Star-Galaxy classification using ALHAMBRA photometry and Machine Learning Algorithms

PRATYAKSH RAJ

Under the guidance of **Dr. Shantanu Desai**
IIT HYDERABAD

Compiled November 21, 2022

1. INTRODUCTION

It is known that classifying an object as star or galaxy, based on their morphology, is a difficult task from earth based telescopes(sometimes even for space based) as their images get blurred due to plenty of reasons. Therefore, Here we have used ML algorithms (Convolutional NN, XG Boost, Artificial NN, Random forest, Ada Boost) to classify an object as star or galaxy over ALHAMBRA-4 field. We used ALHAMBRA survey- isophotal magnitudes (20 Narrow band, 3 NIR broad band, f814, FWHM) over COSMOS field, taking Hubble's Morphology based classification as ground truth. There are a total 31870 objects matched to our reference catalog- 29615 galaxies and 2247 stars. For f814w$\leq$ 22.5, there are 5998 galaxies and 1765 stars. We divide our train and test data in 85:15 ratio.

Here we will find out the best algorithm with least overfitting and best AUC score. XG Boost performed best on datasets- f814w $\leq$ 22.5 with AUC score: 0.995 and on f814w $\leq$ 26 with AUC score: 0.974. However, with only 20 Narrow bands as features, CNN was better than any algorithm with an AUC of 0.983, it captured the relation very well and is not overfitting. Then making correlation heat map of features to find that f814 is highly correlated with 10 other magnitudes (correlation coefficient >0.94).Therefore number of parameters reduced to 13, giving even better AUC scores (CNN: 0.997 and XG Boost: 0.995) for f814w$\leq$ 22.5, also much less overfitting and less CPU time (features should not be too less or it can lead to underfitting, also XG Boost is much faster than CNN). Here,Performance of all algorithms and for different brightness ranges are visualized via ROC Curve. See fig1.

We also ran all algorithms on only narrow optical bands, than gradually adding NIR, f814w and FWHM. We found that most algorithms give bad results on narrow bands except for CNN (0.983 for f814w$\leq$ 22.5 and 0.961 for f814w$\leq$ 26) which increased to 0.989 on addition of fwhm. Also, it is a good practice to consider all objects fainter than 23 as Galaxies, as these objects are mostly galaxies (9000-15000) and very few stars (100-250), they act as contamination or noisy data specially for boosting algos.

Now we calculated the ranking of features towards classifyinng the object (using Randomforest and ADa boost) and found that except FWHM, f814w,K and j magnitude, other bands like 20 narrow bands are only acting to tighten the relation (although CNN captured the relation very well). Therefore instead of 20 narrow band magnitudes, we ran CNN and XG Boost on 19 colors(differnce between consecutive narrorw band magnitudes) + 2 colors(3 broad band) + FWHM + F814w, resulting in AUC score increasing from 0.9860 to 0.994 in CNN and 0.9953 in XG Boost, for f814w$\leq$ 22.5.
Also from ranking of features we see that these colors (f365-f396 , f396-f427 , f458-f459) become equally important as FWHM and f814w in classification.Therefore now using only these 5 magnitudes, we get AUC score 0.994 for f814w$\leq$ 22.5 for both CNN and XG Boost, and much less CPU time.

Therefore we can conclude that only CNN works (very well) with 20 narrow bands(even with contamination) in classification. But XG Boost outperforms it(with negligible margin) by addition of fwhm.

CNN gives the highest AUC score of 0.997 with reduced parameters without being overfitted (13 features of which 8 are narrow bands). it also handeled the fainter objects very well.
Also, we don't need all magnitudes, even better classification can be done using much less features, overfitting and CPU time. We need only six features- FWHM, f814w, (f365-f396), (f396-f427), (f458-f459), (f768-f799). we can see that most of these colors are composed of magnitudes of low wavelengths.

Therefore it can also be concluded (visualised) that CNN was able to perfectly capture the spectrum of stars and galaxies resp. It give very good result with 96.2% accuracy, with nearly no wrongly classified galaxy (only stars fainter than 23 are miss classified as they are very few, can be contamination as these objects are mostly galaxies). On the other hand XG Boost is much faster than CNN and gives a little better results with colors, with an accuracy of 98.6% (very high, as it might have got overfitted to those contaminated stars and correctly classified them as boosting algos are sensitive to outliers). Therefore CNN is better for fainter objects, as it might classify fainter star as galaxy but vice a versa will not happen which is good. Here accuracy are calculated wrt Alhambra's stellar flag. Results of best models and some proofs are given in fig11 and fig12.

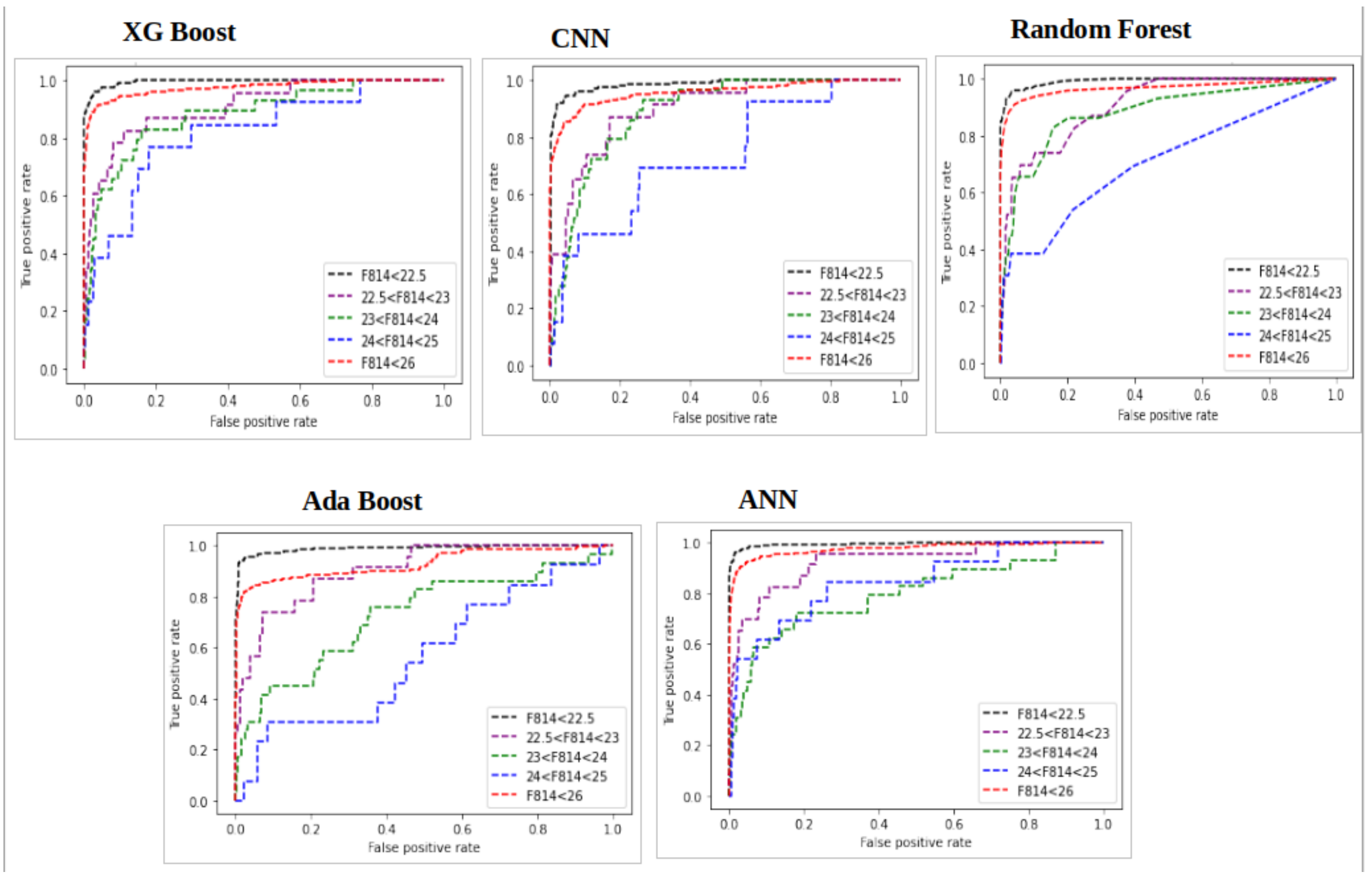


Fig. 1. ROC curves for different ranges of F814W, using all 25 magnitudes as features. In magnitude ranges 22.5-23, 23-24 and 24-25, there are 4000 to 9000 galaxies and 100 to 250 stars in each range

2. ALHAMBRA SURVEY

We have used the ALHAMBRA catalogue [5] over the ALHAMBRA-4 field, which overlaps with COSMOS. It contains 31870 objects matched to our reference COSMOS catalogue with a maximum sepration of 3 arcsecond. Out of which 29615 are galaxies and 2247 are stars. The ALHAMBRA photometric system is characterized by 20 constant width (31 nm), non-overlapping medium band filters covering a wavelength range from 350 to 970 nm. The images were taken using the Calar Alto 3.5m telescope using the wide field optical camera LAICA and the NIR instrument Omega-2000, which are equipped with 20 intermediate width bands and three NIR broad-bands: J, H, K. The catalogue presents multicolour PSF-corrected photometry detected in synthetic F814W images with objects up to a magnitude of F814W 26.5.

3. HUBBLE TELESCOPE SURVEY

Here we will use Hubble's morphology based classification as ground truth. COSMOS space-based imaging catalogue [4], which provides a morphology-based classification (MU_CLASS) for the objects to train and test our methods on. It contains 1.2 × 106 objects to a limiting magnitude of F814W = 26.5 from images observed with the Hubble Space Telescope (HST) using the Advanced Camera for Surveys (ACS)4, therefore its image quality (very deep and unaffected by the atmosphere) can be used as a 'truth' reference. Images were taken through the wide F814W filter (I). The catalogue contains, roughly, 1.1 × 106 galaxies, most towards the faint end, 30 000 stars and the rest are fake detections.

4. DATA WRANGLING

We will take only those objects for which less than 5 band magnitudes is unknown. Here median (rounded off to three decimal places) is used to fill the missing data of respective column. Also, as ALHAMBRA do not provide any classification for most of the objects fainter than 22.5, its a good practice to consider all those objects as galaxies as they generally are. There are very few (100-200)star in this range which produces contamination but here we ran our algorithms on different ranges of F814w >22.5, without doing any changes.

5. MACHINE LEARNING ALGORITHMS AND HYPERPARAMETERIZATION

We have used XG Boost, CNN, ANN, RandomForest, AdaBoost. We have measured the performance of ANN and CNN by making ROC, overfitting curves and AUC scores. For RandomForest and AdaBoost, we can check AUC score and ranking of features (Ada Boost is helpful in getting the complex relations which randomforest might not able to get. It makes weak learners strong). They both are highly important in getting good ranking of features. We will see that CNN will work best and is least overfitted. Then we will also check XG Boost, it is based on gradient boosting and is known to outperform neural nets and Randomforest, quite commonly.
Firstly we hyperparameterized all our algorithms for objects brighter than 22.5, to compare them. ANN, CNN are hypertuned via KerasClassifier and Random forest, AdaBoost and XG Boost via GridSearchCV. These are:-

XG Boost-
gamma= 0.2, learning rate= 0.3, max depth= 4, n estimators =80, reg alpha= 1, scale pos weight= 1

CNN-
Convolution2D(32, kernel size=(5, 1), input shape=(25,1,1), activation= 'LeakyReLU') + (MaxPool2D(pool size=(3,1)))
Convolution2D(32, kernel size=(3, 1),activation= 'LeakyReLU') + MaxPool2D(pool size=(2,1))
Convolution2D(64, kernel size=(2, 1),activation= 'LeakyReLU') + MaxPool2D(pool size=(1,1))
Dropout(0.3) + Flatten() + Dense(1 ,activation="sigmoid")
Loss Function: Binary Cross Entropy, optimizer= Adam(learning rate=0.03)

Random Forest-
estimators: 200, min samples split: 2

Ada Boost-
base estimator: Decision Tree, n estimators: 100, max depth: 4

ANN-
dense layer(neurons: 25, activation func: Relu, BatchNormalization())
dense layer(neurons: 25, activation func: Relu)
dense layer(neurons: 25, activation func: Relu, dropout: 0.5)
dense(neuron: 1, activation func: sigmoid)
Loss Function: Binary Cross Entropy, optimizer= Adam(learning rate=0.03)

6. SELECTION OF BEST ALGORITHM

we ran all parameterized algorithms on data set(15% testing and 85% training) for various ranges of F814w and also by gradually adding features, See fig2 . For only optical bands as features, CNN gives highest AUC 0.983 for f814w$\leq$ 22.5 and 0.96 for f814w$\leq$ 26. followed by XG Boost with an AUC of 0.97 .
Also, we can even see that CNN is not overfitted where as ANN is highly overfitted, See fig3 .
RandomForest also give high AUC scores but performance drops with only optical bands(AUC 0.957). Also ranking of

features shows that it depends highly on FWHM (very less on other magnitude, See fig4), therefore not as useful as CNN . And Ada boost shows the least AUC in every case.
Therefore it can be concluded that all algorithms except CNN and XG Boost are highly dependent on FWHM. CNN is consistent without being overfitted.
Therefore CNN comes out to be the best algorithm without FWHM. And XG Boost while taking fwhm in consideration. AUC scores for different ranges of F814w, for CNN are shown in fig5.
Also in fig2 and fig8 we can see that NIR Broad bands are making a high improvement in classification as already discussed in [1] and [3], but not in case of CNN as classification rate is already too high to improve.

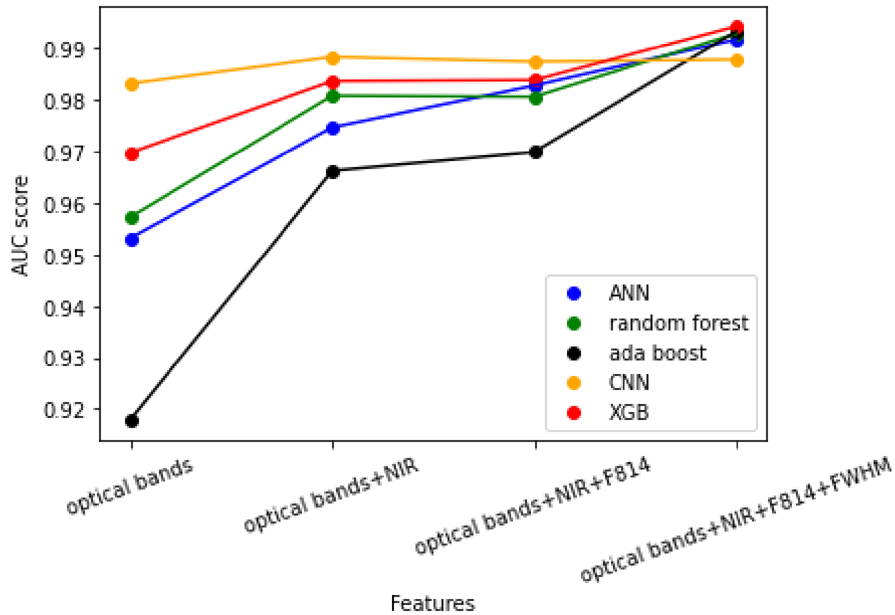


Fig. 2. this graph is for f814w$\leq$ 22.5. CNN shows highest AUC 0.983 with only optical band input and XG boost shows highest AUC 0.994 when FWHM is added. We can see that CNN captured the spectrum successfully, much more than any other algo. Also, with all features and objects brighter than 26, CNN gives an accuracy of 95.7% and XG Boost gives 97.8%

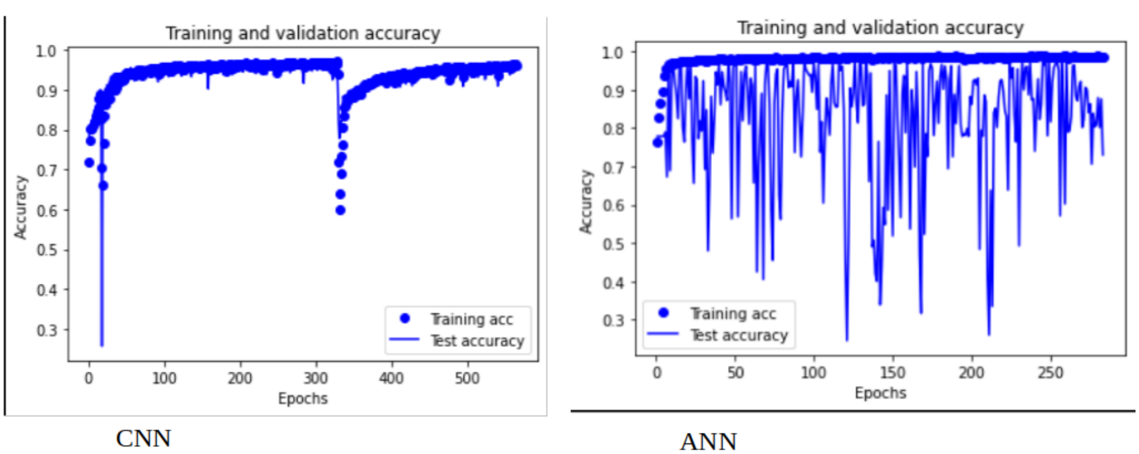


Fig. 3. These are for objects brighter than 22.5. In case of CNN, Training and Test accuracy is nearly same for all epochs, while this is not so in ANN (too many downfalls). Therefore it shows that CNN is not overfitted and captured the relation between feature and target variable very well

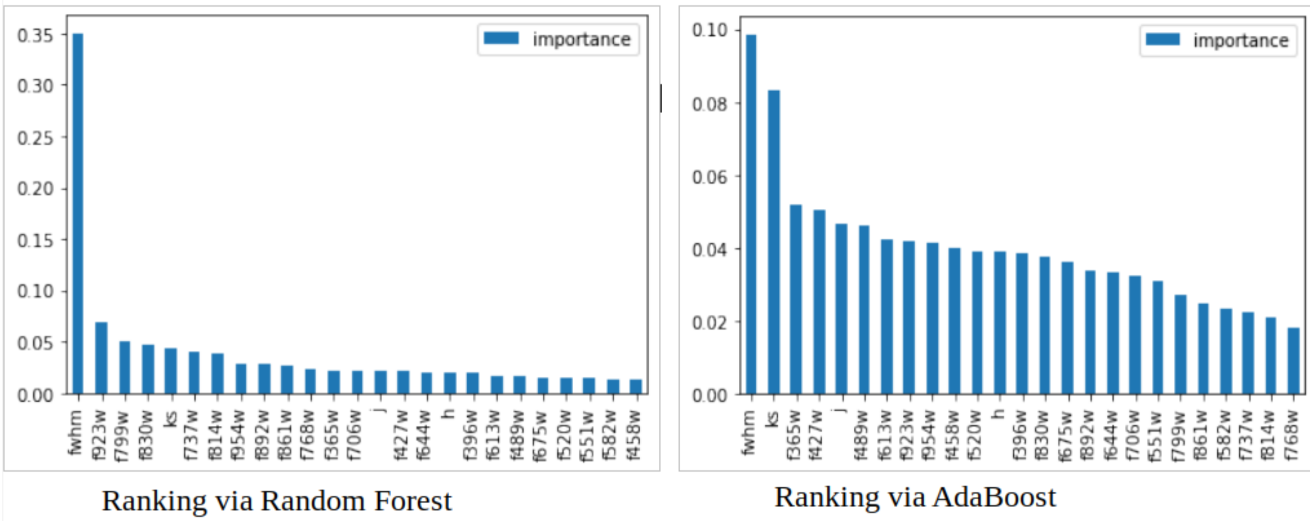
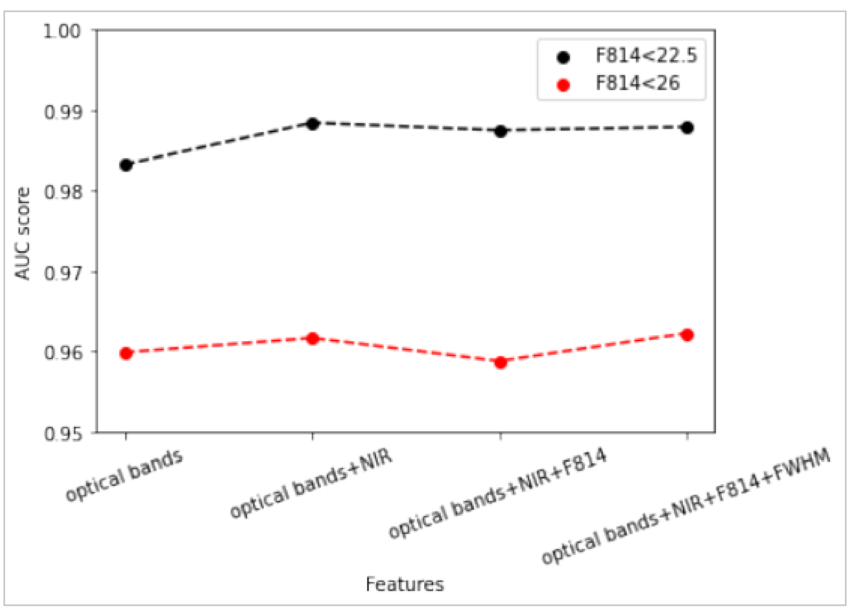


Fig. 4. First graph (randomforest) shows that FWHM is the most important feature in classification, but not others (Narrow bands). Therefore second graph is also used, as Ada Boost make weak learners strong. Therefore, some of these magnitudes, based on correlation heatmap, can be removed as they do nothing except increasing complexity and cpu time in classification. Also, it will be good to see whether colors instead of magnitudes be more useful and important in classification and comparable to fwhm



F814W	AUC SCORE
F814W≤ 22.5	0.989
22.5 <F814W≤ 23	0.902
23 <F814W≤ 24	0.89
24 <F814W≤ 25	0.735
F814W≤ 26	0.96

Fig. 5. Roc Auc for CNN. At right side, we see that AUC scores for fainter objects are low, it can be because there are not enough stars(100-300) in these ranges, they act as noise or contamination to our model as these objects are generally galaxies. Also, as FWHM is most important feature but CNN captured the spectrum well and give the same AUC with nearly no wrongly classified galaxy.

7. FEATURE REDUCTION

Training of CNN can take a lot of time. Therefore feature reduction is important and removal of unnecessary and same kind of features increases the accuracy and lowers the overfitting.
In fig6, We can see that these 12 magnitudes('f923w','f892w','f861w','f830w','f799w','f768w','f737w','f706w','f675w','f644w', 'f613w','f489w') are highly correlated with F814w, with corrl. coefficient > 0.94. Therefore dropping these magnitudes (also in fig4. we can see that their importance is too low and almost same), we are left with only 13 features.

Now running CNN, AUC increases to 0.997 for objects brighter than 22.5, and for XG Boost 0.9958.
Therefore we can conclude that (FWHM, f814w, ks, h, j, f954w, f551, f520, f458, f427, f396, f365) these 13 magnitudes (8-narrow bands) are enough to classify star or galaxy with even high AUC score and much less CPU time for CNN. XG Boost is already quiet fast.

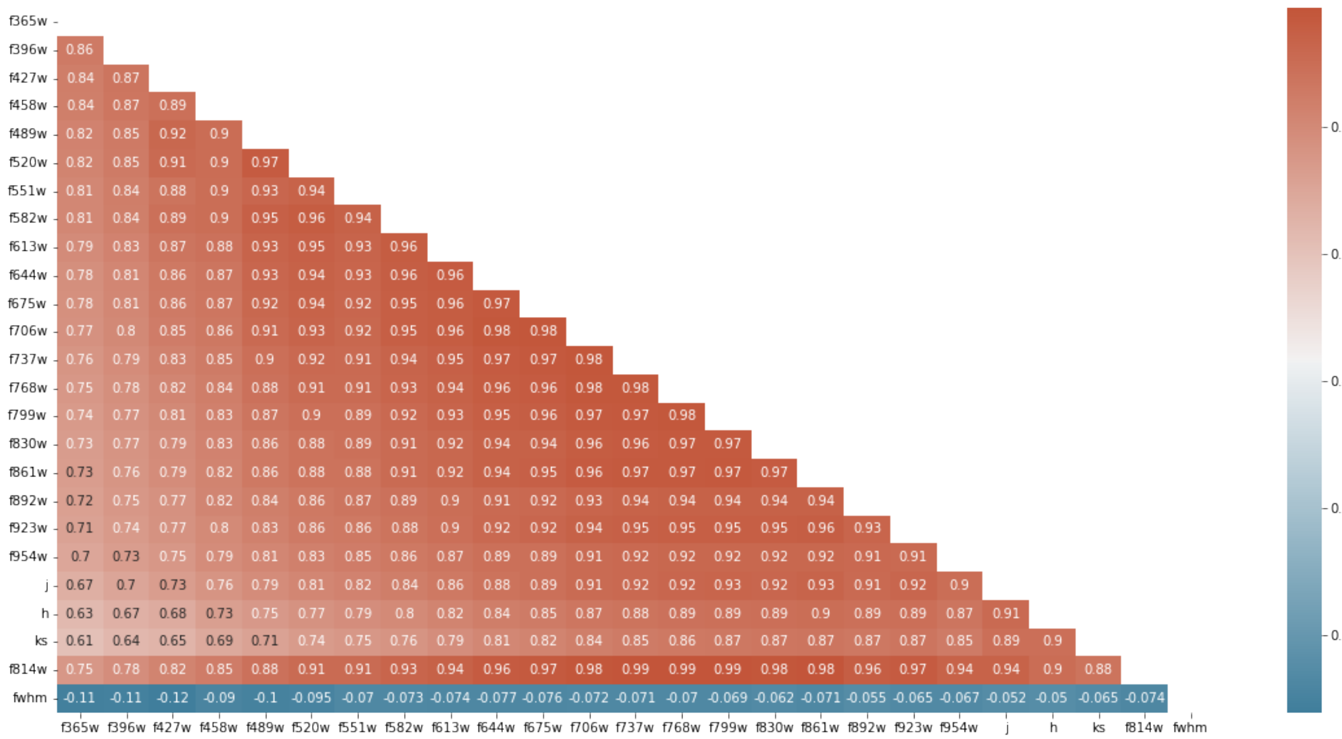


Fig. 6. Correlation Heatmap of features (magnitudes). It shows that Magnitudes, especially of higher wavelengths, are highly correlated to F814w. Therefore, we will drop the features with correl. coefficient >0.94, resulting in just 13 magnitudes. Although, its a common practice to drop all features with corr coefficient > 0.9, but here it may lead to too less features, can result in underfitting

8. INTRODUCING COLORS

From ranking of features in fig4, we see that fwhm is the most important feature followed by K band and 20 narrow bands are quiet low in importance. Therefore it will be good to see that whether colors be more suitable than magnitudes for classification. So now we will use 19 colors (difference of consecutive 20-bands)+ 2 colors(i-h, h-k) + f814w + fwhm on our best algorithms i.e. CNN and XG Boost
See fig7. For CNN, we get an AUC of 0.994 for objects brighter than 22.5 and 0.953 for objects than 26. For XG Boost, we get an AUC of 0.995 and 0.978 resp. Again the later one in both cases a little low because of those 100-200 stars, producing contamination and noise.
In fig8, we see that optical bands (19 colors) are providing good classification, AUC= 0.994 and after using all features, 0.995, for XG Boost. Same results for CNN but is very slow.
Therefore colors are proven to be more important than magnitudes and Fwmh for classification. CNN gives an accuracy of 96.2% and XG Boost 98.6%, compared with Alhambra's classification.
Also, CNN is better for fainter objects,reason given in fig8.

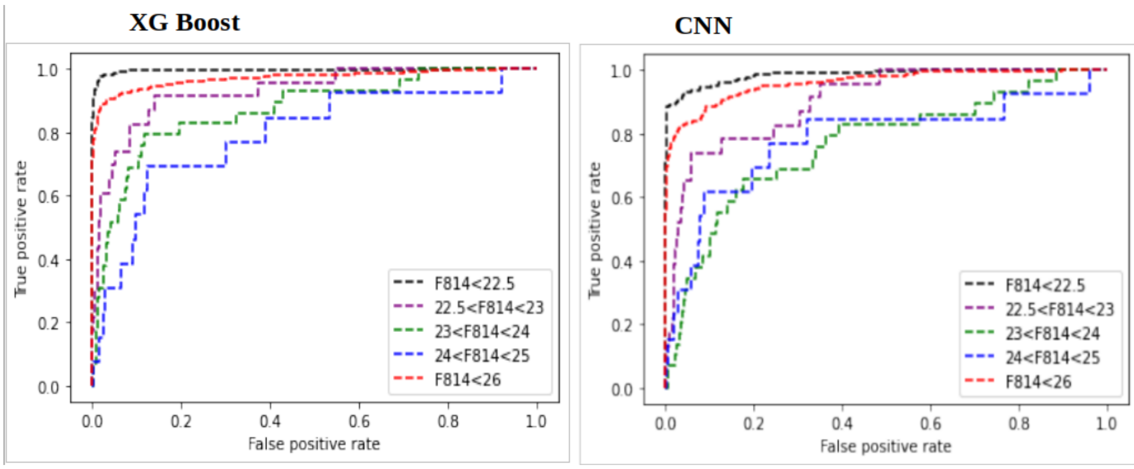


Fig. 7. This time (with colors) Curves are more closer to one and better, corresponding to their Roc's in previous sections. Also XG Boost is performing better than CNN for objects brighter than 22.5 but CNN is better for fainter objects.

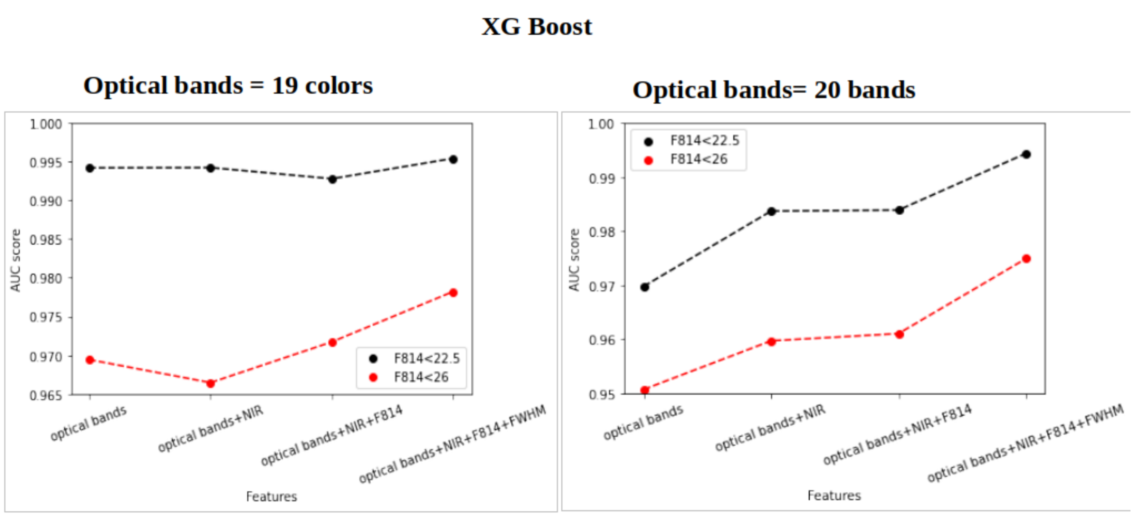


Fig. 8. Here we can see that in right side, auc(0.995) which was obtained by adding FWHM can be obtained just by using colors on left. Also, NIR and specially FWHM seem to very important for fainter objects.

Also, although XG Boost is providing high scores for objects brighter than 26 but CNN is still better for classification as XG Boost is getting overfit for contamination. It is correctly classifying stars which should be taken as galaxies, as a result it might wrongly classify some galaxies as stars, which CNN does not. That's why XG boost is giving 98.6% accuracy comparing with alhambra's stellar flags, which signify overfitting as boosting algos are sensitive to noise.

9. FEATURE IMPORTANCE AND REDUCTION

In fig9, we can see that some colors are very important in classification unlike magnitudes, f369-f427 is even important than FWHM. Most important features are f396-f427, FWHM, f814w, f365-f396, f458-f459, f768-f799. Now we ran CNN and XG Boost on these 6 features only and got an AUC score of 0.995 in both cases, same as when all features were used but overfitting has reduced to a greater extent. See fig10.
Therefore these six features (out of which 4 are colors) are enough to produce better results with same accuracy, less overfitting and much less CPU time.
Here we also saw that most important colors i.e. f369-f427, f396-f427, f365-f396 consists of magnitudes of low wavelengths.

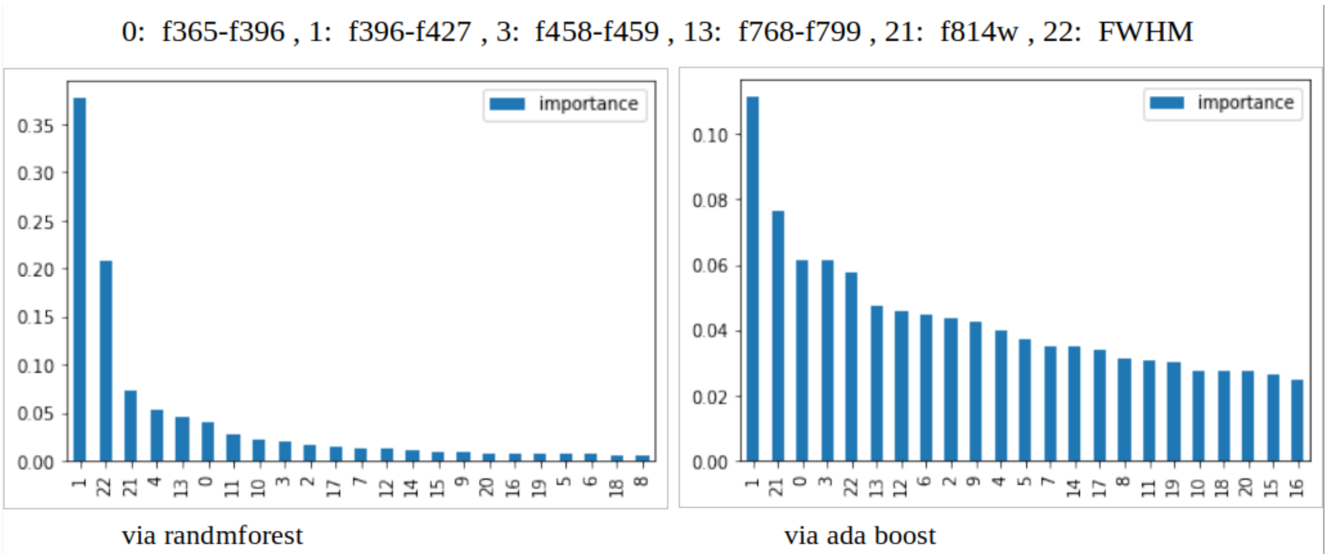
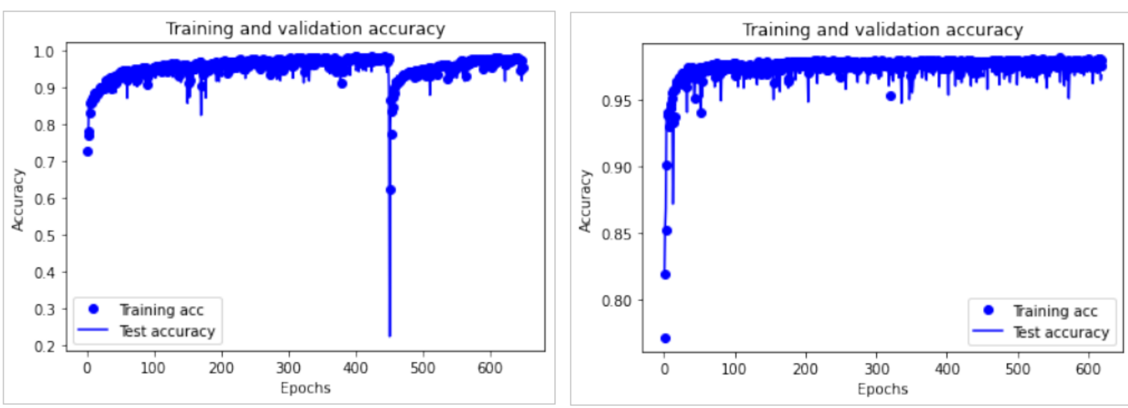


Fig. 9. Now here we see that f396-f427 becomes the most important feature followed by FWHM, f814w, f365-f396, f458-f459 and then f768-f799 at last. we see that magnitudes of lower wavelength are more important in classification.



Using all 21 colors + FWHM + f814w

Using only 4 imp. Colors + FWHM + f814w

Fig. 10. training and test accuracy for CNN. We see that overfitting has reduced significantly in right fig.

10. CONCLUSION

As a result we see that only CNN was perfectly able to capture the spectrum (20-narrow bands) and classify an object as star or galaxy with 96.2% accuracy and AUC score of 0.997 (after feature reduction). Similar results for XG Boost with 98.6% accuracy and very fast but not as suitable as CNN with only optical bands or for fainter objects, as it is sensitive to noise. Here we were perfectly able to classify object as star or galaxy with much less overfitting and CPU time using only 6 features instead of 23. We used only 6 important features, which consists of FWHM, f814w and four colors- f396-f427, f365-f396, f458-f459 and f768-f799, mostly of low wavelengths. This way, we can classify any object, even those for which morphology based classification is not possible (with nearly no misclassified galaxy).

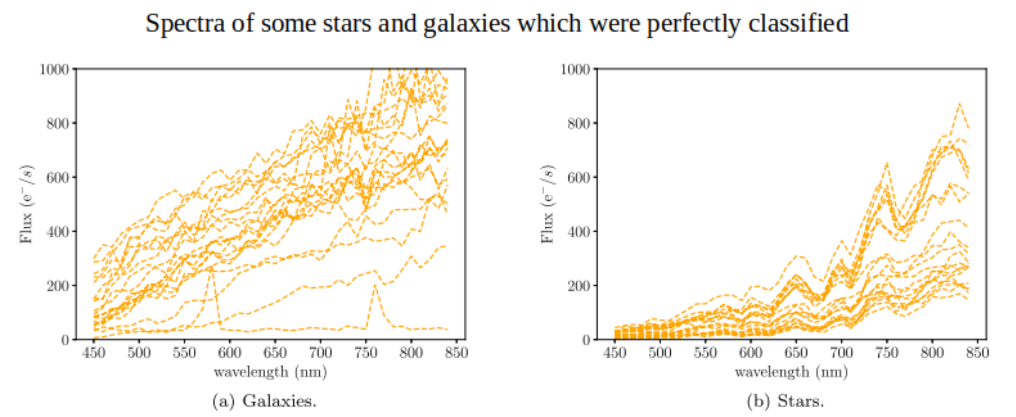


Fig. 11. taken from [2] for PAU survey. Lines are more steep for galaxies in low wavelength ranges. Therefore, fluxes specially of low wavelengths differentiates stars and galaxies better. Similar to what our algorithms showed in above sections

CNN

F814W	AUC SCORE
F814W≤ 22.5	0.997
22.5 <F814W≤ 23	0.902
23 <F814W≤ 24	0.89
24 <F814W≤ 25	0.736
F814W≤ 26	0.962

XG Boost

F814W	AUC SCORE
F814W≤ 22.5	0.995
22.5 <F814W≤ 23	0.93
23 <F814W≤ 24	0.871
24 <F814W≤ 25	0.791
F814W≤ 26	0.978

Fig. 12. Auc scores of CNN are for 8 narrow bands + 3 broad band + f814w + fwhm.
Auc scores of XG Boost are for 19 colors (narrow) + 2 colors (broad) + f814w + fwhm

REFERENCES

1. Banerji, M., Jouvel, S., Lin, H., McMahon, R. G., Lahav, O., Castander, F. J., Abdalla, F. B., Bertin, E., Bosman, S. E., Carnero, A., Kind, M. C., da Costa, L. N., Gerdes, D., Gschwend, J., Lima, M., Maia, M. A. G., Merson, A., Miller, C., Ogando, R., Pellegrini, P., Reed, S., Saglia, R., Sánchez, C., Allam, S., Annis, J., Bernstein, G., Bernstein, J., Bernstein, R., Capozzi, D., Childress, M., Cunha, C. E., Davis, T. M., DePoy, D. L., Desai, S., Diehl, H. T., Doel, P., Findlay, J., Finley, D. A., Flaugher, B., Frieman, J., Gaztanaga, E., Glazebrook, K., González-Fernández, C., Gonzalez-Solares, E., Honscheid, K., Irwin, M. J., Jarvis, M. J., Kim, A., Koposov, S., Kuehn, K., Kupcu-Yoldas, A., Lagattuta, D., Lewis, J. R., Lidman, C., Makler, M., Marriner, J., Marshall, J. L., Miquel, R., Mohr, J. J., Neilsen, E., Peoples, J., Sako, M., Sanchez, E., Scarpine, V., Schindler, R., Schubnell, M., Sevilla, I., Sharp, R., Soares-Santos, M., Swanson, M. E. C., Tarle, G., Thaler, J., Tucker, D., Uddin, S. A., Wechsler, R., Wester, W., Yuan, F., and Zuntz, J. (2014). Combining Dark Energy Survey Science Verification data with near-infrared data from the ESO VISTA Hemisphere Survey. *Monthly Notices of the Royal Astronomical Society*, 446(3):2523–2539.
2. Cabayol, L., Sevilla-Noarbe, I., Fernández, E., Carretero, J., Eriksen, M., Serrano, S., Alarcón, A., Amara, A., Casas, R., Castander, F. J., de Vicente, J., Folger, M., García-Bellido, J., Gaztanaga, E., Hoekstra, H., Miquel, R., Padilla, C., Sánchez, E., Stothert, L., Tallada, P., and Tortorelli, L. (2018). The PAU survey: star–galaxy classification with multi narrow-band data. *Monthly Notices of the Royal Astronomical Society*, 483(1):529–539.
3. Kovács, A. and Szapudi, I. (2015). Star–galaxy separation strategies for WISE-2MASS all-sky infrared galaxy catalogues. *Monthly Notices of the Royal Astronomical Society*, 448(2):1305–1313.
4. Leauthaud, A., Massey, R., Kneib, J.-P., Rhodes, J., Johnston, D. E., Capak, P., Heymans, C., Ellis, R. S., Koekemoer, A. M., Fèvre, O. L., Mellier, Y., Réfrégier, A., Robin, A. C., Scoville, N., Tasca, L., Taylor, J. E., and Waerbeke, L. V. (2007). Weak gravitational lensing with cosmos: Galaxy selection and shape measurements. *The Astrophysical Journal Supplement Series*, 172(1):219.
5. Molino, A., Benítez, N., Moles, M., Fernández-Soto, A., Cristóbal-Hornillos, D., Ascaso, B., Jiménez-Teja, Y., Schoenell, W., Arnalte-Mur, P., Pović, M., Coe, D., López-Sanjuan, C., Díaz-García, L. A., Varela, J., Stefanon, M., Cenarro, J., Matute, I., Masegosa, J., Márquez, I., Perea, J., Del Olmo, A., Husillos, C.,

Alfaro, E., Aparicio-Villegas, T., Cerviño, M., Huertas-Company, M., Aguerri, J., Broadhurst, T., Cabrera-Caño, J., Cepa, J., González, R. M., Infante, L., Martínez, V. J., Prada, F., and Quintana, J. M. (2014). The ALHAMBRA Survey: Bayesian photometric redshifts with 23 bands for 3 deg2. *Monthly Notices of the Royal Astronomical Society*, 441(4):2891–2922.