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

Journal Article

Author(s):
Cabayol, L.; Sevilla-Noarbe, Ignacio; Fernandez, E.; Carretero, Jorge; Eriksen, Martin; Serrano, Santiago; Alarcon, Alex; Amara, Adam; Casas, Ricard; Castander, Francisco J.; De Vicente, Juan; Folger, Martin; García-Bellido, Juan; Gaztañaga, Enrique; Hoekstra, Henk; Miquel, Ramon; Padilla, Cristóbal; Sanchez, E.; Stothert, Lee; Tallada, Pau; Tortorelli, Luca

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]† and L. Tortorelli[5]

[1]*Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, E-08193 Bellaterra (Barcelona), Spain*
[2]*Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), E-28040, Madrid, Spain*
[3]*Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, E-08193 Barcelona, Spain*
[4]*Institut d'Estudis Espacials de Catalunya (IEEC), E-08193 Barcelona, Spain*
[5]*Institute for Particle Physics and Astrophysics, ETH Zürich, Wolfgang-Pauli-Str 27, CH-8093 Zürich, Switzerland*
[6]*Instituto de Fisica Teorica UAM/CSIC, Universidad Autonoma de Madrid, E-28049 Madrid, Spain*
[7]*Leiden Observatory, Leiden University, Leiden, NL-2333CA,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*

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 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, 98.4 per cent purity with a completeness of 98.8 per cent 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: methods: data analysis – techniques: photometric.

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. Soumagnac et al. (2015) to see the impact of this issue in large-scale structure and weak lensing and Sevilla-Noarbe et al. (2018), for the impact in Milky Way studies].

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 of the estimation of an optimal cut on the space of observable image properties, such as a

magnitude-sized 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) and López-Sanjuan et al. (2018).

Other classification methods use Bayesian-based approaches (Sebok 1979; Henrion et al. 2011; Fadely, Hogg & Willman 2012; Rachen 2013). The application of a Bayesian classification approach to multiband data must consider information coming from morphologies and colour: 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 of learning the underlying behaviour 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 (Sevilla-Noarbe & Etayo-Sotos 2015), artificial neural networks (ANN; e.g. Soumagnac et al. 2015), or random forests (RFs; e.g. Morice-Atkinson, Hoyle & Bacon 2017), most of them trained on morphologically based flags from external, deeper data sets 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 limited evidence that this can be improved with colour information (Sevilla-Noarbe et al. 2018), without resorting to infrared data (Banerji et al. 2015; Kovács & Szapudi 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 colour information using Physics of the Accelerated Universe Survey (PAUS; Castander et al. 2012; Martí et al. 2014) and ALHAMBRA (Advanced Large Homogeneous Area Medium Band Redshift Astronomical survey; Moles et al. 2008a, b). 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. SEXTRACTOR provides two star–galaxy classifiers: CLASS_STAR and SPREAD_MODEL. The former relies on a multilayer feed-forward neural network trained with 10 inputs: the object's peak pixel value above the local background, eight isophotal areas and an estimate for the seeing. On the other hand, SPREAD_MODEL indicates whether a model for local PSF or a slightly extended galaxy model best fit the image data. Concerning ALHAMBRA, we will apply our algorithm and compare with the current classification scheme from the Gold catalogue (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 (Serrano et al. in preparation): the fluxes from objects are computed at predefined reference positions from external catalogues, in order to obtain more precise photometric redshifts for these broad-band detections. In the case of COSMOS, the external catalogue is the COSMOS Photometric Redshift Catalogue[1] (Laigle et al. 2016). Therefore, 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 preparation), 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 data sets 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 done in Section 4 and Section 5 provides the results on the ALHAMBRA and PAUS data sets. 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 data sets, PAUS early data and the ALHAMBRA Gold catalogue,[2] in the latter case comparing with the standard classification provided by that survey. We will work on the COSMOS field[3] comparing against the COSMOS space-based imaging catalogue (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),[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, 1.1×10^6 galaxies, most towards the faint end, 30 000 stars and the rest are fake detections.[5]

2.1 PAUS

The PAUS was born in 2008 with the idea of measuring precise redshifts[6] 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 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. 2018), intrinsic alignments of galaxies that can bias cosmological measurements using weak lensing (Abbott et al. 2018) or contributing to the effective modelling of galaxies in image simulations (Tortorelli et al. 2018).

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 to 850 nm and six *ugrizY* broad-band filters (which are not used in this work). The survey covers approximately 0.75 deg^2 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 2018 May, PAUS has been observing approximately 40 nights per year since mid-2015 (with an efficiency below 50 per cent due to bad weather).

---

[1] https://irsa.ipac.caltech.edu/data/COSMOS/gator_docs/cosmos_zphot_mag25_colDescriptions.html

[2] https://cloud.iaa.csic.es/alhambra/

[3] http://cosmos.astro.caltech.edu/

[4] http://www.stsci.edu/hst/acs/documents/handbooks/current/cover.htm

[5] Technically, this classification only separates point-like versus extended objects. QSOs will tend to be mixed with both samples and are neglected in this work, as they are estimated to be ∼3 per cent in the COSMOS field. This can however be a very interesting avenue to explore in the context of CNN narrow-band classification.

[6] ≈0.35 per cent error, meaning a precision of $\sigma(z)/(1 + z) \approx 0.0035$, versus a typical 5 per cent for broad-band measurements

PAUS data are managed by a complex infrastructure which starts at the mountaintop, stores data temporarily there and sends it to the PAUS data centre at the Port d'Informació Científica (PIC) where the nightly and higher level pipelines (Serrano et al. in preparation) are run and data are archived for long-term storage, as well as distributed through a data base for scientific use (Carretero et al. 2017; Tonello et al. 2018).

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)[7] to obtain a spectral energy distribution, 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 preparation, for more information). The photometric error is estimated to be ∼1–2 per cent. It is slightly above 1 per cent for the redder bands, and increasing towards the bluer end.

The PAUS catalogue over this field contains 49 000 astronomical objects, matched to the COSMOS catalogue and with the 40 NB measured: 42 000 galaxies and 7000 stars from magnitudes $I = 16$ to $I = 23$. The flux units of such objects are given in electrons per second.

The training set size for PAUS objects with $I < 22.5$ is 20 000 objects, from which 15 000 are galaxies and 5000 are stars. For validation, the sample size depends on the magnitude range we are testing. For objects with $I < 22.5$, we use 6000 objects, 1000 stars, and 5000 galaxies, decreasing for smaller magnitude ranges. It is worth noting that a small percentage of the training set will include QSOs labelled as 'stars' in our case [around 3 per cent of the total stellar sample in COSMOS, according to the milliQUAS compilation[8] (Flesch 2015)], so a further optimization could be possible by identifying these.

2.2 ALHAMBRA

We have also used the ALHAMBRA photometric redshifts catalogue (Molino et al. 2014) over the ALHAMBRA-4 field, which overlaps with COSMOS. It contains 37 000 objects matched to our reference COSMOS catalogue, from which 34 000 are galaxies and 3000 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 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 *F*814*W* ≈ 26.5.

The catalogue we will work with contains only objects with less than five undetected bands. It contains 29 000 galaxies and 2700 stars. From this catalogue, those with magnitude brighter than 22.5 are 7600: 5900 galaxies and 1700 stars. For these objects, the training size we are using contains 1500 stars and 5000 galaxies. However, as ALHAMBRA goes deeper than PAUS, we could also

train with objects up to magnitude 26. For these objects, the training contains 2300 stars and 15 000 galaxies. In the ALHAMBRA catalogue whenever a source was not detected in a given band, its magnitude was set to a 'sentinel' value of 99.

3 METHODS

Artificial neural networks (ANNs), RFs, and convolutional neural networks (CNNs) have already been used to classify stars and galaxies successfully (as shown, e.g. in Soumagnac et al. 2015; Kim, Brunner & Carrasco Kind 2015; Kim & Brunner 2017), however, as mentioned before, they have never been used solely with photometric measurements of the objects spectra, without additional morphological information. CNNs have also 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 the case of one-dimensional information (Méndez-Jiménez & Cárdenas-Montes 2018), where levels of radon in the environment can very successfully be predicted using CNNs learning from the shape of fluctuations of previous behaviour. 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

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 of 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 of 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.

An ANN consists of a set of layers, an input layer, and an output layer at the beginning and end of the network and a set of hidden layers in the middle, each of them containing a set of trainable weights. The goal of the network is to optimize the set of weights to those that minimize the error in the network prediction.

The network is provided with a loss function that estimates the agreement between the prediction of the network and the truth value. The loss function is evaluated after every iteration on data and the loss value obtained is used to back-propagate the network. Back-propagation is nothing but an optimization technique that modifies the set of weights of the neural network in order to minimize the total loss function. It consists of the loss differentiation with respect to the weights, so that it is estimated how much a change in a given weight affects the total error. These derivatives are the gradients of the loss function with respect to the weights. The gradients are used to update the weights to those that will be used in the following data iteration. Back-propagation starts in the output layer and goes through all the layers of the network until it arrives at the input layer,

---

[7] http://www.stsci.edu/hst/observatory/crds/pickles_atlas.html
[8] http://quasars.org/milliquas.htm

updating all the weights, which is why it is called back-propagation. When the cost function can no longer be minimized by a substantial amount, the weights are saved and the resulting model is used for classification.

An iteration over the full training sample is called epoch. However, it is common practice to divide such training sample in batches and propagate the network over each of these batches instead of using all the training sample simultaneously. This allows the network to update itself more frequently. The size of these batches is called batch_size and it is an important network parameter.

We have used the neural network implementation from the PYTHON scikit-learn package (Pedregosa et al. 2011). The neural network architecture used consists of an input and an output layer with three intermediate hidden layers, with 40 hidden neurons (weights) per layer.

3.1.2 Random forests

An RF (Breiman 2001) is composed of a collection of decision tree predictors, each of them giving an output class when given a set of input features.

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. Each question splits the data as it progresses through the algorithm, forming a hierarchy encoded as a tree. The training set is used to establish the features' hierarchy and the value of the splits in each of the nodes which optimizes the classification. After each iteration over the whole tree, the separation power is evaluated and the splits are selected according to it.

In an RF 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. The term 'with replacement' means that when a given object is sampled for a given decision tree data set, such object is not removed from the complete data set, in such a way that different decision trees may share objects in their data sets. Another difference is in the choice of the question at each node. In the RF 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 an RF implementation provided by the scikit-learn PYTHON package. The architecture of our RF consists of 35 decision trees, with the maximum number of features to consider when looking for the best split equal to the square root of the total number of features and without a predefined maximum depth of each tree, letting the trees expand until each leaf is as pure as possible.

3.1.3 Convolutional neural networks

CNNs (LeCun et al. 1989; Lecun et al. 1998) 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.

One can find different types of layers in a CNN, each of them applying different operations on the data. Particularly, our network is built with three different layers: the convolutional layer, the pooling layer, and the fully connected (or dense) layer. As with ANNs, the input data are propagated through the different layers of the network in batches of a fixed size.

Convolutional layers convolve the input data with a weight matrix, named kernel, which contain the learning parameters. The convolution consists of the multiplication of the kernel by the input map, where the kernel size and the stride between consecutive convolutions are fixed. 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. The layers in an ANN are fully connected, which means that all neurons from a layer are connected to those on the layer below. Conversely, CNNs are locally connected: each neuron only receives input from a small local group of the pixels in the input image, which coincides with the user's defined kernel size. The aim of locally connected layers is to allow for the detection of some subtle nuances of spatial arrangements which are common to the specific spectra we are classifying, independent of their position in wavelength.

Another type of layers are pooling layers. They 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 of 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 pooling matrix size is also a user's defined parameter, as well as the stride between poolings.

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. In fact, these layers are the ANN linear layers. 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.

Fig. 1 shows the architecture of our CNN. The input of the network is a (40 x 1) dimensional array containing the 40 PAUS photometric fluxes. The network contains three convolutional layers with kernel sizes 10, 3, 3, respectively. Such convolutional layers are provided with an activation function, which in our case is a 'LeakyReLu' function (from Keras, see below). The first convolution is larger, so that the algorithm learns about more general features. The following convolutional layers have a smaller kernel to focus on more subtle nuances. Each convolution is followed by a pooling layer. The pooling sizes are 4, 2, 2, respectively. After the last pooling layer, the dimensions of its output are converted into a flat array, an array collapsed into one dimension, in what is called 'Flatten'. The output flattened array is the input of the fully connected layer, which connects directly to the output layer of the network. The fully connected layer and the output layer have dimensions 128 and 2, respectively, corresponding to the dimensions of the last pooling output, 128, and the output of the network, 2 (star and galaxy classes). The final output corresponds to the object's probability of being a star and that 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.

It is also worth mentioning that it is not possible to train the CNN with missing bands: a CNN algorithm cannot be trained if the input contains gaps in some of the bands. There are algorithms that can fill these missing bands with different methods (mean value
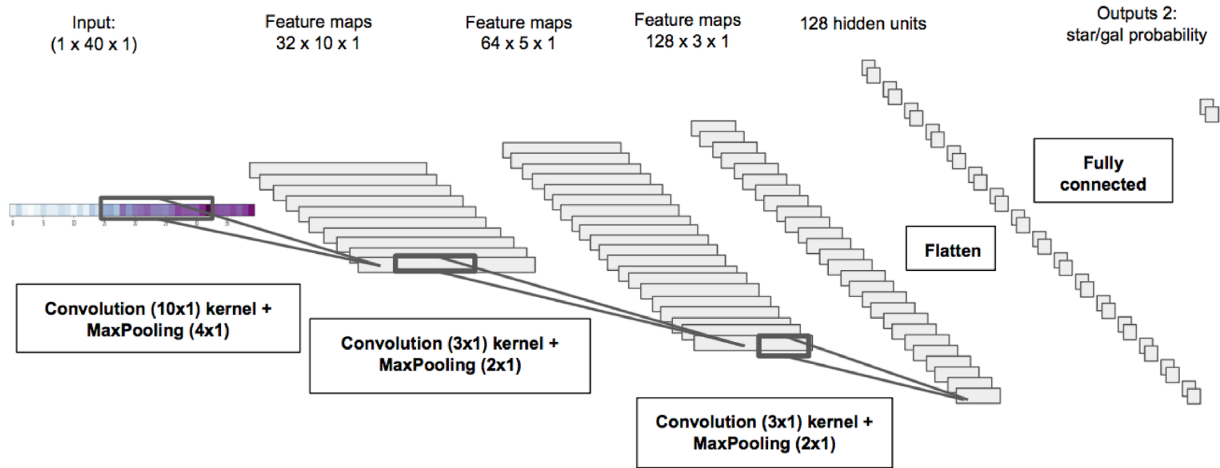
Figure 1. The CNN architecture used for this paper. The input data are a (40x1) dimensional array containing the flux in the 40 PAUS' narrow-band photometric filters. The output is the probability of being a star or a galaxy.

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

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

of the whole input, nearest neighbors, etc.). In case of need, we would fill the gaps with linearly interpolated features based on its contiguous neighbors. In order to have reliable measurements to test our method, we would only keep objects with five or less non-detected bands. We will use this method in the ALHAMBRA section, so that we have a larger data set to train with.

3.2 Analysis

To analyse the performance of the classifiers, we will often refer to *precision* 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 (equation 3).

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

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

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

The performance of the classifiers is generally studied in the ROC space by ROC curves. An 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 object is considered to belong either to a class or another), using the TPR versus 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 true negative rate and the false negative rate, rating the classification/misclassification of stars instead of galaxies.

4 ALGORITHM PERFORMANCE

In this section, we analyse concurrently the performance of the three algorithms defined in Section 3: ANNs, RFs, and CNNs.

We use a training sample over the COSMOS catalogue, 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 catalogue.

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 required.

Fig. 2(a) shows the training size dependence for the neural network, the RF, and the CNN, where the results plotted are those obtained on a distinct validation sample. We only vary the size of the training sample, while maintaining the validation sample size constant.

One can already see here that the CNN is the algorithm yielding the best classification, with a better performance than those of the RF 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, Fig. 2(b) shows that from 10 000 to 30 000 objects the classification is still improving, although the improvement is smaller than from 3000 to 10 000 objects. This
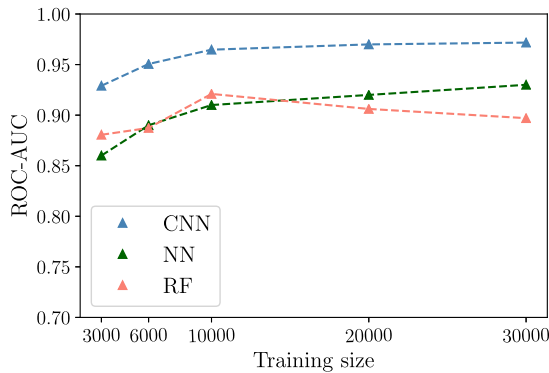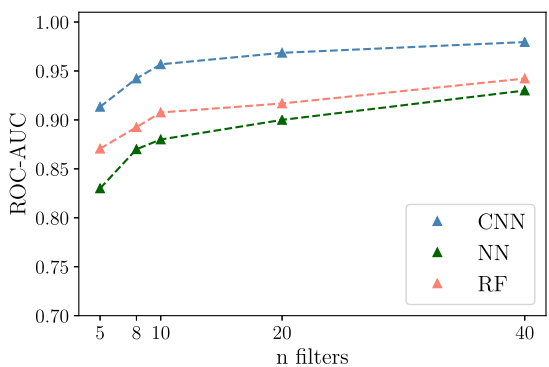
(a) ROC-AUC values for different training sizes.
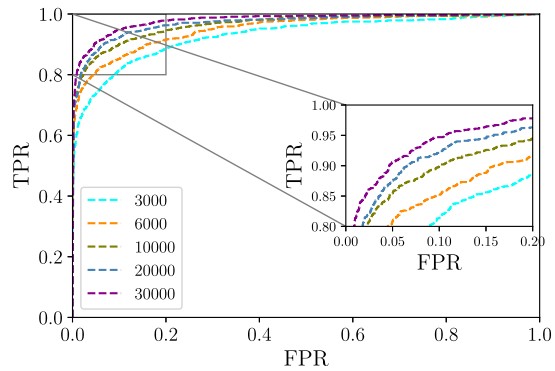
(b) ROC curves corresponding to different number of training objects for the CNN.

Figure 2. (a) ROC-AUC scaling for different training sample sizes for the ANN, RFs, 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.
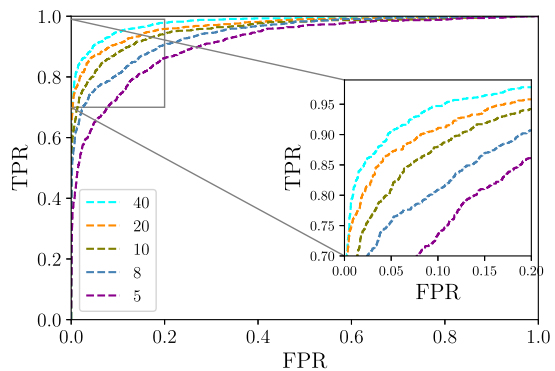
(b) Algorithm performance dependency on input number of bands.

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.

means that the algorithm is more sensitive to training sample size increments when the training data sets 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. Overfitting 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.

Fig. 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.

Fig. 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 in the wavelength range of PAUS, 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
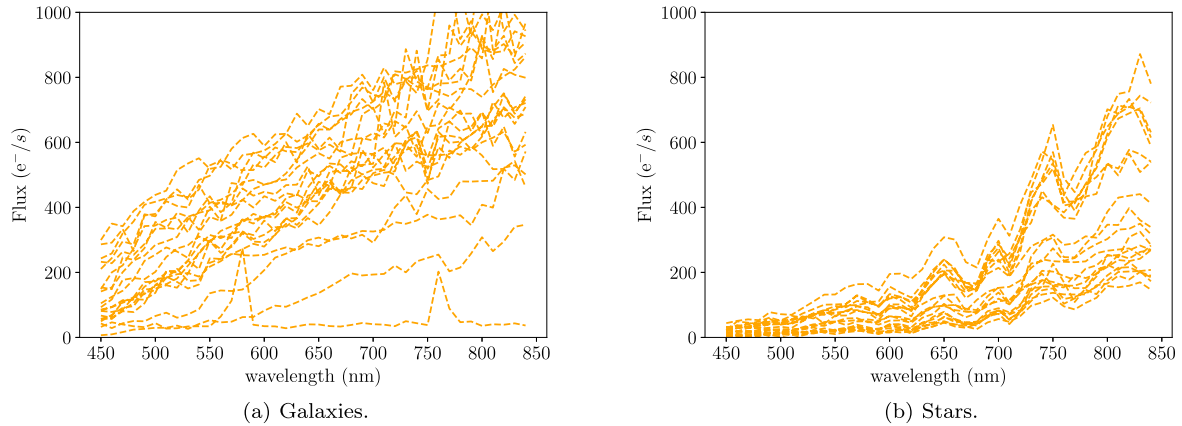
(a) Galaxies.

(b) Stars.

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

redder ones and one would expect that the classification performs worse.

4.3 Algorithm 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 catalogues.

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. Fig. 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 peaks 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 RF 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 data set 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. This result (good performance of CNN on one-dimensional 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 without previous selection (as opposed to what is done in spectroscopic surveys).

5 RESULTS

Previous sections showed that the CNN is the algorithm giving better performances in the star/galaxy classification using PAUS multi narrow-band photometry. In what follows, we will present the results of the application of this algorithm to the PAUS and ALHAMBRA data sets.

5.1 Classification on the PAUS catalogue

As explained in Section 2, the PAUS catalogue in the COSMOS region contains 49 000 objects up to magnitude $I = 23$, from which
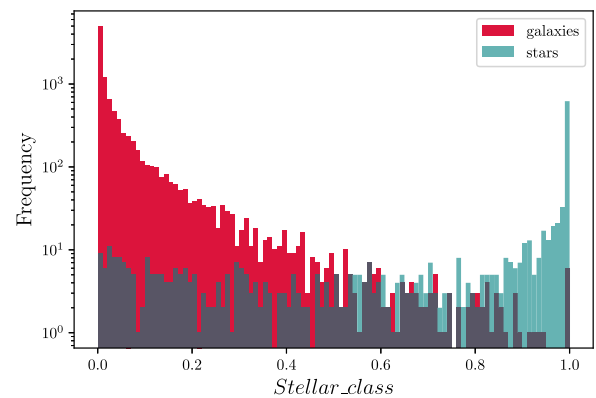

Figure 5. Distribution of *cnn_stellarity* for stars (blue) and galaxies (red), both populations on the validation sample. The darker regions correspond to the overlapping of stars and galaxies.

7000 are stars and 42 000 are galaxies. We will work with a sub-sample of objects with $I < 22.5$, for which the catalogue contains 6000 stars and 28 000 galaxies. As stated in Section 2, 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 5000 stars, whereas the validation sample contains 1000 stars and 5000 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 catalogue 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 catalogue is set to a 'sentinel' value of 99.0. However, in this section we will use the PAUS fluxes as inputs for the algorithm and therefore, the negative counts do not translate into a problem when training the network.

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 98.4 ± 0.1 per cent for a completeness of 98.8 ± 0.1 per cent for objects brighter than $I = 22.5$. This means that the selected galaxy sample still contains a 1.6 per cent of stars contaminating it, while losing a 1.2 per cent of the original true galaxies of the sample.

Analysing in more depth the classification of the PAUS sample, Fig. 5 shows the histogram of such output probability. It exhibits
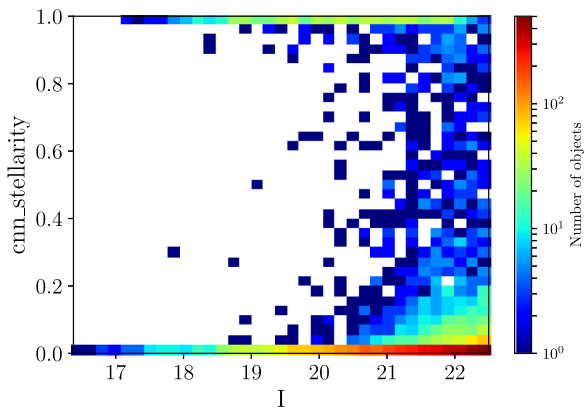
Figure 6. Heatmap for *cnn_stellarity* as as function of *I* magnitude, as measured by the HST-ACS on the COSMOS field.
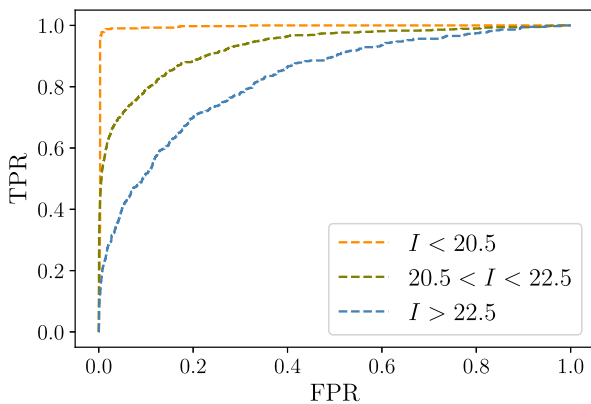

Figure 7. ROC curve for star–galaxy classification on PAUS data given different cuts on the magnitude. The results plotted are those obtained on the validation sample.

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. Fig. 6 shows the same information but as a function of magnitude.

Fig. 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: giving ROC areas of 0.991, 0.930, and 0.822 for ranges $I < 20.5$, $20.5 < I < 22.5$, and $I > 22.5$, respectively. Training sizes have been fixed to 3500 objects for the three cases as the number of objects is limited by the smaller, brightest bin. Considering Fig. 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 data set, 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 Fig. 8, we can see the advantages of using spectral information for classification, versus the standard morphological approach.
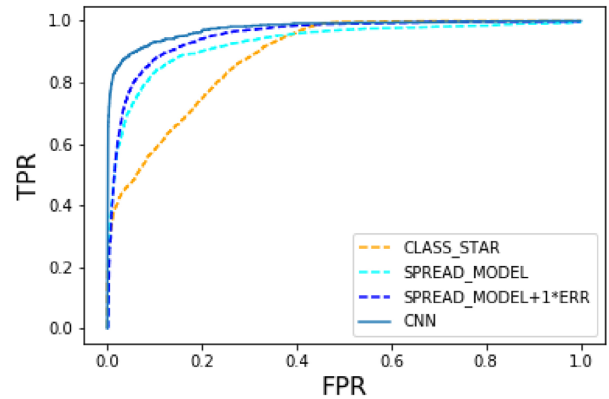

Figure 8. ROC curves for star–galaxy classification in PAUS data using CNN and SEXTRACTOR classifiers. The ERR corresponds to the SPREAD-ERR_MODEL quantity from SEXTRACTOR. Both samples have been selected to have similar signal-to-noise distributions.

5.2 Classification on the ALHAMBRA catalogue

The application of the algorithm on the ALHAMBRA data set should be useful to cross-check 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 width at half-maximum (FWHM) from SEXTRACTOR, a synthetic F814W magnitude, and optical F489W - F814W and near infrared (NIR) J-Ks colours]. The authors derived a probability distribution function 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 catalogues 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 catalogue, 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 catalogue, 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.

To get a classification for the ALHAMBRA catalogue, we are entirely retraining the algorithm using ALHAMBRA's data. Instead of using fluxes, we are using isophotal AB magnitudes, checking the robustness of the classifier 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 colours, the three NIR broad-bands J, H, K magnitudes also included as two colours, the F814W magnitude and the FWHM.

The sample of objects for which ALHAMBRA also provided a classification (hence those with $F814W < 22.5$) represents 20 per cent of the objects, with 5500 galaxies and 1500 stars, in the complete ALHAMBRA Gold catalogue. Taking the COSMOS classification as the 'true' value for classification (admitting some QSO contamination), the ALHAMBRA classification obtains an ROC-AUC of 0.983.

Fig. 9 shows the AUC-ROC of the classification we have performed on objects brighter than 22.5 (blue line) and objects brighter than 26 (orange line). There are different performances, each of them corresponding to the addition of new input features. First, we
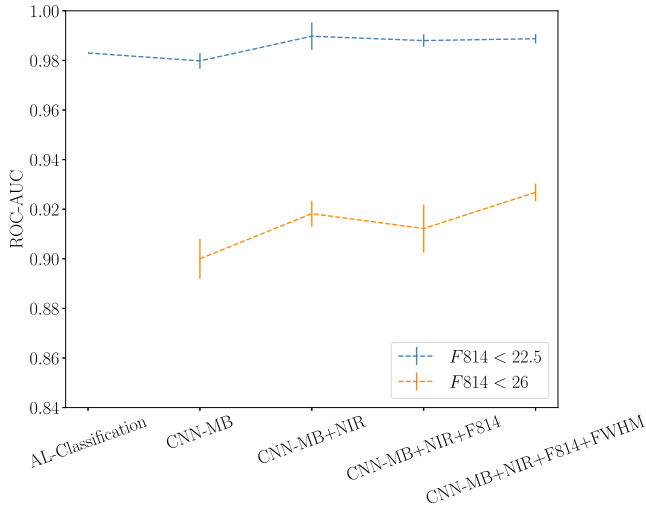
Figure 9. ROC-AUC for the classification of the ALHAMBRA validation sample for different set of input features. In blue, classification for a object's sample with $F814W < 22.5$ in orange, with $F814W < 26$.

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

have run the algorithm with only the optical band information. Then, we have added first the NIR information, then the F814W magnitude, where F814W refers to a synthetic broad-band, 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), and 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. One can also notice that the additional F814W information is not improving the classification. The network's input already contains photometric spectral information with higher resolution than F814W and therefore, the broad-band is not providing any extra information. 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 (Fig. 7), Fig. 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. Fig. 10 illustrates the performance of the algorithm in different binned magnitude ranges and, as expected, for magnitudes fainter than 22.5 the classification performance degrades. Never-
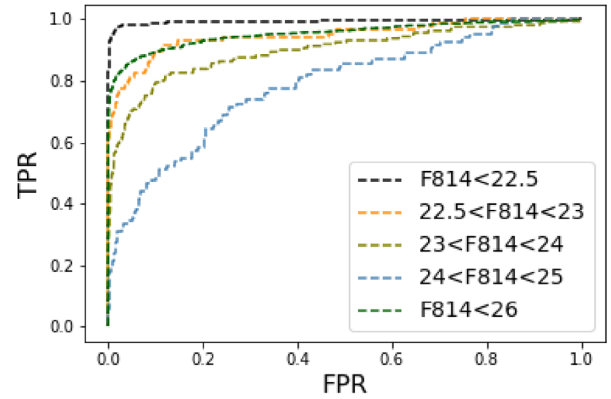
theless, considering all objects brighter than $F814W < 26$, we are able to obtain a 97.4 ± 0.2 per cent purity for a completeness of 98.4 ± 0.1 per cent.

As it was mentioned above, the ALHAMBRA catalogue also provides a classification (Stellar_Flag) for objects brighter than 22.5. For fainter objects, it is a 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.

Considering this approach, for $22.5 < F814W < 23$, the stellar contamination would represent an 8 per cent of the total data set, with a 100 per cent completeness (by definition of the selection). However, with *cnn_stellarity* this contamination can be reduced to only 2.4 per cent for a completeness of 98.4 per cent. Therefore, we significantly improve over this naive classification scheme with minimal loss. The same argument can be applied for fainter bins (keeping the same completeness levels) but with lesser gains in purity as the signal to noise decreases and the stellar sample becomes much smaller in relative terms. For $23 < F814 < 24$, the contamination of stars is 4.8 per cent of the total data set, whereas the algorithm achieves a 3.1 per cent. Finally, within $24 < F814 < 25$, the contamination of the sample would be 2.8 per cent for the naive classifier whereas we obtain 2.4 per cent using *cnn_stellarity*. For fainter objects it is better to consider all sources as galaxies without performing any classification.

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 catalogue used here comes from matching to Hubble Space Catalogue space imaging (Whitmore et al. 2016) making a cut on extendedness of 1.2, which separates cleanly the point-like versus extended sources. The training and validation samples sizes are 20 000 and 5000 objects, respectively, all of them contained in ALHAMBRA-4. For objects with F814W < 22.5, the *cnn_stellarity* gives a ROC area of 0.943, while *Stellar_Flag* is 0.930. We also tested the remaining

Table 2. ROC-AUC values for the classification of stars and galaxies in the ALHAMBRA data set for different input feature maps. ALHAMBRA's Stellar_Flag provides a ROC-AUC of 0.983 up to $F814W < 22.5$.

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

ALHAMBRA fields and obtained and updated classification, see Appendix A.

6 CONCLUSIONS

CNNs have proven to be a real breakthrough in many fields, such as image pattern recognition, image classification, action recognition, or document analysis. 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 per cent and 98 per cent, as shown in Fig. 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 magnitudes and also morphologies by means of a purely statistical method. We have applied our algorithm to their data with the same inputs, leading to a purity and a completeness of 98–99 per cent. Adding the unclassified fainter objects, which contain the bulk of the catalogue up to magnitude 26, leads to a purity of 97 per cent for a completeness of 99 per cent (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 catalogue will be made available upon publication of this work.

The application of CNN on 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 multilabelled 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. With the appropriate band information (blue and infrared) they 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.

REFERENCES

Abbott T. M. C. et al., 2018, Phys. Rev. , 98, 043526
Aparicio Villegas T. et al., 2010, AJ, 139, 1242
Banerji M. et al., 2015, MNRAS, 446, 2523
Bertin E., Arnouts S., 1996, A&AS, 117, 393
Blanton M. R. et al., 2017, AJ, 154, 28
Breiman L., 2001, Mach. Learn., 45, 5
Carretero J. et al., 2017, PoS, EPS-HEP2017, 488
Castander F. J. et al., 2012, Proc. SPIE Conf. Ser. , Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV. SPIE, Bellingham, p. 84466D
Chambers K. C. et al., 2016, preprint (arXiv:1612.05560)
Chollet F. et al., 2015, Keras
Fadely R., Hogg D. W., Willman B., 2012, ApJ, 760, 15
Flesch E. W., 2015, Publ. Astron. Soc. Aust., 32, e010
Henrion M., Mortlock D. J., Hand D. J., Gandy A., 2011, MNRAS, 412, 2286
Hála P., 2014, preprint (arXiv:1412.8341)
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, ApJS, 224, 24
Leauthaud A. et al., 2007, ApJS, 172, 219
LeCun Y., Boser B., Denker J. S., Henderson D., Howard R. E., Hubbard W., Jackel L. D., 1989, Neural Comput., 1, 541
Lecun Y., Bottou L., Bengio Y., Haffner P., 1998, Proc. IEEE, 86, 2278
López-Sanjuan C. et al., 2018, preprint (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
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., 2018, MNRAS, 481, 4194
Méndez-Jiménez I., Cárdenas-Montes M., 2018, in de Cos Juez F. J., Villar J. R., de la Cal E. A., Herrero Á., Quintián H., Sáez J. A., Corchado E.,eds, HAIS'18, Oviedo, Spain, June 20-22, 2018, Proceedings.Springer, p.158
Newman J. A. et al., 2013, ApJS, 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 Evans C. J., Simard L., Takami H., eds, Proc. SPIE Conf. Ser., Vol. 9908, Ground-based and Airborne Instrumentation for Astronomy VI, SPIE, Bellingham, p. 99080Z
Pedregosa F. et al., 2011, J. Mach. Learn. Res., 12, 2825
Pickles A. J., 1998, PASP, 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., AIP Conf. Proc., 1553, Bayesian classification of astronomical objects - and what is behind it. Am. Inst. Phys., New York, p. 254

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

Sevilla-Noarbe I., Etayo-Sotos P., 2015, Astron. Comput., 11, 64

Sevilla-Noarbe I. et al., 2018, MNRAS, 481, 5451

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

Stothert L. et al., 2018, MNRAS, 481, 4221

The Dark Energy Survey Collaboration, 2005, preprint (astro-ph/0510346)

Tonello N. et al. , 2018, preprint (arXiv:1811.02368)

Tortorelli L. et al., 2018, JCAP, 2018, 035T

Werbos P. J., 1982, in Drenick R. F., Kozin F., eds, System Modeling and Optimization, Vol. 38, Springer, Berlin Heidelberg, p. 762

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

APPENDIX A: THE ALHAMBRA CATALOGUE EXTENSION WITH CNN CLASSIFICATION

As a part of this work, we provide an additional column for the AL-HAMBRA Gold data set for which we have computed the stellarity value developed in this paper.

As training, we used the Leauthaud et al. (2007) data set over-lapping 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 catalogue with this classification is available at http://cosmohub.pic.es. In Table A1 we provide the value-added catalogue columns that are being provided (most inherited from the original Gold catalogue, for reference).

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

The catalogue 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 13 659 objects, whereas the validation set has 2096. Table A2 shows the number of objects classified per field.

Table A1. Description of the fields shaping the ALHAMBRA catalogue 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 A2. Number of objects for which we have provided a classification per ALHAMBRA field.

FIELD	Objects
ALHAMBRA-2	25 856
ALHAMBRA-3	27 158
ALHAMBRA-4	14 946
ALHAMBRA-5	15 276
ALHAMBRA-6	27 400
ALHAMBRA-7	26 475
ALHAMBRA-8	27 813

This paper has been typeset from a TeX/LaTeX file prepared by the author.