The PAU Survey: star-galaxy classification with multi narrow-band data

L. Cabayol[1]⋆, I. Sevilla-Noarbe[2]†, E. Fernández[1], J. Carretero[1]‡, M. Eriksen[1]‡, S. Serrano[3], A. Alarcón[3,4], A. Amara[5], R. Casas[3,4], F. J. Castander[3,4], J. de Vicente[2], M. Folger[3,4], J. García-Bellido[6], E. Gaztanaga[3,4], H. Hoekstra[7], R. Miquel[1,8], C. Padilla[1], E. Sánchez[2], L. Stothert[9], P. Tallada[2]‡, L. Tortorelli[5]

[1] *Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, 08193 Bellaterra (Barcelona), Spain*
[2] *Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain*
[3] *Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, 08193 Barcelona, Spain*
[4] *Institut d'Estudis Espacials de Catalunya (IEEC), 08193 Barcelona, Spain*
[5] *Institute for Particle Physics and Astrophysics, ETH Zürich, Wolfgang-Pauli-Str. 27, 8093 Zürich, Switzerland*
[6] *Instituto de Fisica Teorica UAM/CSIC, Universidad Autonoma de Madrid, 28049 Madrid, Spain*
[7] *Leiden Observatory, Leiden University, Leiden, The Netherlands*
[8] *Institució Catalana de Recerca i Estudis Avançats, E-08010 Barcelona, Spain*
[9] *Institute for Computational Cosmology, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

Classification of stars and galaxies is a well-known astronomical problem that has been treated using different approaches, most of them relying on morphological information. In this paper, we tackle this issue using the low-resolution spectra from narrow band photometry, provided by the PAUS (Physics of the Accelerating Universe) survey. We find that, with the photometric fluxes from the 40 narrow band filters and without including morphological information, it is possible to separate stars and galaxies to very high precision, 99% purity with a completeness of 98% for objects brighter than $I = 22.5$. This precision is obtained with a Convolutional Neural Network as a classification algorithm, applied to the objects' spectra. We have also applied the method to the ALHAMBRA photometric survey and we provide an updated classification for its Gold sample.

Key words: techniques: photometric – methods: data analysis

1 INTRODUCTION

A basic step in the extraction of astronomical information from photometric images is the separation of stars from galaxies. This is vital in a photometric survey in order to provide pure samples with minimal systematic contribution from the effect of cross-contamination of the star and galaxy samples, to be used for parameter estimation or model comparison in astrophysical or cosmological analyses (see, e.g., Sevilla-Noarbe et al. (2018) or Soumagnac et al. (2015), where the impact of this issue in large scale structure, weak lensing or Milky Way studies is addressed).

Historically, there have been many different approaches to tackle this problem. The first classification methods were morphology based (MacGillivray et al. 1976; Kron 1980; Shimasaku et al. 2001; Leauthaud et al. 2007) and they consisted in the estimation of an optimal cut on the space of observable image properties, such as a magnitude-size space, or in statistical properties such as measured second-order moments. However, these methods perform poorly when classifying faint objects as morphological information contained in noisy measurements is limited. Improved classification based on morphology relying on more advanced algorithms have been reported in Sevilla-Noarbe et al. (2018); López-Sanjuan et al. (2018).

⋆ E-mail: lcabayol@ifae.es
† E-mail: ignacio.sevilla@ciemat.es
‡ Also at Port d'Informació Científica (PIC), Campus UAB, C. Albareda s/n, 08193 Bellaterra (Cerdanyola del Vallès), Spain

Other classification methods use Bayesian-based approaches (Sebok 1979; Rachen 2013; Henrion et al. 2011; Fadely et al. 2012). The application of a Bayesian classification approach to multi band data must consider information coming from morphologies and color: the morphological features of a galaxy will be correlated with its magnitude. Hence, as the number of photometric bands increases, this approach gets more and more complicated.

Another approach is that provided by machine learning algorithms, which have emerged as an important tool for classification (see e.g. Odewahn et al. 1992; Bertin & Arnouts 1996; Soumagnac et al. 2015). It consists in learning the underlying behavior of a given class sample adaptively from the training data and the later generalization of this learning to samples beyond such training data.

Different possible machine learning algorithms are used in star galaxy classification problems, such as boosted decision trees (BDT) (Sevilla-Noarbe & Etayo-Sotos 2015), neural networks (ANN) (e.g. Soumagnac et al. 2015) or random forests (RF) (e.g. Morice-Atkinson et al. 2017), most of them trained on morphologically-based truth tables from external, deeper datasets or detailed simulations.
Most broad band photometric surveys, such as DES (The Dark Energy Survey Collaboration 2005), SDSS (Blanton et al. 2017) or PANSTARRS (Chambers et al. 2016), rely on morphological information, with just moderate evidence that this can be improved with color information (Sevilla-Noarbe et al. 2018), without resorting to infrared data (Kovács & Szapudi 2015; Banerji et al. 2015).

For the case of narrow-band data one can ask if it is possible to distinguish stars and galaxies only from the fluxes. This way, narrow-band surveys, which do not go as deep as their broad-band counterparts, could provide an accurate classification based on their flux distribution as well. In this work, we examine this question considering several machine learning approaches.

We will discuss the performance of machine learning algorithms on multiple narrow-band color information using PAUS (Physics of the Accelerated Universe Survey, Martí et al. 2014; Castander et al. 2012) and ALHAMBRA (Advanced Large Homogeneous Area Medium Band Redshift Astronomical survey, Moles et al. 2008b,a). In the case of PAUS, the classification can be compared with that provided by `SExtractor` (Bertin & Arnouts 1996), a software that detects, deblends, measures and classifies sources from astronomical images. Concerning ALHAMBRA, we will apply our algorithm and compare with the current classification scheme from the Gold catalog (Molino et al. 2014), which is based on photometric fluxes and morphologies.

The standard processing of PAUS images is carried out by performing forced photometry (S.Serrano et al. in prep.): the fluxes from objects are computed at predefined reference positions from external catalogs, in order to obtain more precise photometric redshifts for these broad band detections. Hence, the objects are already classified from deeper observations, before applying any further method. The results from this paper are meant to demonstrate the efficiency of machine learning algorithms on astronomical classification problems using narrow band photometry spectra, and may be useful to think of implementation of these algorithms to other crucial issues, such as galaxy classification, photo-z or outlier rejection for this kind of data. In addition, the objects used for PAUS photometric calibration are SDSS stars (Castander et al. in prep.), so it would be of great interest for PAUS to have its own classified stars to perform such a calibration, with a more pure selection. Lastly, an algorithm able to classify objects with low-resolution spectra would also be interesting for planned or future narrow-band surveys.

The layout of this paper is as follows; in section 2 we will define the datasets employed in the analysis. In section 3, there is a short definition of all the machine learning algorithms used at some point in this study. The characterization of the algorithms is done in section 4 and section 5 provides the results on the ALHAMBRA and PAUS datasets. In section 6 there is a final discussion of the main results.

2 DATA

In this work we would like to assess the performance of a machine learning classifier over two narrow-band datasets, PAUS early data and the ALHAMBRA Gold catalog[1], in the latter case comparing with the standard classification provided by that survey. We will work on the COSMOS field[2] comparing against the COSMOS space-based imaging catalog (Leauthaud et al. 2007), which provides a morphology-based classification (`MU_CLASS`) for the objects to train and test our methods on. It contains 1.2×10^6 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), 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 catalog contains, roughly, 1.1×10^6 galaxies, most towards the faint end, 30,000 stars and the rest are fake detections[3].

2.1 PAUS

The Physics of the Accelerating Universe Survey was born in 2008 with the idea of measuring precise redshifts[4] for a large number of galaxies using photometric measurements (Martí et al. 2014). The novelty of the project was to carry out a photometric survey with a large number of narrow

[1] `https://cloud.iaa.csic.es/alhambra/`
[2] `http://cosmos.astro.caltech.edu/`
[3] Technically, this classification only separates point-like versus extended objects. QSOs and AGNs will tend to be mixed with both samples and are neglected in this work. This can however be a very interesting avenue to explore in the context of CNN narrow band classification.
[4] $\approx$ 0.35% error, meaning a precision of $\sigma(z)/(1+z) \approx 0.0035$, versus a typical 5% for broad band measurements

filters, in order to obtain a low-resolution spectrum of a large number of cosmological objects. Among other science cases, this will allow the study of clustering at intermediate scales (Stothert et al. in prep.), intrinsic alignments or contributing to the effective modeling of galaxies in image simulations (Tortorelli et al. 2018a).

The PAUS camera, named PAUCam (Padilla et al. 2016), is equipped with 40 narrow band (NB) filters, 13 nm wide and separated by 10 nm, covering a total wavelength range from 450 nm to 850 nm and 6 *ugrizY* broad band filters (which are not used in this work). The survey covers approximately 0.75 square degrees of equivalent 40 NB area per night, delivering low-resolution ($R \approx 50$) spectra for all objects in the field of view. The camera is mounted at the prime focus of the 4.2m William Herschel Telescope. As of May 2018, PAUS has been observing for approximately 40 nights per year since mid-2015 (with an efficiency below 50% due to bad weather).

PAUS data is managed by a complex infrastructure which starts at the mountaintop, stores data temporarily there and sends it to the PAUS data center at the Port d'Informació Científica (PIC) where the nightly and higher level pipelines (Serrano et al. in prep.) are run and data is archived for long term storage, as well as distributed through a database for scientific use (Tonello et al. in prep., Carretero et al. (2017)).

Photometric calibration is tied to the Sloan Digital Sky Survey (SDSS, Smith et al. 2002) stellar photometry. Each PAUS image is separately calibrated using high signal-to-noise detected stars that are matched to the SDSS catalogues. The SDSS broad band photometry for these stars is fit to the Pickles stellar templates (Pickles 1998)[5] to obtain a spectral energy distribution (SED), which is used to synthesize the expected NB fluxes in PAUS. Single image zero points are then determined by comparing the modelled and observed fluxes (see Castander et al. in prep. for more information).

The PAUS catalog over this field contains 49,000 astronomical objects, matched to the COSMOS catalog: 42,000 galaxies and 7,000 stars from magnitudes $I = 16$ to $I = 23$.

2.2 ALHAMBRA

We have also used the ALHAMBRA photometric redshifts catalog (Molino et al. 2014) over the ALHAMBRA-4 field, which overlaps with COSMOS. It contains 37,000 objects matched to our reference COSMOS catalog, from which 34,000 are galaxies and 3,000 are stars. The ALHAMBRA photometric system (Aparicio Villegas et al. 2010) is characterized by 20 constant width (31 nm), non-overlapping medium band filters covering a wavelength range from 350 nm 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 3 NIR broad bands: J, H, K.

The catalog presents multicolor PSF-corrected photometry detected in synthetic F814W images with objects up to a magnitude of $F814W \approx 26.5$.

3 METHODS

Neural networks and random forests have already been used to classify stars and galaxies successfully (as shown, e.g., in Kim et al. 2015; Soumagnac et al. 2015), however, as mentioned before, they have never been used solely with band fluxes inputs. On the other hand, Convolutional Neural Networks (CNN) have been applied to different fields with excellent results, for instance in medical imaging (Qayyum et al. 2017), and they have proven to be very powerful in image processing and pattern recognition, also for one dimensional information (Méndez-Jiménez & Cárdenas-Montes 2018). These algorithms have also been applied to spectral classification (Hála 2014) and to tackle the star-galaxy classification problem using whole CCD images as input feature map (Kim & Brunner 2017).

3.1 Machine Learning algorithms

In this section, we describe the three machine learning algorithms for which we have compared performances in our case of study.

3.1.1 Artificial Neural Networks (ANN)

Neural networks (Werbos 1982) are a biologically-inspired programming paradigm that enables a computer to learn from observed data. They can be applied to difficult classification tasks, where a training sample already classified by other means is used to 'teach' the network. The learning process consists in recursively weighting the input features (the fluxes on the different bands in our case) by some factors, the weights, chosen in order to optimize the classification algorithm. This consists in the evaluation of a 'cost function', which is a measure of the overall agreement between the actual nature of the objects in the training sample and that inferred from the weighted inputs. In every iteration each weight is updated (back-propagated) in proportion to the corresponding gradient of the cost function. We have used the neural network implementation from the Python `scikit-learn` package (Pedregosa et al. 2011). The combination of weights which minimizes the cost function is considered to be a solution of the learning problem.

3.1.2 Random Forests

As with neural networks, a random forest algorithm is also a supervised classification algorithm (Breiman 2001). It is composed of a collection of simple decision tree predictors, each of them giving an output class when given a set of input features.

[5] `http://www.stsci.edu/hst/observatory/crds/pickles_atlas.html`

A decision tree classifies data items executing step-by-step choices by posing a series of questions about the features associated with the objects. Each question is contained in a node and each node leads to children nodes, one per possible answer to the parents' node question. The questions therefore form a hierarchy encoded as a tree. The training set is used to establish the features' hierarchy and the value of the cuts in each of the nodes which optimizes the classification. After each iteration over the whole tree, the separation power is evaluated and the cuts are selected according to it.

In a random forest approach, many different decision trees are created. The training set is sampled with replacement so as to produce a training set for each of the decision trees taking part of the forest. Another difference is in the choice of the question at each node. In the random forest approach, only a random subset of the features is considered. Therefore, each decision tree shaping the forest may give a different classification output for the same sample. The prediction output is a combination of all the particular results by taking the most common prediction.
As with the case with neural networks, we have also used a random forest implementation provided by the `scikit-learn` Python package.

3.1.3 Convolutional Neural Networks (CNN)

Convolutional Neural Networks (Lecun et al. 2015) are a category of neural networks that have proven very effective in areas such as image recognition and classification. One characteristic of CNN compared to its predecessors is its ability to recognize patterns based on local features.

The architecture of the network consists basically of two types of layers: the convolutional layer and the pooling layer. In the convolutional layer, the algorithm takes a batch of the input feature map (the photometric spectrum in our case) and convolves it with a set of filters. This filtering consists of a weight vector (kernels, learning parameters) that multiplies the input map in stacks such that all the input features are covered at least once. The output of this layer is a set of feature maps resulting from the convolution of the initial input. A remarkable aspect of CNN is its local connectivity. While in ANN the different layers are fully connected, which means that all neurons from a layer are connected to those on the layer below, CNN are locally connected, meaning exactly the opposite. Each neuron only receives input from a small local group of the pixels in the input image, where the inputs that go into a given neuron are actually close to each other. The aim of locally connected layers is to avoid the loss of some subtle nuances of spatial arrangements.

On the other hand, pooling layers are used to reduce the dimensionality of the feature maps. There are different pooling methods carried out by different functions across local regions of the input. One usual pooling function is the maximum, which consists in grouping features together and keeping only that containing the largest value, although another typical alternative uses the mean function instead. This layer reduces the number of operations required for all the following layers while still passing on the valid information from the previous layer. The trainable parameters are located in the convolutional layer and typically, the pooling layer does not contain any. The final CNN output is generated through a fully connected layer (also called dense layer). It applies a linear operation in which every input is connected to every output by the weight to generate an output with dimensionality equal to the number of output classes we need. The output layer contains again a cost function that evaluates the error in the prediction. Similar to the ANN, once the forward pass is complete the back-propagation begins to update the weights for loss reduction.

Table 1. TP stands for 'True positive'. Alike, FP stands for 'False positive', FN for 'False negative' and TN for 'True negative', for a given threshold.

	Classified galaxy	**Classified star**
True Galaxy	TP	FN
True star	FP	TN

Figure 1 shows the architecture of our CNN. It contains three convolutional layers provided with an activation function. The first convolution is larger, so that the algorithm learns about more general features. The following layers have a smaller kernel, to focus on more subtle nuances. Between convolutions, there are also two pooling layers and again, the first one is the largest, 4x1, whereas the other two are 2x1. The third pooling layer connects the last convolution with the fully connected layers. The last two layers are fully connected layers (dense layers), where the second one has the dimension of the output, which in our case is two, one giving the object's probability of being a star and the other of being a galaxy (technically, however, both add up to unity in our case). We have used the `Keras` Python library (Chollet et al. 2015) to build our algorithm.

3.2 Analysis

To analyze the performance of the classifiers, we will often refer to *precision*, *completeness* or *recall*, and receiver operating characteristic curves, *ROC* curves. In the context of this paper, a positive result means an object classified as a galaxy whereas a negative result refers to any object classified as a star. With such terminology, Table 1 defines the concepts of true and false positive and true and false negative contextualized to our problem. With such parameters, we can define the true positive rate (TPR) and the false positive rate (FPR) (Equations (1) and (2)) and also, the precision and the recall (completeness) (Equations (3), (4)).

$$TPR = \frac{TP}{TP + FN}, \tag{1}$$

$$FPR = \frac{FP}{FP + TN}, \tag{2}$$

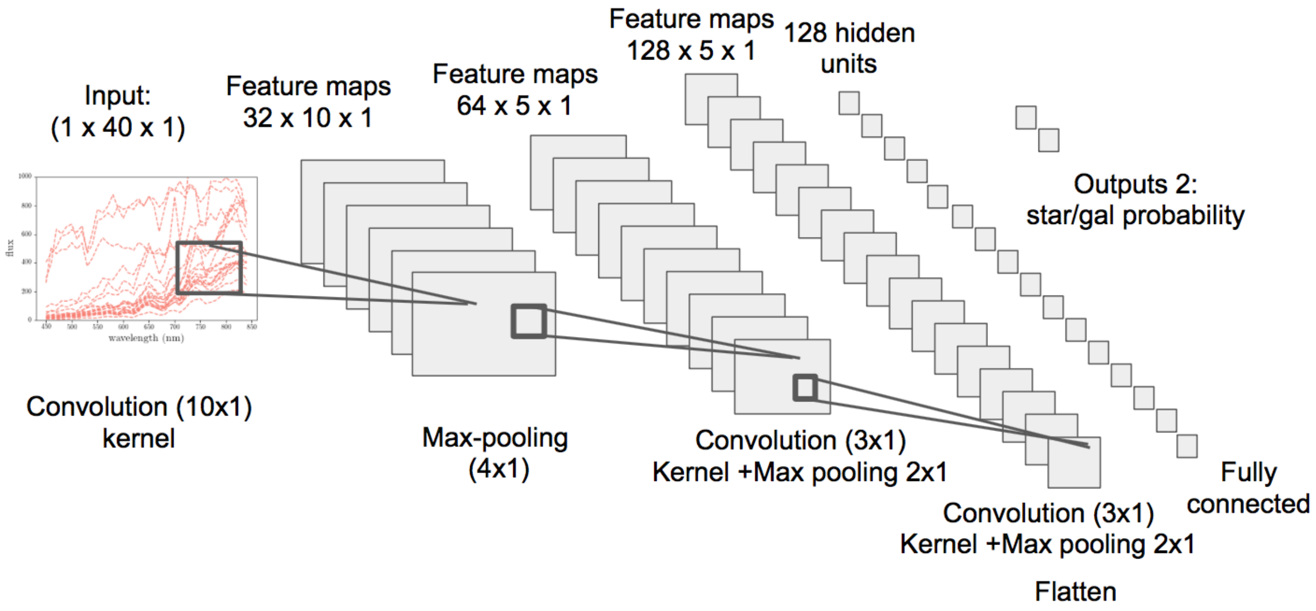


Figure 1. The CNN structure used for this paper. It is provided with 3 convolutional layers, with a larger kernel for the first convolution, and with 3 intercalated pooling layers. The last two layers are fully connected (dense) layers. NB that the input information is a one dimensional spectrum and not an image.

.

$$Precision = \frac{TP}{TP + FP}, \tag{3}$$

$$Recall = \frac{TP}{TP + FN} = TPR, \tag{4}$$

The performance of the classifiers is generally studied in the ROC space by ROC curves. A ROC curve is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied (the limit on a given classifier for which an objects is considered to belong either to a class or another), using the TPR vs FPR values typically. The area under the curve (AUC) gives a measurement of the performance of the classifier, where an area of 1.0 would mean a perfect classifier. A diagonal through the plot would indicate a random performance (therefore with an AUC ~ 0.5).

For our case of study, the algorithms output is the object's probability of being a galaxy. The ROC curve shows the TPR (the number of galaxies classified as galaxies over the total number of galaxies) against the FPR, (the number of stars classified as galaxies over the total number of objects classified as galaxies) when the probability threshold for which an object is considered either a star or a galaxy is varied. The ROC curve could also be represented with the TNR and the FNR, rating the classification/misclassification of stars instead of galaxies.

4 ALGORITHM PERFORMANCE

In this section, we analyze concurrently the performance of the three algorithms defined in section 3: artificial neural networks, random forests and convolutional neural networks. We use a training sample over the COSMOS catalog, matched to PAUS objects, where their 40 narrow-band fluxes have been used as the input data vector, up to magnitude $I = 22.5$ as defined by our reference COSMOS catalog.

4.1 Training set size dependence

The performance of any machine learning algorithm is related to the number of samples used in the training phase. However, using too many training samples may be self-defeating; the training could take much longer than really needed.

Figure 2(a) shows the training size dependence for the neural network, the random forest and the convolutional neural network, where the results plotted are those obtained on a distinct validation sample. We only vary the size of the training sample, while the validation sample size keeps constant in all cases.

One can already see here that the CNN is the algorithm yielding the best classification, with a better performance than those of the random forest or the ANN (see section 4.3 for discussion). It also showcases that for this sample, 10,000 objects are enough to get high classification rates with the CNN. Nevertheless, Figure 2(b) shows that from 10,000 to 30,000 objects the classification is still improving, although the improvement is smaller than from 3,000 to 10,000 objects. This means that the algorithm is more sensitive to training sample size increments when the training datasets are small.

4.2 Number of input bands dependency

In any classification problem, the more information is available about each class, the easier it is to identify particular patterns useful to differentiate between them. For the case of astronomical object classification, any spectral related or morphological information may be meaningful. However, a large number of input features can also have its drawbacks. Over-fitting or scaling problems may arise from such a large dimensionality. When the input's dimension increases, the hypervolume in input feature space increases so fast that the available data becomes sparse. Also, the data needed to provide reliable results increases exponentially.

For PAUS, the 40 available optical bands are probably not enough to encounter such problems. However, it is also of interest to check this and study how the performance scales with the number of bands (i.e., with the spectral resolution). To see how the algorithm scales with the loss of resolution, we have merged the 40 PAU NB in groups of 2,4,5 and 8 summing the flux of contiguous bands and therefore providing data samples with 20,10,8 and 5 bands, respectively.

Figure 3(a) presents the scaling with the number of bands for the three different algorithms, exhibiting the same pattern in all cases: as the spectral resolution increases, the performance of the algorithms improves. Such improvement is not linear; it has a more significant slope from 5 to 10 bands in all cases. One can see that with the CNN the photometry does not need to have 40 narrow bands to already give a good classification of stars and galaxies: 20 bands already result into high classification rates, as the shapes from the spectrum used to differentiate these two cases are already evident at such resolutions.

Figure 3(b) shows the different performances in the ROC space for the CNN. It exhibits the results we have already mentioned: there are important gaps between the curves from 5 to 20 bands, whereas increasing from 20 to 40 bands translates into a smaller improvement.

We have also studied the difference between using the 20 bluest bands versus the 20 reddest. The ROC-AUC for the bluer bands is 0.913 ± 0.005 whereas for those redder bands, it is 0.950± 0.004. Therefore, we find that star-galaxy separation is therefore more sensitive to the information contained in the redder bands, as many of the stars are typically red dwarfs with characteristic absorption features. There is also another effect one could consider: the bluer bands have lower S/N than the redder ones and naively one would expect that the classification performs worse.

4.3 Algorithms comparison

We have shown that the CNN is exhibiting the best performance so henceforth it will be the fiducial algorithm applied to the classification on the PAUS and ALHAMBRA catalogs.

There are many effects that are contributing to this result. As was mentioned above, CNN are provided with locally connected layers that are capable of recognizing subtle nuances of spatial arrangements. Figure 4 shows objects classified as galaxies (a) or stars (b) with high probability by the CNN. In the case of galaxies, one can notice that many of them contain gaps in one or two consecutive bands that most likely correspond to emission lines, meaning that the CNN is able to learn from these characteristic traits. We have made the same check with the ANN or the random forest algorithms and none of them present clear emission line patterns in the best classified galaxies. For stars, one can see that there are many objects with the same spectral shape, including certain patterns (peaks, valleys) usually in approximately nearby sections of the spectrum. Most of these correspond to red stars which in general are more commonplace in the dataset at the considered magnitudes. The algorithm in these cases is able to recognize these objects from the training set so that they can be identified readily as stars. It is worthwhile noting that a small percentage of the training set will include QSOs labeled as 'stars' in our case, so a further optimization could be possible by identifying these. This result (good performance of CNN on 1D quasi-spectral data for classification) is a true finding of this work, which opens up possibilities only available to this kind of photometric surveys, in which object types could be classified for large sets of objects simultaneously.

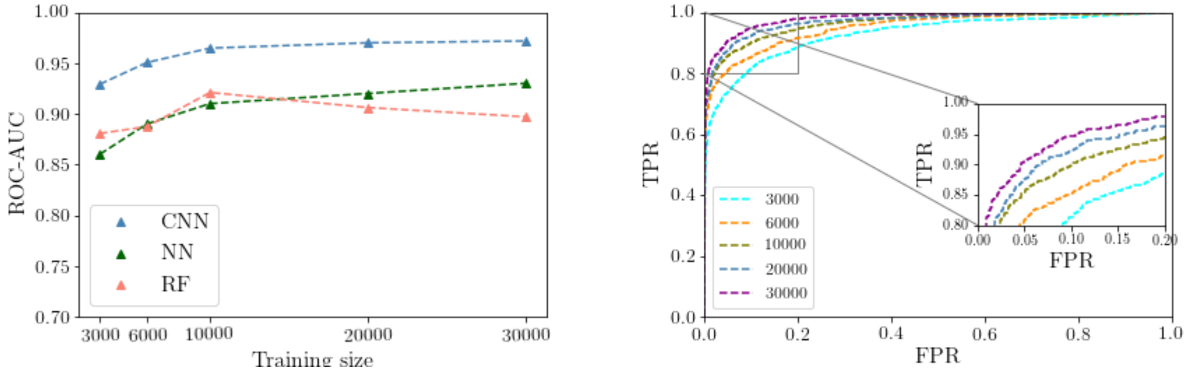
5 RESULTS

5.1 Classification on the PAUS catalog

As explained in section 2, the PAUS catalog in the COSMOS region contains 49,000 objects up to magnitude $I = 23$, from which 7,000 are stars and 42,000 are galaxies. We will often work with a subsample of objects brighter than $I = 22.5$, for which the catalog contains 6,000 stars and 28,000 galaxies. In section 4, we already noted that CNN is the best choice for classifying stars and galaxies using band fluxes input, and therefore we will use it by default in the rest of this work.

The PAUS catalog contains objects with negative flux measurements. This may happen for sources with a very low signal-to-noise in a given band and for which the background has been overestimated. In these cases, it is not possible to estimate a magnitude, and the corresponding value in the catalog is set to a 'sentinel' value of 99.0. It is not possible to train the algorithm with random measurements set to 99, thus we will use the PAUS fluxes as inputs for the algorithm, as it works well despite the negative flux counts.

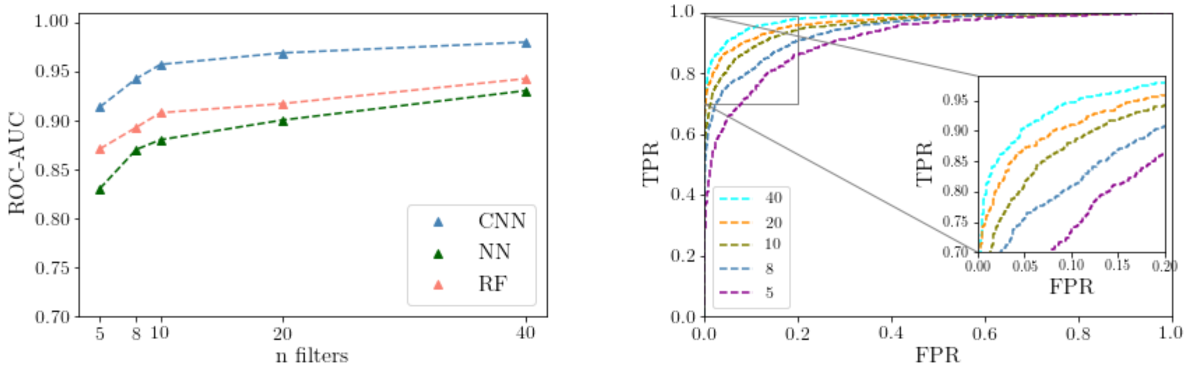
The training sample employed to carry out the classification is composed of 20,000 objects up to magnitude $I = 22.5$, 15,000 galaxies and 5,000 stars, whereas the



(a) ROC-AUC values for different training sizes.

(b) ROC curves corresponding to different number of training objects for the CNN.

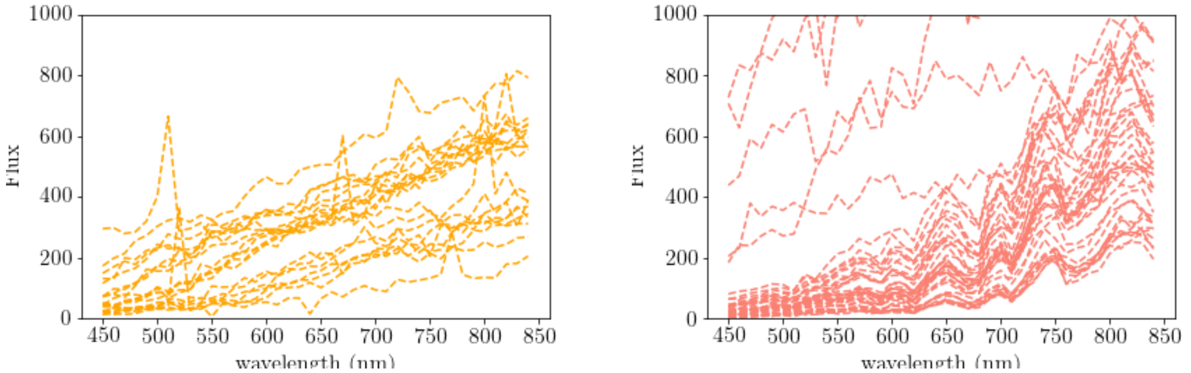
Figure 2. (a) ROC-AUC scaling for different training sample sizes for the ANN, random forests and the CNN using PAUS data with $i_{auto} < 22.5$ and 40 NB inputs. The results plotted are those obtained on the validation sample. (b) ROC curve showing the scaling of the CNN performance with the number of training objects



(a) ROC-AUC values for different number of inputs bands.

(b) Algorithm performance dependency on input number of bands.

Figure 3. (a) ROC-AUC scaling for different number of bands for the ANN, the RF and the CNN using PAUS data with $i_{auto} < 22.5$ and 10,000 training objects. The results plotted are those obtained on the validation sample. (b) ROC curves showing the scaling with the spectral resolution with the CNN algorithm.



(a) Galaxies.

(b) Stars.

Figure 4. (a) Spectra of objects classified as galaxies with the CNN with probability ~ 1. (b) Same for stars.

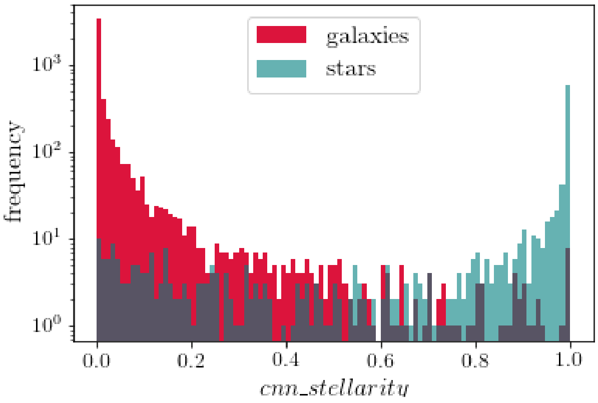


Figure 5. Distribution of *cnn_stellarity* for stars (blue) and galaxies(red), both populations on the validation sample.

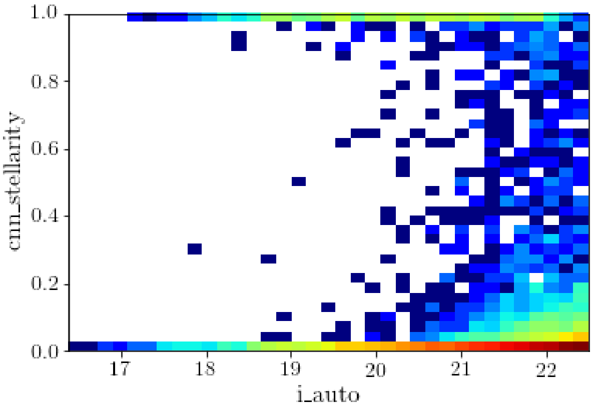


Figure 6. Heatmap for *cnn_stellarity* as as function of I magnitude, as measured by the HST-ACS on the COSMOS field.

test sample contains 1,000 stars and 5,000 galaxies. The algorithm outputs a probability of the object being either a star or a galaxy, which we will call *cnn_stellarity*. The resulting ROC-AUC is 0.973 ± 0.001, leading to a purity of 99% for a completeness of 98% for objects brighter than $I = 22.5$. This means that, the selected galaxy sample still contains a 1% of stars contaminating it, while losing a 2% of the original true galaxies of the sample.

Analyzing in more depth the classification of the PAUS sample, Figure 5 shows the histogram of such output probability. It exhibits two clearly differentiated peaks in 0 and 1, which correspond to stars and galaxies classified without any ambiguity. For probabilities far from 0 or 1, it presents some noisy measurements, coming mainly from faint galaxies. Figure 6 shows the same information but as a function of magnitude.

Figure 7 shows the performance of the algorithm for training sets in three different magnitudes ranges: for $I < 20.5$, $20.5 < I < 22.5$ and for $I > 22.5$. As expected, it shows a degradation as the sample becomes fainter. Concretely, the brighter range gives a ROC area of 0.991, worsening to 0.930 and 0.822, respectively. Training sizes have been fixed to 3,500 objects for the three cases as the number of objects is limited by the smaller, brightest bin. Considering figure 2(b), the performances could still improve with a larger training set, specially for $20.5 < I < 22.5$ and $I > 22.5$.

We can compare with morphological measurements on the same dataset, so a `SExtractor` run was executed over the same field to obtain the `CLASS_STAR` and `SPREAD_MODEL` estimates of the shape of the object (see Sevilla-Noarbe et al. (2018)). We used as an example the measurements in the 615 nm narrow band and compared with the CNN results for a flux limited sample $I < 22.5$ adjusting both samples to have the same signal to noise distributions. In Figure 8 we can see the advantages of using spectral information for classification, versus the standard morphological approach.

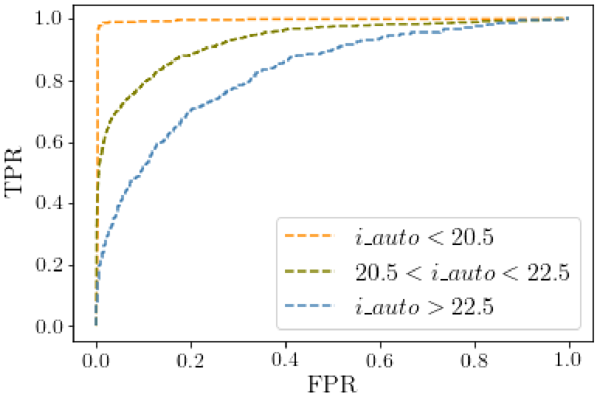


Figure 7. ROC curve for star-galaxy classification on PAUS data given different cuts on the magnitude. The results plotted are those obtained on the validation sample.

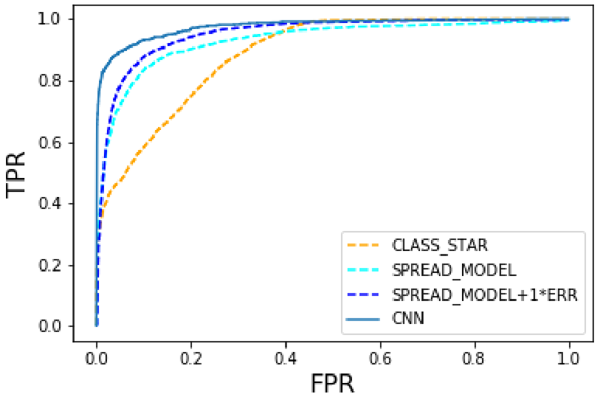


Figure 8. ROC curves for star-galaxy classification in PAUS data using CNN and `SExtractor` classifiers. ERR corresponds to the `SPREADERR_MODEL` quantity from `SExtractor`. Both samples have been selected to have similar signal to noise distributions.

5.2 Classification on the ALHAMBRA catalog

The application of the algorithm on the ALHAMBRA dataset should be useful to crosscheck the algorithm itself and also to test its power against an alternative classification scheme. The ALHAMBRA survey also performed star-galaxy classification (Molino et al. 2014) assigning a probability to every detection given its apparent geometry (the Full With at Half Maximum (FWHM) from `SExtractor`, a synthetic F814W magnitude, and optical F489W - F814W and near infrared (NIR) J-Ks colors). The authors derived a probability distribution function (PDF) based on the typical distribution of stars and galaxies for each of the variables cited above. The final probabilities, the star-galaxy classifier, are included in the catalogs as the statistical variable 'Stellar Flag'.

However, the ALHAMBRA images do not provide reliable morphological information for magnitudes $F814W > 22.5$, therefore the classification scheme is only applied up to this flux limit. For the rest of the catalog, they assigned a probability of 0.5.

It is of interest to see if by applying our algorithm based on low-resolution spectra on the ALHAMBRA catalog, we are able to match the purity provided (or even improve it) for objects brighter than $I = 22.5$. It is also of interest to see whether the algorithm is also able to classify faint objects for which ALHAMBRA did not provide any classification.

The ALHAMBRA catalog we are working with is smaller than that of PAUS. It contains 36,000 objects in the COSMOS region, from which, according to the COSMOS catalog, 3,000 are stars and 33,000 are galaxies. As with the PAUS catalog, in the ALHAMBRA catalog whenever a source was not detected in a given band, its magnitude was set to a 'sentinel' value of 99. As was the case for the PAUS catalog, it is not possible to train the CNN with random band measurements of 99.0: a CNN algorithm cannot be trained if the input contains gaps in some of the bands and there are algorithms that can fill these with different methods (mean value of the whole input, nearest neighbors, etc.). To fill in the ALHAMBRA catalog, we have changed the 99.0 with linearly interpolated magnitudes based on its contiguous neighbors. To have reliable measurements to test our method, we have only kept objects with 5 or less non-detected bands.
We run the CNN on magnitudes this time, therefore also checking the robustness of the code with a different range of inputs.

The input features for the CNN are a total of 23 parameters distributed as follows: the 20 mid-band optical magnitudes introduced as 19 colors, the 3 NIR broad bands J, H, K magnitudes also included as 2 colors, the F814W magnitude and the Full Width at Half Maximum (FWHM).

The sample of objects for which ALHAMBRA also provided a classification (hence those with $F814W < 22.5$) represents 20% of the objects, with 5,500 galaxies and 1,500 stars, in the complete ALHAMBRA Gold catalog. Taking the COSMOS classification as the 'true' value for classification (admitting some QSO contamination), the ALHAMBRA classification obtains a ROC-AUC of 0.983.

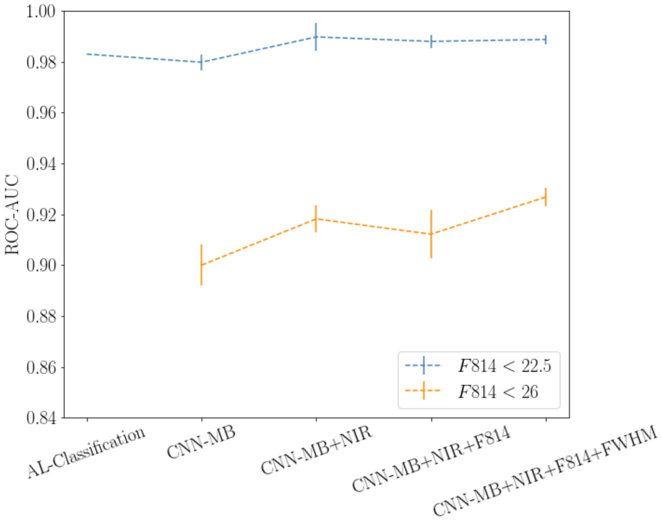


Figure 9. ROC AUC for the classification of the ALHAMBRA validation sample for different set of input features. In blue, classification for a object's sample with $F814W < 22.5$ in orange, with $F814W < 26$.

Figure 9 shows the AUC-ROC of the classification we have performed on objects brighter than 22.5 (blue) and objects brighter than 26 (orange). There are different performances, each of them corresponding to the addition of new input features. Firstly, we have run the algorithm with only the optical band information. Then, we have added first the NIR information, then the F814W magnitude and finally the FWHM. Each line corresponds to the performance with a concrete CNN feature set. This way, we can study how the algorithm scales as we add new features. The curve shows that by means of only the optical band information, the classification we get is similar to the original ALHAMBRA classifier *Stellar_Flag*.

The addition of the NIR data makes the most difference and implies an important improvement in the classification performance (the power of the addition of infrared bands was already explored in (Banerji et al. 2015; Kovács & Szapudi 2015; Sevilla-Noarbe et al. 2018)). The best classification obtained is that with a ROC AUC of 0.99 corresponding to the performance with all the input features. The FWHM seems to be improving the classification only for fainter objects (orange line). However, it may be that the brighter ones already have a classification rate too high to be improved with an additional parameter. Table 2 contains the ROC-AUC values for the classification with the different input feature maps for both cases, brighter than 22.5 and 26.

As we did for PAUS (Figure 7), Figure 10 shows how the classification scales with the objects' magnitude. For $F814W < 22.5$, we have already seen that the algorithm leads to a high ROC-AUC. Figure 10 exhibits the performance of the algorithm in different binned magnitude ranges and, as expected, for magnitudes fainter than 22.5 the classification

Table 2. ROC-AUC values for the classification of stars and galaxies in the ALHAMBRA dataset for different input feature maps. ALHAMBRA's exhisting classification provides a ROC-AUC of 0.983.

Information used	F814 <22.5	F814 <26
Optical Bands	0.980±0.003	0.901±0.001
Optical bands + NIR	0.987±0.002	0.918±0.006
Optical bands + NIR + F814	0.988±0.002	0.910±0.007
Optical bands + NIR + F814 + FWHM	0.989±0.002	0.927±0.004

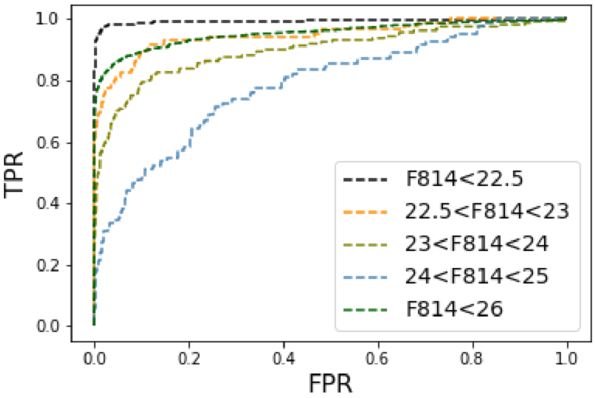


Figure 10. ROC curves obtained in the ALHAMBRA for different magnitude cuts.

performance degrades. Nevertheless, considering all objects brighter than $F814W < 26$, we are able to get 97% purity for completeness of 99%.

As it was mentioned above, ALHAMBRA provides a classification for objects brighter than 22.5. For fainter objects, it is common practice to consider every object to be a galaxy. This is also a good approach as there are relatively far fewer stars fainter than 22.5. In the catalog, for $F814W < 22.5$ the misclassified stars represent a 3% of the total dataset, meaning that one would have a 3% of contamination without nearly any misclassified galaxy. In the same magnitude range, we are able to obtain a purity of 98%.
If we move to $22.5 < F814W < 23$, considering all objects to be galaxies implies, according to the COSMOS classification, that the stellar contamination represents a 7.6% of the total dataset. The CNN is able to achieve a purity of 98%, therefore, we are able to improve this naive classification scheme for this magnitude range.
For $23 < F814 < 24$, the contamination of stars is also a 3.7% of the total dataset, whereas the algorithm obtains a 97.5% of purity. Finally, within $24 < F814 < 25$, the contamination of the samples is of a 2.2% whereas we get also a 98% purity. Therefore, we are able to improve the ALHAMBRA classification up to $F814W < 25$. For fainter objects on the other hand, it is better to consider all sources as galaxies.

In order to further validate our algorithm, we have tested on a different field, ALHAMBRA-2, corresponding to DEEP2 observations (Newman et al. 2013) training on ALHAMBRA-4 (COSMOS field). The reference catalog use here comes from matching to Hubble Space Catalog space imaging (Whitmore et al. 2016) making a cut on extendedness of 1.2, which separates cleanly the point-like versus extended sources. For objects with F814w < 22.5, the *cnn_stellarity* gives a ROC area of 0.943, while *Stellar_Flag* is 0.930.

6 CONCLUSIONS

Convolutional neural networks have proven to be a real breakthrough in many fields, such as image pattern recognition. We have shown here that they can be used as a powerful object classification tool using the shape of low resolution spectra from photometric data.

With such an algorithm, we have been able to classify stars and galaxies from the PAU survey by means of solely the object fluxes, without resorting to morphology, which in absence of a deeper detection image, can degrade significantly in the fainter end. This is done with a purity and a completeness of 99% and 98%, as shown in Figure 7, using the COSMOS field as our training and testing grounds.

These results demonstrate the power of both the PAUS photometric quality, as the CNN is able to detect subtle nuances in the spatial arrangement, such as characteristic of stellar spectra or emission lines, and use them to differentiate both populations.

In addition, using the same framework we have expanded and improved the ALHAMBRA classification. This survey also performed a star galaxy classification for objects brighter than 22.5, using fluxes, color and morphology as input features. We have applied our algorithm also to their data with the same inputs, leading to a purity and a completeness of 98% – 99%. Adding also the unclassified fainter objects which contain the bulk of the catalog up to magnitude 26 leads to a purity of 97% for a completeness of 99% (nearly no misclassified galaxies). Under the assumption that all sources fainter than 22.5 are galaxies, we are able to improve the classification from objects with F814W < 25. This classification for the ALHAMBRA Gold catalog will be made available upon publication of this work.

The application of CNN in such a fashion, using it to discover features in low-resolution spectra from this kind of surveys, opens up the possibilities beyond star-galaxy classification, such as for the identification of other families of objects (e.g. adding a representative sample of quasars or AGNs in multi-labeled classification) or photometric redshift determination. This expands on their current astronomical applications which up to now where mainly for image processing and extraction of information from them directly.

An interesting avenue to explore is the comparison with template fitting methods, which have not shown optimal results but might prevail over machine learning methods in circumstances where the training set is poor (Fadely et al. 2012). Templates could additionally be used to augment the training set and improve classification when a wider range of labels is required.

ACKNOWLEDGEMENTS

ISN would like to thank Miguel Cárdenas-Montes, Iván Méndez and Ben Hoyle for useful suggestions and discussions concerning the use of CNNs. Funding for PAUS has been provided by Durham University (via the ERC StG DEGAS-259586), ETH Zurich, Leiden University (via ERC StG ADULT-279396 and Netherlands Organisation for Scientific Research (NWO) Vici grant 639.043.512) and University College London. The PAUS participants from Spanish institutions are partially supported by MINECO under grants CSD2007-00060, AYA2015-71825, ESP2015-88861, FPA2015-68048, SEV-2016-0588, SEV-2016-0597, and MDM-2015-0509, some of which include ERDF funds from the European Union. IEEC and IFAE are partially funded by the CERCA program of the Generalitat de Catalunya. The PAU data center is hosted by the Port d'Informació Científica (PIC), maintained through a collaboration of CIEMAT and IFAE, with additional support from Universitat Autònoma de Barcelona and ERDF. CosmoHub has been developed by PIC and was partially funded by the "Plan Estatal de Investigación Científica y Técnica y de Innovación" program of the Spanish government.

REFERENCES

Aparicio Villegas T., et al., 2010, AJ, 139, 1242

Banerji M., et al., 2015, MNRAS, 446, 2523

Benítez N., et al., 2009, ApJ, 691, 241

Benítez N., et al., 2015, in Cenarro A. J., Figueras F., Hernández-Monteagudo C., Trujillo Bueno J., Valdivielso L., eds, Highlights of Spanish Astrophysics VIII. pp 148–153

Bertin E., Arnouts S., 1996, A&AS, 117, 393

Bertin E., Arnouts S., 2010, SExtractor: Source Extractor, Astrophysics Source Code Library (ascl:1010.064)

Blanton M. R., et al., 2017, AJ, 154, 28

Breiman L., 2001, Machine Learning, 45, 5

Capak P., et al., 2007, ApJS, 172, 99

Carretero J., et al., 2017, PoS, EPS-HEP2017, 488

Castander F. J., et al., 2012, in Ground-based and Airborne Instrumentation for Astronomy IV. Proceedings of the SPIE, Volume 8446, article id. 84466D, 9 pp. (2012).. p. 84466D, doi:10.1117/12.926234

Cenarro A. J., et al., 2018, preprint, p. arXiv:1804.02667 (`arXiv:1804.02667`)

Chambers K. C., et al., 2016, preprint, (`arXiv:1612.05560`)

Chawla N. V., Bowyer K. W., Hall L. O., Kegelmeyer W. P., 2011, preprint, (`arXiv:1106.1813`)

Chollet F., et al., 2015, Keras, `https://github.com/keras-team/keras`

Fadely R., Hogg D. W., Willman B., 2012, ApJ, 760, 15

Hála P., 2014, preprint, p. arXiv:1412.8341 (`arXiv:1412.8341`)

Henrion M., Mortlock D. J., Hand D. J., Gandy A., 2011, MNRAS, 412, 2286

Kim E. J., Brunner R. J., 2017, MNRAS, 464, 4463

Kim E. J., Brunner R. J., Carrasco Kind M., 2015, MNRAS, 453, 507

Kovács A., Szapudi I., 2015, MNRAS, 448, 1305

Kron R. G., 1980, ApJS, 43, 305

Laigle C., et al., 2016, The Astrophysical Journal Supplement Series, 224, 24

Leauthaud A., et al., 2007, ApJS, 172, 219

Lecun Y., Bengio Y., Hinton G., 2015, Nature, 521, 436

López-Sanjuan C., et al., 2018, preprint, p. arXiv:1804.02673 (`arXiv:1804.02673`)

MacGillivray H. T., Martin R., Pratt N. M., Reddish V. C., Seddon H., Alexander L. W. G., Walker G. S., Williams P. R., 1976, MNRAS, 176, 265

Martí P., Miquel R., Castander F. J., Gaztañaga E., Eriksen M., Sánchez C., 2014, MNRAS, 442, 92

Méndez-Jiménez I., Cárdenas-Montes M., 2018, in HAIS'18, Oviedo, Spain, June 20-22, 2018, Proceedings (accepted for publication). Springer, pp 158–170

Moles M., et al., 2008a, AJ, 136, 1325

Moles M., et al., 2008b, AJ, 136, 1325

Molino A., et al., 2014, MNRAS, 441, 2891

Morice-Atkinson X., Hoyle B., Bacon D., 2017, preprint, (`arXiv:1712.03970`)

Newman J. A., et al., 2013, The Astrophysical Journal Supplement Series, 208, 5

Odewahn S. C., Stockwell E. B., Pennington R. L., Humphreys R. M., Zumach W. A., 1992, AJ, 103, 318

Padilla C., et al., 2016, in Ground-based and Airborne Instrumentation for Astronomy VI. p. 99080Z, doi:10.1117/12.2231884

Pedregosa F., et al., 2011, Journal of Machine Learning Research, 12, 2825

Pickles A. J., 1998, Publications of the Astronomical Society of the Pacific, 110, 863

Qayyum A., Anwar S. M., Majid M., Awais M., Alnowami M., 2017, preprint, (`arXiv:1709.02250`)

Rachen J. P., 2013, in von Toussaint U., ed., American Institute of Physics Conference Series Vol. 1553, American Institute of Physics Conference Series. pp 254–261 (`arXiv:1302.2429`), doi:10.1063/1.4820007

Richards G. T., et al., 2004, ApJS, 155, 257

Sebok W. L., 1979, AJ, 84, 1526

Sevilla-Noarbe I., Etayo-Sotos P., 2015, Astronomy and Computing, 11, 64

Sevilla-Noarbe I., et al., 2018, preprint, (`arXiv:1805.02427`)

Shimasaku K., et al., 2001, AJ, 122, 1238

Smith J. A., et al., 2002, AJ, 123, 2121

Soumagnac M. T., et al., 2015, MNRAS, 450, 666

Taniguchi Y., et al., 2007, ApJS, 172, 9

The Dark Energy Survey Collaboration 2005, pp astro–ph/0510346

Tortorelli L., et al., 2018a, preprint, p. arXiv:1805.05340 (`arXiv:1805.05340`)

Tortorelli L., et al., 2018b, preprint, p. arXiv:1805.05340 (`arXiv:1805.05340`)

Werbos P. J., 1982, in Drenick R. F., Kozin F., eds, System Modeling and Optimization. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 762–770

Whitmore B. C., et al., 2016, AJ, 151, 134

APPENDIX: THE ALHAMBRA CATALOG EXTENSION WITH CNN CLASSIFICATION

As part of this work, we provide an additional column for the ALHAMBRA Gold dataset for which we have computed the stellarity value developed in this paper.

As training, we used the Leauthaud et al. (2007) dataset overlapping with ALHAMBRA-4, and we have updated the classification to cover objects up to $F814W < 26.5$. The fields covered are from ALHAMBRA-2 to ALHAMBRA-8, in correspondence to DEEP-2, SDSS, COSMOS, HDF-N, GROTH, ELAIS-N1 and SDSS, respectively. The catalog with this classification is available at `http://cosmohub.pic.es`. In Table 3 we provide the value added catalog columns that are being provided (most inherited from the original Gold catalog, for reference).

We only provide a classification for those objects with all bands measured (20 optical, 3 NIR and F814W). For those without, the class is set to a 'sentinel' value of -1.

The catalog is constructed training and validating on ALHAMBRA-4, in which 'truth' classification is obtained from (Leauthaud et al. 2007), for those objects with the 24 bands measured. The training set counts with 13659 objects, whereas the validation set has 2096. Table 4 shows the number of objects classified per field.

This paper has been typeset from a TEX/LATEX file prepared by the author.

Table 3. Description of the fields shaping the ALHAMBRA catalog where we provide *cnn_stellarity*.

FIELD	**Objects**
ID	ALHAMBRA's objects unique identifier.
RA	Right Ascension in decimal degrees.
DEC	Declination in decimal degrees.
Stellar_Flag	ALHAMBRA's Statistical STAR/GALAXY Discriminator (0:Pure-Galaxy,0.5:Unknown,1:Pure-Star)
F814W	Isophotal magnitude [AB]
dF814W	Isophotal magnitude uncertainty [AB]
cnn_stellarity	CNN star/galaxy discriminator probability: [0:Pure-Galaxy,1:Pure-Star]

Table 4. Number of objects for which we have provided a classification per ALHAMBRA field.

FIELD	**Objects**
ALHAMBRA-2	25856
ALHAMBRA-3	27158
ALHAMBRA-4	14946
ALHAMBRA-5	15276
ALHAMBRA-6	27400
ALHAMBRA-7	26475
ALHAMBRA-8	27813