FINAL REPORT

Deep Learning for Amazon Rainforest Monitoring

Written By

Daianne Starr

Can a deep learning model identify and categorize environmental changes in the Amazon rainforest using multi-spectral satellite imagery?

DATASET

train.csv *file*

IMAGE NAME

LABELS

SINGLE IMAGES

	image_name	tags
0	train_0	haze primary
1	train_1	agriculture clear primary water
2	train_2	clear primary
3	train_3	clear primary
4	train_4	agriculture clear habitation primary road
...	...	...
40474	train_40474	clear primary
40475	train_40475	cloudy
40476	train_40476	agriculture clear primary
40477	train_40477	agriculture clear primary road
40478	train_40478	agriculture cultivation partly_cloudy primary

train.csv *file description*

NUMBER OF IMAGES

NUMBER OF LABELS

NUMBER OF LABEL COMBINATIONS

MOST FREQUENT LABEL COMBINATION

THE MOST FREQUENT LABEL COMBINATION TOTAL

	image_name	tags
count	40479	40479
unique	40479	449
top	train_0	clear primary
freq	1	13636

JPEG satellite images

LABELS

IMAGE NAME

cloudy

train_413

primary, clear, water

train_47

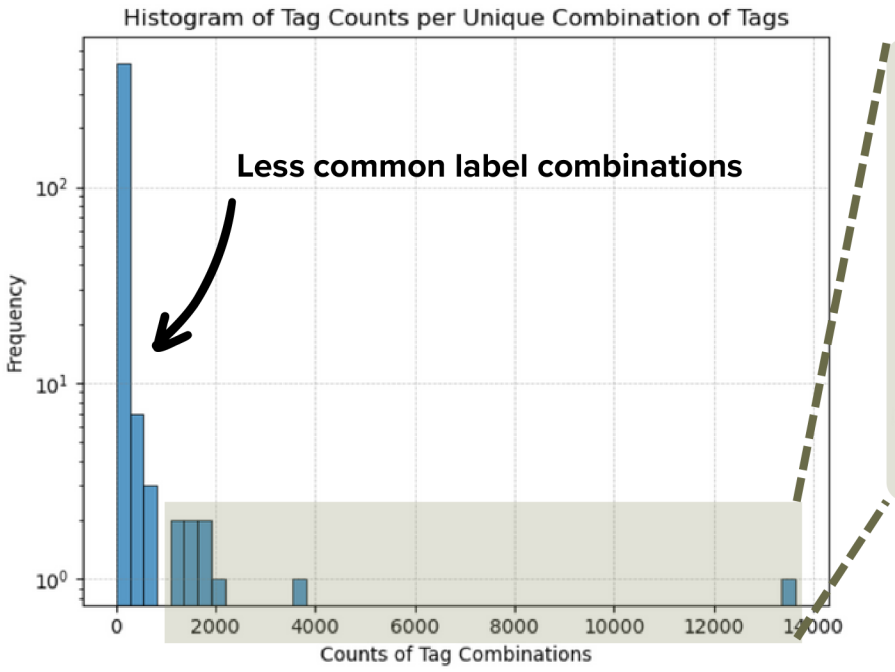
clear, habitation, road

train_533

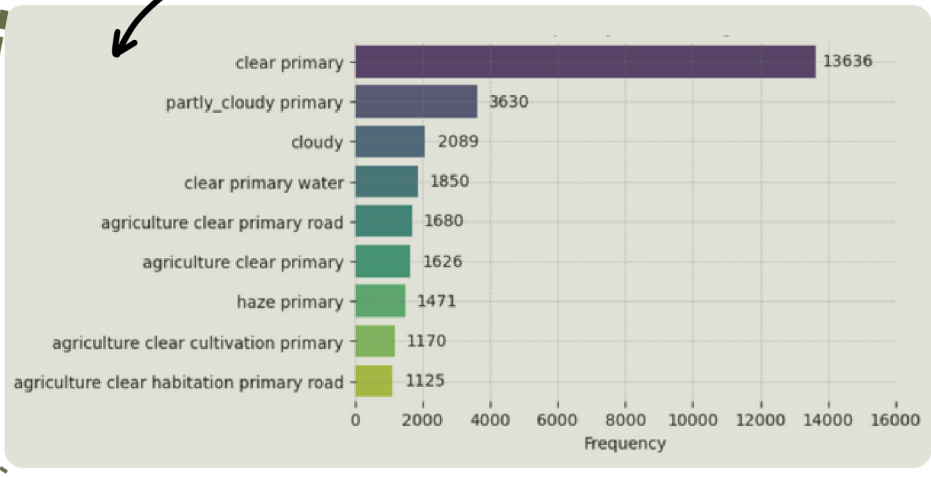
primary, partly_cloudy conventional_mine

train_25

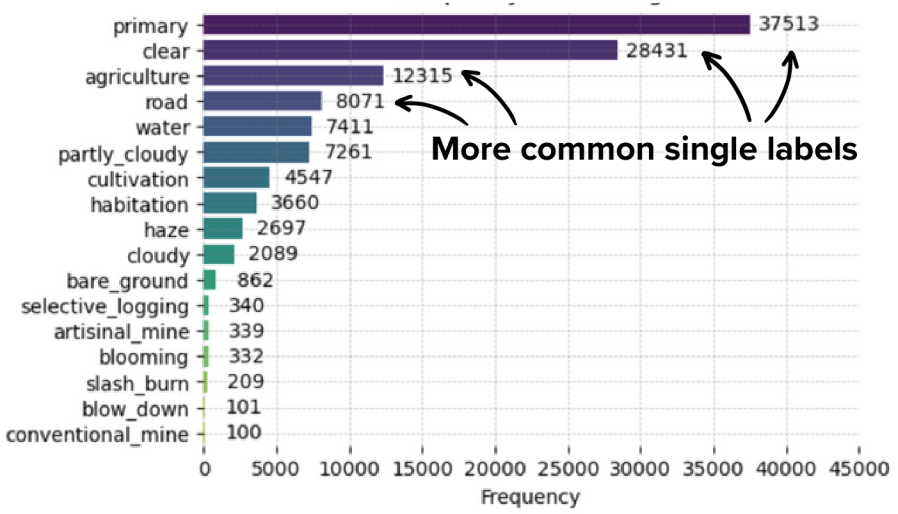
EXPLORATORY DATA ANALYSIS



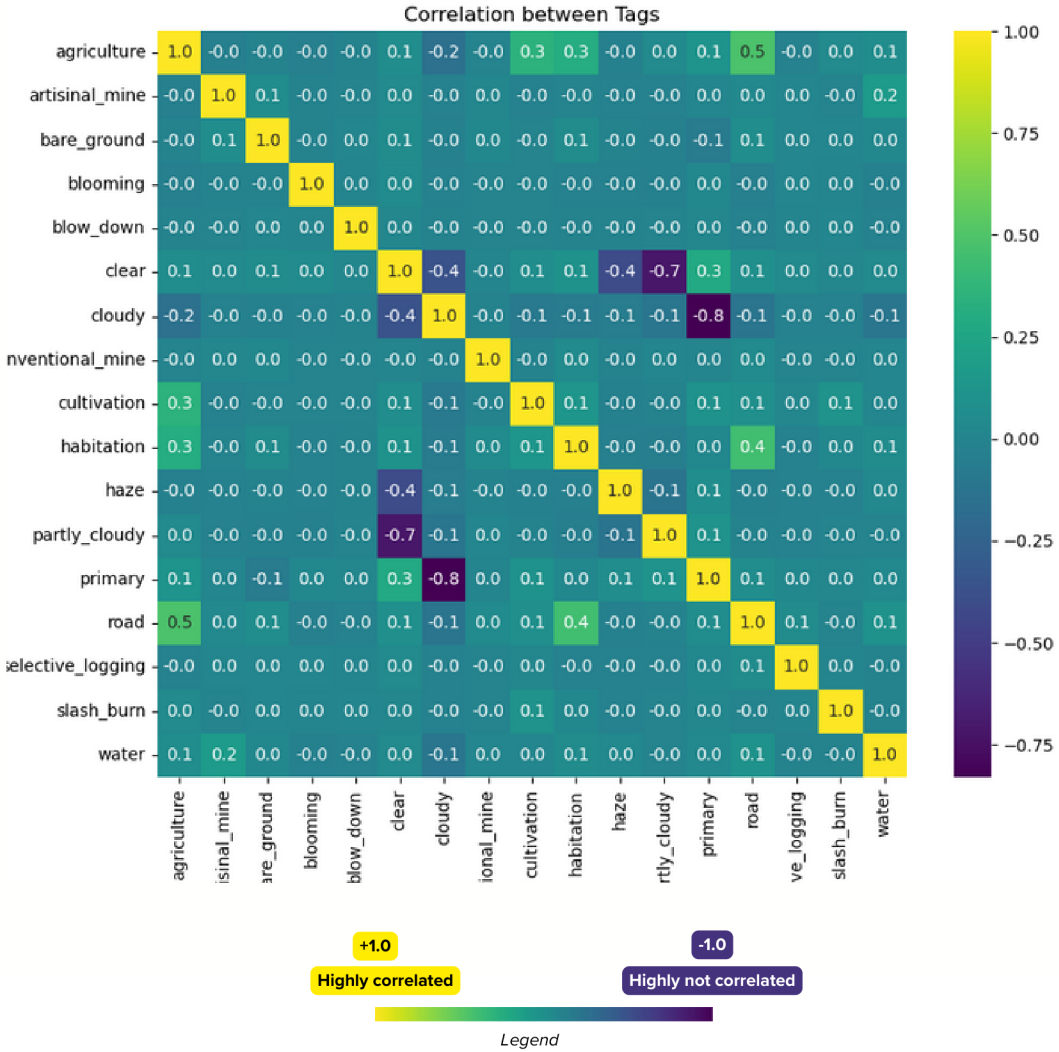
More common label combinations



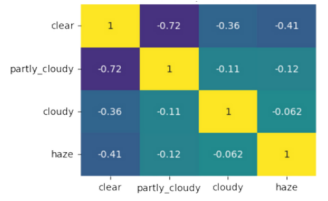
CLASS IMBALANCE



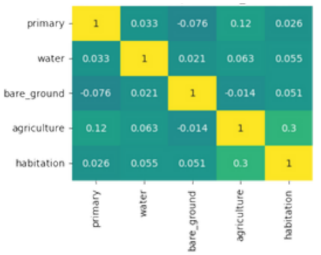
EXPLORATORY DATA ANALYSIS



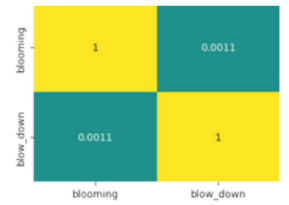
WEATHER



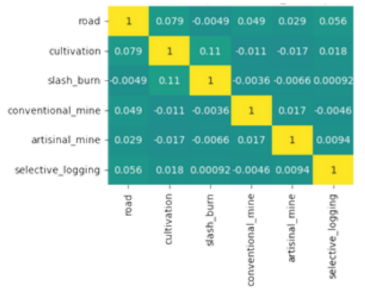
LAND COVER



PHENOMENON

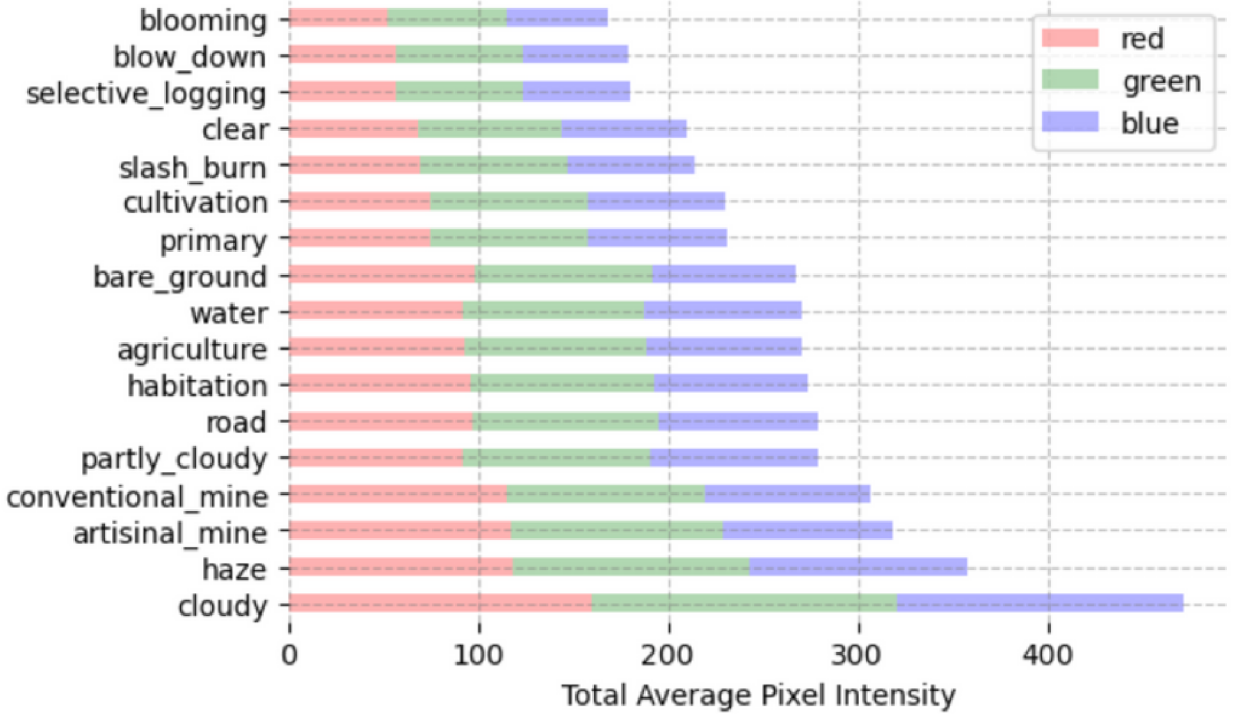


ACTIVITY



EXPLORATORY DATA ANALYSIS

AVERAGE RBG



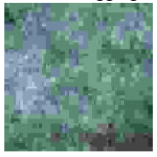
EXPLORATORY DATA ANALYSIS

AVERAGE IMAGES OF EACH LABEL

EXPLORATORY DATA ANALYSIS

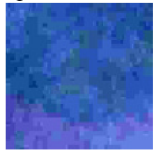
CONTRAST BETWEEN AVERAGE IMAGES

slash_burn vs. selective_logging



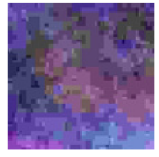
slash_burn vs. habitation

slash_burn vs. agriculture



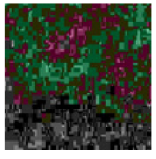
slash_burn vs. primary

slash_burn vs. haze



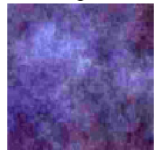
selective_logging vs. primary

selective_logging vs. blow_down



selective_logging vs. haze

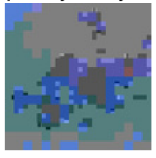
conventional_mine vs. bare_ground



conventional_mine vs. haze

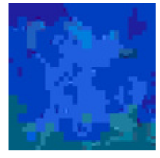
road vs. partially_cloudy

primary vs. partially_cloudy



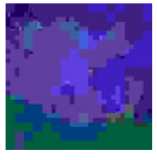
primary vs. water

water vs. road

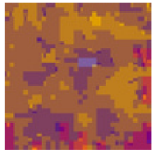


road vs. haze

agriculture vs. road



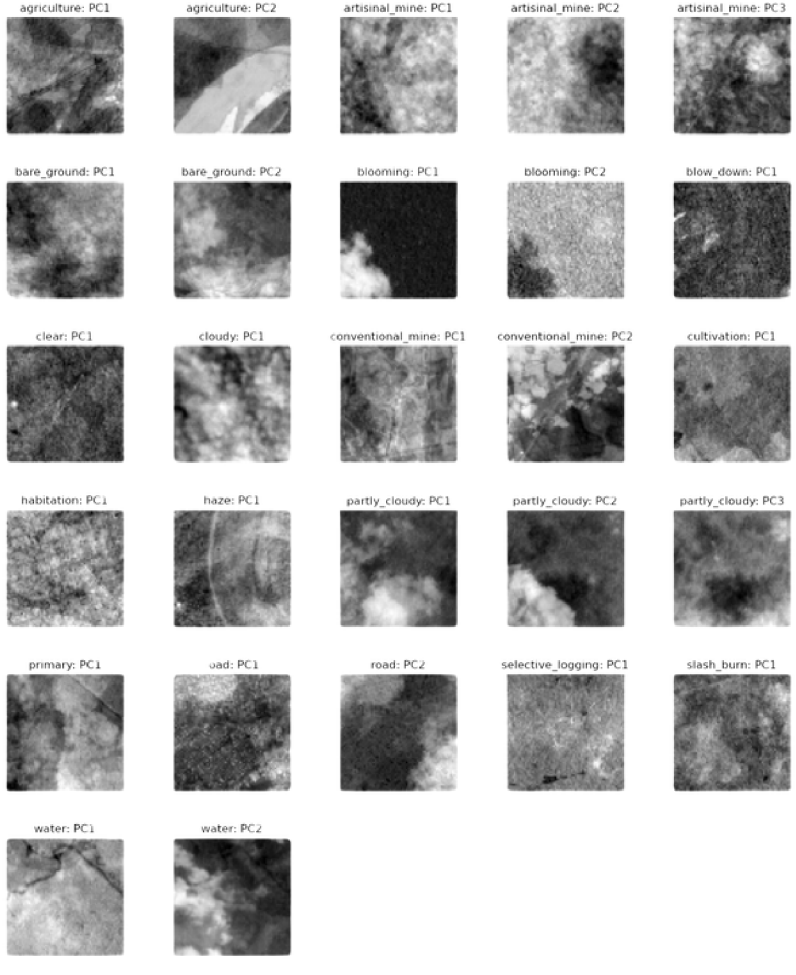
bare-ground vs. habitation



clear vs. cloudy

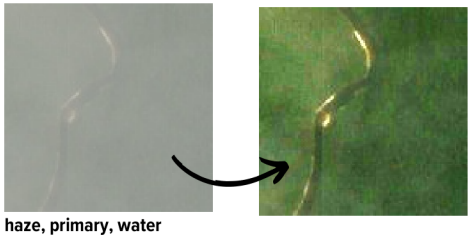
EXPLORATORY DATA ANALYSIS

EIGENIMAGES



PREPROCESSING

RESIZING & DEHAZING



haze, primary, water

train_104.jpg

haze, primary

train_315.jpg

agriculture, haze,
primary, road

train_518.jpg

agriculture, haze,
primary, road, water

PREPROCESSING

40,478 satellite images → **449 label combinations** → **17 single labels** → **4 main categories**

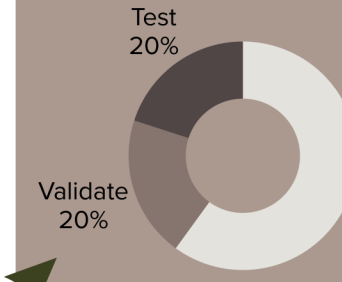
clear primary
haze primary
partly_cloudy primary
cloudy
clear primary water
agriculture clear primary
bare_ground clear habitation primary
blooming partly_cloudy primary
agriculture clear cultivation primary
agriculture clear primary road
agriculture clear habitation primary road
conventional_mine partly_cloudy primary

agriculture
artisanal_mine
bare_ground
blooming
blow_down
clear
cloudy
conventional_mine
cultivation
habitation
haze
partly_cloudy
primary
road
selective_logging
slash_burn
water

Weather
Land Cover
Activity
Phenomenon

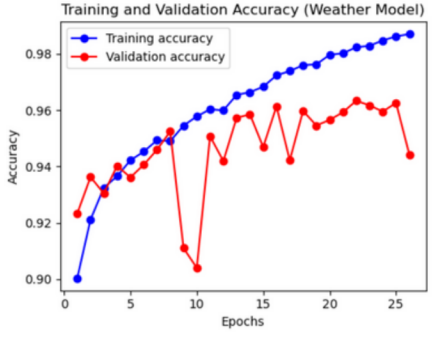
Image Dehazing

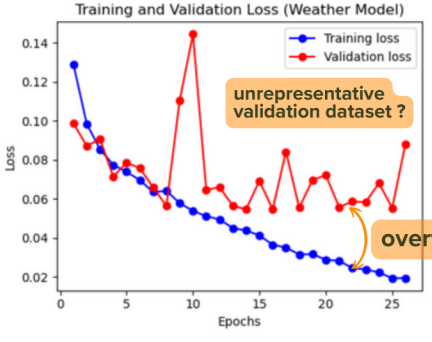
Data & Image split



MODELING

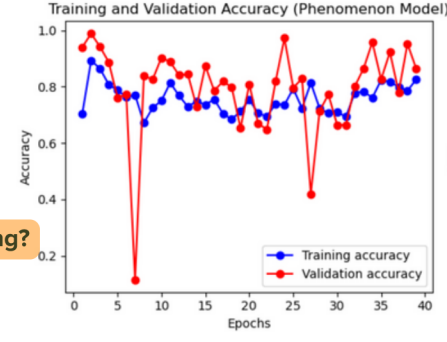
WEATHER

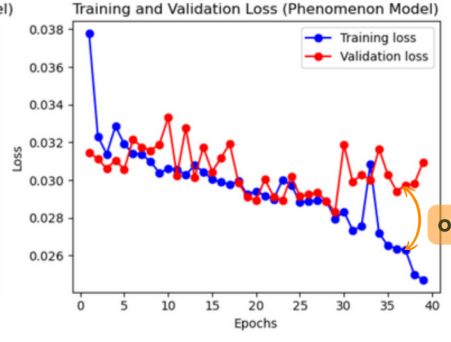




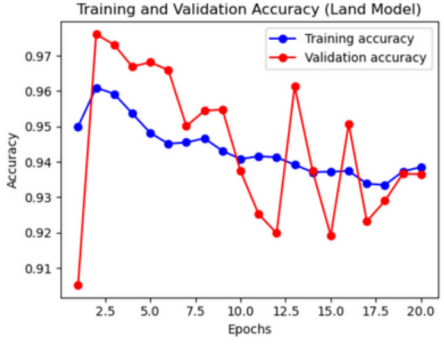
overfitting?

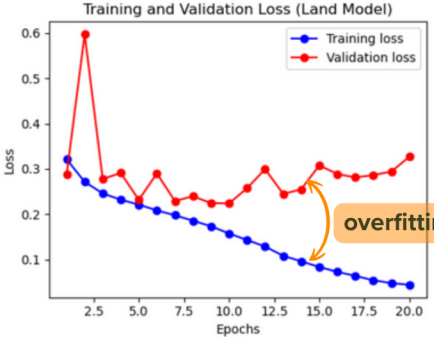
PHENOMENON





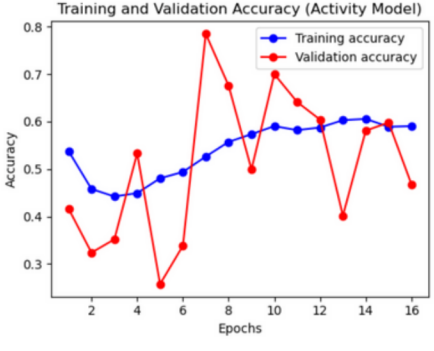
overfitting?

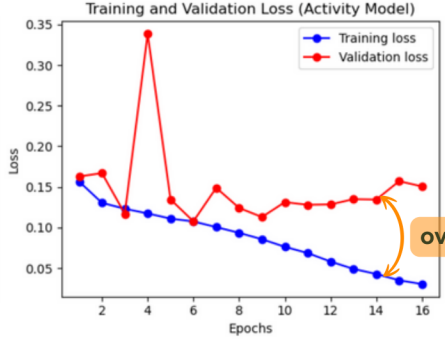




overfitting?

LAND COVER

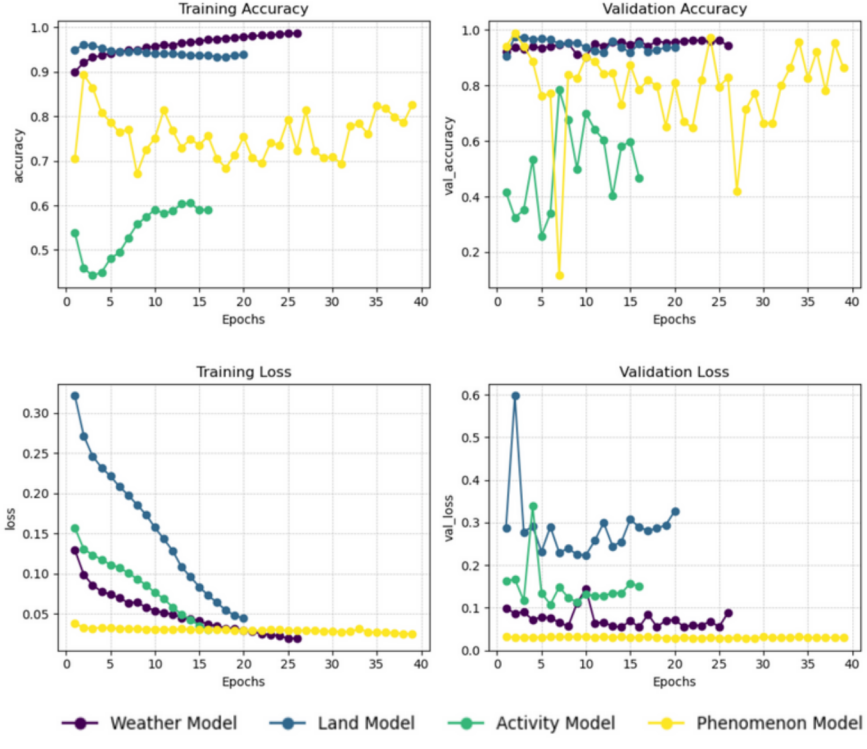


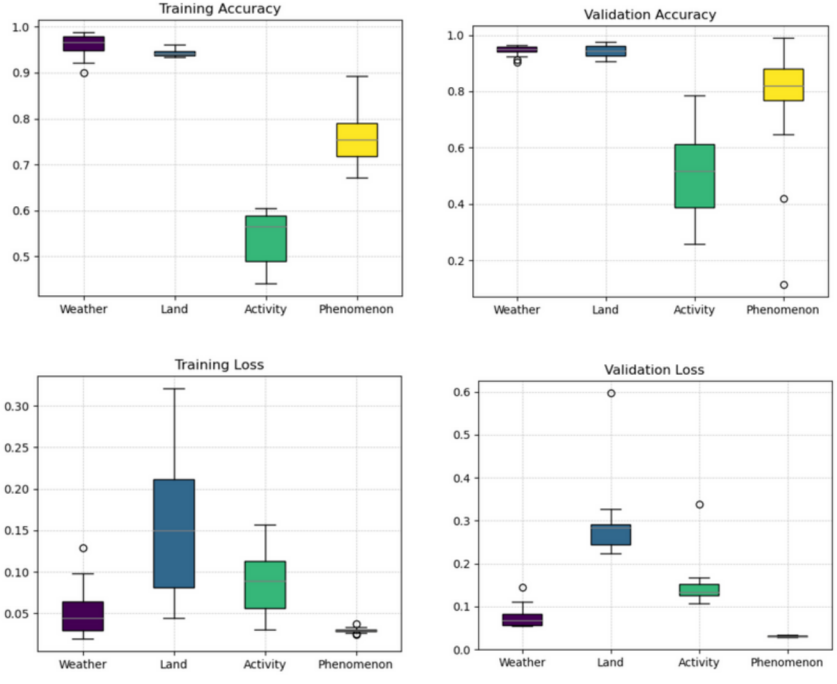


overfitting?

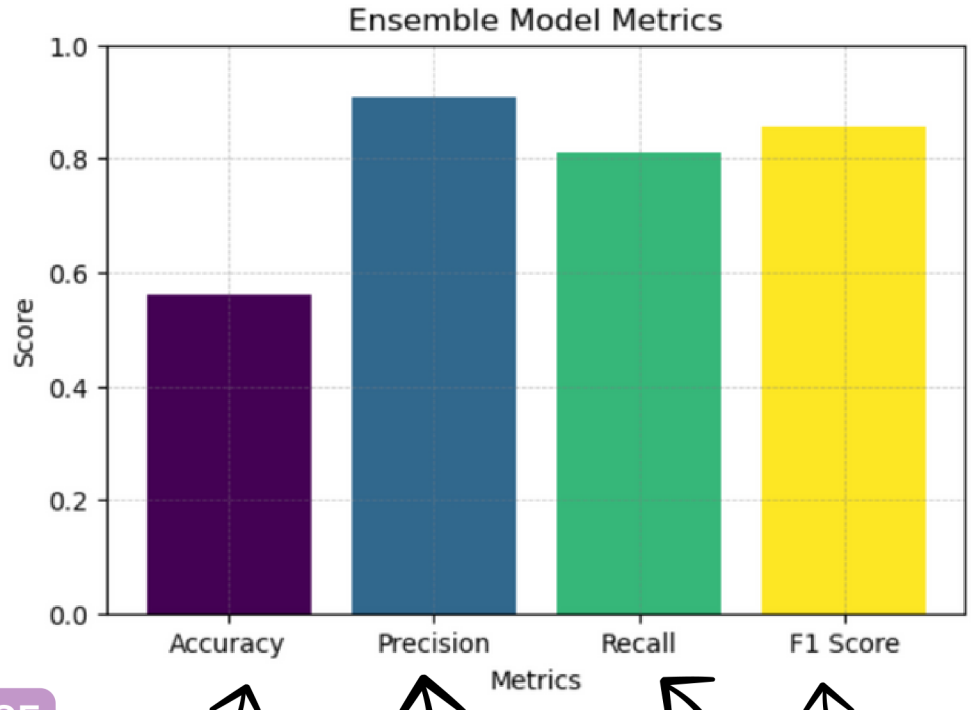
ACTIVITY

MODELING


Training Accuracy
Validation Accuracy
Training Loss
Validation Loss
accuracy
val_accuracy
loss
val_loss
Epochs
Weather Model
Land Model
Activity Model
Phenomenon Model


Training Accuracy
Validation Accuracy
Training Loss
Validation Loss
Weather
Land
Activity
Phenomenon

MODELING



CLASS IMBALANCE
Accuracy is not the best Metric for the model

True positive results divided by the total number of positive results predicted by the classifier

Model's ability to detect positive instances across all actual positives

Harmonic mean of precision and recall

TAKEAWAYS

- **Class imbalance in multi-label dataset**
  - EDA
  - Weights
- **Large number of images**
  - Batches
  - Generators
- **CNN**
  - Several iterations
  - Batch size, epochs, early stopping
  - Several models ensemble perform well: high F1 score