Can Deep Learning help us to unveil LAEs through broadband filters?

BRUNO CERQUEIRA

This work was supported by Fundação para a Ciência e a Tecnologia (FCT) through the research grants UIDB/04434/2020 and UIDP/04434/2020, and in the form of an exploratory project with the EXPL/FIS-AST/1085/2021 reference (PI: Paulino-Afonso).

DATA

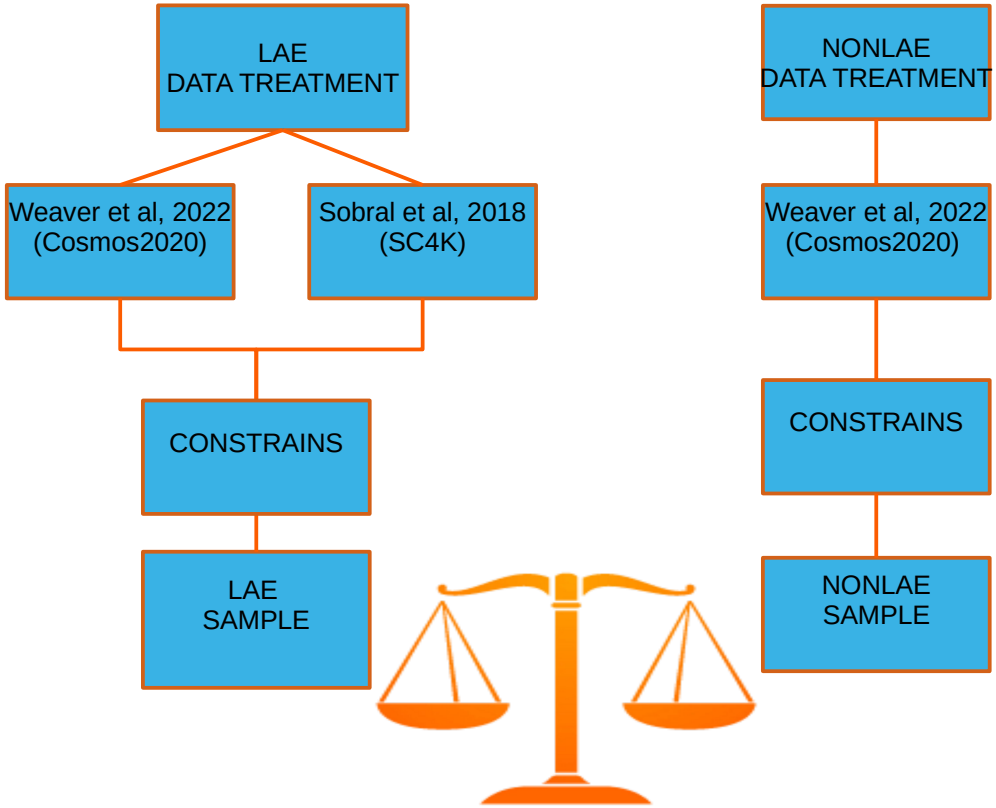
ONO ET AL, 2021

- Redshift: 2.2-7.0
- 8 CLASSES
- 6 NarrowBands + 5 BroadBands
- Motivation: identify LAEs in the middle of everything, noise, artifacts

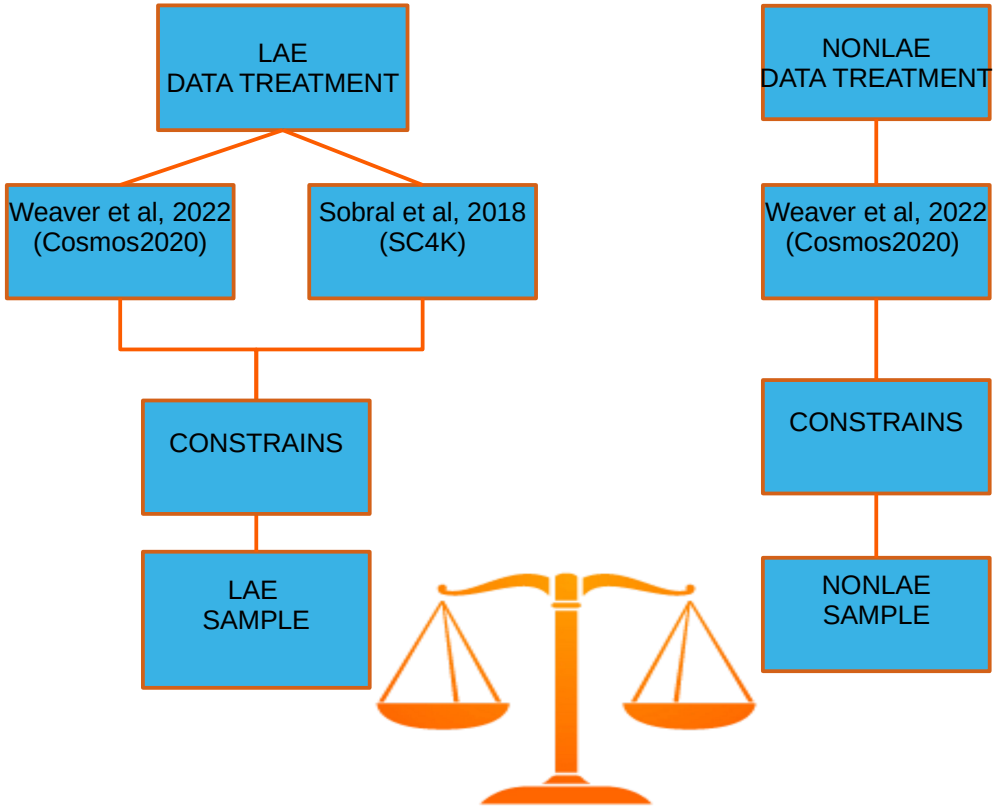
FLAEMING – Deep Learning

- Redshift: 2.0-6.0
- 2 CLASSES (LAE - NONLAE)
- BroadBands 3 or 4*
- Motivation: Distinguish LAES from other galaxies

DATA


LAE
DATA TREATMENT
Weaver et al, 2022
(Cosmos2020)
Sobral et al, 2018
(SC4K)
CONSTRAINS
LAE
SAMPLE
NONLAE
DATA TREATMENT
Weaver et al, 2022
(Cosmos2020)
CONSTRAINS
NONLAE
SAMPLE
BALANCED DATA

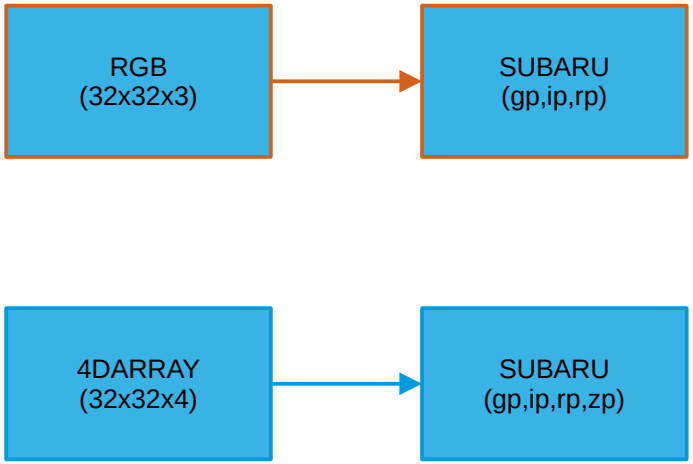
DATA

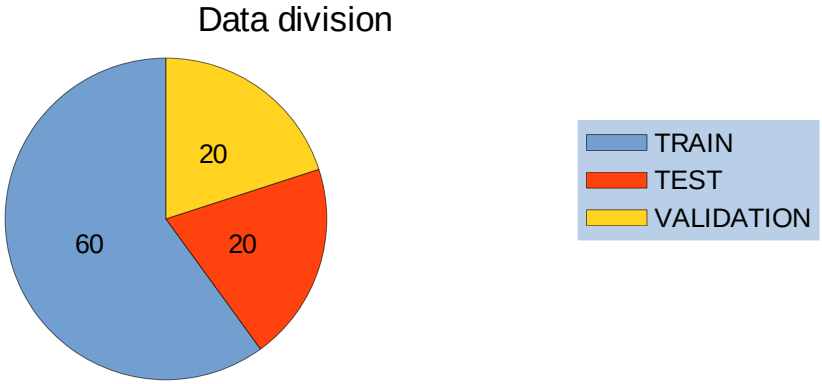


BALANCED DATA

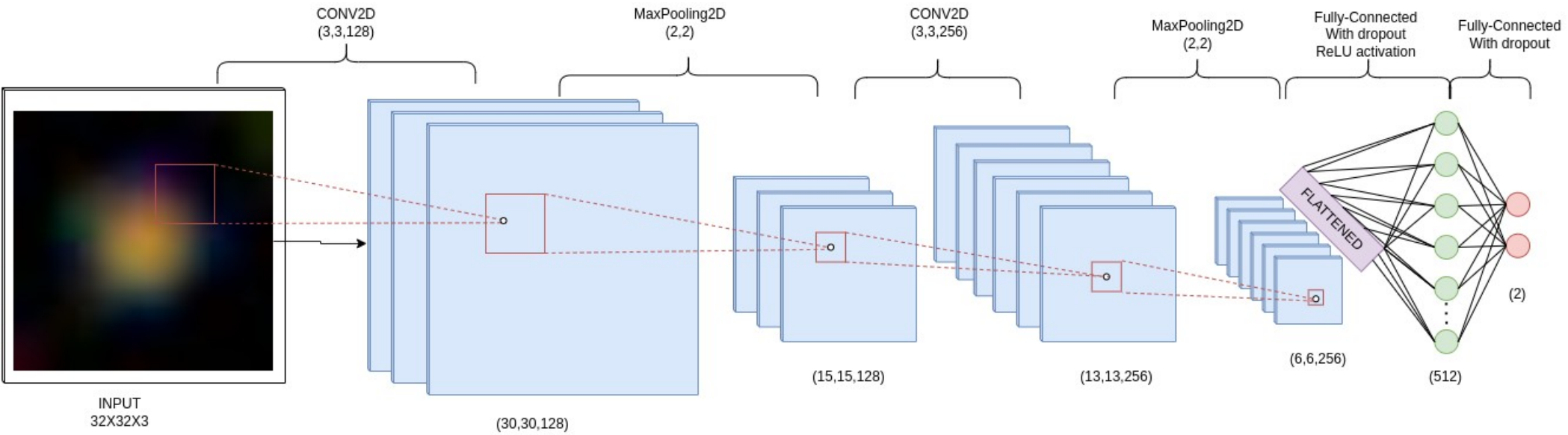
- CONSTRAINS
  - Redshift 2-6 for both datasets
  - ip_Mag 22-40 for SC4K
  - ip_Mag 0-40 for cosmos
  - Remove LAE SAMPLE from cosmos
  - Mimic the redshift and ip_mag from LAE to NONLAE

DATA


RGB
(32x32x3)
SUBARU
(gp,ip,rp)
4DARRAY
(32x32x4)
SUBARU
(gp,ip,rp,zp)


Data division
20
60
20
TRAIN
TEST
VALIDATION

CNN STRUCTURE



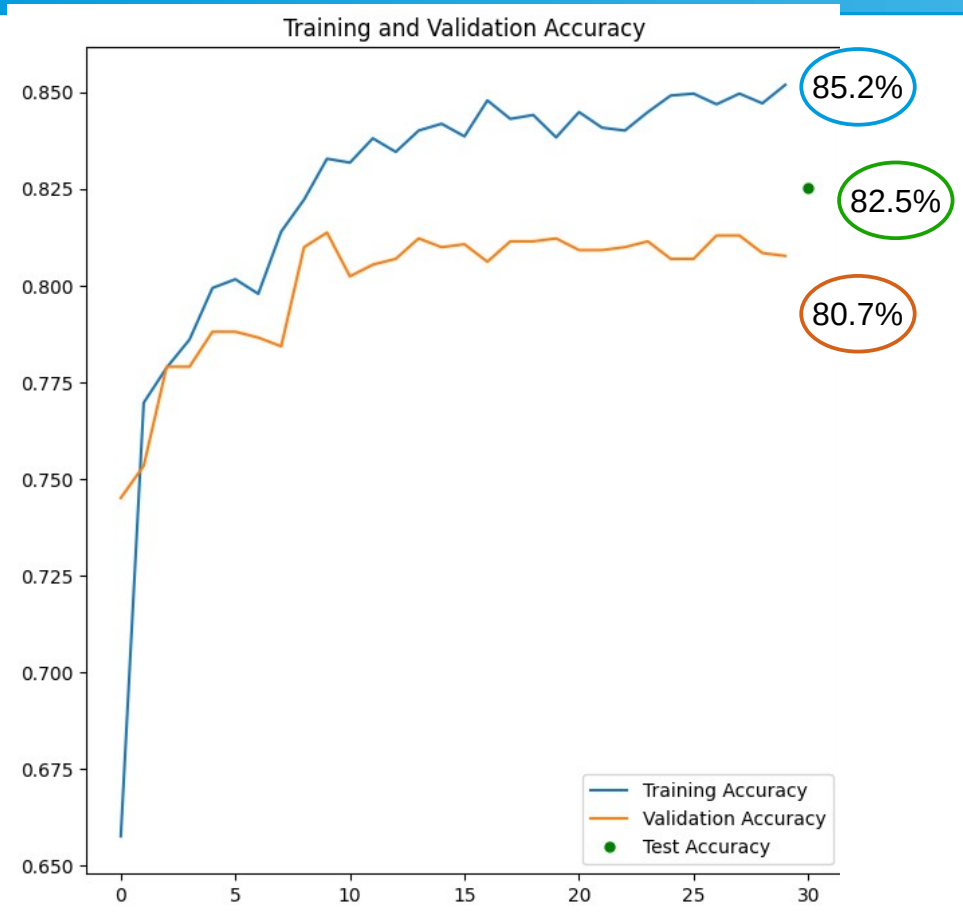
Duration: ~10 min

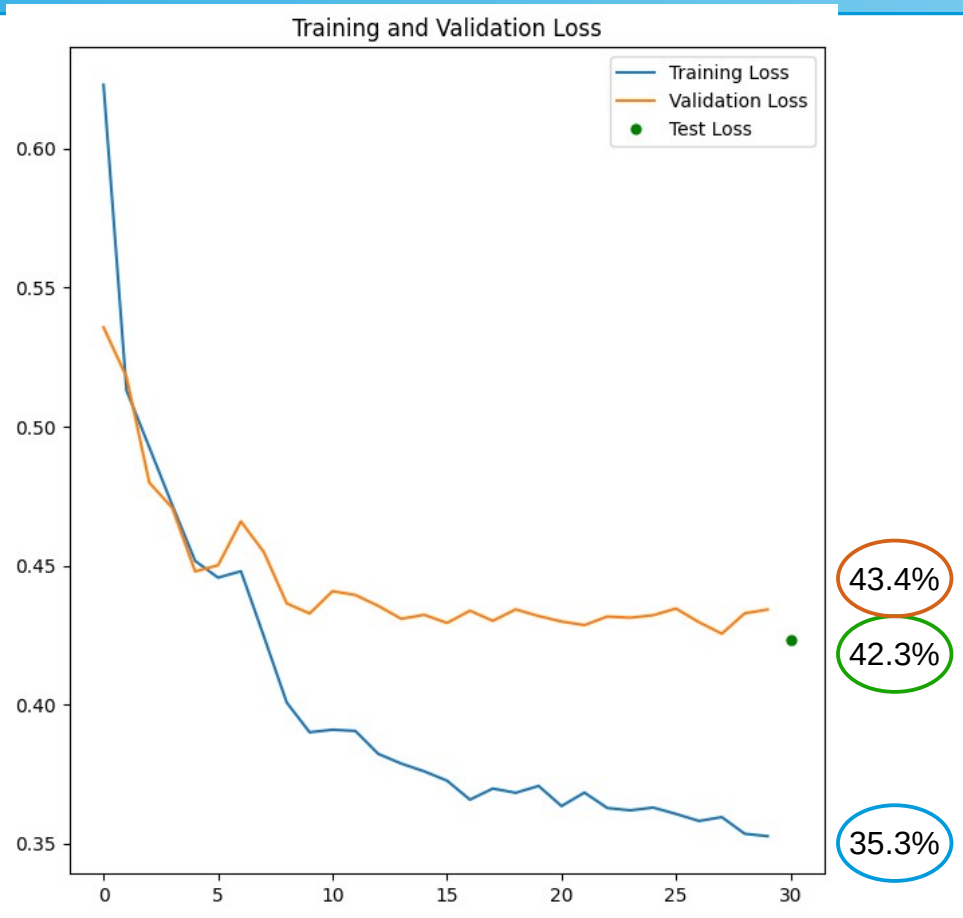
CONV2d CONV2d_1

Total params: (((3x3x3)+1)x128)+((3x3x128)+1)x256)+((512x9216)+1x512)+((2x512)+1x2)= 5,018,882 Parameters

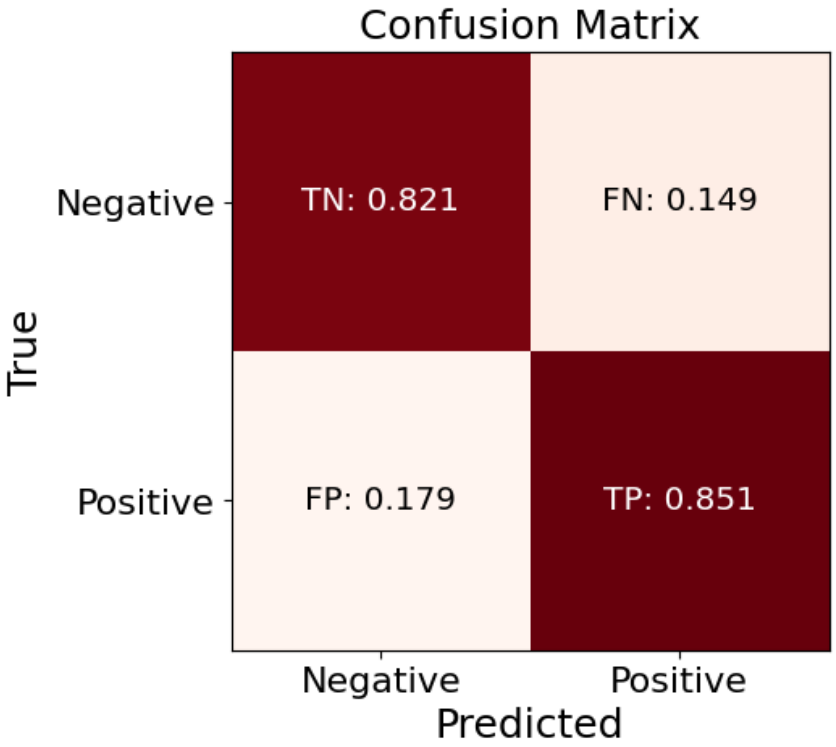
dense dense_1

Results – RGB


Training and Validation Accuracy
0.850
0.825
0.800
0.775
0.750
0.725
0.700
0.675
0.650
0
5
10
15
20
25
30
Training Accuracy
Validation Accuracy
Test Accuracy
85.2%
82.5%
80.7%

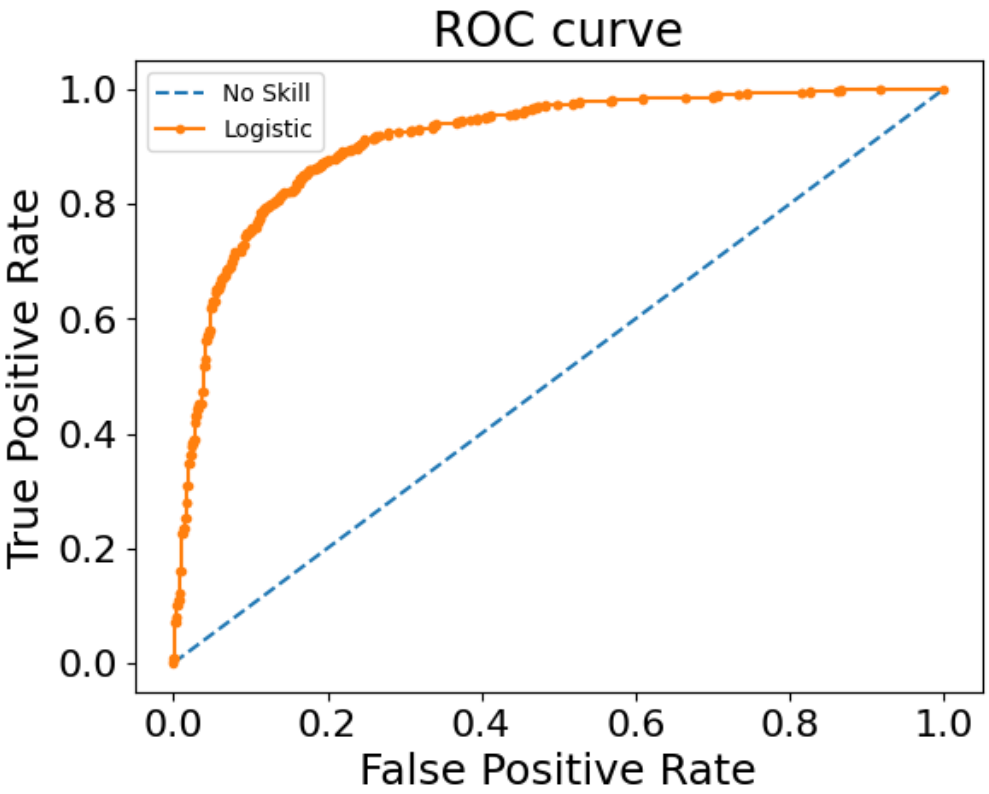

Training and Validation Loss
0.60
0.55
0.50
0.45
0.40
0.35
0
5
10
15
20
25
30
Training Loss
Validation Loss
Test Loss
43.4%
42.3%
35.3%

MODEL EVALUATION



- Evaluate the perfomance of a classification
  - TN= true negative
  - FN= false negative
  - FP= false positive
  - TP= True positive

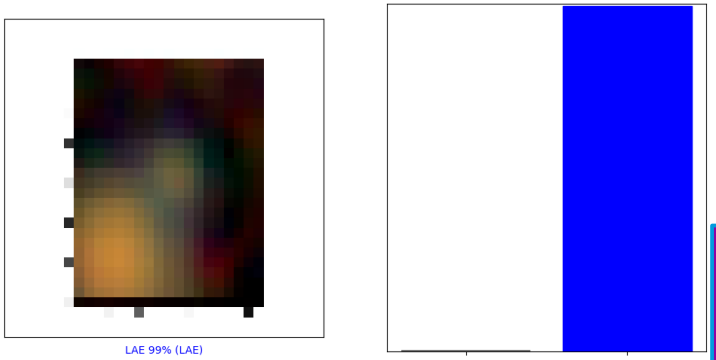
MODEL EVALUATION



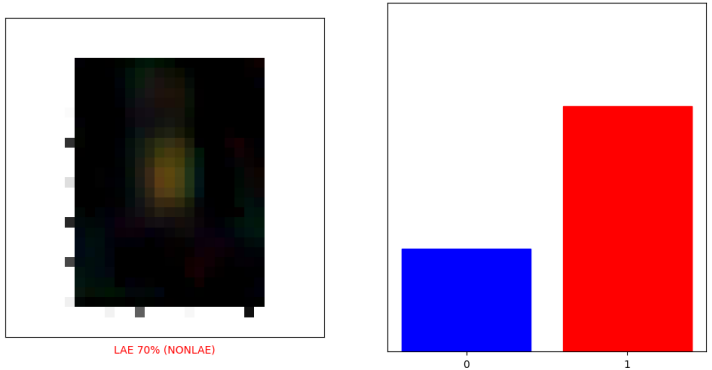
- ROC curve: Receiver operating characteristic curve
- Shows the perfomance of a classification

Predictions over test

LAE 99% (LAE)

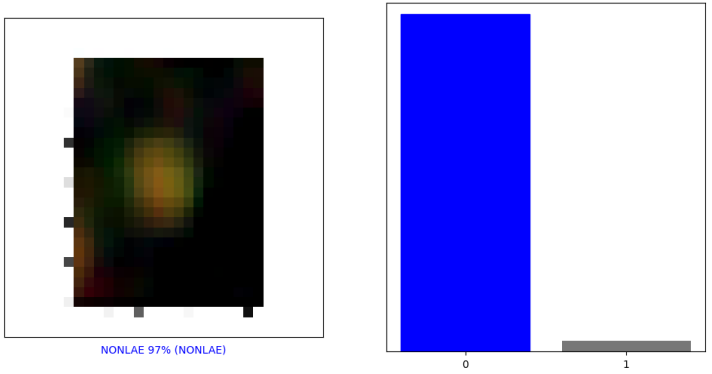


LAE 70% (NONLAE)

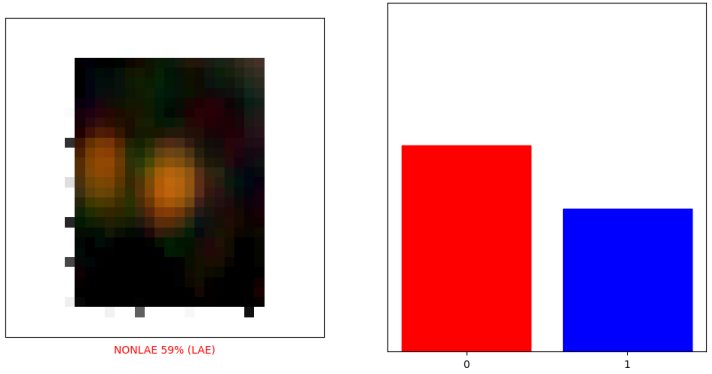


Accuracy:82.5%

NONLAE 97% (NONLAE)



NONLAE 59% (LAE)



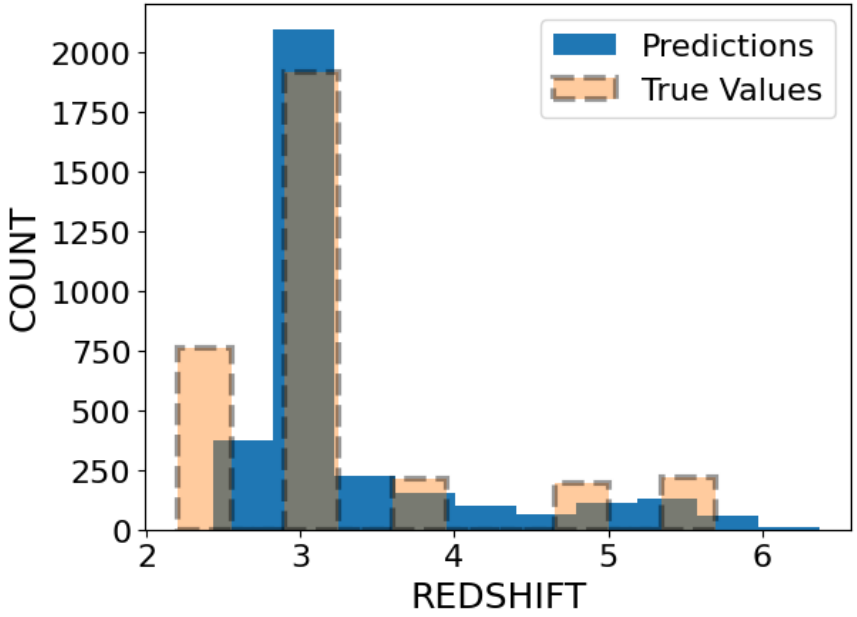
Predictions over all sources

- Over all Cosmos 2020, our model predict 53958 possible LAEs
- Among them
    - Comparing to HETDEX has 25 True Positive matches
    - Comparing to DEIMOS has 84 True Positive matches

REGRESSION

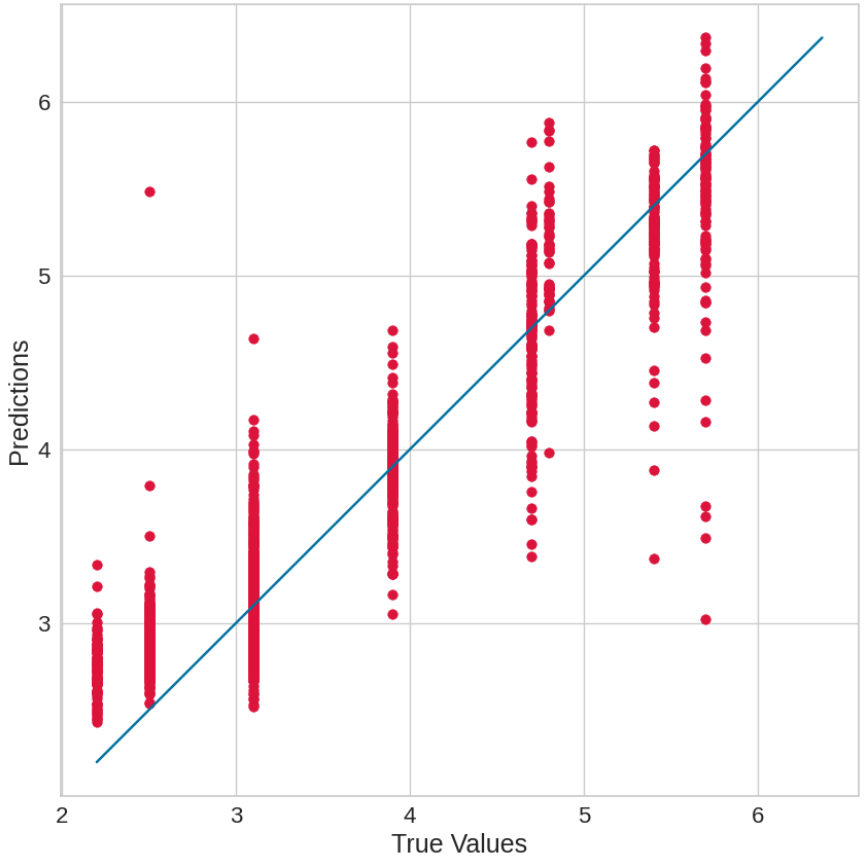
- Change the CNN structure to fit regression, - in classification we want the output to be one of the wanted class, while in regression we want a single value.

- Also despite the CNN structure Classification model was functional in our data, that does not mean the same CNN structure would work for the regression



REGRESSION

- Predictions for redshift using only rgb input
- Applied over the LAEs Sample



SUMMARY

- Classification: Create two samples one for LAE the other NONLAE, use subaru matched_psf to generate rgb with gp,rp,ip, divide them by 60 train 20 test 20 validation, the input are only images, train the model, make predictions over test set and later over the not used NONLAE.

- Regression: Same steps for rgb generation, and data division, change the CNN structure to fit a regression, and the data input should have the rgb images and the true labels, after train, make prediciton over test sample and compare the true label with predict label.

- 53958 Possible LAEs among them 25 true positives matches with HETDEX and 84 matches with DEIMOS

TAKE HOME MESSAGE

- Yes it is possible to identify LAE with only broadbands, it is open to improvements such as:
  - Parameters fine tuning
  - Combination with predictions over 1d data (e.g. UV morphology, Paulino-Afonso et al, 2018)
- We aim to generalize to other extragalatic fields and surveys.
- THANK YOU!!