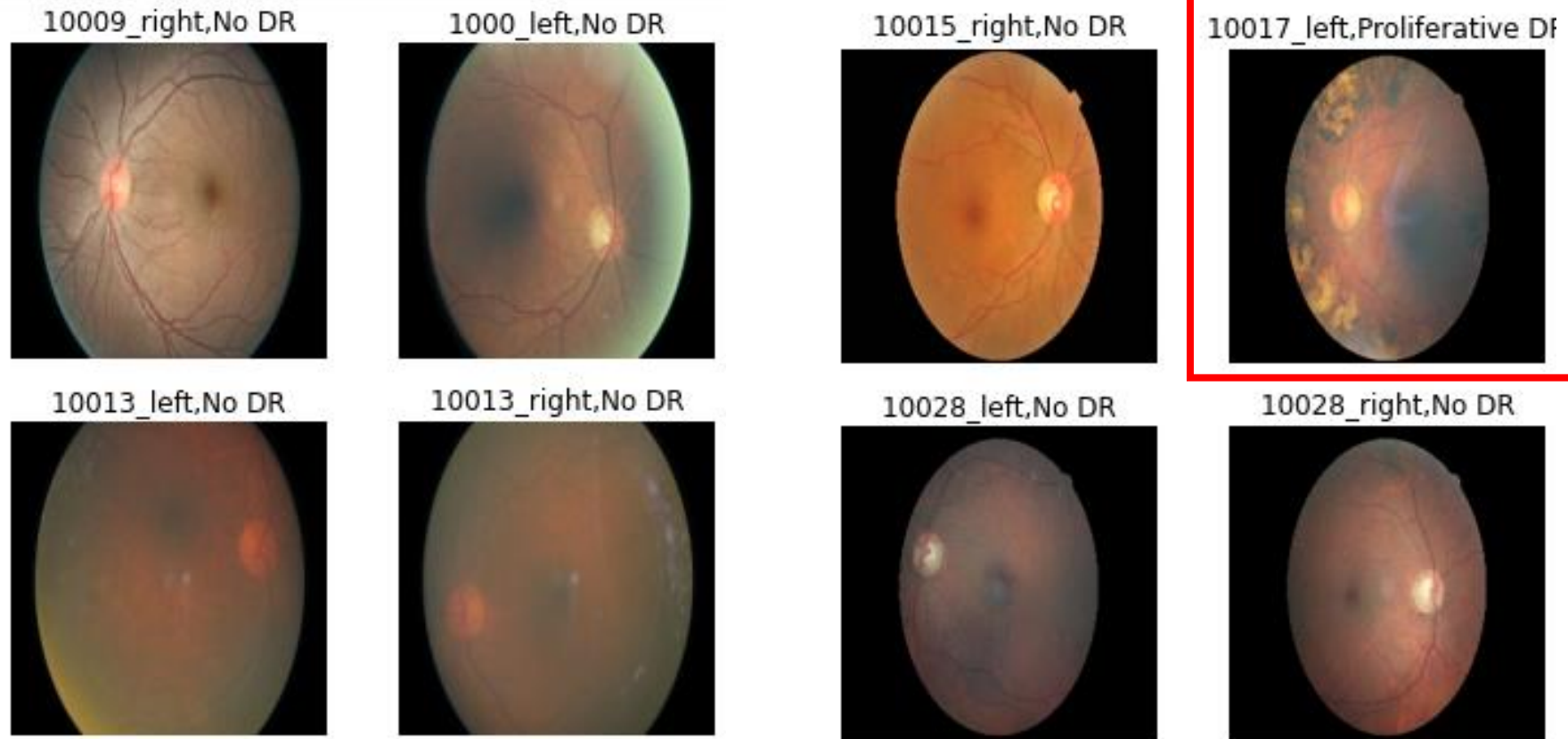# Diabetic Retinopathy – Automated Detection

Diabetic Retinopathy (DR) is a complication resulting from diabetes that affects the eyes. It is caused by the damage of the blood vessels in the tissue at the back of the eye called retina.

Problem Statement:
- To build a model to detect the condition successfully with as high recall as possible

- To build a model that can detect the severity of condition

Diabetic Retinopathy – Automated Detection

10009_right,No DR

10000_left,No DR

10015_right,No DR

10017_left,Proliferative DR

10013_left,No DR

10013_right,No DR

10028_left,No DR

10028_right,No DR

https://www.kaggle.com/competitions/diabetic-retinopathy-detection/overview

Data:

- Kaggle dataset

- Left and Right fundus colored images
- Labels:
  - 0: No DR
  - 1: Mild
  - 2: Moderate
  - 3: Severe
  - 4: Proliferative
- Images under variety of conditions:
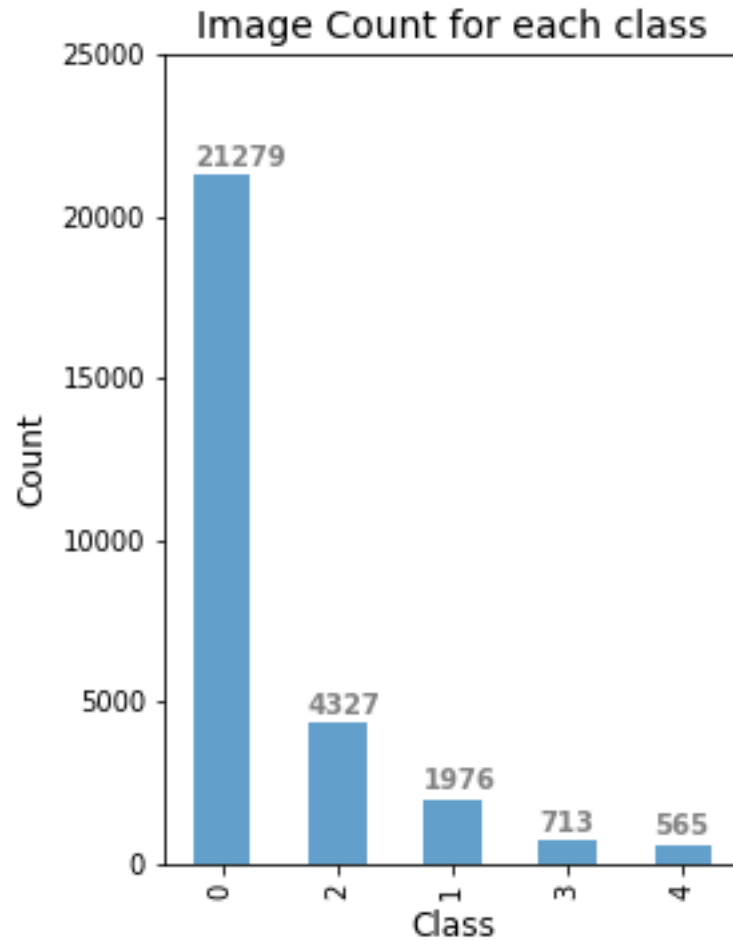 - Overexposed
  - Underexposed
  - Flipped
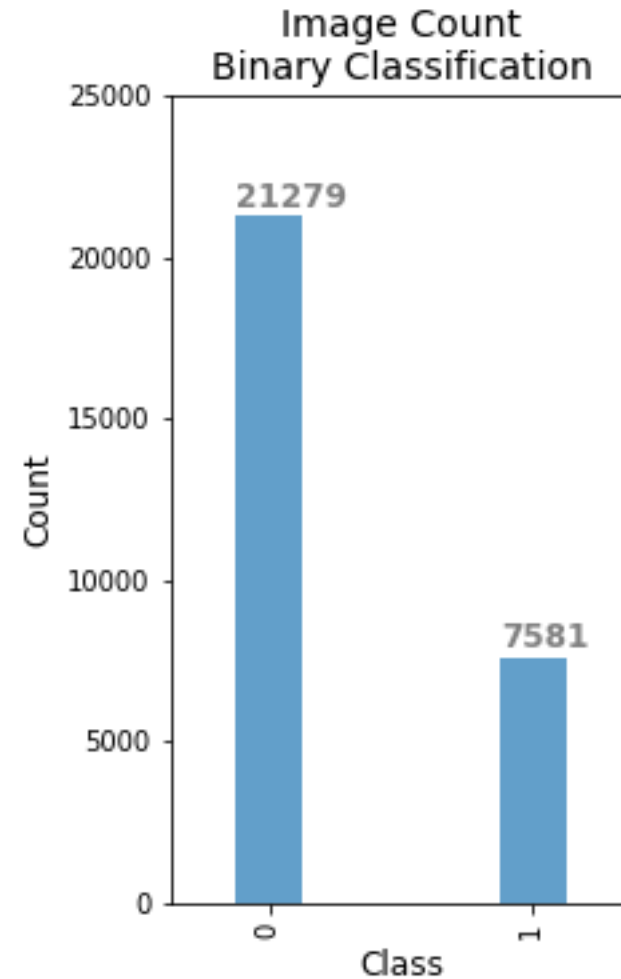 - noisy

Diabetic Retinopathy – Automated Detection

Data

Highly imbalanced data

Whole Dataset

Binary classification

Binary Classification
Approach

Binary classification on whole dataset

Metrics- Recall

1 CNN

CNN models with increasing depth and complexity (with **initial bias**, with and without **Class weights** and **HSV conversion**)

2 Transfer Learning

Mobile Net V2

with and without **Class weights**
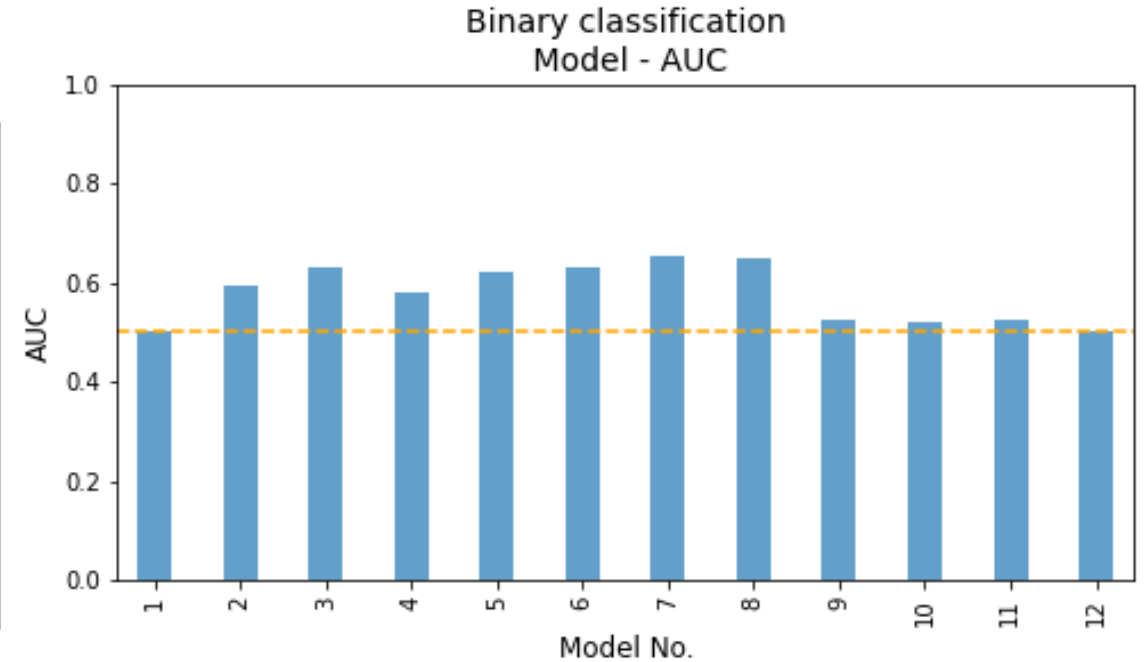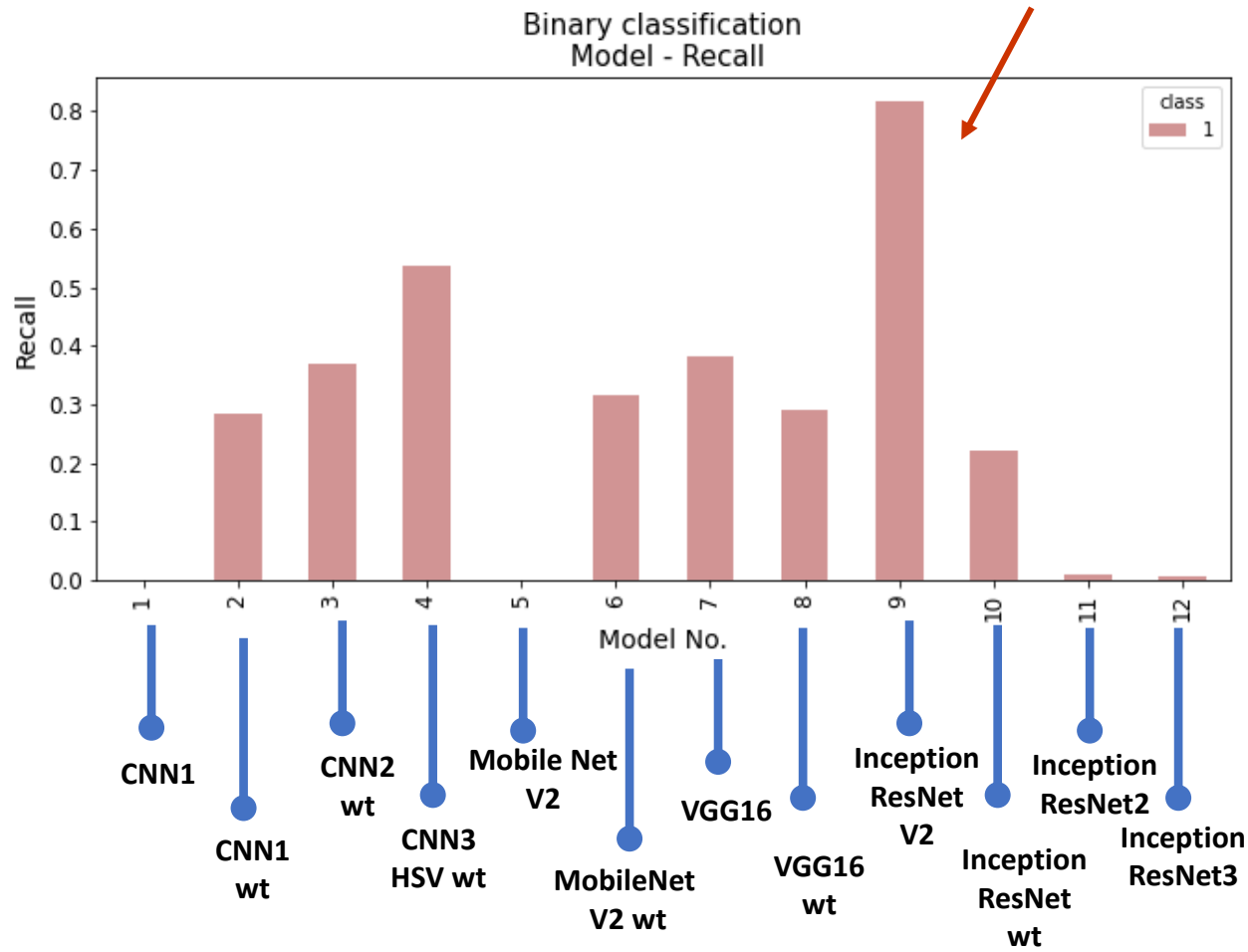
VGG 16

with and without **Class Weights**

Inception ResNet V2

with and without **Class Weights**, and **hidden layer**
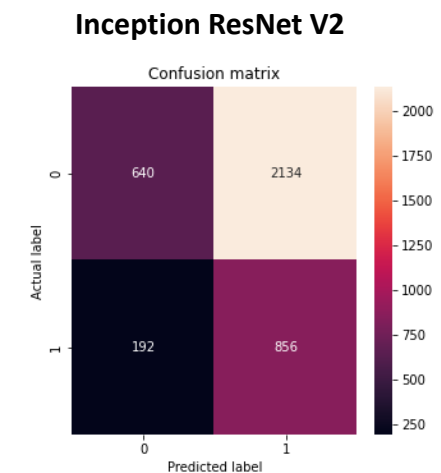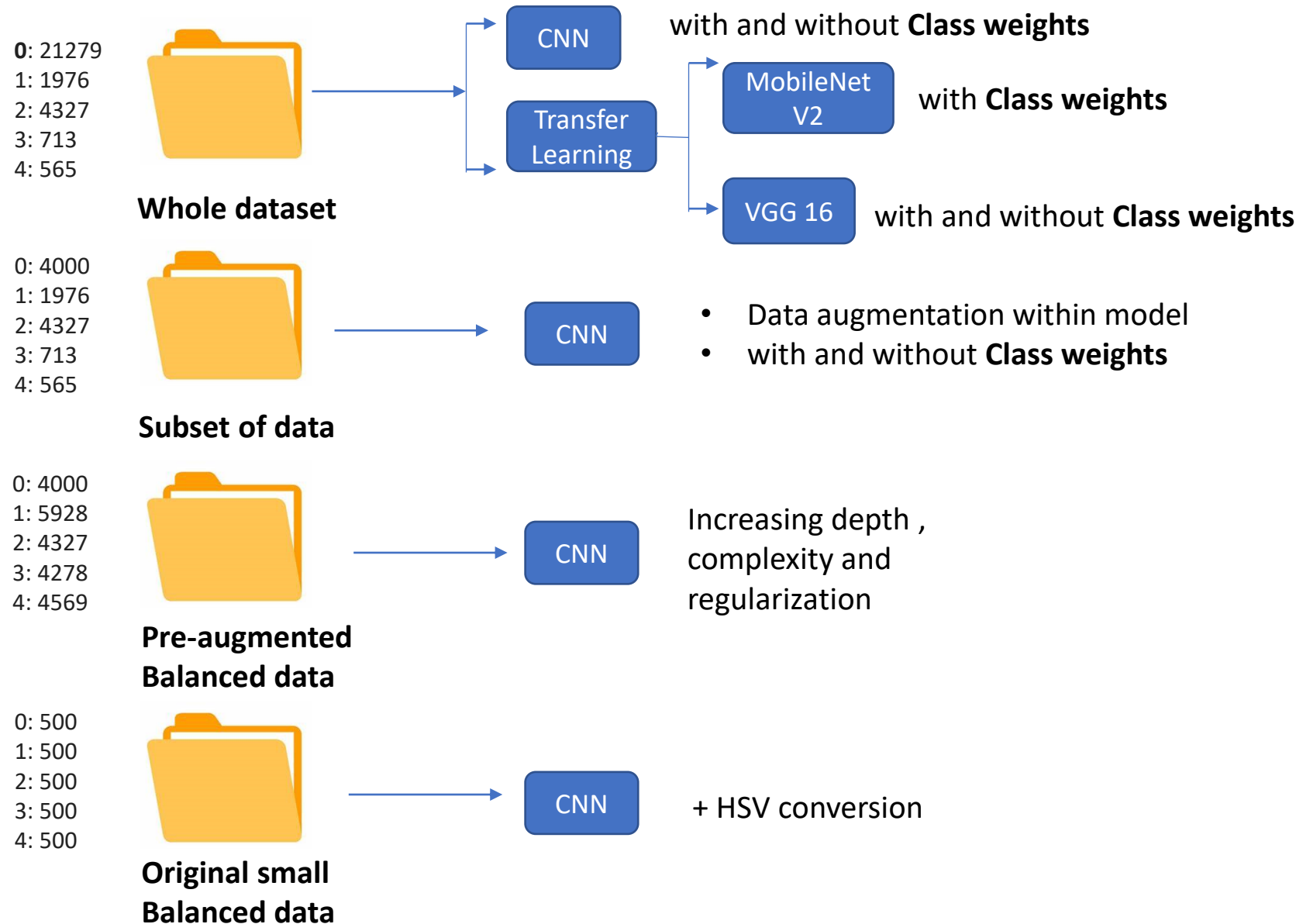
Binary Classification
Model Evaluation
12 models

Binary classification
Model - Recall

Binary classification
Model - AUC

Model 9
Inception ResNet V2

Winner:

Model 9 with highest Recall

Multiclass Classification

Data and Approach

0: 21279
1: 1976
2: 4327
3: 713
4: 565

Whole dataset

CNN → with and without **Class weights**

Transfer Learning → MobileNet V2 → with **Class weights**

Transfer Learning → VGG 16 → with and without **Class weights**

0: 4000
1: 1976
2: 4327
3: 713
4: 565

Subset of data

CNN
- Data augmentation within model
- with and without **Class weights**

0: 4000
1: 5928
2: 4327
3: 4278
4: 4569

Pre-augmented Balanced data

CNN → Increasing depth, complexity and regularization

0: 500
1: 500
2: 500
3: 500
4: 500

Original small Balanced data

CNN → + HSV conversion

Multiclass Classification

Model Evaluation

22 models

Multiclass classification
Model SCAccuracy
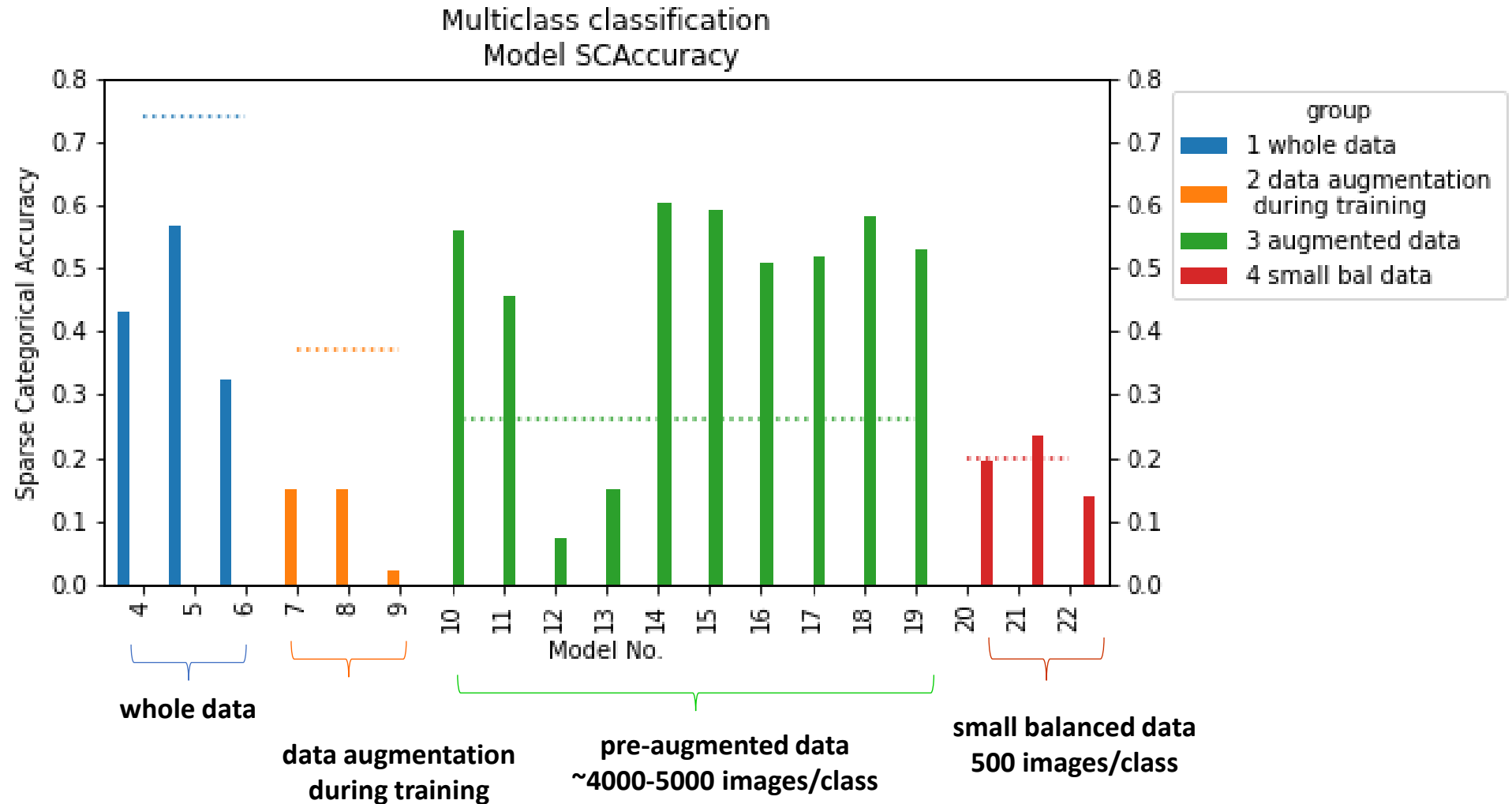
Multiclass Classification

Multiclass classification
Model - Recall for each class
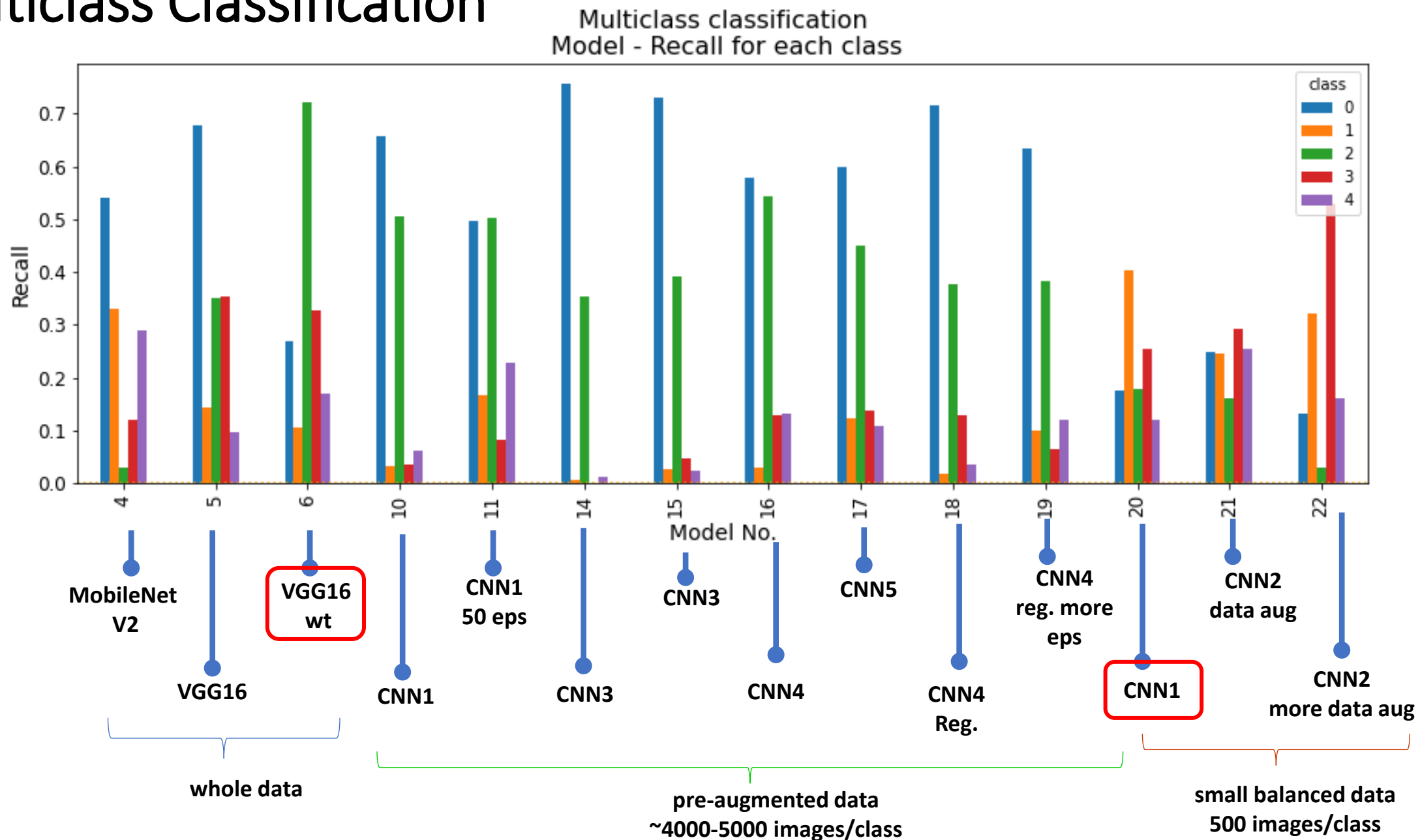
Multiclass Classification

Model Evaluation

Winner:

Model 6 – VGG16 wt trained on the whole dataset

Model 6- VGG16 wt

Confusion matrix
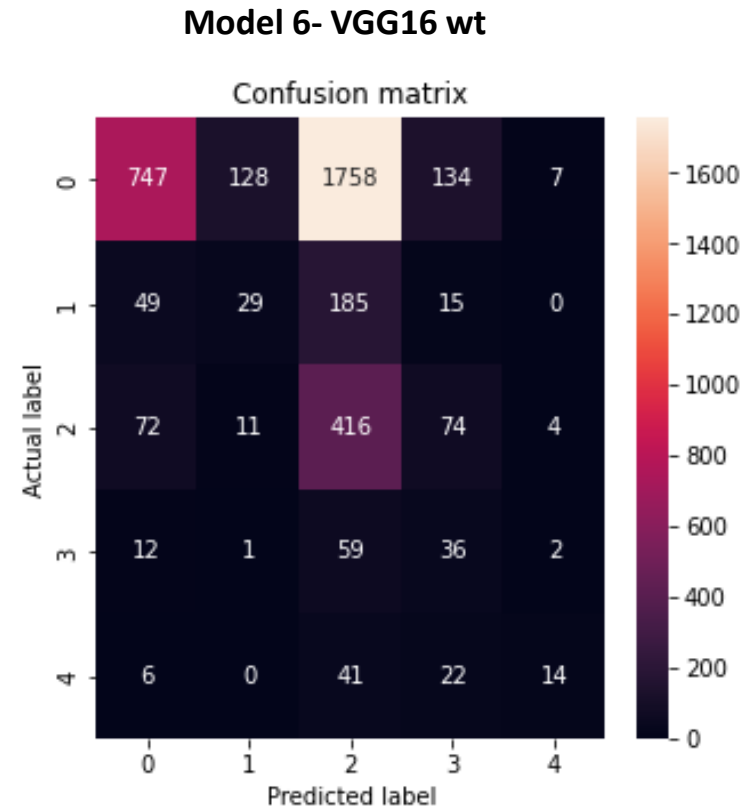
Conclusions and Recommendations

- **DR detection:** it is recommended to move forward with model 9 (transfer learning-InceptionResNet V2) as production model candidate.

    - the model has high recall for class 1 but needs to be tuned to improve classification and reduce false positives.

- **DR, detection of severity:** it is recommended to move forward with (VGG16 wt) as production model candidate.

    - It gives good recall (0.7) for class 1 and detects other classes, but needs to be more fine tuned to improve recall for other classes.

Next Steps

- **DR detection:**

    - Fine tune transfer learning
    - Image pre-processing, data augmentation or training on more balanced data set

- **DR, detection of severity:**

    - Fine tune transfer learning
    - Training with more balanced data with or without augmentation, and with more pre-processing of data
    - Training just on the positive classes for detection of severity.

Thank you!!

Acknowledgements:

- Hank
- Alanna
- Eric
- Devin
- TAs

- Classmates!!