



HAL
open science

Pushing automated morphological classifications to their limits with the Dark Energy Survey

J. Vega-Ferrero, H. Domínguez Sánchez, M. Bernardi, M. Huertas-Company, R. Morgan, B. Margalef, M. Aguena, S. Allam, J. Annis, S. Avila, et al.

► To cite this version:

J. Vega-Ferrero, H. Domínguez Sánchez, M. Bernardi, M. Huertas-Company, R. Morgan, et al.. Pushing automated morphological classifications to their limits with the Dark Energy Survey. Monthly Notices of the Royal Astronomical Society, 2021, 506 (2), pp.1927-1943. 10.1093/mnras/stab594 . hal-03122292

HAL Id: hal-03122292

<https://hal.science/hal-03122292>

Submitted on 5 May 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Pushing automated morphological classifications to their limits with the Dark Energy Survey

J. Vega-Ferrero^{1,2}★, H. Domínguez Sánchez^{1,3}★, M. Bernardi¹, M. Huertas-Company^{4,5,6,7,8}, R. Morgan⁹, B. Margalef¹, M. Aguena^{10,11}, S. Allam¹², J. Annis¹², S. Avila¹³, D. Bacon¹⁴, E. Bertin^{15,16}, D. Brooks¹⁷, A. Carnero Rosell^{7,18}, M. Carrasco Kind^{19,20}, J. Carretero²¹, A. Choi²², C. Conselice^{23,24}, M. Costanzi^{25,26}, L. N. da Costa^{11,27}, M. E. S. Pereira²⁸, J. De Vicente²⁹, S. Desai³⁰, I. Ferrero³¹, P. Fosalba^{3,32}, J. Frieman^{12,33}, J. García-Bellido¹³, D. Gruen^{34,35,36}, R. A. Gruendl^{19,20}, J. Gschwend^{11,27}, G. Gutierrez¹², W. G. Hartley^{17,37,38}, S. R. Hinton³⁹, D. L. Hollowood⁴⁰, K. Honscheid^{22,41}, B. Hoyle^{42,43,44}, M. Jarvis¹, A. G. Kim⁴⁵, K. Kuehn^{46,47}, N. Kuropatkin¹², M. Lima^{10,11}, M. A. G. Maia^{11,27}, F. Menanteau^{19,20}, R. Miquel^{21,48}, R. L. C. Ogando^{11,27}, A. Palmese^{12,33}, F. Paz-Chinchón^{20,49}, A. A. Plazas⁵⁰, A. K. Romer⁵¹, E. Sanchez²⁹, V. Scarpine¹², M. Schubnell²⁸, S. Serrano^{3,32}, I. Sevilla-Noarbe²⁹, M. Smith⁵², E. Suchyta⁵³, M. E. C. Swanson²⁰, G. Tarle²⁸, F. Tarsitano³⁸, C. To^{34,35,36}, D. L. Tucker¹², T. N. Varga^{43,44} and R. D. Wilkinson⁵¹

Affiliations are listed at the end of the paper

Accepted 2021 February 22. Received 2021 February 5; in original form 2020 October 21

ABSTRACT

We present morphological classifications of ~ 27 million galaxies from the Dark Energy Survey (DES) Data Release 1 (DR1) using a supervised deep learning algorithm. The classification scheme separates: (a) early-type galaxies (ETGs) from late-type galaxies (LTGs); and (b) face-on galaxies from edge-on. Our convolutional neural networks (CNNs) are trained on a small subset of DES objects with previously known classifications. These typically have $m_r \lesssim 17.7$ mag; we model fainter objects to $m_r < 21.5$ mag by simulating what the brighter objects with well-determined classifications would look like if they were at higher redshifts. The CNNs reach 97 per cent accuracy to $m_r < 21.5$ on their training sets, suggesting that they are able to recover features more accurately than the human eye. We then used the trained CNNs to classify the vast majority of the other DES images. The final catalogue comprises five independent CNN predictions for each classification scheme, helping to determine if the CNN predictions are robust or not. We obtain secure classifications for ~ 87 per cent and 73 per cent of the catalogue for the ETG versus LTG and edge-on versus face-on models, respectively. Combining the two classifications (a) and (b) helps to increase the purity of the ETG sample and to identify edge-on lenticular galaxies (as ETGs with high ellipticity). Where a comparison is possible, our classifications correlate very well with Sérsic index (n), ellipticity (ϵ), and spectral type, even for the fainter galaxies. This is the largest multiband catalogue of automated galaxy morphologies to date.

Key words: methods: observational – catalogues – galaxies: structure.

1 INTRODUCTION

Morphology is a fundamental property of a galaxy. It is intimately related to galaxy mass, star formation rate (SFR), stellar kinematics, and environment (e.g. Pozzetti et al. 2010; Wuyts et al. 2011; Huertas-Company et al. 2016). Galaxy structure changes across cosmic time and this mutation is intertwined with formation channels and evolutionary paths. Whether galaxy morphology determines the fate of a galaxy or, conversely, morphological transformations are driven by their stellar population content is still a matter of debate (e.g. Lilly & Carollo 2016). Distinguishing between the two options is

one of the main challenges in understanding galaxy formation and evolution.

Having large samples of galaxies with morphological classifications is crucial for studying the relation between shapes and star formation histories or mass assembly. Traditionally, morphological classification was done by visual inspection. This method has the great inconvenience of being very expensive in terms of time (limiting the available samples to a few thousands – e.g. Nair & Abraham 2010), but it is also affected by the subjectivity of the classifier. An alternative is people-powered research like the Galaxy Zoo project (e.g. Lintott et al. 2008; Willett et al. 2013; Walmsley et al. 2020), where volunteers are asked to classify galaxies. The large number of classifiers significantly reduces the task time and allows for a statistical analysis of the answers. However, these methods

* E-mail: vegaf@sas.upenn.edu (JV-F); dominguez@ice.csic.es (HDS)

still present important biases (see fig. 24 in Fischer, Domínguez Sánchez & Bernardi 2019). More importantly, they will be unable to keep up with the enormous amount of data (millions of galaxy images) that the next generation of surveys such as the Vera Rubin Observatory Legacy Survey of Space and Time or Euclid will deliver: about a hundred years would be needed to classify all data from the Euclid mission with a Galaxy Zoo-like approach. Therefore, applying automated classification methods to such large surveys is mandatory.

An alternative common approach is the quantitative estimation of galaxy morphology. In this methodology, the galaxy light is described in terms of structural quantities (e.g. magnitude, size, ellipticity, asymmetry, concentration, etc.). For Dark Energy Survey (DES), such measurements are available for ~ 50 million galaxies up to $m_i < 23$ mag (Tarsitano et al. 2018, which also provides a comprehensive overview). This technique can rely either on the parametrization of galaxy light profile (e.g. Sérsic function) or on the analysis of the distribution of galaxy light without accounting for the point spread function (PSF). Therefore, it requires specific calibrations.

Recent studies in machine learning and deep learning (DL) techniques, in particular, present an attractive way forward. The use of convolutional neural networks (CNNs) for the analysis of galaxy images has proven to be extremely successful for classifying galaxy images (e.g. Dieleman, Willett & Dambre 2015; Aniyán & Thorat 2017; Tuccillo et al. 2018; Huertas-Company et al. 2018; Domínguez Sánchez et al. 2018; Metcalf et al. 2019; Pasquet et al. 2019; Ntampaka et al. 2019; Hausen & Robertson 2020; Ghosh et al. 2020 – and many others). CNNs have overtaken other supervised automated methods such as Logistic Regression, Support Vector Machines, random forest, decision trees, etc., in terms of both accuracy and computation time (see Cheng et al. 2020 for a detailed comparison), specially for image-based (or array-based) data. However, supervised algorithms rely on large samples of pre-labelled objects on which to train. Complex classification, such as the separation between early-type galaxies (ETGs) and late-type galaxies (LTGs), requires deep CNNs with a large number of free parameters. These training samples should come from the same data domain (e.g. instrument, depth) as the sample to be classified. This ideal is particularly challenging for new surveys where the overlap with available morphological catalogues may be limited. Transfer learning between surveys is an alternative that helps reduce the required training sample size by almost one order of magnitude (see Domínguez Sánchez et al. 2019), but a set of labelled objects with a similar distribution to the main target sample is still needed.

In this context, we aim to provide morphological classifications for galaxies from the DES public data release DR1 (Abbott et al. 2018a). Although the scope of DES is to probe the nature of dark energy, the survey has observed the sky over 6 yr mapping ~ 300 million galaxies in the first 3 yr as a by-product of DR1 – which will become ~ 600 million galaxies for DR2 (Diehl 2019). This data set is one of the largest and deepest among modern galaxy surveys, reaching up to 24 mag in the r -band. Although there are enough DES galaxies in common with previous morphological catalogues (in particular, with Domínguez Sánchez et al. 2018) to properly train CNNs, they are limited to bright magnitudes ($m_r < 17.7$ mag).

To reach the fainter magnitudes that are necessary to probe the redshift evolution of morphological transformations, in Section 2, we use DES galaxies with well-known classifications and simulate what they would look like if they were at higher redshifts. This dramatically reduces the quality of the redshifted images (while keeping track of the original *true* labels). We then check if the CNNs are able to recover features hidden to the human eye. In Section 3, we

use the original and simulated samples to train our CNNs to classify images as ETGs or LTGs, and edge-on or face-on. We then compare our CNN classifications with the corresponding *true* labels of a subsample that was reserved for testing, as well as with the properties of faint DES galaxies from other available catalogues (see Section 5).

This is the largest catalogue of galaxy morphological classification to date (as detailed in Section 4), along with the independent catalogue produced by the companion DES paper presented in Cheng et al., where CNNs are used to classify DES galaxies into elliptical or spirals on the basis of the i -band image. The multiband morphological catalogue presented in this work provides reliable ETG/LTG classifications for 27 million galaxies from the DES survey with $m_r < 21.5$ mag. Our catalogue also includes an edge-on classification, which can be useful for other science analyses (e.g. probing self-interacting dark matter, Secco et al. 2018; estimating dust attenuation, Li et al. 2020; Masters et al. 2010; or studying diffuse ionized gas, Levy et al. 2019; among others).

2 DATA SETS

In this section, we describe the data set used for training and testing, as well as the final sample to which we apply our models in the construction of our catalogue.

2.1 Dark Energy Survey science DR1

The main objective of this work is to provide morphological classifications for a large set of galaxies in the public release of the first 3 yr of DES data¹ (DES DR1, Abbott et al. 2018a). The DES DR1 covers the full DES footprint (about 5000 square degrees) and includes roughly 40 000 exposures taken with the Dark Energy Camera (Flaugher et al. 2015). The co-add images, taken in *griz*-bands, are available along with catalogues of roughly 300 million galaxies, reaching a median co-added catalogue depth of $g = 24.33$, $r = 24.08$, and $i = 23.44$ mag at signal-to-noise ratio $S/N = 10$, with a pixel size of 0.263 arcsec (see Abbott et al. 2018a for technical details).

We selected a high-quality galaxy sample based on a classification that separates PSF-like objects [such as stars and quasi-stellar objects (QSOs)] and extended objects (i.e. galaxies). The classifier is denoted as EXTENDED_CLASS_COADD in the DES data base and it is derived using the *spread_model* quantity from SExtractor photometry (see Abbott et al. 2018b, for more details); its value should be greater than 1 in order to select medium and high confidence galaxies. We also excluded regions of the sky with missing data or bright stars in any of the observed bands by employing the masks described in DES data base, since that could affect our model predictions. We used photometric data in the *gri*-bands and selected galaxies brighter than $m_r = 21.5$ mag, where m_r denotes the magnitude in an elliptical aperture shaped by Kron radius in the r -band (MAG_AUTO_R in the DES data base), and with a half-light radius in the r -band (FLUX_RADIUS_R in the DES data base; denoted as r_r throughout the paper) larger than 2.8 pixels (or 0.74 arcsec; see Section 3.1). This selection produces a final catalogue of 26 971 945 (i.e. nearly 27 million) galaxies. We provide morphological classifications for these galaxies and describe our catalogue in Section 4.

¹DES data base is publicly accessible at: <https://des.ncsa.illinois.edu/desacc/ess/>

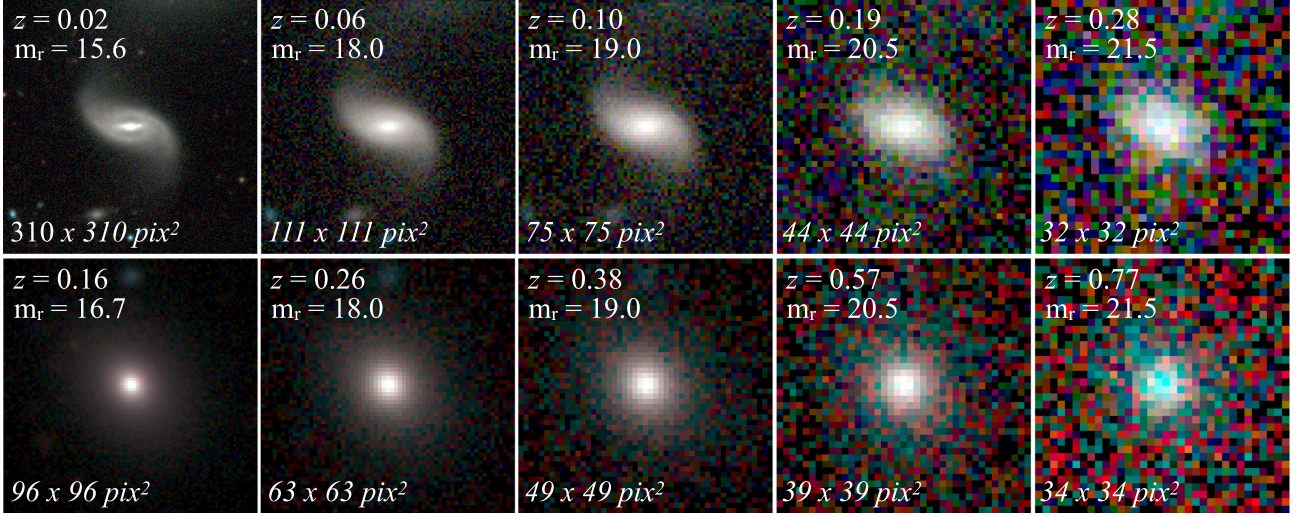


Figure 1. Cutouts of an LTG (upper panels) with T-Type = 5.0 from DS18 observed at $z_0 = 0.02$ with $m_r = 15.6$ mag and for an ETG (lower panels) with T-Type = -2.4 observed at $z_0 = 0.16$ and $m_r = 16.7$ mag. Cutouts from left to right show the original galaxies redshifted to an apparent magnitude of $m_r = (18.0, 19.0, 20.5, 21.5)$ mag, respectively. For each panel, the redshift (z) and the apparent magnitude (m_r) are shown on the upper left corner, while the size of each image (in pixels) is indicated in the lower left corner. The features that distinguish ETG and LTG galaxies become less evident at fainter magnitudes.

2.2 SDSS morphological catalogue

To derive morphological classifications of galaxies within the DES footprint, we have used the morphological catalogue published by Domínguez Sánchez et al. (2018, DS18 hereafter), which partially overlaps with the DES DR1 (see Section 3.2 for details). The DS18 is a publicly available catalogue that provides morphologies for $\sim 670\,000$ galaxies in the Sloan Digital Sky Survey (SDSS) with $m_r < 17.7$ mag. These were obtained by combining accurate existing visual classification catalogues (Nair & Abraham 2010) and Galaxy Zoo 2 (Willett et al. 2013; hereafter GZOO) with DL algorithms using CNNs. The DS18 catalogue provides several GZOO-like classifications, as well as a T-Type (i.e. the numerical Hubble stage, de Vaucouleurs 1959) and a separation between elliptical and S0 galaxies. Although these classifications are automated (i.e. derived without any visual inspection), they are good enough (accuracy > 97 per cent) to provide reliable labels for our training sample.

2.3 Simulating DES galaxies at higher redshifts

Although the number of DES galaxies that are also in the DS18 catalogue is large ($\sim 20\,000$), it is unlikely that a CNN trained on a galaxy sample brighter than $m_r < 18$ mag would accurately classify the vast majority of considerably fainter galaxies in DES. One way to remedy this issue is to visually classify a sample of fainter galaxies. That would have been a tedious task, subject to biases, since different classifiers (i.e. observers) would almost certainly assign different classifications (as can be seen in the GZOO catalogue). In addition, some of the features that distinguish ETGs from LTGs are not so evident for faint distant galaxies (e.g. spiral arms, bar; see Fig. 1), which would complicate the classification. An alternative to visual inspection is to simulate what actual DES galaxies with a well-known classification would look like if they were at higher redshift. We simulate the effects of observing a galaxy at a higher z , given its original DES cutout at z_0 . To do so, we use GALSIM² (Rowe et al.

2015) and assume a Λ CDM cosmology with $\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$, and $h = 0.7$. We perform the following steps:

(i) a de-convolution of the original image by the corresponding PSF in each band. As an approximation for the PSF,³ we assume a Moffat surface brightness profile ($\beta = 2$) with full width at half-maximum (FWHM) equal to the FWHM values for the DES PSF presented in Abbott et al. (2018b), (1.12, 0.96, 0.88) arcsec for the *gri*-bands, respectively;

(ii) a change in angular size due to the cosmological dimming as follows:

$$\frac{s(z)}{s(z_0)} = \frac{D_L(z_0)}{D_L(z)} \frac{(1+z)^2}{(1+z_0)^2}, \quad (1)$$

where s denotes the size of the image, D_L corresponds to the luminosity distance, and z_0 and z are the observed and the simulated redshift, respectively. Note that we keep constant the DES pixel size of 0.263 arcsec and, therefore, it is the size of the image in pixels that changes (it shrinks with increasing z);

(iii) a change in the apparent magnitude. This change includes the cosmological dimming effect, the k -correction, and the evolution of the intrinsic brightness of the galaxies. For the ETGs, we assume that the evolution corresponds to that of a single stellar population model, while for the LTG class, we assume a constant SFR (taking the value from SDSS spectroscopy). In summary, we express the change in apparent magnitude as:

$$m(z) - m(z_0) = \Delta m_{\text{evo}} + 5 \log \left[\frac{D_L(z)}{D_L(z_0)} \right], \quad (2)$$

where $m(z_0)$ and $m(z)$ indicate the observed and redshifted apparent magnitude, respectively, and Δm_{evo} corresponds to the change in magnitude according to the k -correction and the evolutionary models;

(iv) a convolution of the resulting image by the above-mentioned PSF in each band; and

³This approximation has no impact in the performance of the classifications on real DES images, as demonstrated in Section 5.

²<https://github.com/GalSim-developers/GalSim>

(v) the addition of Gaussian noise to the final image. In order to avoid contamination from the central galaxy, we estimate the noise from a set of cutouts with a larger field of view (~ 20 times the half-light radius). We use a robust wavelet-based estimator of the Gaussian noise standard deviation (`estimate_sigma` function available for the `scikit-image`⁴ package in PYTHON).

We apply this procedure to each band independently and then we combine the three bands into an RGB image. We simulate each galaxy satisfying the following conditions: (a) the final apparent magnitude (in the r -band) below $m_r(z) < 22.5$; (b) the final size of the image, $s(z)$, larger than 32×32 pixels; and (c) the final redshift $z < 1.0$. The first condition ensures that the CNN is learning from images of galaxies that are even fainter than the limiting brightness of our project ($m_r < 21.5$) but are still bright enough to pass the DES detection threshold. The second condition avoids extreme interpolations when constructing the input matrix (see Section 3.1). As mentioned above, the pixel size is kept constant, while the size of the image decreases with increasing z . This choice ensures that both the original and the simulated images of the galaxy have the same physical size.

In Fig. 1, we show the original cutout at z_0 along with four simulated cutouts for two galaxies (one LTG and one ETG). For the LTG galaxy (T-Type = 5.0 from DS18), it can be clearly seen how the spiral arms and the bar are almost indistinguishable (by eye) when the galaxy is simulated at a $m_r \geq 20.5$ mag. The original size of the LTG image is 310×310 pixels, while its simulation at $m_r = 21.5$ mag is only 32×32 pixels. For the ETG galaxy (T-Type = -2.4), the original size of the image is 96×96 pixels, while the size of its simulation at $m_r = 21.5$ mag is 34×34 pixels.

3 DEEP LEARNING MORPHOLOGICAL CLASSIFICATION MODEL

We apply DL algorithms using CNNs to morphologically classify galaxy images. DL is a methodology that automatically learns and extracts the most relevant features (or parameters) from raw data for a given classification problem through a set of non-linear transformations. The main advantage of this methodology is that no pre-processing needs to be done: the input to the machine are the raw RGB cutouts for each galaxy (i.e. *gri*-bands, respectively). The main disadvantage is that, given the complexity of extracting and optimizing the features and weights in each layer, a large number of already classified (or labelled) images need to be provided to the machine. As explained in Section 2, we combine a previous morphological catalogue that overlaps with the DES data set with simulations of the original DES images for training and testing how our model performs.

Since we want to apply the morphological classification to the DES sample that covers a much fainter range of objects than the SDSS sample, we limit the morphological classification to a simpler scheme than that presented in DS18: we classify the DES galaxies according to two different schemes: (a) ETGs versus LTGs; and (b) face-on versus edge-on galaxies.

3.1 Network architecture

Given the success of previous studies that have used CNNs for galaxy classification (Huertas-Company et al. 2015; Dieleman et al. 2015; Domínguez Sánchez et al. 2018), we adopt a similar (but not

identical) CNN configuration. Testing the performance of different network architectures is beyond the scope of this paper. We use the same input images and CNN configuration for each classification task. We use the KERAS library,⁵ a high-level neural network application programming interface, written in PYTHON.

The input to the CNN are the RGB cutouts (i.e. *gri*-bands) downloaded from the DES DR1, with a varying size that is function of the half-light radius of the galaxy in the r -band. The cutouts have a size of ~ 11.4 times the half-light radius centred on the target galaxy to guarantee that no galaxy light is missed. The algorithm reads the images that are re-sampled into $(64, 64, 3)$ matrices, with each number representing the flux in a given pixel at a given band. Down-sampling the input matrix is necessary to reduce the computing time and to avoid over-fitting in the models, as commonly done in the literature (Dieleman et al. 2015; Domínguez Sánchez et al. 2018; Walmsley et al. 2020). The flux values are normalized to the maximum value in each filter for each galaxy to eliminate colour information. For the smaller galaxies, the fixed pixel size can lead to cutout sizes that are below the 64×64 pixels for which the CNN has been designed. For these, the cutouts are up-sampled to 64×64 matrices by interpolating between pixels. Since this could create some artefacts and affect the spatial resolution of the images, we require all cutouts to be at least 32×32 pixels in size. This condition leads to a minimum galaxy half-light radius of 2.8 ($32/11.4$) pixels (as mentioned in Section 2.1).

The network architecture, shown in Fig. 2, is composed of four convolutional layers with a linear activation function (denoted as *ReLU*) and squared filters of different sizes (6, 5, 2, and 3, respectively), followed by a fully connected layer. Dropout is performed after each convolutional layer to avoid over-fitting, and a 2×2 max-pooling follows the second and third convolutional layers. The number of weights (i.e. free parameters) in each layer – before dropout – are also indicated. (See Goodfellow, Bengio & Courville 2016, for a comprehensive review on DL concepts).

We train the models in binary classification mode for both classification schemes. For each, the output is a single value between 0 and 1 and can be interpreted as the probability p of being a positive example (labelled as $Y = 1$ in our input matrix) or as the probability $1 - p$ of being a negative example (labelled as $Y = 0$ in our input matrix). We use 50 training epochs, with a batch size of 30 and an *adam* optimization (default *learning rate* of 0.001). In the training process, we perform *data augmentation*, allowing the images to be zoomed in and out (0.75–1.3 times the original size), flipped and shifted both vertically and horizontally (by 5 per cent). This ensures that our model does not suffer from over-fitting since the input is not the same in every training epoch. *Early stopping* is also switched on with a maximum of 10 epochs after convergence is reached. The best model, defined as the optimal one for the validation sample during the training, is then saved and applied to cutouts from the full DES DR1 galaxy catalogue, generated in the same way as for the training sample.

3.2 Training sample

3.2.1 Primary training sample

Our primary training sample is the result of cross-matching the sources in the DS18 and the DES DR1 catalogues presented in Section 2. We identified sources in both catalogues as those with

⁴<https://scikit-image.org/>

⁵<https://keras.io/>

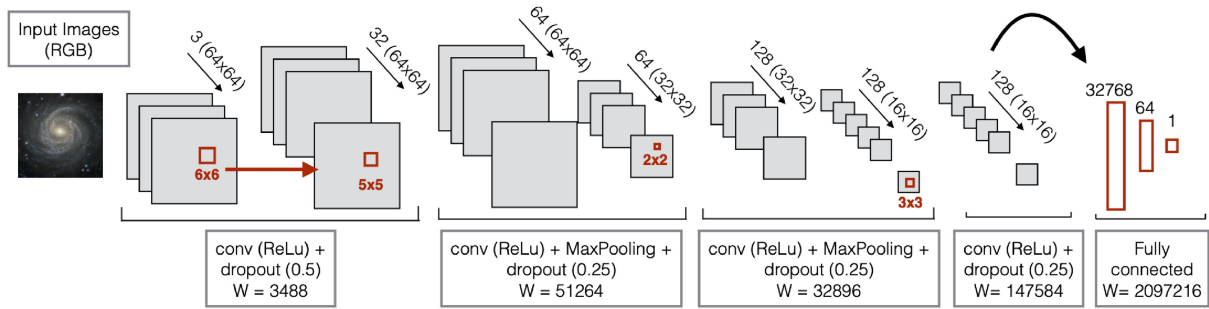


Figure 2. Network architecture used for training the models, consisting of four convolutional layers with a linear activation function (denoted as *ReLU*) and different kernel sizes represented by the red squares, followed by a fully connected layer. The numbers above each layer correspond to the size of the output convoluted images, while the number of weights at each level are indicated below and denoted by *W*.

a separation in the sky of less than 2 arcsec, after removing multiple identifications. We remove those objects missing spectra (or having bad spectroscopy according to SDSS flags) and with relative differences in z of more than 5 per cent between the photo- z for DES Y3 Gold catalogue (De Vicente, Sánchez & Sevilla-Noarbe 2016 and Sevilla-Noarbe et al. 2020) and spec- z for SDSS data. Only 50 galaxies are excluded according to these criteria. The resulting catalogue consists of 19 913 galaxies with good-quality imaging and secure spectroscopic z .

3.2.2 Simulated training sample

The DS18 catalogue (described in Section 2.2) reaches only an observed magnitude of $m_r < 17.7$ mag. Since we aim to push the limits of the morphological classification of galaxies to fainter magnitudes, we extend our primary training sample by simulating each galaxy at higher redshift z and, consequently, making it look fainter and smaller.

Following the pipeline described in Section 2.3, we generate two sets of simulations: (a) one at a random z (hereafter *rnd*) chosen from a uniform distribution between the observed z_0 and the maximum z_{\max} to which the galaxy can be redshifted according to the criteria mentioned in Section 2.3 (i.e. brighter than $m_r(z) < 22.5$; cutout larger than 32×32 pixels and $z < 1$); and (b) a second one at the maximum z (hereafter *max*) under the three given conditions above. By combining the primary training sample and the two sets of simulations, we obtain a more uniform distribution of the apparent magnitude as can be seen in Fig. 3 for the r -band. Note that the primary training set is limited in apparent magnitude to $m_r < 17.7$ mag, while the two sets of simulations extend our training sample to $m_r < 22.5$ mag. As there are indications that close to the limits of the training sample, the CNN results are not as accurate (see e.g. Yan et al. 2020), we extend the training sample one magnitude fainter than the test sample and the final catalogue to avoid such effects. There are almost 6000 galaxies that can be simulated to the maximum apparent magnitude of $m_r < 22.5$ mag.

3.3 ETG versus LTG classification scheme

In this section, we present our CNN predictions for differentiating between ETGs and LTGs. We denote the ETG as the negative class ($Y = 0$), while the LTG is considered as the positive class ($Y = 1$). The T-Type parameter derived by DS18 is a continuous variable ranging from -3 to 10 , where values of T-Type < 0 correspond to ETGs and values of T-Type > 0 designate LTGs. Unfortunately, the quality of the galaxy images, especially at fainter magnitudes, prevents us

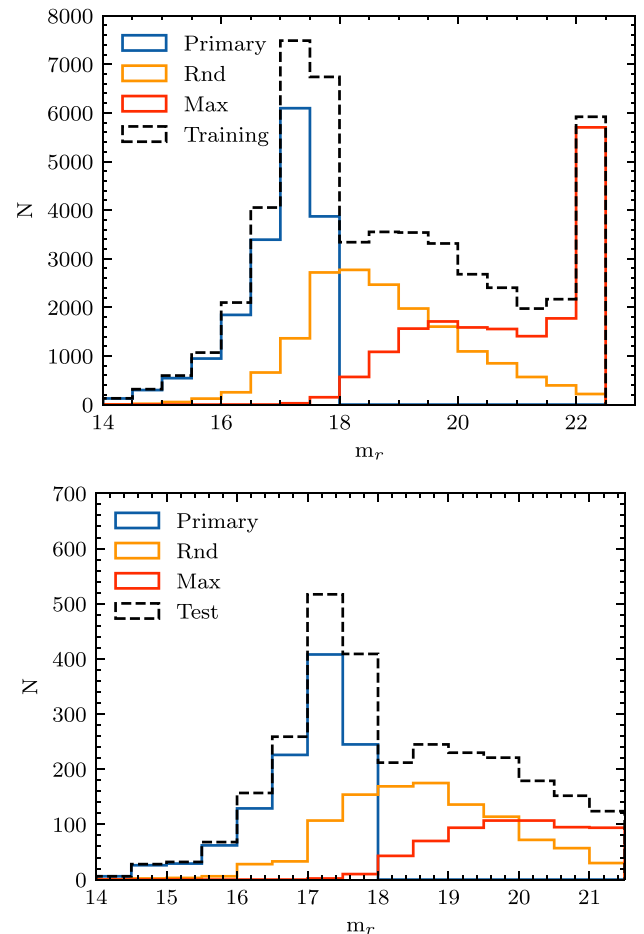


Figure 3. Top panel: Distribution of apparent magnitude in the r -band (m_r) for the primary training sample (solid blue); the two sets of simulations (solid orange and red show the *rnd* and *max* sets); and the training sample distribution (dashed black) used for the ETG versus LTG classification scheme. Bottom panel: Same as top but only for the test sample (see Section 3.3). Note that the CNN predictions are trained with a sample that extends to $m_r < 22.5$, while the model is tested only to $m_r < 21.5$ (the limit of the DES catalogue presented in this work).

from providing such a fine classification for DES galaxies. Separating the sample in two main subclasses (ETGs and LTGs) seems like a reasonable goal for the present catalogue. However, the transition between ETGs and LTGs is smooth and continuous, where intermediate

T-Type values are usually assigned to lenticular galaxies (also known as S0s). Given that this classification is trained in binary mode, we select a galaxy sample of LTGs and ETGs, therefore, not including intermediate T-Types ($-0.5 < \text{T-Type} < 0.5$). According to these criteria, a total of 1293 galaxies (~ 6 per cent) were excluded. Since the DES observations are deeper than the SDSS ones, the classification based on SDSS imaging could differ for some of the DES galaxies. To improve the quality of our training sample, we also excluded 1488 galaxies (~ 7 per cent) with wrong labels in DS18 identified after a visual inspection of the miss-classifications for the predictions of a preliminary model. Then, we re-trained the model without those objects. In summary, our primary training sample consists of 17 132 galaxies with $|\text{T-Type}| > 0.5$ and accurate spec- z and is magnitude-limited with $m_r < 17.7$ mag (as the original DS18 catalogue).

3.3.1 Training

As described in Section 3.2, we use a combination of the primary and the simulated training samples. Nevertheless, from the primary training sample of 17 132 galaxies, we randomly select a subset of 1132 galaxies and their corresponding *rnd* and *max* simulated samples that we never use for training. We denote this subset as the ‘test sample’ and we use it to check the models’ performances. Since none of these galaxies (neither the original nor the simulated ones) have been shown to the CNN, using this subset as a test sample is a secure way to check the results of our classification scheme. Since we want to test our models to $m_r < 21.5$, we only show results for galaxies up to that apparent magnitude threshold. In Fig. 3, we show the apparent magnitude distribution of the test sample to $m_r < 21.5$. The primary test sample consists of 1132 galaxies, and the *rnd* and *max* test samples include 1088 and 623 galaxies, respectively. Therefore, the test sample includes a total 2843 galaxies to $m_r < 21.5$, of which 1557 (55 per cent) are labelled as ETGs and 1286 (45 per cent) are labelled as LTGs.

After removing the galaxies belonging to the test sample, we end up with a training sample of 48 000 galaxy images ($16\,000 \times 3$), with roughly 50 per cent of each class (ETGs and LTGs). We randomly shuffle the galaxies in the training sample and apply a k -fold cross-validation technique (with $k = 5$) for which the original training sample is randomly partitioned into k equal-sized sub-samples. Of the k sub-samples, a single sub-sample is retained as the validation data while training the model, and the remaining $k - 1$ sub-samples (38 400) are used as training data. By doing so, we ensure that each of the five CNN models derived is trained with a different set of images and a different initialization; this provides a (rough) estimate of the classification uncertainty.

3.3.2 Testing

One way to check the reliability of the model predictions made for the ETG versus LTG classification scheme (hereafter interpreted as a predicted probability, p_i) is to compute the difference in the maximum and minimum values for the predicted probabilities of the five models (expressed as Δp). We find that 92.3 per cent of the galaxies from the test sample have $\Delta p < 0.3$, and we designate these as secure classifications. The remaining 7.7 per cent of the galaxies within the test sample have less secure classifications. In Fig. 4, we show the distribution of the predicted probabilities for the test sample along with the distribution of their median probability value (for the full and for the secure classifications). Note that the majority of the insecure classifications are clustered around intermediate values of p .

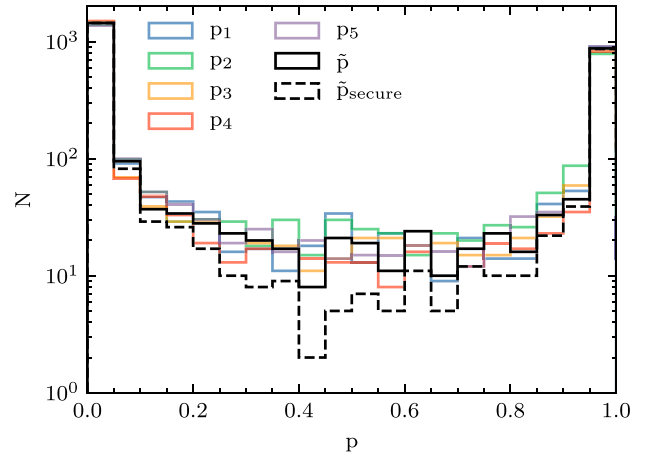


Figure 4. Distribution of the predicted probabilities (p_i) for the test sample i in the ETG versus LTG classification scheme. Black solid histogram corresponds to the distribution of the median probability of the five models (\tilde{p}), while the dashed black histogram shows the distribution of the median probability of the five models only for the secure classifications (i.e. those with a $\Delta p < 0.3$).

As extensively done in literature (see Powers & Ailab 2011, for instance), we also check the accuracy of our models by computing the area under the ROC curve (ROC AUC) for the different predicted probabilities. The ROC curve is a representation of the false-positive rate ($\text{FPR} = \text{FP}/N$, i.e. the ratio of the number of false positives to negative cases) versus the true-positive rate ($\text{TPR} = \text{TP}/P$, i.e. the ratio of the number of true positives to positive cases) for different probability thresholds. A good classifier maximizes TP and minimizes FP values. The top panel of Fig. 5 shows the ROC curves for the five models and for their median value. Good classifiers should be as close as possible to the left hand and upper boundaries of this panel, and this is clearly true for all the curves, which all have ROC AUC values of 0.99.

A complementary way to test the model performance is the precision (Prec) and recall (R) scores (e.g. Dieleman et al. 2015), which can be defined as follows:

$$\text{Prec} = \frac{\text{TP}}{\text{TP} + \text{FP}}; \quad (3)$$

$$\text{R} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \text{TPR}, \quad (4)$$

where the separation between positive and negative samples is determined with respect to a probability threshold. The precision is intuitively the ability of the classifier not to label as positive a sample that is negative (or a purity/contamination indicator). The recall is intuitively the ability of the classifier to find all the positive samples (i.e. a completeness indicator). Additionally, the accuracy of the model prediction is defined as the fraction of correctly classified instances:

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{P} + \text{N}}. \quad (5)$$

We derive the probability threshold (p_{thr}) that optimizes the ROC curve (i.e. maximizes TPR and minimizes FPR), but depending on the user purpose, one can vary the p_{thr} to obtain a more complete or less contaminated sample.

In Table 1, we present a summary of these different estimators for the five independent model predictions (p_i) and for their median value (\tilde{p}). We have already noted that the ROC AUC equals 0.99 for

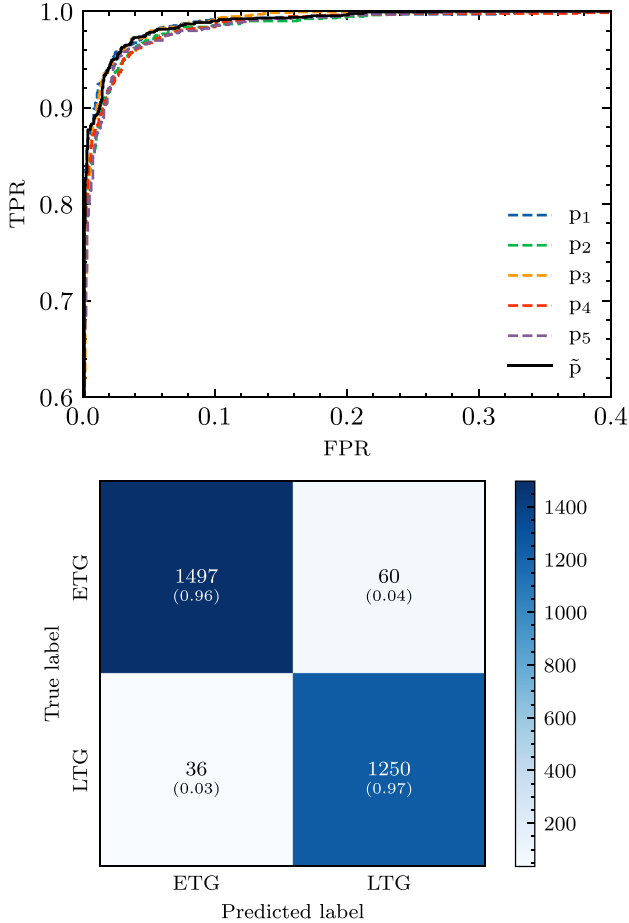


Figure 5. Top panel: ROC curves for the five predicted probabilities (dashed coloured lines). Solid black line corresponds to the ROC curve when the predicted probability is equal to the median value (\bar{p}) of the five predicted probabilities (p_i). Bottom panel: Confusion matrix for the ETG versus LTG classification scheme. In each cell, we show the number of candidates and the fraction of candidates (in brackets) of TN (top left), FN (top right), FP (bottom left), and TP (bottom right).

Table 1. Summary of the ETG versus LTG model performance for the five runs: optimal threshold (p_{thr}), area under the ROC curve, precision, recall, and accuracy values. The last row shows the values obtained for the median probability of the five runs, \bar{p} , which we use throughout the paper as the standard model.

Model	p_{thr}	ROC AUC	Prec	R	Acc
p_1	0.49	0.99	0.97	0.96	0.97
p_2	0.46	0.99	0.95	0.96	0.96
p_3	0.39	0.99	0.96	0.97	0.97
p_4	0.35	0.99	0.95	0.97	0.96
p_5	0.54	0.99	0.96	0.96	0.96
\bar{p}	0.40	0.99	0.95	0.97	0.97

all the cases: Prec, R , and Acc range from 0.95 to 0.97. Besides, if we assume \bar{p} as our fiducial probability, there are only 36 FN and 60 FP in the test sample (3 per cent and 4 per cent, respectively), as shown in the confusion matrix in the bottom panel of Fig. 5, which translates into an Acc = 0.97. If we restrict attention to the subset of secure classifications, the number of FN and FP decreases to 20 and 28,

respectively. This restriction leads to an accuracy classification score for the secure subset of Acc = 0.98. Nevertheless, ~ 80 per cent of the insecure classifications are still valid (FN and FP are $16 + 32 = 48$ out of 219 galaxies that comprise 7.7 per cent of the test sample).

Even for the subset of secure candidates (92 per cent of the test sample), there are some galaxies with intermediate probabilities. We define a *robust* sub-sample of ETGs and LTGs as those with $\max(p_i) < 0.3$ and $\min(p_i) > 0.7$, respectively. Note that *robust* classifications are (by definition) within the secure subset. We find that 1391 galaxies are classified as *robust* ETGs (they are 53 per cent of the secure and a 49 per cent of the whole test sample). On the other hand, 1077 galaxies are *robust* LTGs (41 per cent of the secure and a 38 per cent of the whole test sample). The remaining 6 per cent of the secure sample (156 galaxies) are intermediate (but still secure) candidates. Nevertheless, if we use the median of the probabilities \bar{p} and the optimal threshold of 0.40, most of the galaxies from the secure sample (92 per cent) are still correctly classified. These results demonstrate that the model is able to separate ETGs and LTGs even for the intermediate candidates.

Additionally, we apply our models to the subset of 1293 galaxies (and their simulated counterparts) with $-0.5 < \text{T-Type} < 0.5$ that we did not include in the training of our models. In total, we classified 3879 galaxies with intermediate values of T-Type (3×1293 , the original galaxies and the two simulated counterparts), of which 2004 are ETGs (i.e. $-0.5 < \text{T-Type} < 0.0$) and 1875 are LTGs (i.e. $0.0 < \text{T-Type} < 0.5$). We find that 62 per cent of them are secure classifications (i.e. $\Delta p < 0.3$), significantly lower than the same value for the whole test sample. The number of this subset of galaxies that are classified as ETGs and LTGs is 2026 (52 per cent of the total) and 1853 (48 per cent of the total), respectively. We also find that 1348 (35 per cent of the total) and 874 (23 per cent of the total) galaxies are classified as *robust* ETGs and *robust* LTGs, respectively. Only 177 galaxies (5 per cent of the total) are classified as intermediate but still secure candidates. In terms of accuracy, 75 per cent of these galaxies are correctly classified as ETGs, while 88 per cent of these galaxies are correctly classified as LTGs. Therefore, our classifications are reliable (although more uncertain) even for those objects with intermediate values of T-Type (i.e. $-0.5 < \text{T-Type} < 0.5$) that are a priori difficult to classify.

Note that the fraction of galaxies with intermediate T-Types is very small (6 per cent in the primary training sample). Including these galaxies in the test set (assuming the same fraction as in the primary training sample) would reduce the accuracy from 97 per cent to 96 per cent. The fraction of such objects in the full DES catalogue (Section 4) is unknown and their labels are uncertain (see the large scatter in fig. 11 from DS18). As a result, it is difficult to quantify how they impact the overall accuracy. Therefore, we quote only the final accuracy of the models after such objects have been removed.

One of the key questions we would like to answer is how much are the results of our classification affected by the galaxy brightness. In Fig. 6 and in Table 2, we show how the metrics used to test the model performance change with apparent magnitude. In general, there are very small variations, with the AUC ROC being the most stable parameter (> 0.99 always). The accuracy range is also small ($0.96 < \text{Acc} < 0.98$), while the precision and recall show variations of ~ 5 per cent. There is no clear trend with apparent magnitude, i.e. the models seem to be able to distinguish between ETGs and LTGs regardless of the faintness of the images. We did the same exercise by dividing the test sample in bins of half-light radius, finding accuracy values above 94 per cent, even for the smallest galaxies. Evidently, CNNs detect features hidden to the human eye and therefore classify significantly better than visual inspection.

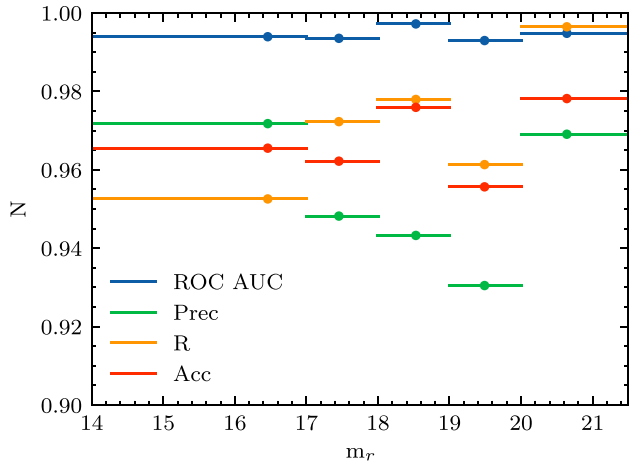


Figure 6. ETG versus LTG model performance (in terms of ROC AUC, precision, accuracy, and recall) as a function of magnitude for the test sample. The values are calculated using the p_{thr} of the median model \tilde{p} from Table 1. The lack of dependence of the metrics with magnitude demonstrates that the model is able to correctly classify even the fainter galaxies.

Table 2. Summary of the ETG versus LTG performance in magnitude bins. The values are calculated using the $p_{\text{thr}} = 0.40$ obtained for the full test sample and the median model \tilde{p} .

Mag bin	ROC AUC	Prec	R	Acc
$14 < m_r < 21.5$	0.99	0.95	0.97	0.97
$14 < m_r < 17$	0.99	0.97	0.95	0.97
$17 < m_r < 18$	0.99	0.95	0.97	0.96
$18 < m_r < 19$	1.00	0.94	0.98	0.98
$19 < m_r < 20$	0.99	0.93	0.96	0.96
$20 < m_r < 21.5$	0.99	0.97	1.00	0.98

3.4 Face-on versus Edge-on classification scheme

In this section, we present our CNN predictions for the second classification scheme to distinguish face-on versus edge-on galaxies. Whereas what we mean by ‘edge-on’ is intuitively obvious, and we treat these as the positive class ($Y = 1$), we use the term ‘face-on’ to refer to the objects that are not edge-on (i.e. the negative class, $Y = 0$), i.e. this classification does not aim to return a continuous output, such as galaxy inclination or ellipticity, but rather to select galaxies that are clearly viewed edge-on.

3.4.1 Training

As for the first classification scheme, our training sample is a combination of the original and the simulated samples. We use the information provided by DS18 on the probability of being edge-on (see Section 2.2) to select a reliable sample of galaxies with which to train our CNNs. We define face-on galaxies as those with $p_{\text{edge-on}} < 0.1$ and edge-on galaxies as those with $p_{\text{edge-on}} > 0.9$, corresponding to 11 783 galaxies. We randomly select 2783 galaxies (and their simulated versions) as the test sample and the remaining 9000 for the training sample. The training sample consists of 27 000 galaxies (3×9000 , the originals and their simulated versions) with 23 424 (87 per cent) face-on galaxies and 3576 (13 per cent) edge-on galaxies. As for the ETG versus LTG model, we train five different models with k -folding. We have reserved a total of 8349 original and simulated galaxies for testing. However, as for the

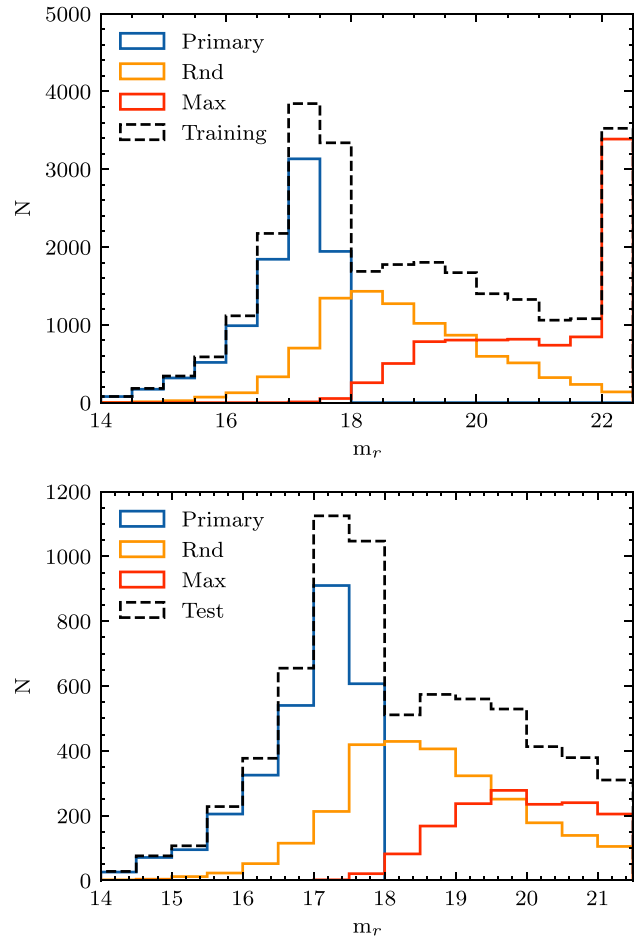


Figure 7. Top panel: Same as Fig. 3 but for the face-on versus edge-on classification scheme. Bottom panel: Same as top but only for the test sample (see Section 3.4). Note that the CNN predictions are trained with a sample that extends up to $m_r < 22.5$, while we test the model predictions only up to $m_r < 21.5$ (the limit of the DES catalogue presented in this work).

first classification scheme (Section 3.3), we show only results for galaxies to $m_r < 21.5$: all the 2783 galaxies within the primary test sample, 2673 galaxies from the *rnd* test sample, and 1477 galaxies from the *max* test sample. Therefore, the test sample includes a total 6933 galaxies, of which 6066 (87 per cent) are face-on and 876 (13 per cent) are edge-on. In Fig. 7, we show the distribution of (m_r) for the different data sets that make up the training and the test samples for the face-on versus edge-on classification scheme.

Since the fractions of face-on and edge-on galaxies are so unequal, we use balanced weights during the training phase of our CNN. In other words, the algorithm compensates for the lack of examples of one class by dividing the loss of each example by the fraction of objects of that particular class.

3.4.2 Testing

As described in Section 3.3.2, we check the accuracy of our model predictions by means of the ROC AUC, Prec, R , and Acc estimators. In Table 3, we show these values for the five face-on versus edge-on models (denoted as p_i^e) and the median one (denoted as \tilde{p}_e). In Fig. 8, we show the distribution of the predicted probabilities for the test

Table 3. Summary of the edge-on versus face-on model performance for the five runs: optimal threshold (p_{thr}), area under the ROC curve, precision, recall, and accuracy values. The last row shows the values obtained for the median probability of the five runs, \tilde{p}_e , which we use throughout the paper as the standard model.

Model	p_{thr}	ROC AUC	Prec	R	Acc
p_1^e	0.26	1.00	0.86	0.97	0.98
p_2^e	0.21	1.00	0.83	0.98	0.97
p_3^e	0.32	1.00	0.92	0.97	0.98
p_4^e	0.35	1.00	0.80	0.98	0.97
p_5^e	0.29	0.99	0.79	0.97	0.96
\tilde{p}_e	0.33	1.00	0.90	0.98	0.98

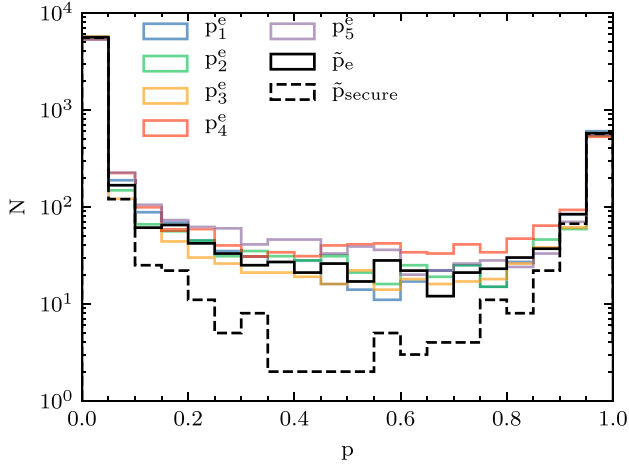


Figure 8. Distribution of the predicted probabilities (p_i^e) for the test sample i in the face-on versus edge-on classification scheme. Black solid histogram corresponds to the distribution of the median probability of the five models (\tilde{p}_e), while the dashed black histogram shows the distribution of the median probability of the five models only for the secure classifications (i.e. those with a $\Delta p_e < 0.3$).

sample along with the distribution of their median probability value (for the full and for the secure classifications). Note that the majority of the insecure classifications are clustered around intermediate values of p_e . The top panel of Fig. 9 shows the ROC curve for the different models while the bottom panel summarizes the results of \tilde{p}_e in a confusion matrix showing the number of TN (5977), FP (89), FN (21), and TP (846) along with their respective fractions within the two classes. The median model \tilde{p}_e model has a better performance than the five individual models with $\text{Acc} = 0.98$ and $R = 0.98$, while $\text{Prec} = 0.90$ is slightly smaller. This is in part due to the unbalanced test sample: the total number of FP is about 1/10 of the TP, although the FP are only ~ 1 –2 per cent of the face-on galaxies. On the other hand, the number of FP (only 89, or 1 per cent of the predicted edge-on galaxies) is considerably lower than the number of TP, which translates into an excellent R value.

Analogously to the ETG versus LTG model, we define a *secure* sub-sample of galaxies for the edge-on classification where $\Delta p_e < 0.3$. There are 93 per cent of *secure* galaxies in the test sample. The *robust* edge-on are galaxies with $\min(p_i^e) > 0.7$. We find 668 galaxies classified as *robust* edge-on (10 per cent of the secure sample and 12 per cent of the whole test sample).

The dependence of the edge-on classification with apparent magnitude is highlighted in Fig. 10, which plots the performance of the \tilde{p}_e model in the same magnitude bins (summarized in Table 4). There

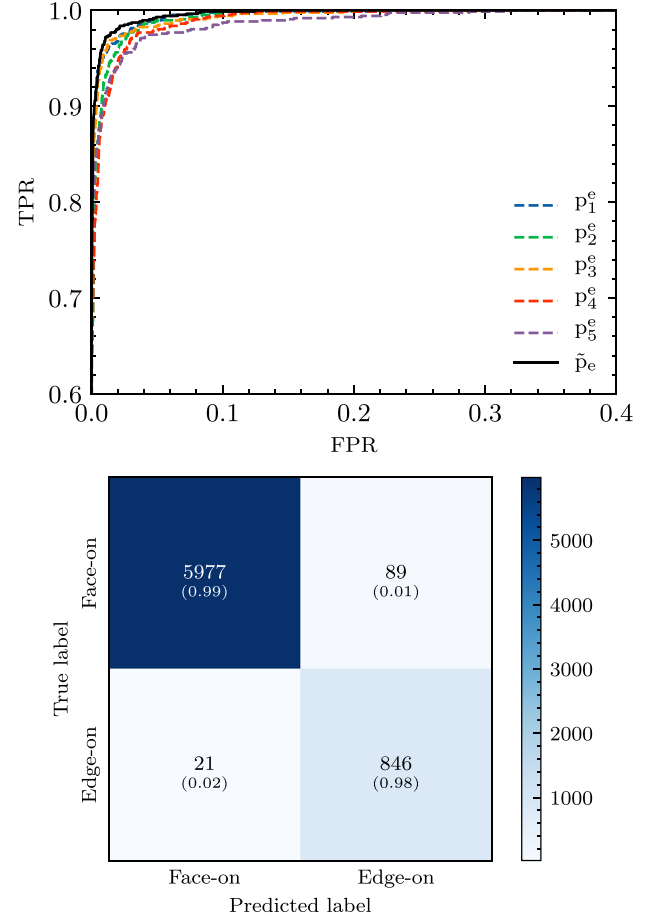


Figure 9. Same as Fig. 5 but for the face-on versus edge-on classification. Note the better performance of the average model \tilde{p}_e .

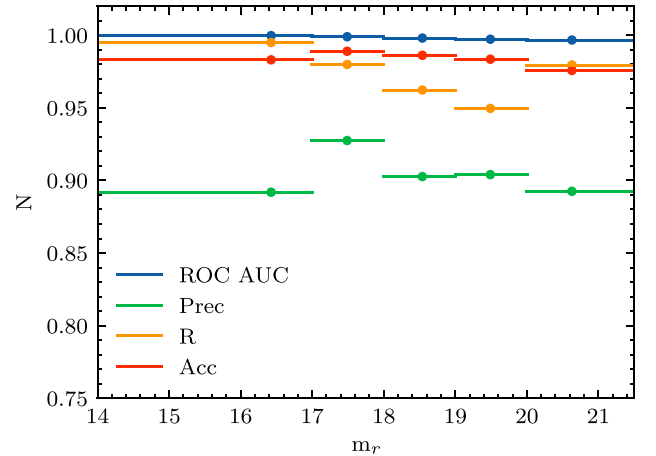


Figure 10. Same as Fig. 6 but for the edge-on model performance. The lack of dependence of the metrics with magnitude demonstrates that the model is able to correctly classify even the fainter galaxies.

is a very small variation with apparent magnitude: the most affected quantity decreases from 0.99 at $14.0 < m_r < 17.0$ to 0.95 at $19.0 < m_r < 20.0$. In the same table, we show the values obtained for a balanced test sample, robust against class representation. In this

Table 4. Summary of the edge-on model performance in magnitude bins. The values are calculated using the $p_{\text{thr}} = 0.33$ obtained for the full test sample and the median model \tilde{p}_e (in brackets for a balanced test sample).

Mag bin	ROC AUC	Prec	R	Acc
$14 < m_r < 21.5$	1.00 (1.00)	0.90 (0.98)	0.98 (0.98)	0.98 (0.98)
$14 < m_r < 17$	1.00 (1.00)	0.89 (0.97)	0.99 (0.99)	0.98 (0.98)
$17 < m_r < 18$	1.00 (1.00)	0.93 (1.00)	0.98 (0.98)	0.99 (0.99)
$18 < m_r < 19$	1.00 (1.00)	0.90 (0.99)	0.96 (0.96)	0.99 (0.98)
$19 < m_r < 20$	1.00 (1.00)	0.90 (0.97)	0.95 (0.95)	0.98 (0.96)
$20 < m_r < 21.5$	1.00 (1.00)	0.89 (0.99)	0.98 (0.98)	0.98 (0.98)

case, the precision values are significantly improved (from $\text{Prec} = 0.90$ to $\text{Prec} = 0.98$ for the full test sample) while the other indicators are almost unchanged. We did the same exercise by dividing the test sample in bins of half-light radius, finding accuracy values above 96 per cent, even for the smallest galaxies.

The face-on versus edge-on classification is useful for different scientific purposes (see Section 1) but might also help as an additional test for the ETG versus LTG classification presented in Section 3.3. Since only discs can be seen edge-on, a galaxy should not be classified simultaneously as an ETG and edge-on. We find 91 (predicted) ETGs in the ETG versus LTG test sample with $\tilde{p}_e > 0.33$, corresponding to ~ 3 per cent of the test sample. This fraction is reduced to 0.7 per cent when only *robust* ETG and *robust* edge-on are considered. This small fraction reassures us about the performance of the two models. A visual inspection of these galaxies confirms that most of them look like edge-on lenticulars, with a clear bulge and disc but no signs of spiral arms. This is especially evident for *robust* edge-on ETGs (see Fig. A1). Thus, including the additional information provided by the edge-on versus face-on classification helps to increase the purity of the ETG sample and is an efficient way to identify edge-on lenticulars.

4 DES DR1 MORPHOLOGICAL CATALOGUE

In this section, we present the results of applying the classification schemes described in Sections 3.3 and 3.4 to the DES DR1 galaxy catalogue presented in Section 2.1. We briefly summarize the overall results here but address a more exhaustive comparison with other observed galaxy properties in Section 5. Table 5 summarizes the statistics for the full DES morphological catalogue, as well as for three comparison samples, while the magnitude distribution of each sub-sample is shown in Fig. 11. In Table 6, we describe the content of the full DES DR1 morphological catalogue, which will be released along with the paper. Examples of each class at different magnitudes are shown in Appendix A.

For the ETG versus LTG classification scheme, 87 per cent of the 26 971 945 galaxies in the DES DR1 morphological catalogue are secure classifications, i.e. $\Delta p < 0.3$ (where Δp corresponds to the maximum difference between the five predicted probabilities). Within this subset of secure classifications, 10 per cent of the galaxies are classified as *robust* ETGs (i.e. $\max(p_i) < 0.3$), while 85 per cent are classified as *robust* LTGs (i.e. $\min(p_i) > 0.7$). The remaining 5 per cent of the galaxies may be considered as intermediate (but still secure) candidates. Being less conservative, ~ 12 per cent of the galaxies from the subset of secure classifications are classified as ETGs (i.e. $\tilde{p} < p_{\text{thr}} = 0.4$) while, consequently, ~ 88 per cent are classified as LTGs (i.e. $\tilde{p} > p_{\text{thr}} = 0.4$). The much larger fraction of LTGs with respect to ETGs in a magnitude-limited sample is consistent with previous work (see e.g. Pozzetti et al. 2010).

Fig. 12 shows how the galaxies of the whole DES DR1 morphological catalogue populate the apparent magnitude (m_r) and photometric redshift (z_{photo}) plane, colour coded by density, secure fraction, and predicted LTG fraction. The predicted LTG fraction is computed as the average of the predicted labels (i.e. 0 for ETGs, 1 for LTGs) of the galaxies in each bin. As expected, the brightest galaxies at low z_{photo} are dominated by ETGs, while the faint galaxies are predominantly LTGs. The fraction of secure classified galaxies is relatively constant for the (observed) bright galaxies and there is an interesting trend with redshift for galaxies fainter than $m_r > 19$: the fraction of insecure galaxies increases with z_{photo} , as expected. In any case, note that the average fraction of secure galaxies is 87 per cent; it remains greater than 50 per cent even in the more uncertain regions of the m_r - z_{photo} plane.

Although most of the faintest (observed) galaxies are classified as LTGs, the classification model is able to retrieve a significant fraction of ETGs (~ 50 per cent) at intermediate $z_{\text{photo}} \sim 0.5$ and $m_r \gtrsim 20.0$ mag. Note that the faint, low redshift population corresponds to intrinsically faint (and therefore low mass) galaxies, that are, in general, LTGs. Unfortunately, there are no additional parameters with which to further test the full DES DR1 catalogue. We cross-correlate this sample with other available measurements in the following sections.

For the edge-on versus face-on classification scheme, 18 per cent of the galaxies have values of $\tilde{p}_e > 0.33$ (9 per cent if the limit is $\min(p_i^e) > 0.70$). The fraction of *robust* ETGs with $\tilde{p}_e > 0.33$ is less than ~ 3 per cent (0.3 per cent if $\min(p_i^e) > 0.70$). This small fraction is reassuring since, as explained in Section 3.4.2, edge-on galaxies should only be discs (and therefore LTGs). We strongly recommend that users combine the two classifications since many of these galaxies could actually be miss-classified LTGs or edge-on lenticulars. Some examples are shown in Fig. A1.

5 VALIDATION OF THE CLASSIFICATION ON REAL DES GALAXIES

Although the results presented in Sections 3.3.2 and 3.4.2 show that the CNNs perform well, it may be argued that the tests were done on a similar set of simulated images as the ones used to train the CNNs. To further test the goodness of the morphological classification on real DES DR1 galaxy images, we now present a comparison with other available data (both photometric and spectroscopic).

5.1 DES DR1 stellar mass catalogue

The DES DR1 stellar mass catalogue is the result of running the LePhare code (Arnouts & Ilbert 2011) on the DES DR1 galaxy catalogue using Bruzual & Charlot (2003) templates, three different metallicities (including solar), Chabrier Initial Mass Function (IMF), and exponentially declining star formation histories (similarly to Palmese et al. 2020). The redshift of each galaxy is assumed to be equal to the mean photo- z mean statistic obtained from multiobject fitting photometry (Drlica-Wagner et al. 2018). The resulting catalogue contains estimates of the stellar mass and the absolute magnitude for ~ 184 million galaxies.

To select galaxies in the DES DR1 stellar mass catalogue for which the stellar mass and absolute magnitude estimates are reliable (given the large uncertainties associated with the z_{photo}), we cross-match the above-mentioned catalogue with a spectroscopic compilation from

Table 5. Comparison of the test samples discussed in Section 5. Columns show the total number of galaxies and the corresponding fraction of secure ones ($\Delta p < 0.3$). Also given are the fractions of the secure galaxies classified as (*robust*) ETGs, LTGs, and edge-on. The last column contains the fraction of *robust* ETGs that are also classified as *robust* edge-on. These cases should be taken with care since only discs should be edge-on.

Sample	# galaxies	Secure ($\Delta p < 0.3$) % from total	ETGs (<i>robust</i>) % from secure	LTGs (<i>robust</i>) % from secure	Secure ($\Delta p_c < 0.3$) % from total	Edge-on (<i>robust</i>) % from secure	ETGs + Edge-on % from <i>robust</i> ETGs
DES DR1	26971945	87	12 (10)	88 (85)	73	9 (6)	0.3
DES struct. param.	6060018	89	9 (7)	91 (88)	79	7 (6)	0.3
DES stellar mass	137956	83	48 (44)	52 (47)	86	7 (6)	0.5
VIPERS	7384	81	22 (20)	78 (76)	77	2 (2)	0.1

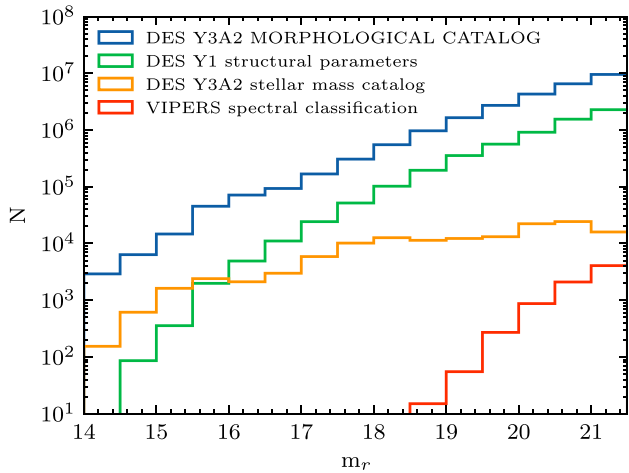


Figure 11. Normalized apparent magnitude distribution for the full DES morphological catalogue presented in this work, as well as for the three catalogues used for comparison.

several surveys⁶ and select galaxies with $(|z_{\text{photo}} - z_{\text{spec}}|)/(1 + z_{\text{spec}}) < 0.05$. The sample used for the comparison with our work consists of 137 956 galaxies covering a redshift range of $0 < z < 1$ and the magnitude shown in Fig. 11. The summary of the statistics is shown in Table 5.

For this sub-sample, 86 per cent of the galaxies show secure classifications (i.e. $\Delta p < 0.30$). The fraction of *robust* ETGs and LTGs is 44 and 47 per cent, i.e. this is a much more balanced sample compared to the full DES DR1 (and also to the other sub-samples shown in Table 5). We note that the magnitude distribution of this subset of galaxies is relatively flat, meaning that a large fraction of the faint LTGs may be missing. Regarding the second classification, 7 per cent of the galaxies are edge-on and less than 0.5 per cent of the *robust* ETGs are classified as *robust* edge-on.

In Fig. 13, we show how the galaxies populate the absolute magnitude – redshift plane (M_r and z_{spec} , respectively). There is a clear separation between the ETG and LTG populations, with the (intrinsically) brightest galaxies at each redshift dominated by the ETGs, as expected. The lack of ETGs at the lowest redshifts ($z \lesssim 0.2$) is due to the scarcity of massive ETGs in such a small volume.

5.2 DES Y1 structural parameters

The DES Y1 structural and morphological catalogue presented in Tarsitano et al. (2018) consists of ~ 50 million objects selected from the first year of the DES. For a comparison with our predicted

morphologies, we use the single Sérsic index (n_r) and the ellipticity (ϵ_r) obtained with GALFIT for the r -band. Following appendix B3.2 of Tarsitano et al. (2018), we extract a clean sample of validated and calibrated objects by applying the recommended cuts $\text{FIT_STATUS_R} = 1$ and $\text{SN_R} > 10$ in the r -band. We also select objects with realistic values for the Sérsic index and ellipticity within $0 < n_r < 10$ and $0 < \epsilon_r < 1$, respectively. These criteria are fulfilled by a 54 per cent of the objects in the catalogue. Then, we cross-match the resulting catalogue with our DES DR1 catalogue to $m_r < 21.5$ mag and excluded (~ 600) objects with unreliable redshifts (i.e. $z_{\text{photo}} < 0.0$). Finally, we construct a catalogue for comparison with 6 060 018 (~ 12 per cent of the original catalogue) galaxies for which accurate n_r , ϵ_r and apparent magnitudes are available. The magnitude distribution is shown in Fig. 11 and the median of the apparent magnitude of the selected sub-sample is $\tilde{m}_r = 20.8$ mag.

As detailed in Table 5, 89 per cent of the galaxies in this subset show secure ETG/LTG classifications (i.e. $\Delta p < 0.30$). While the fraction of edge-on, 7 per cent, is very similar to the other sub-samples, the fraction of *robust* ETGs and LTGs (7 and 88 per cent, respectively) is very uneven. The r -band magnitude distribution is similar to the DES DR1 morphological catalogue, although missing some galaxies at the very bright end), which can explain the larger fraction of LTG for this subset.

We check the reliability of this sub-sample by using the structural parameters derived by Tarsitano et al. (2018). It is well known that the Sérsic index correlates well with galaxy morphologies: large n_r is a good proxy for ETGs and vice versa for LTGs (e.g. Fischer et al. 2019). On the other hand, edge-on galaxies should have large ellipticity values, ϵ_r . In Fig. 14, we show how this sub-sample populates the $n_r - \epsilon_r$ and $r_r - \epsilon_r$ planes, as well as cumulative distribution functions (CDFs) for the Sérsic index and the ellipticity for the two classifications schemes.

For the ETG versus LTG classification, we find an evident separation of each class around $n_r \sim 2$ and almost no ETGs with $\epsilon_r < 0.5$, as expected. Although the fraction of galaxies with high Sérsic index classified as LTGs is ~ 30 per cent, we note that this is due to the much larger fraction of LTGs in this sub-sample. According to the CDF, 88 per cent of the *robust* ETGs have $n_r > 2$, while 87 per cent of the *robust* LTGs have $n_r < 2$. Although the transition between ETGs and LTGs is not exactly at $n_r = 2$, the very different distribution of the ETGs and LTG samples is an indicator of the accuracy of our model predictions in real DES galaxy images. It is also interesting to note that galaxies not classified within the previous two classes, i.e. the secure intermediate candidates, show a CDF that places them in between the CDFs for the ETGs and the LTGs.

For the face-on versus edge-on classification scheme, we find an even sharper separation at $\epsilon_r \sim 0.5$ at all radius, except for the smallest galaxies ($r_r \lesssim 1.0$ arcsec), which indicates that in those

⁶J. Gschwend, private communication (see also Gschwend et al. 2018).

Table 6. Content of the full DES DR1 morphological catalogue.

COADD_OBJECT_ID	Unique object ID for Y3 co-add processing
RA	Right ascension (J2000.0 in degrees)
DEC	Declination (J2000.0 in degrees)
MAG_AUTO_R	Apparent magnitude in an elliptical aperture shaped by the Kron radius (m_r throughout the paper)
FLUX_RADIUS_R	Radius (in pixels) of the circle containing half of the flux of the object (r_r throughout the paper)
P_i _LTG	Probability of being LTG for each of the five models (with $i = [1, 5]$)
MP_LTG	Median probability of the five models of being LTG
P_i _EdgeOn	Probability of being edge-on for each of the five models (with $i = [1, 5]$)
MP_EdgeOn	Median probability of the five models of being edge-on
FLAG_LTG	Classification for ETG versus LTG model; 0 = ETG, 2 = secure ETG, 4 = robust ETG; 1 = LTG, 3 = secure LTG, 5 = robust LTG
FLAG_EdgeOn	Classification for edge-on model; 0 = no edge-on, 1 = edge-on, 2 = secure edge-on, 3 = robust edge-on

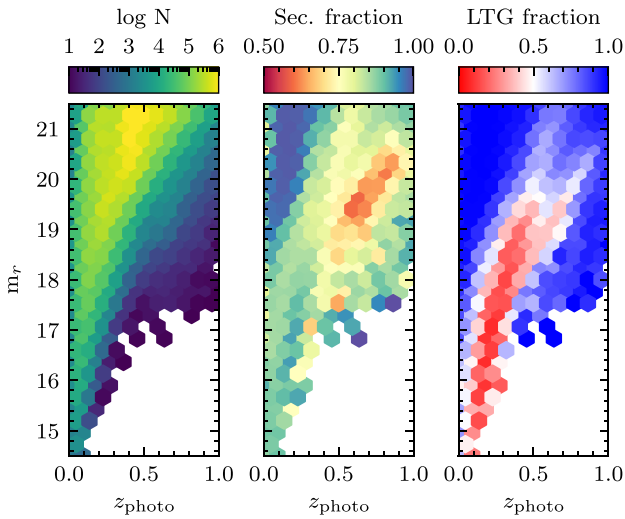


Figure 12. Apparent magnitude (m_r) versus photometric redshift (z_{photo}) for the secure subset ($\Delta_p < 0.3$) within the DES DR1 morphological catalogue. Left-hand panel shows the number of galaxies (in log scale) in each hexagonal bin. Middle panel shows the fraction of secure galaxies. Right-hand panel indicates the predicted ETG/LTG fraction. A predicted LTG fraction of 1 means that 100 per cent of the galaxies in a particular hexagonal bin are LTGs, while a predicted LTG fraction of 0 indicates that 100 per cent of the objects in the bin are ETGs. The brightest galaxies at low z_{photo} are dominated by ETGs, while the faint galaxies are predominantly LTGs. The fraction of secure classified galaxies is relatively constant for the (observed) bright galaxies and the fraction of insecure galaxies increases with z_{photo} . The average fraction of secure galaxies is 87 per cent and greater than 50 per cent even in the more uncertain regions of the m_r - z_{photo} plane.

cases the spatial resolution is not enough for identifying edge-on galaxies. Regarding the CDF, 87 per cent of the *robust* face-on galaxies ($\max(p_i^e) < 0.30$) have $\epsilon_r < 0.5$, while ~ 100 per cent of the *robust* edge-on ($\min(p_i^e) > 0.70$) have $\epsilon_r > 0.5$, thus allowing us to be confident about our model predictions (Fig. 14). The fact that only 0.3 per cent of the *robust* ETGs are classified as *robust* edge-on is also a good sanity check.

5.3 VIPERS spectral classification

In this section, we compare the predictions made by our ETG versus LTG classification with an unsupervised machine-learning classification extracted from the VIMOS Public Extragalactic Redshift Survey (VIPERS, Guzzo et al. 2014) presented in Siudek et al. (2018). The data release provides spectroscopic measurements and photometric properties for 86 775 galaxies. The galaxy classification

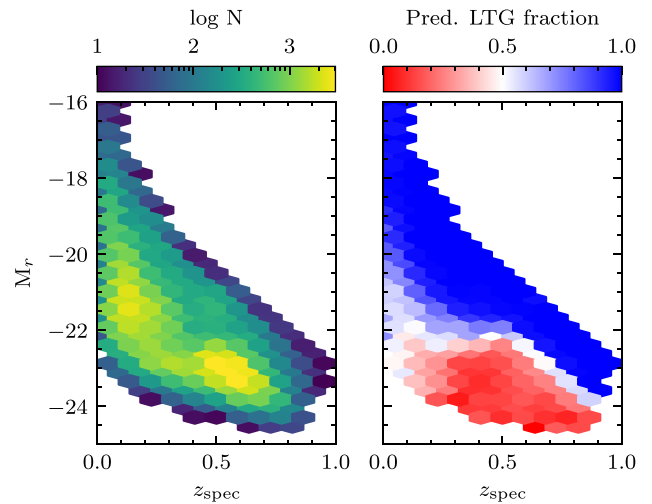


Figure 13. Absolute magnitude (M_r) versus spectroscopic redshift (z_{spec}) for the secure subset ($\Delta_p < 0.3$) for the comparison sample of the DES DR1 stellar mass catalogue. Left-hand panel shows the number of galaxies (in log scale) in each bin. Right-hand panel indicates the predicted ETG/LTG fraction. There is a clear separation between the ETG and LTG populations, with the (intrinsically) brightest galaxies at each redshift dominated by the ETGs, as expected. The lack of ETGs at the lowest redshifts ($z \lesssim 0.2$) is due to the scarcity of massive ETGs in such a small volume.

is based on a Fisher Expectation-Maximization (FEM) unsupervised algorithm working in a parameter space of 12 rest-frame magnitudes and spectroscopic redshift.

The FEM unsupervised algorithm is able to distinguish 12 classes (11 classes of galaxies and an additional class of broad-line active galactic nuclei, AGNs). In particular, classes 1–3 host the reddest spheroidal-shape galaxies showing no sign of star formation activity and dominated by old stellar populations; classes 7–11 contain the blue star-forming galaxies. Classes 4–6 host intermediate galaxies whose physical properties (such as colours, sSFR, stellar masses, and shapes) are intermediate between those of red, passive, and blue, active galaxies. These intermediate galaxies have more concentrated light profiles and lower gas contents than star-forming galaxies. Class 11 may consist of low-metallicity galaxies, or AGNs according to its localization on the Baldwin, Phillips & Telervich (BPT) diagram. See colour-colour diagrams in fig. 2 of Siudek et al. (2018) for further details.

We include in this comparison VIPERS galaxies that are also present in the DES DR1 morphological catalogue with an accurate spectroscopic redshift estimate and the highest membership probability to one of the classes. This subset includes 7384 galaxies with

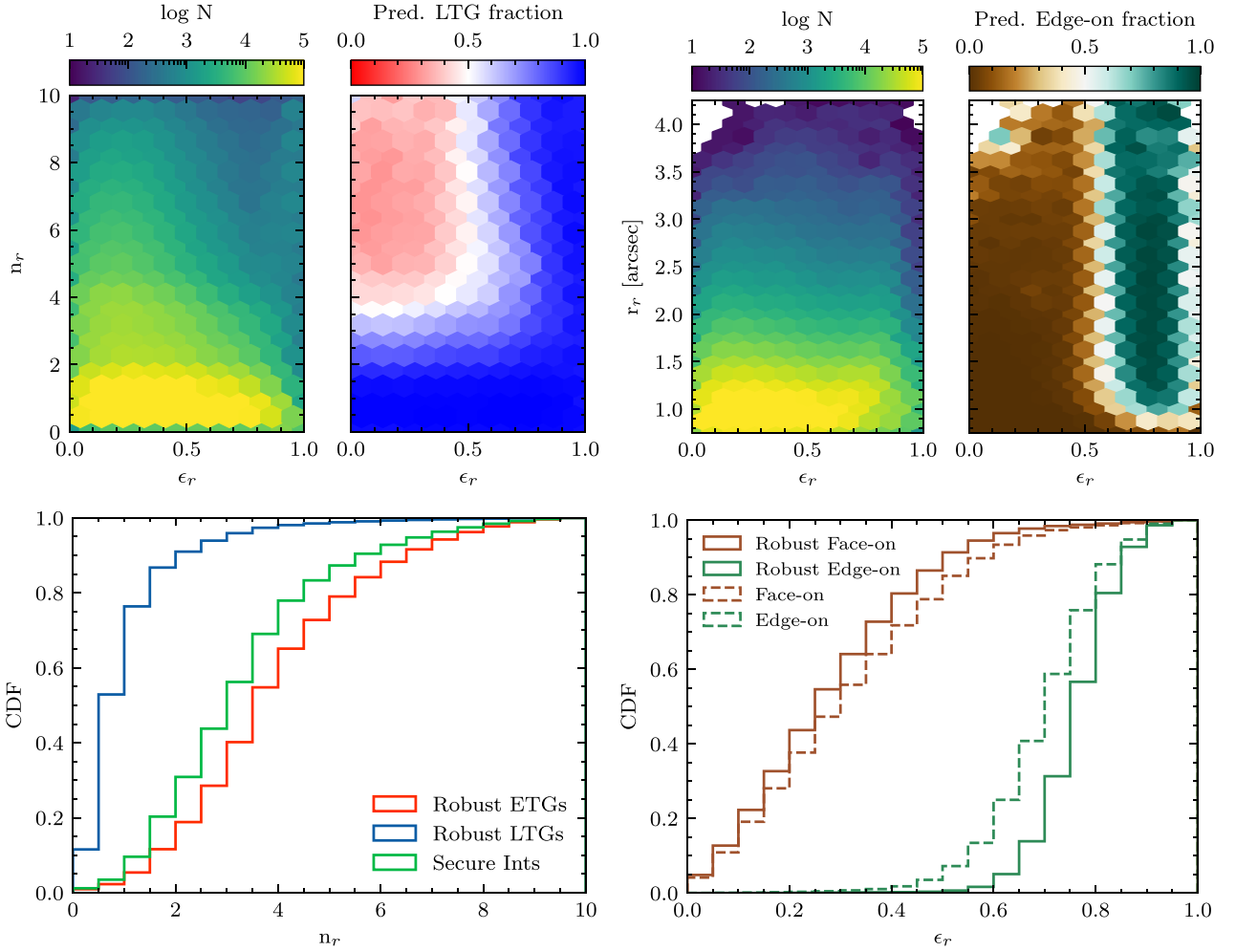


Figure 14. Top-left panels: Sérsic index versus ellipticity for the sample in common with the DES Y1 structural parameters catalogue. The bins are colour coded by number density and fraction of LTG over the total. Bottom-left panel: Cumulative distribution function (CDF) of the Sérsic index for the ETG versus LTG classification scheme: red histogram corresponds to the *robust* ETGs, those with $\max(p_i) < 0.30$; blue histogram shows the *robust* LTGs, those with $\min(p_i) > 0.70$; and green histogram corresponds to the intermediate but secure candidates. Top right-hand panels: Observed radius versus ellipticity colour coded by number density and fraction of edge-on galaxies over the total. Bottom right-hand panel: CDFs for the ellipticity. Black histograms show the CDFs for the face-on galaxies with $\tilde{p}_e < 0.33$ (dashed) and $\max(p_i^e) < 0.30$ (solid). Orange histograms correspond to the CDFs for the edge-on galaxies with $\tilde{p}_e > 0.33$ (dashed) and $\min(p_i^e) > 0.70$ (solid). The very different distribution of the sub-samples is an indicator of the accuracy of our model predictions in real DES galaxy images.

an apparent magnitude of $m_r \in [18.0, 21.5]$ mag (see Fig. 11) and with a spectroscopic redshift distribution ranging from $0.04 < z_{\text{spec}} < 1.46$ with a median value of $\tilde{z}_{\text{spec}} \approx 0.55$. Note that this is the faintest of the comparison samples.

Table 5 shows statistics for this sub-sample, for which 81 per cent of the galaxies show secure classifications (i.e. $\Delta p < 0.30$). Of the secure subset, 20 per cent of the galaxies are classified as *robust* ETGs while 76 per cent are *robust* LTGs. Although the LTGs still dominate the number counts, the two classes are much more balanced than for the full DES morphological catalogue or the Tarsitano et al. (2018) sub-sample. On the other hand, the fraction of edge-on galaxies (2 per cent) is smaller than for the other sub-samples, and only 0.1 per cent of the *robust* ETGs are classified as *robust* edge-on.

In Fig. 15, we show the number of galaxies belonging to each of the classes derived by Siudek et al. (2018) for the ETGs and LTGs sub-samples according to our model predictions. The ETGs clearly

dominate at classes below 4 and are negligible for classes above 6. On the other hand, the LTGs dominate for classes above 6, with a very small fraction with classes 1–3. The intermediate classes are composed of a mix of ETGs and LTGs but mainly populated by LTGs. This strong correlation nicely demonstrates that our model is able to correctly classify original DES images, even at the fainter magnitudes.

To quantify these trends, we can consider as negatives (N) the galaxies belonging to classes 1–3 and as positives (P) the galaxies falling with the classes 4–11. By doing so, we find 89 per cent of TN and a 97 per cent of TP. This translates into an accuracy classification score of $\text{Acc} \approx 0.95$. We have visually inspected the FN images within the VIPERS data set (i.e. ETGs with classes 4–11) finding that for most of them there are neither clear signs of features (such as spiral arms) nor edge-on morphologies, indicating that the ETG morphological classification might be correct regardless of their spectral classification. In the case of the FP, we noticed that

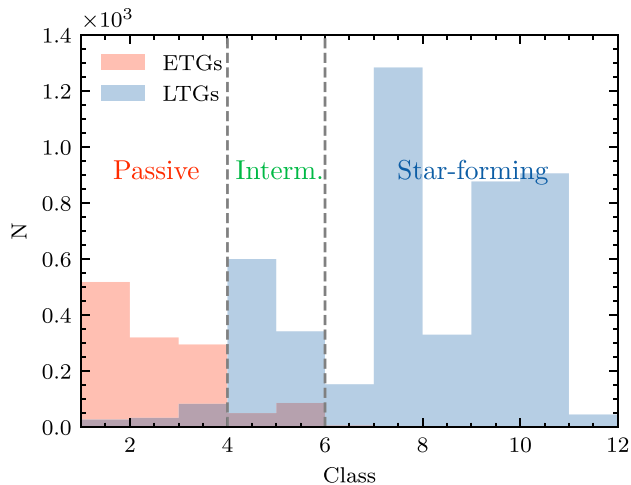


Figure 15. Number of galaxies belonging to each VIPERS class derived by Siudek et al. (2018) ranging from 1 to 12. The red histogram corresponds to the number of ETGs with $\bar{p} < 0.4$. The blue histogram shows the number of LTGs with $\bar{p} > 0.4$. The ETGs clearly dominate at classes below 4 and are negligible for classes above 6, while the LTGs dominate for classes above 6. This strong correlation demonstrates that our model is able to correctly classify original DES images, even at the fainter magnitudes.

for a large fraction of them there is (at least) one close companion within the field of view of the cutout that might lead to an inaccurate classification.

6 CONCLUSIONS

(i) We present a morphological classification according to two schemes: (a) ETGs versus LTGs and (b) edge-on versus face-on for ~ 27 million DES galaxies to $m_r < 21.5$ mag. The classifications are based on the predictions of supervised DL models using CNNs (Section 3.1).

(ii) The training sample consists of bright ($m_r < 17.7$ mag) DES galaxies with a previously known morphological classification (from Domínguez Sánchez et al. 2018) as well as their artificially redshifted counterparts described in Section 2.3.

(iii) Although some of the features that distinguish ETGs and LTGs almost disappear for the fainter galaxies (Fig. 1), the models are able to correctly classify galaxies according to the two schemes with excellent results (accuracy > 97 per cent) even at the fainter magnitude bins (Figs 6 and 10).

(iv) We train five different models using k -folding to obtain a measurement of the classification uncertainty. About 87 per cent of the galaxies in the final catalogue have secure labelling for the ETG versus LTG classification (i.e. $\Delta p < 0.30$). This fraction is 73 per cent for the edge-on classification.

(v) The classifications on real DES faint images are consistent with other available observables, such as absolute magnitude, Sersic index n , ellipticity ϵ , or spectral classification (Section 5).

(vi) Our work demonstrates that machines are able to recover features hidden to the human eye and so can reliably classify faint galaxy images. The methodology adopted in this work to overcome the lack of faint labelled samples can be applied to future big data surveys such as Euclid or Vera Rubin Observatory Legacy Survey of Space and Time.

(vii) The exceptional amount of data provided by DES DR1 has allowed us to construct the largest automated morphological

catalogue to date (along with the companion DES morphological catalogue presented in Cheng et al.) by several orders of magnitude compared to previous works (e.g. Domínguez Sánchez et al. 2018). This classification will be a fundamental tool for our understanding of morphological transformations across cosmic time.

(viii) The complete DES data set DR2, including observations for 600 million galaxies, will be made public in early 2021. The DL models presented in this work can be directly applied to DR2, providing accurate morphological classification for a great fraction of the galaxies with very little effort. In addition, the existence of deep fields within the DES DR2 will allow us to extend this classification to even fainter magnitude limits and to carry out crucial scientific analysis for galaxy formation and evolution.

ACKNOWLEDGEMENTS

This work was supported in part by National Science Foundation (NSF) grant AST-1816330. HDS acknowledges support from the Centro Superior de Investigaciones Científicas PIE2018-50E099. We are grateful to R. Sheth for a careful reading of the manuscript.

Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, the Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministério da Ciência, Tecnologia e Inovação, the Deutsche Forschungsgemeinschaft, and the Collaborating Institutions in the Dark Energy Survey.

The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas – Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenössische Technische Hochschule (ETH) Zürich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciències de l’Espai (IEEC/CSIC), the Institut de Física d’Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig-Maximilians Universität München and the associated Excellence Cluster Universe, the University of Michigan, NFS’s NOIRLab, the University of Nottingham, The Ohio State University, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, Texas A&M University, and the OzDES Membership Consortium.

Based in part on observations at Cerro Tololo Inter-American Observatory at NSF’s NOIRLab (NOIRLab Prop. ID 2012B-0001; PI: J. Frieman), which is managed by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation.

The DES data management system is supported by the National Science Foundation under grant numbers AST-1138766 and AST-1536171. The DES participants from Spanish institutions are partially supported by MICINN under grants ESP2017-89838, PGC2018-094773, PGC2018-102021, SEV-2016-0588, SEV-2016-0597, and MDM-2015-0509, some of which include ERDF funds

from the European Union. IFAE is partially funded by the CERCA program of the Generalitat de Catalunya. Research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Program (FP7/2007-2013) including ERC grant agreements 240672, 291329, and 306478. We acknowledge support from the Brazilian Instituto Nacional de Ciência e Tecnologia (INCT) do e-Universo (CNPq grant 465376/2014-2).

This manuscript has been authored by Fermi Research Alliance, LLC under Contract no. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.

DATA AVAILABILITY

The DES Y1 morphological catalogue is available in the Dark Energy Survey Data Management (DESDM) system at the National Center for Supercomputing Applications (NCSA) at the University of Illinois and can be accessed at: <https://des.ncsa.illinois.edu/releases/other/morphCNN>. The pipeline used to construct the DES Y1 morphological catalogue will be shared on request to the corresponding author.

REFERENCES

- Abbott T. M. C. et al., 2018a, *ApJS*, 239, 18
 Abbott T. M. C. et al., 2018b, *ApJS*, 239, 18
 Aniyani A. K., Thorat K., 2017, *ApJS*, 230, 20
 Arnouts S., Ilbert O., 2011, LePHARE: Photometric Analysis for Redshift Estimate. ascl:1108.009
 Bruzual G., Charlot S., 2003, *MNRAS*, 344, 1000
 Cheng T.-Y. et al., 2020, *MNRAS*, 493, 4209
 de Vaucouleurs G., 1959, *Handbuch der Physik*, 53, 275
 De Vicente J., Sánchez E., Sevilla-Noarbe I., 2016, *MNRAS*, 459, 3078
 Diehl H. T. e. a., 2019, available at: <https://www.osti.gov/biblio/1596042>
 Dieleman S., Willett K. W., Dambre J., 2015, *MNRAS*, 450, 1441
 Domínguez Sánchez H., Huertas-Company M., Bernardi M., Tuccillo D., Fischer J. L., 2018, *MNRAS*, 476, 3661
 Domínguez Sánchez H. et al., 2019, *MNRAS*, 484, 93
 Drlica-Wagner A. et al., 2018, *ApJS*, 235, 33
 Fischer J.-L., Domínguez Sánchez H., Bernardi M., 2019, *MNRAS*, 483, 2057
 Flaugher B. et al., 2015, *AJ*, 150, 150
 Ghosh A., Urry C. M., Wang Z., Schawinski K., Turp D., Powell M. C., 2020, *ApJ*, 895, 112
 Goodfellow I., Bengio Y., Courville A., 2016, Deep Learning. MIT Press, <https://www.deeplearningbook.org/>
 Gschwend J. et al., 2018, *Astron. Comput.*, 25, 58
 Guzzo L. et al. 2014, *A&A*, 566, A108
 Hausen R., Robertson B. E., 2020, *ApJS*, 248, 20
 Huertas-Company M. et al., 2015, *ApJS*, 221, 8
 Huertas-Company M. et al., 2016, *MNRAS*, 462, 4495
 Huertas-Company M. et al., 2018, *ApJ*, 858, 114
 Levy R. C. et al., 2019, *ApJ*, 882, 84
 Lilly S. J., Carollo C. M., 2016, *ApJ*, 833, 1
 Li N., Li C., Mo H., Hu J., Zhou S., Du C., 2020, *ApJ*, 896, 38
 Lintott C. J. et al., 2008, *MNRAS*, 389, 1179
 Masters K. L. et al., 2010, *MNRAS*, 404, 792
 Metcalf R. B. et al., 2019, *A&A*, 625, A119
 Nair P. B., Abraham R. G., 2010, *ApJS*, 186, 427
 Ntampaka M. et al., 2019, *ApJ*, 876, 82
 Palmese A. et al., 2020, *MNRAS*, 493, 4591
 Pasquet J., Bertin E., Treyer M., Arnouts S., Fouchez D., 2019, *A&A*, 621, A26
 Powers D., Ailab, 2011, *J. Mach. Learn. Technol*, 2, 2229
 Pozzetti L. et al., 2010, *A&A*, 523, A13
 Rowe B. T. P. et al., 2015, *Astron. Comput.*, 10, 121
 Secco L. F., Farah A., Jain B., Adhikari S., Banerjee A., Dalal N., 2018, *ApJ*, 860, 32
 Sevilla-Noarbe I. et al., 2020, preprint (arXiv:2011.03407)
 Siudek M. et al., 2018, *A&A*, 617, A70
 Tarsitano F. et al., 2018, *MNRAS*, 481, 2018
 Tuccillo D., Huertas-Company M., Decencière E., Velasco-Forero S., Domínguez Sánchez H., Dimauro P., 2018, *MNRAS*, 475, 894
 Walmsley M. et al., 2020, *MNRAS*, 491, 1554
 Willett K. W. et al., 2013, *MNRAS*, 435, 2835
 Wuyts S. et al., 2011, *ApJ*, 742, 96
 Yan Z., Mead A. J., Van Waerbeke L., Hinshaw G., McCarthy I. G., 2020, preprint (arXiv:2005.11819)

APPENDIX A: EXAMPLES OF GALAXY IMAGES

In the appendix, we show some examples of real DES galaxies classified by our morphological catalogue in different magnitude bins. Note that our models are able to correctly classify the fainter galaxies, despite the fact that the noise significantly degrades their morphological features.

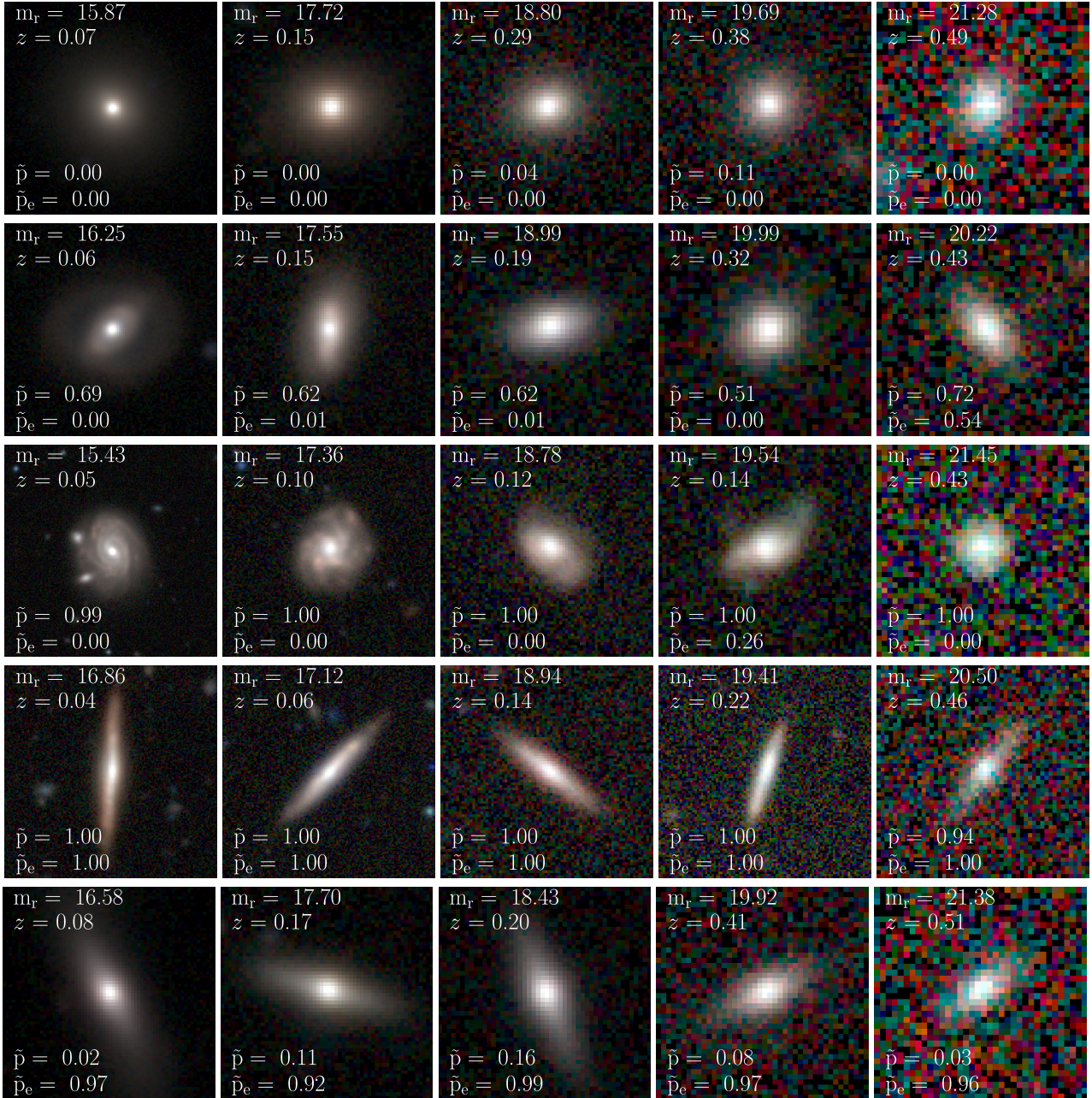


Figure A1. Examples of DES galaxies classified as *robust* ETGs, secure intermediates, *robust* LTGs, *robust* edge-on, and *robust* edge-on ETGs (from top to bottom). Each column corresponds to a magnitude bin, fainter towards the right. The cutouts include the redshift and apparent magnitude of the galaxies from DES DR1 catalogue, as well as our median CNN-derived probabilities of being LTG (\tilde{p}) and edge-on (\tilde{p}_e).

¹Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA

²IFCA, Instituto de Física de Cantabria (UC-CSIC), Av. de Los Castros s/n, E-39005 Santander, Spain

³Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, E-08193 Barcelona, Spain

⁴LERMA, PSL Research University, Observatoire de Paris, CNRS, Sorbonne Universités, UPMC Univ. Paris 06, F-75014 Paris, France

⁵Univeristé de Paris, 5 Rue Thomas Mann, F-75013, Paris, France

⁶Departamento de Astrofísica, Universidad de La Laguna, E-38206 La Laguna, Tenerife, Spain

⁷Instituto de Astrofísica de Canarias, E-38205 La Laguna, Tenerife, Spain

⁸SCIPP, University of California, Santa Cruz, CA 95064, USA

⁹Physics Department, University of Wisconsin-Madison, 2320 Chamberlin Hall, 1150 University Avenue Madison, WI 53706, USA

¹⁰Departamento de Física Matemática, Instituto de Física, Universidade de São Paulo, CP 66318, São Paulo, SP, 05314-970, Brazil

¹¹Laboratório Interinstitucional de e-Astronomia - LIneA, Rua Gal. José Cristiano 77, Rio de Janeiro, RJ - 20921-400, Brazil

¹²Fermi National Accelerator Laboratory, PO Box 500, Batavia, IL 60510, USA

¹³Instituto de Física Teórica UAM/CSIC, Universidad Autónoma de Madrid, E-28049 Madrid, Spain

¹⁴Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth PO1 3FX, UK

- ¹⁵CNRS, UMR 7095, Institut d'Astrophysique de Paris, F-75014, Paris, France
- ¹⁶Sorbonne Universités, UPMC Univ Paris 06, UMR 7095, Institut d'Astrophysique de Paris, F-75014, Paris, France
- ¹⁷Department of Physics & Astronomy, University College London, Gower Street, London WC1E 6BT, UK
- ¹⁸Universidad de La Laguna, Dpto. Astrofísica, E-38206 La Laguna, Tenerife, Spain
- ¹⁹Department of Astronomy, University of Illinois at Urbana-Champaign, 1002 W. Green Street, Urbana, IL 61801, USA
- ²⁰National Center for Supercomputing Applications, 1205 West Clark St., Urbana, IL 61801, USA
- ²¹Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, E-08193 Bellaterra (Barcelona) Spain
- ²²Center for Cosmology and Astro-Particle Physics, The Ohio State University, Columbus, OH 43210, USA
- ²³Jodrell Bank Center for Astrophysics, School of Physics and Astronomy, University of Manchester, Oxford Road, Manchester M13 9PL, UK
- ²⁴University of Nottingham, School of Physics and Astronomy, Nottingham NG7 2RD, UK
- ²⁵INAF-Osservatorio Astronomico di Trieste, via G. B. Tiepolo 11, I-34143 Trieste, Italy
- ²⁶Institute for Fundamental Physics of the Universe, Via Beirut 2, I-34014 Trieste, Italy
- ²⁷Observatório Nacional, Rua Gal. José Cristino 77, Rio de Janeiro, RJ - 20921-400, Brazil
- ²⁸Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA
- ²⁹Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), E-28040 Madrid, Spain
- ³⁰Department of Physics, IIT Hyderabad, Kandi, Telangana 502285, India
- ³¹Institute of Theoretical Astrophysics, University of Oslo. P.O. Box 1029 Blindern, NO-0315 Oslo, Norway
- ³²Institut d'Estudis Espacials de Catalunya (IEEC), E-08034 Barcelona, Spain
- ³³Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA
- ³⁴Department of Physics, Stanford University, 382 Via Pueblo Mall, Stanford, CA 94305, USA
- ³⁵Kavli Institute for Particle Astrophysics & Cosmology, P. O. Box 2450, Stanford University, Stanford, CA 94305, USA
- ³⁶SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA
- ³⁷Département de Physique Théorique and Center for Astroparticle Physics, Université de Genève, 24 quai Ernest Ansermet, CH-1211 Geneva, Switzerland
- ³⁸Department of Physics, ETH Zurich, Wolfgang-Pauli-Strasse 16, CH-8093 Zurich, Switzerland
- ³⁹School of Mathematics and Physics, University of Queensland, Brisbane QLD 4072, Australia
- ⁴⁰Santa Cruz Institute for Particle Physics, Santa Cruz, CA 95064, USA
- ⁴¹Department of Physics, The Ohio State University, Columbus, OH 43210, USA
- ⁴²Faculty of Physics, Ludwig-Maximilians-Universität, Scheinerstr. 1, D-81679 Munich, Germany
- ⁴³Max Planck Institute for Extraterrestrial Physics, Giessenbachstrasse, D-85748 Garching, Germany
- ⁴⁴Universitäts-Sternwarte, Fakultät für Physik, Ludwig-Maximilians Universität München, Scheinerstr. 1, D-81679 München, Germany
- ⁴⁵Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
- ⁴⁶Australian Astronomical Optics, Macquarie University, North Ryde, NSW 2113, Australia
- ⁴⁷Lowell Observatory, 1400 Mars Hill Rd, Flagstaff, AZ 86001, USA
- ⁴⁸Institució Catalana de Recerca i Estudis Avançats, E-08010 Barcelona, Spain
- ⁴⁹Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK
- ⁵⁰Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, NJ 08544, USA
- ⁵¹Department of Physics and Astronomy, University of Sussex, Pevensey Building, Brighton BN1 9QH, UK
- ⁵²School of Physics and Astronomy, University of Southampton, Southampton SO17 1BJ, UK
- ⁵³Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

This paper has been typeset from a \TeX/L\AA\TeX file prepared by the author.