Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **509,** 4024–4036 (2022) https://doi.org/10.1093/mnras/stab3089
Advance Access publication 2021 October 27

SDSS-IV DR17: final release of MaNGA PyMorph photometric and deep-learning morphological catalogues

H. Domínguez Sánchez [1,2]★ B. Margalef [3] M. Bernardi[3] and M. Huertas-Company [4,5,6,7]

[1]*Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, E-08193 Barcelona, Spain*
[2]*Institut d'Estudis Espacials de Catalunya (IEEC), Carrer Gran Capità, E-08034 Barcelona, Spain*
[3]*Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA*
[4]*LERMA, Observatoire de Paris, PSL Research University, CNRS, Sorbonne Universités, UPMC Univ. Paris 06, F-75014 Paris, France*
[5]*University of Paris Denis Diderot, University of Paris Sorbonne Cité (PSC), F-75205 Paris Cedex 13, France*
[6]*Instituto de Astrofísica de Canarias, E-38200 La Laguna, Tenerife, Spain*
[7]*Departamento de Astrofísica, Universidad de La Laguna, E-38206 La Laguna, Tenerife, Spain*

Accepted 2021 October 18. Received 2021 October 18; in original form 2021 September 3

ABSTRACT

We present the MaNGA PyMorph photometric Value Added Catalogue (MPP-VAC-DR17) and the MaNGA Deep Learning Morphological VAC (MDLM-VAC-DR17) for the final data release of the MaNGA survey, which is part of the SDSS Data Release 17 (DR17). The MPP-VAC-DR17 provides photometric parameters from Sérsic and Sérsic+Exponential fits to the two-dimensional surface brightness profiles of the MaNGA DR17 galaxy sample in the *g*, *r*, and *i* bands (e.g. total fluxes, half-light radii, bulge-disc fractions, ellipticities, position angles, etc.). The MDLM-VAC-DR17 provides deep-learning-based morphological classifications for the same galaxies. The MDLM-VAC-DR17 includes a number of morphological properties, for example, a T-Type, a finer separation between elliptical and S0, as well as the identification of edge-on and barred galaxies. While the MPP-VAC-DR17 simply extends the MaNGA PyMorph photometric VAC published in the SDSS Data Release 15 (MPP-VAC-DR15) to now include galaxies that were added to make the final DR17, the MDLM-VAC-DR17 implements some changes and improvements compared to the previous release (MDLM-VAC-DR15): Namely, the low end of the T-Types is better recovered in this new version. The catalogue also includes a separation between early or late type, which classifies the two populations in a complementary way to the T-Type, especially at the intermediate types ($-1 <$ T-Type < 2), where the T-Type values show a large scatter. In addition, *k*-fold-based uncertainties on the classifications are also provided. To ensure robustness and reliability, we have also visually inspected all the images. We describe the content of the catalogues and show some interesting ways in which they can be combined.

Key words: catalogues – surveys – galaxies: disc – galaxies: elliptical, lenticular, cD – galaxies: photometry – galaxies: structure.

1 INTRODUCTION

As we enter the age of large galaxy samples for which spatially resolved spectroscopic information is available thanks to integral field spectroscopic surveys like ATLAS^{3D} (Cappellari et al. 2011), CALIFA (Sánchez et al. 2012), or SAMI (Allen et al. 2015), it is useful to have accompanying analyses of the associated photometry. Fischer, Domínguez Sánchez & Bernardi (2019) describe a step in this direction: They provide imaging-based morphological information, as well as one- and two-component fits to the two-dimensional (2D) surface brightness distributions of the galaxies in an early release (SDSS DR15, Aguado et al. 2019) of the MaNGA (Mapping Nearby Galaxies at Apache Point Observatory; Bundy et al. 2015) Survey. Now that the survey is complete, the main goal of this work is to extend that analysis to all the $\sim 10^4$ nearby ($z \sim 0.03$) galaxies in it. This has culminated in the production of two 'value-added' catalogues (VACs), which are part of the SDSS-DR17 release (SDSS Collaboration, in preparation): the MaNGA PyMorph photometric Value Added Catalogue (hereafter MPP-VAC-DR17) and the MaNGA Deep Learning Morphology Value Added Catalogue (hereafter MDLM-VAC-DR17), which summarize the photometric and deep-learning-based morphological information for the MaNGA galaxies.

MaNGA is a component of the Sloan Digital Sky Survey IV (SDSS IV; Blanton et al. 2017). Wake et al. (2017) describe how the MaNGA galaxies were selected from the SDSS footprint. Integral field unit (IFU) technology allows the MaNGA survey to obtain detailed kinematic and chemical composition maps of these galaxies (e.g. Gunn et al. 2006; Smee et al. 2013; Drory et al. 2015; Law et al. 2015, 2016; Yan et al. 2016a,b; Greene et al. 2017; Graham et al. 2018)

For reasons discussed in Fischer, Bernardi & Meert (2017), we do not use the SDSS pipeline photometry of these objects. Rather, we use the significantly more accurate PyMorph analysis described in a series of papers (Vikram et al. 2010; Meert, Vikram & Bernardi 2013, 2015, 2016; Bernardi et al. 2014). PyMorph provides one- and two-component fits to the 2D surface brightness distributions

★ E-mail: dominguez@ice.csic.es

of MaNGA galaxies, and was used to produce the MPP-VAC of the galaxies in DR15 (Fischer et al. 2019). The MPP-VAC-DR17, which we describe below, extends this to include all the objects in the completed MaNGA survey.

We also provide the MDLM-VAC-DR17, which includes deep-learning-based morphological classifications (the methodology is described in detail by Domínguez Sánchez et al. 2018) for the same galaxies. In contrast to the photometric MPP-VACs, in which the main difference between the DR15 and DR17 versions is sample size, the MDLM-VAC-DR17 includes some improvements in methodology and content with respect to DR15, which we describe below.

Note that the MaNGA data were only used for the identification of the sources included in the two VACs presented in this paper. Both the 2D fits to the surface brightness distributions and the morphological classifications are based on the SDSS imaging data (DR15 for the MPP-VAC and DR7 for the MDLM-VAC).

Section 2 describes the minor changes we have made when reporting the photometric parameters listed in the MPP-VAC: See Fischer et al. (2019) for a detailed discussion of how these PyMorph-based parameters were determined, and how they compare with previous work. Section 3 describes our morphological classification scheme and the MDLM-VAC-DR17. Section 4 combines our MPP- and MDLM-VACs to show how the photometric parameters correlate with morphology. A final section summarizes.

2 MANGA PYMORPH PHOTOMETRIC VALUE ADDED CATALOGUE (MPP-VAC-DR17)

The MPP-VAC-DR17[1] is one of the VACs available online of the completed MaNGA survey, which is part of the SDSS-DR17 release.[2] It is similar to the MPP-VAC-DR15 (Fischer et al. 2019, hereafter F19) published as part of the SDSS-DR15 release (Aguado et al. 2019). The MPP-VAC-DR17 is updated to include all the galaxies in the final MaNGA release. Some PLATE-IFU entries are reobservations of the same galaxy so the catalogue also provides three variables that identify galaxies with multiple MaNGA spectroscopic observations (see DUPL-GR, DUPL-N, and DUPL-ID). Although the number of entries is 10 293, the actual number of different galaxies in this catalogue is 10 127. The structural parameters and morphological classifications included in the VACs are identical for the duplicate observations.

The MPP-VAC-DR17 also includes one minor technical change regarding how the position angle (PA) of each image is reported. The PA (from PyMorph) given in this catalogue is with respect to the camera columns in the SDSS 'fpC' images (which are not aligned with the north direction); to convert to the usual convention where north is up, east is left,[3] set PA(MaNGA) = (90 − PA) − SPA, where SPA is the SDSS camera column position angle with respect to north reported in the primary header of the 'fpC' SDSS images. PA (MaNGA) is defined to increase from east towards north. In contrast to the MPP-VAC-DR15 release, where the SPA angles were provided in a separate file, the MPP-VAC-DR17 catalogue includes the SPA angles.

Except for this change, MPP-VAC-DR17 is similar in format to MPP-VAC-DR15. In particular, table 1 in F19 describes the content of the catalogue, which is in the FITS file format and includes three Header Data Units (HDUs). Each HDU lists the parameters measured in the *g*, *r*, and *i* bands, respectively. These include the luminosity, half-light radius, Sérsic index, etc. for single Sérsic (Ser) and two-component Sérsic+Exponential (SerExp) profiles – from fitting the 2D surface brightness profiles of each galaxy. Although for most galaxies the Exponential component is a disc, for the most luminous galaxies, it represents a second component,which is not necessarily a disc.

Table 1. Top: fraction of galaxies that do not have PyMorph parameters from Sérsic (FLAG_FAILED_S = 1), SerExp (FLAG_FAILED_SE = 1), or both (FLAG_FIT = 3) in the SDSS *g*, *r*, and *i* bands. Bottom: fraction of galaxies that are better described by one-component Sérsic fits (FLAG_FIT = 1), two-component SerExp fits (FLAG_FIT = 2), or for which both fits are equally acceptable (FLAG_FIT = 0).

Band	Sérsic fit failed (FLAG_FAILED_S = 1)	SerExp fit failed (FLAG_FAILED_SE = 1)	Both fits failed (FLAG_FIT = 3)
g	0.069	0.065	0.038
r	0.065	0.058	0.034
i	0.077	0.062	0.037

Galaxies with successful fits (FLAG_FIT ≠ 3) better described by:

Band	One component (FLAG_FIT = 0)	Two components (FLAG_FIT = 1)	Both (FLAG_FIT = 2)
g	0.103	0.586	0.312
r	0.106	0.567	0.327
i	0.104	0.569	0.327

None of the algorithms has changed since DR15, so the discussion in F19 about how photometric parameters were determined remains appropriate. In particular, we still use the fitting algorithm called PyMorph (Vikram et al. 2010; Meert et al. 2013, 2015, 2016; Bernardi et al. 2014), a PYTHON-based code that uses Source Extractor (SExtractor; Bertin & Arnouts 1996) and GALFIT (Peng et al. 2002) to estimate the structural parameters of galaxies. Likewise, decisions about refitting (section 2.1.1 in F19), when to 'flip' the two components of a SerExp fit (section 2.1.3 in F19), and how to truncate the profiles (section 2.1.4 in F19) are all the same as before, as is the (visual-inspection-based) flagging system, which indicates which fit is to be preferred for scientific analyses (see discussion in section 2.2 of F19). We urge users to pay attention to the preferences expressed by FLAG_FIT: FLAG_FIT = 1 means that only the Sérsic fit is preferred (the SerExp fit may be unreliable), FLAG_FIT = 2 means that only the SerExp fit is preferred (the Sérsic fit may be unreliable), FLAG_FIT = 0 means that both Sérsic and SerExp fits are acceptable, and FLAG_FIT = 3 means that none of the fits were reliable and so no parameters are provided. Table 1 lists the fraction of objects for each FLAG_FIT type in the SDSS *g*, *r*, and *i* bands.

The flags FLAG_FAILED_S = 1 or FLAG_FAILED_SE = 1 indicate failed Sérsic or SerExp fits, respectively. Failures can happen for several reasons: contamination, peculiarity, bad image, or bad model fit. The numbers in the top half of Table 1 give the fraction of objects without photometric measurements for the different bands. About 7 per cent of the objects do not have parameters from the Sérsic and SerExp fits, respectively. About 4 per cent of these objects do not have any PyMorph photometric parameters (i.e. FLAG_FIT = 3).

Figs 1 and 2 show the distributions of the Sérsic index *n* and $n_{\rm bulge}$ in our single- and two-component fits. These are very similar to figs 12 and 14 in F19, illustrating that other than the factor of 2 increase in sample size (from DR15 to DR17) the trends are unchanged. In

[1] https://www.sdss.org/dr17/data_access/value-added-catalogs/
[2] www.sdss.org/dr17/data_access/value-added-catalogs/
[3] Note that the MaNGA data cubes have north up and east right.

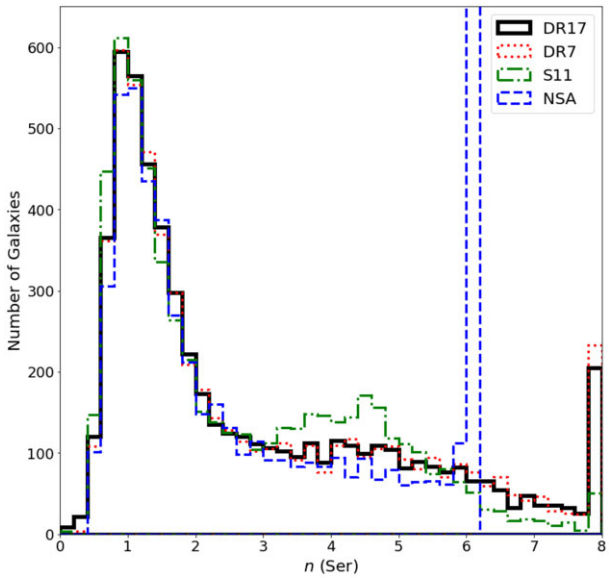

Figure 1. Distribution of n from only Sérsic fits to the r-band surface brightness profiles of the objects in our sample (DR17), compared to the corresponding distribution from Meert et al. (2015, hereafter DR7), Simard et al. (2011, hereafter S11), and the NASA-Sloan Atlas catalogue (NSA; `nsatlas.org`). Our analysis limits $n \leq 8$, whereas the S11 analysis allows $0.5 \leq n \leq 8$, and NSA does not allow $n > 6$. This explains the spike at $n = 6$, where NSA has 1709 galaxies.

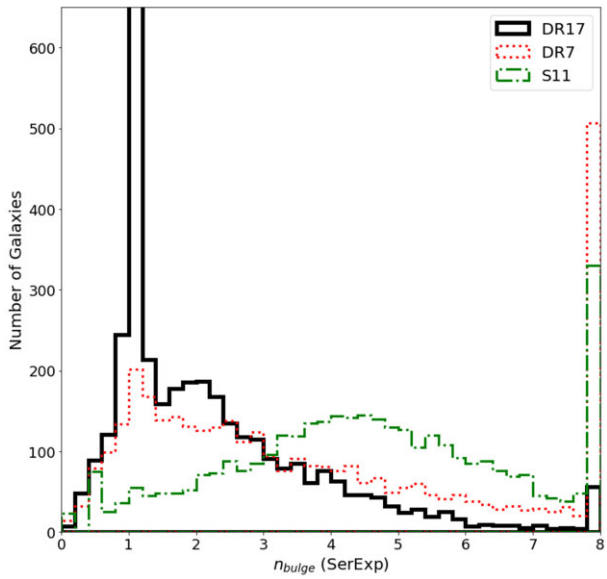

Figure 2. Same as the previous figure, but for $n_{\rm bulge}$ of the two-component SerExp fits. See F19 for a discussion of the obvious differences with respect to S11. Similarly to DR15, our DR17 analysis has several more galaxies with $n_{\rm bulge} = 1$ but many fewer $n_{\rm bulge} = 8$ compared to the DR7 analysis, as a result of our eyeball-motivated refitting and flipping. The spike $n_{\rm bulge} = 1$ for DR17 extends to 735 galaxies.

particular, our reductions do not show a preference for $n = 6$ (in contrast to NSA, which does not allow $n > 6$), or for $n = 4$ or $n_{\rm bulge} = 4$ (in contrast to S11). Similarly to DR15, our DR17 analysis has several more galaxies with $n_{\rm bulge} = 1$ but many fewer $n_{\rm bulge} = 8$ compared to the DR7 analysis, as a result of our eyeball-motivated refitting and flipping. Likewise, we have repeated all the other tests and comparisons shown in F19, but now for the full DR17 sample, finding consistent results with our DR15 analysis, so we do not show them here.

3 MANGA DEEP LEARNING MORPHOLOGY VALUE ADDED CATALOGUE (MDLM-VAC-DR17)

The MDLM-VAC-DR17 provides morphological classifications for the final MaNGA galaxy sample (which is part of the SDSS-DR17 release) using an automated classification based on supervised deep learning. It extends the 'MaNGA Deep Learning Morphology DR15 VAC', described in F19, to now include galaxies which were added to make the final DR17. In addition, as we describe in the following sections, it incorporates some changes and improvements with respect to the DR15 version.

The morphological classifications were obtained following the methodology explained in detail in Domínguez Sánchez et al. (2018, hereafter DS18). Briefly, for each classification task, we trained a convolutional neural network (CNN) using as input the RGB cutouts downloaded from the SDSS-DR7 server[4] with a variable size that is proportional to the Petrosian radius of the galaxy ($5\times R_{90}$[5]). The cutouts are then resampled to 69 × 69 pixels, which are the dimensions used to feed the CNN – note that by doing this the pixel scale varies from one galaxy to another. The counts in each pixel are normalized by the maximum value in that cutout for that particular colour band. As this value is different in each band, this step prevents colour information from playing a role in the morphological classifications, and potentially biasing studies of colour–morphology relations. We refer the reader to DS18 for further details.

3.1 Classification scheme

The classification scheme of the morphological catalogue is presented in Fig. 3. We provide a T-Type value, which ranges from −4 to 9, and was obtained by training the CNN in regression mode based on the T-Types from Nair & Abraham (2010, hereafter N10) catalogue. N10 presents visual classifications for 14 034 galaxies from SDSS up to $m_g < 16$ mag. We only use galaxies with confident classifications for training (T-Type flag = 0, i.e. ∼ 96 per cent of the sample). While the N10 T-Type values are integers within the range [−5, 10], none of their galaxies have T-Type values of −4, −2, or −1. We reassigned T-Type values by shifting them and filling the gaps in our training labels, as doing so helps the model to converge.

In general, T-Type < 0 corresponds to early-type galaxies (ETGs), while T-Type > 0 corresponds to late-type galaxies (LTGs). Following F19, we sometimes subdivide LTGs into S1 (0 ≤ T-Type ≤ 3) and S2 (T-Type > 3) – see Section 3.3.2.

The catalogue provides two other binary classifications, which were trained with N10-based labels:

(i) $P_{\rm LTG}$, which separates ETGs from LTGs; and
(ii) $P_{\rm S0}$, which separates pure ellipticals (E) from S0s.

For the $P_{\rm LTG}$ model, we labelled positive examples those with T-Type > 0 and negative examples those with T-Type ≤ 0 (from N10). This classification complements the T-Type by providing a cleaner

[4] http://casjobs.sdss.org/ImgCutoutDR7/
[5] R_{90} from the NSA catalogue.

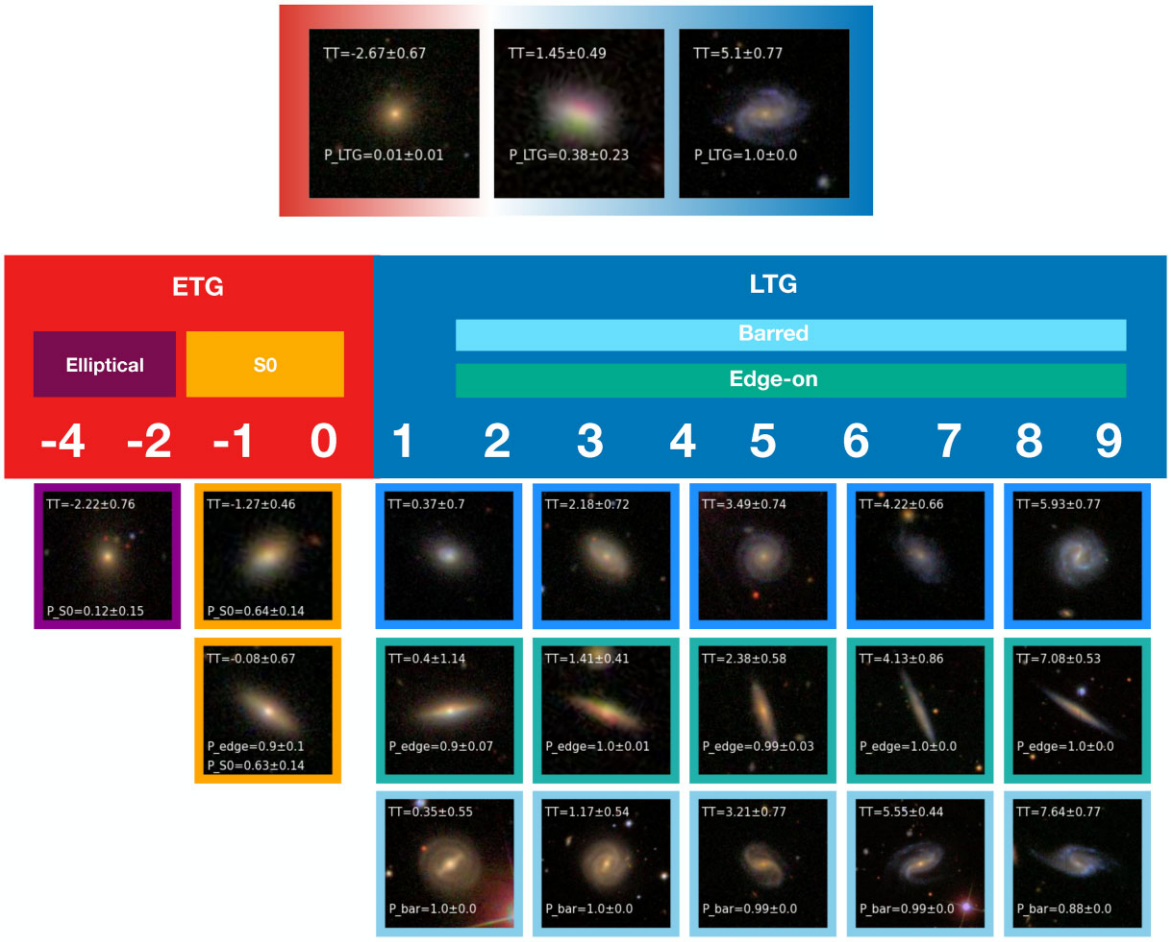

Figure 3. Schematic representation of the morphological classification presented in the MDLM-VAC. It includes a T-Type ranging from −4 to ∼9, where the transition between early and late types happens around T-Type ∼ 0. A complementary binary classification separates ETGs and LTGs (see discussion in the text for the differences between these two classifications). Three further binary classifications include (a) separating E from S0 – this separation is only meaningful for galaxies with T-Type < 0, (b) identifying edge-on galaxies, and (c) identifying galaxies with bar features. The cutouts show examples of galaxies of different types, with classification values shown in white and according to their frame colours. The cutouts are proportional to the size of each galaxy (∼5 $\times R_{90}$). The topmost cutouts show a galaxy classified as ETG by both the T-Type and $P_{\rm LTG}$ models (left-hand side), a galaxy classified as LTG by both the T-Type and $P_{\rm LTG}$ models (right-hand side), and a galaxy with T-Type > 0 and $P_{\rm LTG} < 0.5$ – see discussion related to these galaxies in Section 3.3.2.

separation between ETGs and LTGs, especially at intermediate T-Types, where the scatter of the T-Type model is larger (see discussion in Section 3.3.1). For the $P_{\rm S0}$ model, we used as training sample only galaxies with T-Type < 0 and labelled positive examples those with −5 < T-Type < 0 and negative examples those with T-Type = −5 (i.e. pure E according to N10).

The catalogue also provides two additional binary classifications:

(i) $P_{\rm edge\text{-}on}$, which identifies edge-on galaxies; and
(ii) $P_{\rm bar}$, which identifies barred galaxies.

The value reported in the catalogue is the probability of being a positive example (edge-on or barred galaxy, respectively). These are based on the Galaxy Zoo 2 (GZ2; Willett et al. 2013) labels. GZ2 is a citizen science project with morphological classifications of 304 122 galaxies drawn from SDSS up to $m_r < 17$. Following DS18, the training sample was composed of galaxies with robust classifications, i.e. at least five votes and weighted fraction values greater than 0.8 (for the 'yes' or 'no' answers in each classification task). See DS18 for further details.

3.2 Training methodology

The CNN architecture used for the morphological classifications of the binary models (i.e. $P_{\rm LTG}$, $P_{\rm S0}$, $P_{\rm edge\text{-}on}$, $P_{\rm bar}$) is identical to that of DS18 (see fig. 1 therein for a schematic representation). Namely, the inputs are arrays of dimension (3, 69, 69), and the CNN consists of four convolutional layers with *relu* activation, filter sizes of 6 × 6, 5 × 5, 2 × 2, and 3 × 3; 32, 64, 128, and 128 channels; and dropouts of 0.5, 0.25, 0.25, and 0.25, respectively; followed by a fully connected layer of 64 neurons with 0.5 dropout, *sigmoid* activation, and *adam* optimizer. The output of the model is one single value, which can be interpreted as the probability of being a positive example. The total number of trainable parameters is 2602 849. The CNN was trained for 50 epochs with binary cross-entropy as the loss function.

Due to the complexity of the T-Type classification, we used a slight variation of the CNN architecture described in DS18 to train the T-Type model: The convolutional layers remain as explained above, but the model includes two fully connected layers of 128 and 64 neurons each, with 0.5 dropout. The total number of trainable parameters increases up to 4978 657. The CNN was trained for 100 epochs in regression mode and mean squared error as the loss function.

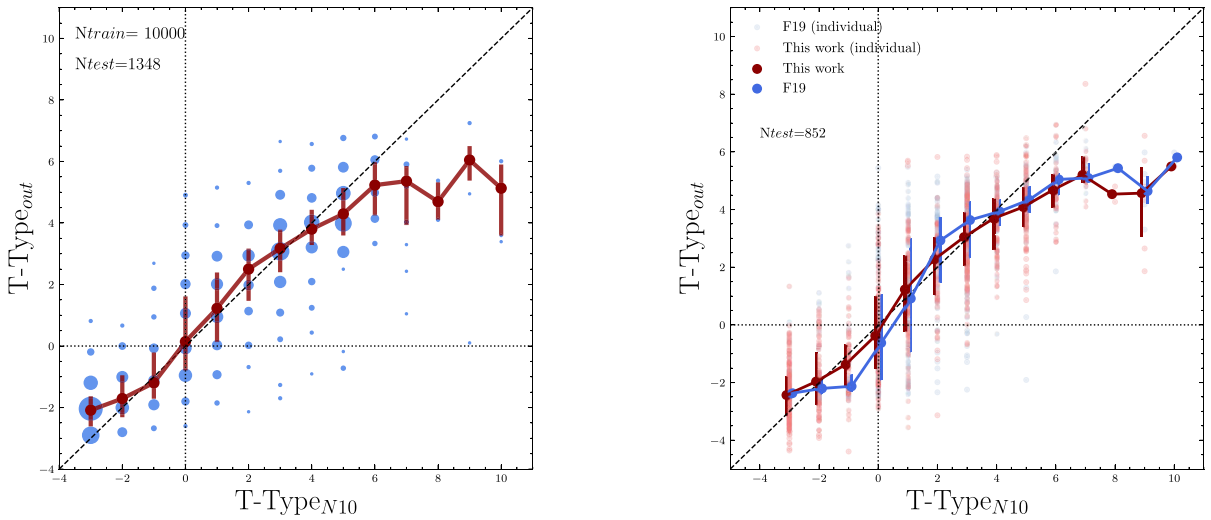

Figure 4. Left-hand panel: comparison of the T-Type derived from the CNN (the average of 15 models trained with k-folding) and the original T-Type (from N10) for the test sample of 1348 galaxies. To better visualize it, we plot average binned values, where the symbol size is proportional to the number of objects in each bin. The red dots (joined by a solid line) show the median value at each T-Type, while the error bars show the inter-quartile ranges (i.e. the difference between 75th and 25th percentiles). The predicted T-Types follow the one-to-one relation (dashed line) very well up to T-Type ~ 5. Right-hand panel: same as the left-hand panel but comparing the new results (red) with the models presented in F19 (blue) for 852 individual galaxies. There is an improvement in the bias, especially at T-Type < 0 and $1 <$ T-Type < 4. (The average values and their error bars have been shifted $\pm$ 0.1 T-Types for better visualization.)

One of the main improvements with respect to the MDLM-VAC-DR15 is that the new catalogue includes model uncertainties provided by the standard deviation obtained with k-folding (with $k = 10$, except for the T-Type model where $k = 15$). This methodology is very close to deep ensembles – which are formally demonstrated to be a Bayesian uncertainty quantification (see Lakshminarayanan, Pritzel & Blundell 2016) – and accounts for variations in the initialization of the CNN weights as well as variations due to the training sample. The value reported in the catalogue is the average of the k models and the uncertainty is their standard deviation. This methodology has been demonstrated to improve the performance with respect to the value of the individual models (see Vega-Ferrero et al. 2021). We reserved a sample that has never gone through the CNN and used it as test sample from which to measure the performance of the k models (see Section 3.3).

3.3 Models' performance

In this section, we show the performance of the models, i.e. how the predicted morphological classifications compare to the original ones. For this purpose, we define a 'test sample', which is a set of galaxies that were not used to train in any of the k-folds. The results shown in this section are obtained by applying the deep-learning models to these test samples.

3.3.1 T-Type

The T-Type model is a linear regression, and the best way to test its performance is to make a one-to-one comparison with the 'true' value. Fig. 4 shows that there is excellent agreement between the input and predicted T-Types up to T-Type ~ 5, where the predicted T-Type underestimates the correct value. We attribute this to the small number of such objects in the training and test samples (the symbol sizes are proportional to the number of objects in each T-Type bin). In addition, there is also a slight flattening at the lowest T-Type values. This was evident, and more pronounced, in the older models presented in F19 (as can be seen in the right-hand panel, where the results of the two models are compared) and is the main reason why we also provide a classification between pure ellipticals and lenticulars (P_{S0}). If we limit the analysis to T-Type < 5, the average bias values (T-Type$_{in}$ $-$ T-Type$_{out}$) are $b = -0.1$ and -0.4 for this work and for the F19 models, respectively. This is smaller than typical differences between the visual classifications of different professional astronomers ($b \sim 1$). The scatter is also smaller for the new model at intermediate T-Types (-1, -3) but larger otherwise. In part because the flattening of the F19 model, especially around T-Type $= -2$ and 3, reduces the F19 scatter.

To summarize: Our new T-Type model shows a smaller bias compared to the one presented in F19 (especially at T-Type < 0 and $1 <$ T-Type < 4) and includes an uncertainty value (determined from the standard deviation of the T-Type predicted by each of the 15 models trained with k-folding).

3.3.2 P_{LTG} and P_{S0} models

In addition to the T-Type model, the MDLM-VAC-DR17 provides two binary classifications trained with the N10 catalogue.

The first one, $P_{\rm LTG}$, separates ETGs from LTGs.

The model does an excellent job at separating the E and S2 populations, recovering 98 per cent of the ellipticals (defined as true negatives, TN; i.e. $P_{\rm LTG} < 0.5$ and labelled negative in the training sample) and 97 per cent of the S2 (defined as true positives, TP; i.e. $P_{\rm LTG} > 0.5$ and labelled positive in the training sample). The separation for the intermediate populations, S0 and S1, is less clean, as expected. For the S0s, 59 per cent are classified as ETGs and 41 per cent as LTGs, while for the S1s, the fractions of galaxies classified as ETGs and LTGs are 27 per cent and 73 per cent, respectively. This classification is very useful for making a broad separation between ETGs and LTGs,

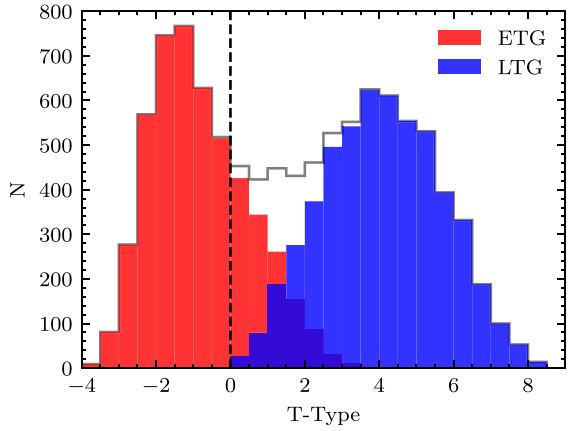

Figure 5. Bimodal distribution of the predicted T-Type is well described by our binary ETG and LTG classifier (recall that LTGs are defined as having $P_{\rm LTG} > 0.5$, while ETGs have $P_{\rm LTG} < 0.5$). The black dashed line at T-Type = 0 marks the separation between E/S0 and S, based on the T-Type model. Note that there are some galaxies with T-Type > 0 classified as ETGs, while the opposite is negligible (see discussion in Section 3.3.2).

especially at intermediate T-Types, where the T-Type scatter is large.

Fig. 5 compares our T-Type and $P_{\rm LTG}$ predictions for the full MaNGA DR17. The bimodality distribution of T-Types is very well traced by the ETG and LTG populations. Only two galaxies with T-Type < 0 are classified as LTGs ($P_{\rm LTG} > 0.5$). On the other hand, 1315 galaxies with T-Type > 0 are classified as ETGs ($P_{\rm LTG} < 0.5$).

Fig. 6 shows the distribution in apparent magnitude, angular size, and central velocity dispersion,[6] of the galaxies having inconsistent T-Type and $P_{\rm LTG}$ classifications. These galaxies (open histograms) occupy the faint end of the magnitude distribution (top), have small angular sizes (middle), and low central velocity dispersions (bottom) (similar to the LTGs). That is, they are probably too faint or small to clearly show spiral structure. We conclude that these galaxies are the most difficult to classify: The T-Type classification might be correct, while the $P_{\rm LTG}$ model is actually separating galaxies with evident spiral features from galaxies that look smoother. An example of that kind of galaxies can be seen in the top cutout of Fig. 3.

The second binary classification trained with the N10 catalogue separates S0s from pure ellipticals (E). This model, $P_{\rm S0}$, is trained with galaxies having T-Type < 0 (from N10) and therefore is only meaningful for galaxies with negative values of the predicted T-Type. The reason for constructing this model is, again, the large scatter around intermediate T-Types, where the transition between Es and S0s occurs. Fig. 7 shows that, in the test sample, the model classifies as ellipticals 95 per cent of the galaxies from N10 with T-Type = −5 and as S0 83 per cent of the galaxies classified as S0/a from N10 (with a T-Type = 0). The predicted $P_{\rm S0}$ for the galaxies with T-Types in between is distributed around intermediate values, as expected. The performance of this $P_{\rm S0}$ is not as good as $P_{\rm LTG}$, which is reasonable given that what separates Es from S0s is rather subtle compared to the differences between ETGs and LTGs.

[6]Defined as the velocity dispersion at 0.25 arcsec derived from the MaNGA data-analysis pipeline (following the methodology described in Domínguez Sánchez et al. 2019).

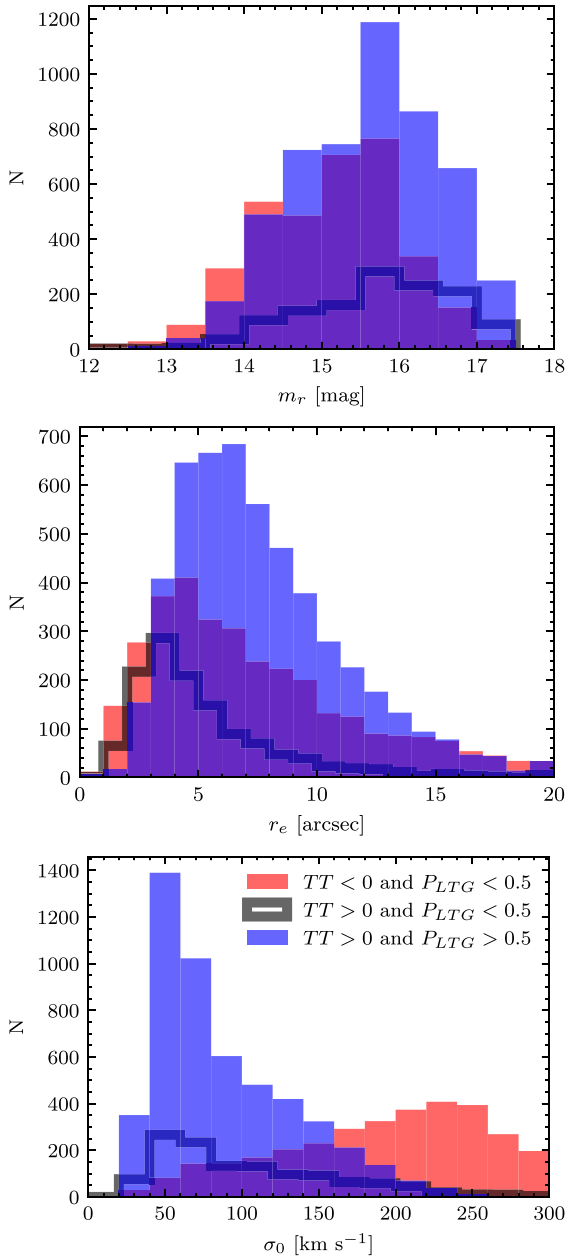

Figure 6. Galaxies having inconsistent T-Type and $P_{\rm LTG}$ classifications (open histograms) tend to be faint (top), have small angular sizes (middle), and small central velocity dispersions (bottom).

Hereafter – as done in F19, we classify the galaxies into three broad categories (E, S0, and S) by combining the T-Type and $P_{\rm S0}$ as follows:

(i) E: T-Type < 0 and $P_{\rm S0}$ < 0.5;
(ii) S0: T-Type < 0 and $P_{\rm S0}$ > 0.5;
(iii) S: T-Type > 0.

In some sections, we further subdivided the S galaxies into two subsamples:

(i) S1: 0 < T-Type < 3;
(ii) S2: T-Type > 3.

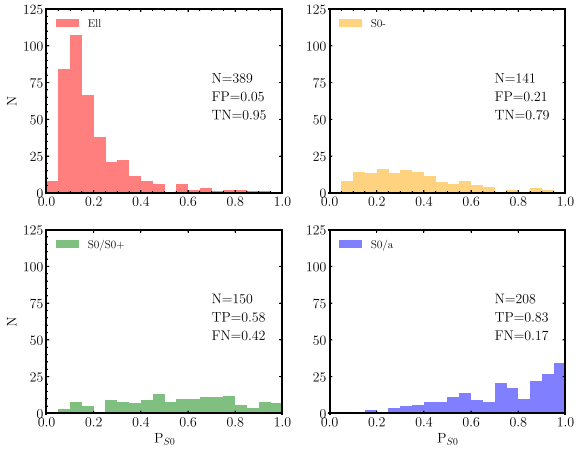

Figure 7. Distribution of P_{S0} separating Es from S0s for different classes from N10. Each panel also provides the total number of galaxies and the false-positive and true-negative rates (top) or the true-positive and false-negative rates (bottom).

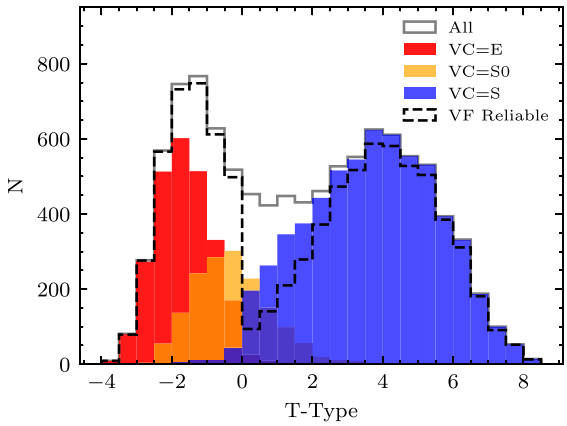

Figure 8. Predicted T-Types for galaxies according to their visual classification. The black dashed line shows galaxies with certain visual classifications (VF = 0). There is an evident drop of certain visual classifications for the intermediate T-Types (∼ 0), where the distinction between E/S0 and S is very subtle.

3.3.3 Visual classification

To have a more robust classification, and since the sample size allows it, we have also carried out a visual inspection of all the galaxies. The catalogue includes two columns reporting the results:

(i) Visual class, which corresponds to the visual classification assigned to each galaxy (VC = 1 for elliptical, VC = 2 for S0, VC = 3 for S/Irr, and VC = 0 for unclassifiable);
(ii) Visual flag, which reports the level of confidence in our visual class (VF = 0 for reliable and VF = 1 for uncertain classifications).

The visual classifications were based on the models (i.e. they were not blind) to spot evident misclassifications. Fig. 8 shows that the visual classifications correlate very well with the predicted T-Types. Galaxies with VC = 1 (E) peak around T-Type ∼ 2 and barely extend beyond T-Type > 0, galaxies with VC = 3 (S) peak around T-Type ∼ 4 and barely extend below T-Type < 0, while galaxies with VC = 2 (S0) tend to have intermediate T-Types with a tail that extends

Table 2. Comparison of automated (by combining the T-Type and P_{S0} models or according to P_{LTG}) and visual classifications (VC = 1 for elliptical, VC = 2 for S0, VC = 3 for S/Irr, and VC = 0 for unclassifiable).

TT+P_{S0}		VC_E (VC = 1)	VC_S0 (VC = 2)	VC_S (VC = 3)	VC_unc (VC = 0)
All					
E	2632	95	4	1	<1
S0	963	< 1	96	4	<1
S	6698	< 1	9	90	< 1
VF = 0					
E	2598	95	4	1	<1
S0	922	< 1	97	3	<1
S	5320	< 1	1	99	< 1
P_{LTG}		VC_E (VC = 1)	VC_S0 (VC = 2)	VC_S (VC = 3)	VC_unc (VC = 0)
All					
ETG	4908	52	32	16	< 1
LTG	5385	0	< 1	99	< 1
VF = 0					
ETG	3700	67	27	5	< 1
LTG	5140	0	< 1	99	< 1

In the leftmost column, we report the number of galaxies of each type, while the other columns show the percentage of each type with the corresponding visual classification for all galaxies (top) and for galaxies with reliable visual classifications (VF = 0, bottom).

to T-Type > 0. Thus, the reader should be aware that selecting a sample of S0 galaxies based on T-Type (i.e. T-Type < 0 and P_{S0} > 0.5) produces a pure but not complete sample. It also shows that galaxies with low-confidence visual classifications are mostly those with intermediate T-Types.

To quantify the comparison, Table 2 shows the number of galaxies classified as E, S0, or S according to the combination of the T-Type and P_{S0} values (as defined in Section 3.3.2) and the visual classification. Only 5 per cent of galaxies with elliptical morphologies (according to the models) are visually classified as S0 (4 per cent) or S (1 per cent). Similarly, for the galaxies classified as S0, there is 96 per cent agreement, with most of the discrepancies coming from galaxies visually classified as S. The more important mismatch is for the S sample, where 9 per cent of galaxies classified as S by the models are assigned type S0 after visual inspection. Note that, since the P_{S0} classification is only meaningful for galaxies with T-Type < 0, there is no 'model' to distinguish between S0 and S for galaxies with T-Type > 0.

We also note that one third of the galaxies with T-Type > 0 and VC = 2 (S0) have a large probability of being edge-on ($P_{\text{edge-on}}$ > 0.5, see Section 3.3.4), which explains the discrepancies (it is almost impossible to distinguish an S from an S0 when seen edge-on). In fact, if we focus on galaxies with reliable visual classifications (VF = 0), the agreement is significantly improved up to 99 per cent (see the upper panel with VF = 0 in Table 2).

The bottom part of Table 2 compares the visual classifications with the separation between ETGs and LTGs according to the P_{LTG} model. The LTG sample is very pure: 99 per cent of the galaxies with P_{LTG} > 0.5 were visually classified as S. On the other hand, the ETG

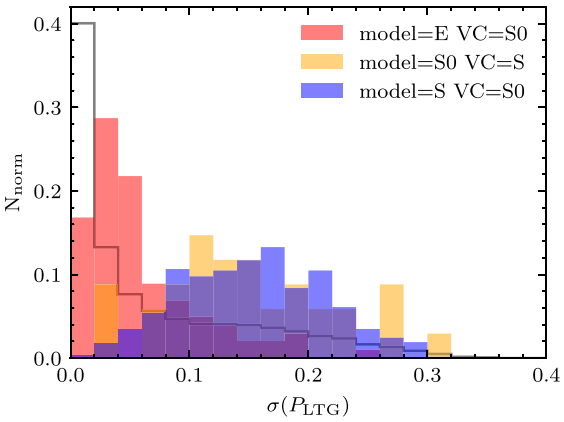

Figure 9. $P_{\rm LTG}$ uncertainty for the full sample (grey open histogram) and for galaxies whose visual classification is in disagreement with the classification obtained by combining the T-Type and $P_{\rm S0}$ models (red for Es visually classified as S0, orange for S0s visually classified as S, and blue for Ss visually classified as S0 – the other combinations are not shown due to their small number, as detailed in Table 2). The uncertainties for the 'misclassified' galaxies are significantly larger.

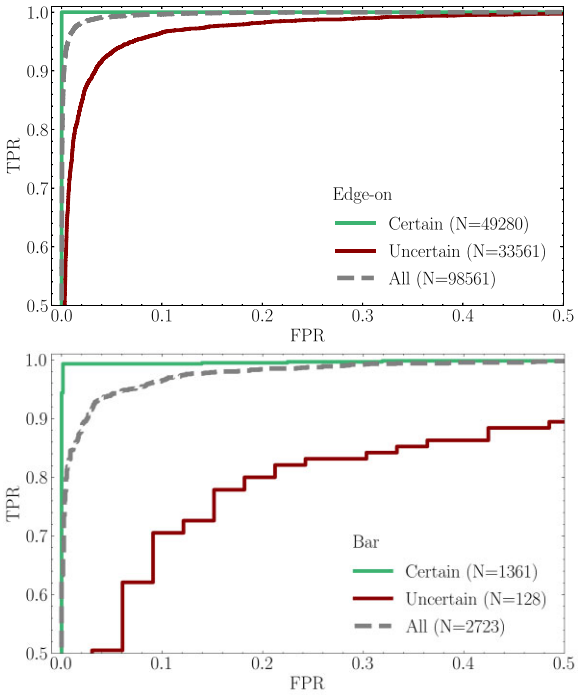

Figure 10. ROC curve – TPR versus FPR – for the edge-on (top) and bar (bottom) classifications. Grey, green, and red curves show the full test sample, the subset for which the model uncertainties are below the average, and the subset for which the uncertainties are larger than 3σ, with N representing the size of each subsample. Galaxies with more certain classifications (green) show better performance.

population ($P_{\rm LTG} \leq 0.5$) is composed of 52 per cent Es, 32 per cent S0s, and 16 per cent Ss, according to the visual classification. Considering only galaxies with reliable visual classifications, the fraction of E increases up to 67 per cent, while the fractions of S0 and S become 27 per cent and 5 per cent, respectively.

Interestingly, the percentage of galaxies classified as ETGs with reliable visual classifications is only 75 per cent (3700/4908), comparable to the same fraction for galaxies classified as S (5320/6698 = 79 per cent). While this may seem contradictory, it is because most of the galaxies with VF = 1 are those with T-Type > 0 (classified as S by the T-Type models) and $P_{\rm LTG} < 0.5$ (classified as ETGs by the $P_{\rm LTG}$ model), i.e. they are the faint and small galaxies difficult to classify and for which the T-Type and the $P_{\rm LTG}$ classifications disagree: 86 per cent of the galaxies with T-Type > 0 and $P_{\rm LTG} < 0.5$ have VF = 1 (1134/1315) and 78 per cent of galaxies with VF = 1 have T-Type > 0 and $P_{\rm LTG} < 0.5$ (1134/1453).

To test whether the model uncertainties correlate with the 'misclassifications', Fig. 9 shows the standard deviation of the value returned by $k = 10$ models separating ETGs from LTGs for galaxies whose visual classification is different from the classification obtained by combining T-Type and $P_{\rm S0}$ models. The uncertainties are significantly larger than those for the overall population. The same is true for the $P_{\rm S0}$ uncertainties for Es visually classified as S0s (not shown here).

On the contrary, the 'misclassified' galaxies do not show larger T-Type uncertainties than the full sample. Although that might seem unexpected, we must take into account that the T-Type model is a linear regression and is not aware of our 'artificial' separation between E/S0s and Ss at T-Type = 0 defined in Section 3.3.2. What happens for these galaxies is that they are close to that limit, with average values of T-Type = −0.3 and 0.67 for the 'misclassified' S0 and S, respectively. To quantify this uncertainty, we have generated 100 classifications based on the average T-Type by bootstrapping one of the $k = 15$ models in each realization. The percentage of galaxies which changes class (i.e. has median T-Type > or < 0 in a different realization) more than 10 times is ∼ 5 per cent for the overall population, while this happens for 40 per cent of the 'misclassified' galaxies, demonstrating that the T-Type scatter is consistent with these subsamples being more difficult to separate into the broad E/S0 and S classes.

Table 3. Accuracy, precision, recall, and F1 score for $P_{\rm edge\text{-}on}$ and $P_{\rm bar}$, as well as the number of galaxies in the test sample and the fraction of those labelled positive (according to GZ2).

Model	$N_{\rm test}$	% Positives	Accuracy	Precision	Recall	F1
$P_{\rm edge\text{-}on}$	98 561	14	0.98	0.87	0.98	0.93
$P_{\rm bar}$	2723	50	0.93	0.92	0.90	0.93

3.3.4 *Edge-on and bar classifications*

The catalogue includes two binary classifications based on the GZ2 catalogue (Willett et al. 2013): identification of edge-on galaxies ($P_{\rm edge\text{-}on}$) and identification of galaxies with bar signatures ($P_{\rm bar}$). The receiver operating characteristic (ROC) curve – the true-positive rate (TPR) versus false-positive rate (FPR) – is commonly used to assess the performance of binary classifications. Fig. 10 shows the ROCs for $P_{\rm edge\text{-}on}$ and $P_{\rm bar}$. The models perform well, with accuracy of 98 per cent and 93 per cent, respectively. The precision, recall, and F1 scores, given in Table 3, are defined as

$$\text{Precision} = \text{TP}/(\text{TP}+\text{FP});$$

$$\text{Recall} = \text{TP}/(\text{TP}+\text{FN});$$

$$\text{F1 score} = 2\times(\text{Recall} \times \text{Precision})/(\text{Recall} + \text{Precision}).$$

These values reach 100 per cent in all cases when computed for the subsample of galaxies with certain classifications, defined as those with standard deviations of the predicted models below the average (see also green lines in Fig. 10). This is reassuring, not just regarding the quality of our classifications, but also on the meaning of the reported model uncertainties.

Combining these classifications with the previous ones, only 9 galaxies classified as E (according to T-Type and P_{S0}) have $P_{edge\text{-}on} > 0.8$ (compared to 1551 for the full catalogue), while 53 have $P_{bar} > 0.8$ (compared to 1300 for the full catalogue). The number of ETGs (i.e. $P_{LTG} < 0.5$) with $P_{edge\text{-}on}$ or $P_{bar} > 0.8$ is 224 and 246, respectively (out of 4908 ETGs). These small numbers are expected, because E galaxies should not have bar features or disc shapes. We want to highlight that it is very difficult to distinguish between S0 and S galaxies when seen edge-on, and therefore, the separation between these two families for galaxies with large $P_{edge\text{-}on}$ is not accurate. In fact, one third of the galaxies with T-Type $>$ 0 and VC $=$ S0 have $P_{edge\text{-}on} > 0.5$ (see discussion related to Table 2).

3.4 The MDLM-VAC-DR17 catalogue

Table 4 shows the format of the MDLM-VAC-DR17 catalogue. The catalogue is released with the SDSS DR17 and is available online. It includes the classifications discussed in the previous sections plus additional information on the galaxies, such as their coordinates, redshifts, or duplicates.

In contrast to the DR15 version, MDLM-VAC-DR17 does not include P_{disc} or P_{bulge} values since the B/T and b/a values from the MPP-VAC are sufficient for providing such estimates. In addition, MDLM-VAC-DR17 no longer reports P_{merger} because it does not properly identify true (3D) mergers but rather projected neighbours (also see discussion in DS18).

3.4.1 A note on the selection of S0s

There are several ways to combine the morphological classifications provided in the MDLM-VAC-DR17 to construct samples of E, S0, and S galaxies. Depending on the scientific purpose, users can be more (or less) restrictive in order to obtain more pure (or complete) samples. The more restrictive selection would be to combine all the information included in the catalogue as follows:

(i) E: ($P_{LTG} < 0.5$) and (T-Type < 0) and ($P_{S0} < 0.5$) and (VC $=$ 1) and (VF $= 0$);
(ii) S0: ($P_{LTG} < 0.5$) and (T-Type < 0) and ($P_{S0} > 0.5$) and (VC $=$ 2) and (VF $= 0$);
(iii) S: ($P_{LTG} > 0.5$) and (T-Type > 0) and (VC $= 3$) and (VF $=$ 0).

This selection will return 2467, 891, and 5125 galaxies classified as E, S0, and S, respectively. However, there would be 1810 galaxies ($\sim$18 per cent of the sample) which do not belong to any of the classes.

If the selection is based on the reliable visual classifications (i.e. VF $= 0$), there will be 2474 Es, 1031 S0s, and 5325 Ss. But again, there is a large fraction of galaxies ($\sim$14 per cent) without a class.

Alternatively, the classification could be based on the combination of P_{LTG} and P_{S0} (which selects 2774 Es, 2134 S0s, and 5385 Ss) or on the combination of T-Type and P_{S0} (which selects 2632 Es, 963 S0s, and 6698 Ss). It is evident that the S0 galaxies are most affected by the classification criteria. Out of the 1315 galaxies with T-Type > 0 and $P_{LTG} < 0.5$, 541 (44 per cent) are visually classified as S0s, but only 33 of these have reliable visual classifications (VF $= 0$). As already noted throughout the text (see Section 3.3.2), S0 galaxies are sometimes very difficult to distinguish from Sa, and it is practically impossible to identify them when seen edge-on. Therefore, we strongly recommend that catalogue users test the effects different selection criteria may have on their scientific conclusions, especially when dealing with S0s.

4 COMBINING THE TWO CATALOGUES

In this section, we consider the benefits of combining the MPP-VAC-DR17 with the MDLM-VAC-DR17. Table 5 shows the frequency of FLAG_FIT for galaxies separated by morphological classes. E and S2 galaxies tend to be better described by a one-component fit, while S0 and S1 show a mixture of one- and two-component fits. This does not depend on whether we use T-Type or visual classifications to define the morphological class. On the other hand, more than half of the galaxies classified as ETGs or LTGs are better described by a one-component fit (this is due to the ETGs being a mix of E and S0 and LTGs a mix of S1 and S2).

Figs 11–14 show how the distributions of n, B/T, luminosity, central velocity dispersion, and $\epsilon \equiv 1 - b/a$ depend on morphology and FLAG_FIT. The morphological classes shown in the following are based on the VC values, with an additional separation between S1 and S2 at T-Type $= 3$.

The figures show that Es and S2s tend to be dominated by galaxies better described by a one-component fit (FLAG_FIT $= 1$), with a Sérsic index peaking around $\sim$4–6 and $\sim$1, respectively. The S2s with FLAG_FIT $= 1$ also tend to have larger ϵ.

The S0s and S1s tend to have more similar numbers of galaxies better described by a one- or two-component fit (FLAG_FIT $= 1$ or 2), with one-component objects tending to be less luminous and to have smaller σ_0. The B/T distributions of galaxies better described by a two-component fit (FLAG_FIT $= 2$) also show the expected trends: As one goes to later types, the peak of the distribution (and its skewness) shifts towards lower B/T. Since the PyMorph fits played no role in the morphological classification, the correspondence between FLAG_FIT and morphology in these figures is remarkable, and this is why we believe FLAG_FIT should be used in scientific analyses of our photometric catalogue.

5 COMPARISON WITH GALAXY ZOO

We now compare our MDLM Deep Learning morphologies with those of the GZ2 provided by Willett et al. (2013), in the same format as figs 24–26 in F19. We use the 'weighted fraction' GZ2 probability P_{smooth}, which is sometimes used as a proxy for ETGs and LTGs.

∗The top panel of Fig. 15 shows that objects with $P_{LTG} \leq 0.5$ – i.e. that are unlikely to be LTGs – tend to have large $P_{smooth} \geq 0.6$ (their images are smooth, with no disc features). Although we do not show it here, objects with $P_{LTG} \leq 0.5$ tend to have $P_{disc} \leq 0.3$ (i.e. they are unlikely to be discs), as expected. The bottom panel shows that, although most galaxies with $P_{smooth} \geq 0.6$ are dominated by Es or S0s, there is a significant fraction ($\sim$ 30 per cent) of objects that are Ss.

To check if our S1 and S2 classifications at $P_{smooth} \geq 0.6$ are incorrect, Fig. 16 shows the distribution of n for one-component galaxies (FLAG_FIT $= 1$) and B/T for two-component galaxies (FLAG_FIT

Table 4. Content of the Deep Learning Morphological catalogue for the DR17 MaNGA sample. This catalogue is available online.[7]

Column name	Data type	MDLM-VAC: The MaNGA Deep Learning Morphological VAC Description
INTID	int	Internal identification number
MANGA-ID	string	MaNGA identification
PLATEIFU	string	MaNGA PLATE-IFU
OBJID	long64	SDSS-DR14 photometric identification number
RA	double	Object right ascension (degrees)
DEC	double	Object declination (degrees)
Z	double	NSA redshift (from SDSS when not available)
DUPL_GR	int	Group identification number for a galaxy with multiple MaNGA spectroscopic observations
DUPL_N	int	Number of multiple MaNGA spectroscopic observations associated with DUPL_GR
DUPL_ID	int	Identification number of the galaxy in the group DUPL_GR
TType	double	T-Type value trained with the N10 catalogue; the value in the catalogue is the average of 15 k-fold models; T-Type < 0 for ETGs and T-Type > 0 for LTGs
TT_std	double	Standard deviation of the value returned by the $k = 15$ T-Type models; can be used as a proxy of the T-Type uncertainty
P_LTG	double	Probability of being LTG rather than ETG; trained with the N10 catalogue
P_LTG_std	double	Standard deviation of the value returned by the $k = 10$ $P_{\rm LTG}$ models; can be used as a proxy of the $P_{\rm LTG}$ uncertainty
P_S0	double	Probability of being S0 rather than pure elliptical, trained with the N10 catalogue; only meaningful for galaxies with T-Type ≤ 0 and not seen edge-on
P_S0_std	double	Standard deviation of the value returned by the $k = 10$ $P_{\rm S0}$ models; can be used as a proxy of the $P_{\rm S0}$ uncertainty
P_edge-on	double	Probability of being edge-on, trained with the GZ2 catalogue
P_edge-on_std	double	Standard deviation of the value returned by the $k = 10$ $P_{\rm edge\text{-}on}$ models; can be used as a proxy of the $P_{\rm edge\text{-}on}$ uncertainty
P_bar	double	Probability of having a bar signature, trained with GZ2 catalogue; edge-on galaxies should be removed to avoid contamination
P_bar_std	double	Standard deviation of the value returned by the $k = 10$ $P_{\rm bar}$ models; can be used as a proxy of the $P_{\rm bar}$ uncertainty
Visual_Class	int	Visual classification: VC = 1 for ellipticals, VC = 2 for S0, VC = 3 for S (including irregulars), and VC = 0 for unclassifiable
Visual_Flag	int	Visual classification flag: VC = 0 certain visual classification and VC = 1 uncertain visual classification

Table 5. Leftmost columns: fraction of galaxies of a given morphological type, which have PyMorph parameters (from Sérsic and/or SerExp, i.e. FLAG_FIT $\neq 3$) and reliable visual classification (VF = 0). Rightmost columns: fraction of galaxies with FLAG_FIT $\neq 3$, VF = 0 and having two components (FLAG_FIT = 2), one component (FLAG_FIT = 1), or for which both descriptions are equally acceptable (FLAG_FIT = 0).Please, keep the format of the table as close as possible to the original one. At least separate the two parts of the table with a line, in a similar way as in the original manuscript. The content of the table is less clear in this new version.

	Good fits (Sérsic and/or SerExp) (FLAG_FIT $\neq 3$)	Good fits and reliable visual class (FLAG_FIT $\neq 3$ + VF = 0)	Good Sérsic and SerExp fits and reliable visual class (FLAG_FIT = 0 + VF = 0)	Good Sérsic fits and reliable visual class (FLAG_FIT = 1 + VF = 0)	Good SerExp fits and reliable visual class (FLAG_FIT = 2 + VF = 0)
Class			Based on T-Type + $P_{\rm S0}$		
E	0.968	0.956	0.182	0.618	0.200
S0	0.964	0.930	0.123	0.450	0.427
S1	0.966	0.550	0.077	0.390	0.532
S2	0.966	0.920	0.048	0.640	0.312
			Based on visual classification		
E	0.966	0.946	0.184	0.630	0.186
S0	0.973	0.626	0.125	0.438	0.437
S1	0.966	0.692	0.078	0.392	0.529
S2	0.966	0.921	0.048	0.640	0.312
			Based on $P_{\rm LTG}$		
ETG	0.967	0.729	0.161	0.561	0.277
LTG	0.966	0.925	0.056	0.575	0.369

= 2). The Es clearly have larger n and B/T, and the Ss clearly have $n \sim 1$ and lower B/T (neither n nor B/T played a role in determining T-Type or $P_{\rm smooth}$). This strongly suggests that our classifications are appropriate, so (a) the presence of Ss with $P_{\rm smooth} > 0.6$ implies that conclusions about Es that are based on GZ2 $P_{\rm smooth}$ should be treated with caution; (b) selecting Es based on our MDLM T-Type classifications is much more robust than selecting on GZ2 $P_{\rm smooth}$.

In their analysis of DR15, F19 showed that selecting objects with $P_{\rm disc} < 0.3$ produces almost identical results as Fig. 16. This remains true in DR17, so we have not shown it explicitly.

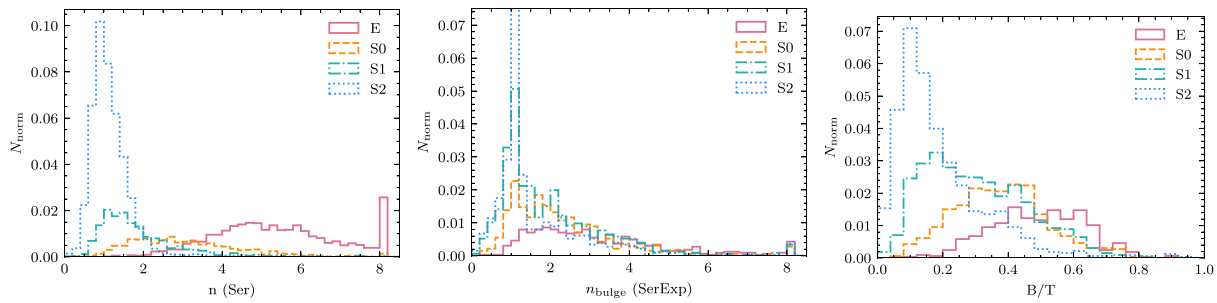

Figure 11. Distribution of *r*-band n and n_{bulge} for galaxies better described by a one-component fit (FLAG_FIT = 1, left-hand panel) and a two-component fit (FLAG_FIT = 2, middle panel). For galaxies with FLAG_FIT = 2, the distribution of B/T is also shown (right-hand panel). Galaxies are colour coded by morphology based on the VC values, with an additional separation between S1 and S2 at T-Type = 3. There is a strong correlation between morphological classes and photometric parameters. (The peak height of the S2 histogram is 0.153.)

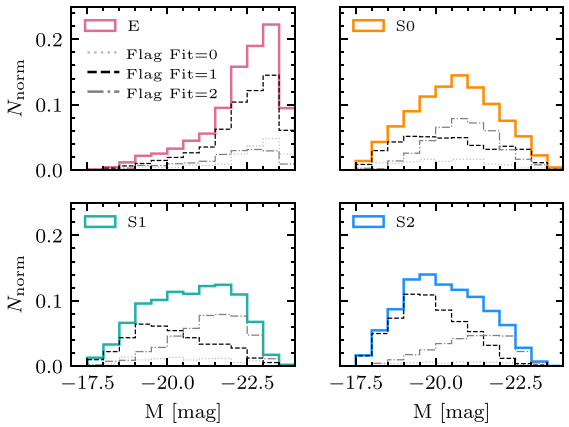

Figure 12. Distribution of *r*-band absolute magnitude for galaxies selected on the basis of their morphology and FLAG_FIT. Es are the brightest, while Ss peak at fainter absolute magnitudes, especially those with a preferred one-component fit (FLAG_FIT = 1 – i.e. without a bulge component).

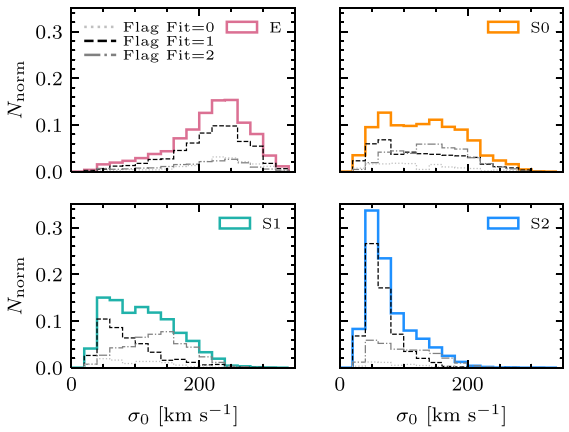

Figure 13. Same as the previous figure, but for central velocity dispersion σ_0. Es tend to have large σ_0, whereas for S2s σ_0 tends to be very small. S1s with a preferred one-component fit (FLAG_FIT = 1) have smaller σ_0.

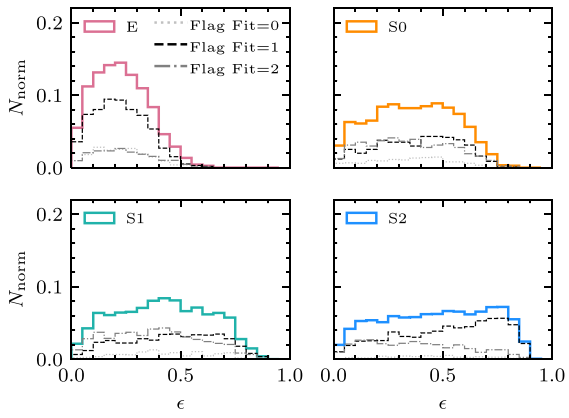

Figure 14. Same as the previous figure, but for observed $\epsilon \equiv 1 - b/a$. Es tend to be round, whereas S2s have a wide range of ϵ as expected for inclined discs. This correlation with morphology is striking because PyMorph *b/a* played no role in the classification.

6 SUMMARY AND CONCLUSIONS

We have presented the MPP-VAC-DR17 and MDLM-VAC-DR17 for the final data release of the MaNGA survey (which is part of the SDSS Data Release 17 – DR17).

The MPP-VAC-DR17 is an extension of the MPP-VAC-DR15 to include all the galaxies in the final MaNGA release. It provides photometric parameters for 2D surface brightness profiles for 10 293 observations (of which 10 127 are unique galaxies) in the *g*, *r*, and *i* bands. The MPP-VAC is identical to the one presented in F19, and its content is detailed in table 1 of F19. The only difference with the MPP-VAC-DR15 is the definition of the PA, given in this catalogue with respect to the camera columns in the SDSS 'fpC' images. The 2D light profile fittings are derived both for Sérsic and SerExp models. The catalogue contains a flagging system that indicates which fit is to be preferred for scientific analyses (FLAG_FIT = 1 for Sérsic, FLAG_FIT = 2 for SerExp, FLAG_FIT = 0 when both are acceptable). We urge users to pay attention to the preferences expressed by this flag since some fits may be unreliable.

The MDLM-VAC-DR17 is also an extension of the MDLM-VAC-DR15 presented in F19 and includes exactly the same entries as MPP-VAC-DR17. The MDLM-VAC-DR17 implements some changes

[7]https://www.sdss.org/dr17/data_access/value-added-catalogs/

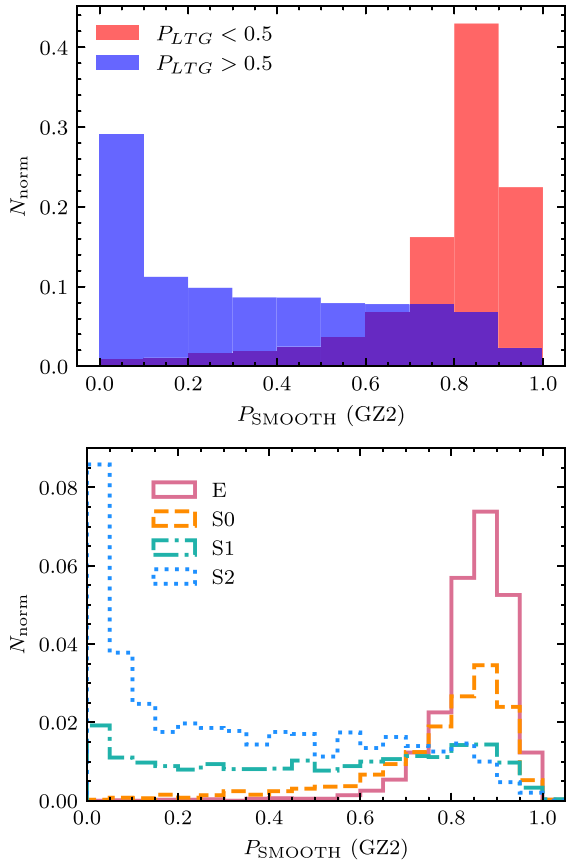

Figure 15. Distribution of $P_{\rm smooth}$ from GZ2 for DR17 galaxies divided according to our $P_{\rm LTG}$ (top) and T-Type + $P_{\rm S0}$ (bottom). Most galaxies with $P_{\rm smooth} \geq 0.6$ have $P_{\rm LTG} \leq 0.5$ and tend to be Es or S0s, although there is a non-negligible fraction of Ss.

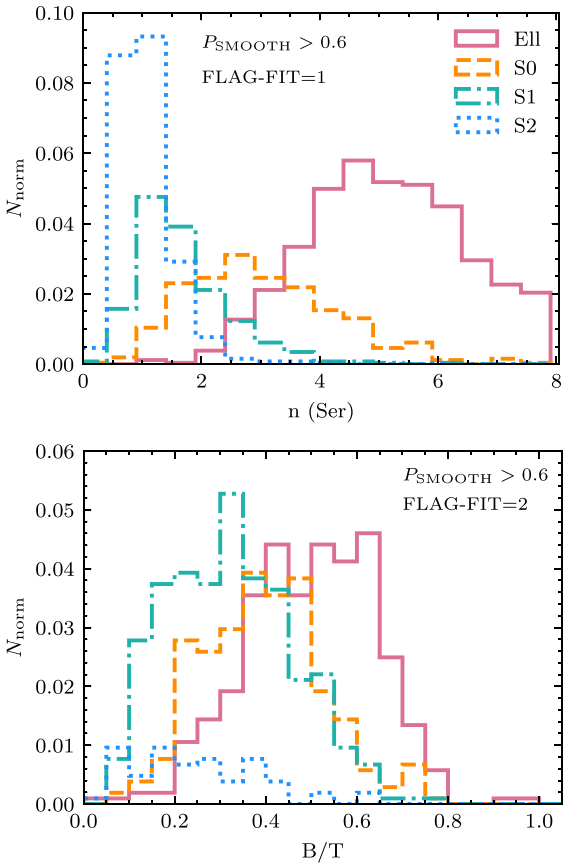

Figure 16. Distribution of r-band Sérsic index n for galaxies better described by a one-component fit (FLAG_FIT = 1, top) and B/T for galaxies better described by a two-component fit (FLAG_FIT = 2, bottom), for galaxies with GZ2 $P_{\rm smooth} > 0.6$. Objects with small n or B/T tend to be S, confirming that our morphological classification is correct. Results for $P_{\rm disc} \leq 0.3$ from GZ2 are nearly identical, so we have not shown them here.

compared to the previous release, and its content is detailed in Table 4. The main improvements of the new release are as follows:

(i) The low end of the T-Types is better recovered, with a smaller bias b compared to the previous version, thanks to a change in the CNN architecture (an additional dense layer was added, see Section 3.2).

(ii) A new binary model $P_{\rm LTG}$, which separates ETGs from LTGs in a complementary way to the T-Type, especially at intermediate T-Types where the scatter is larger.

(iii) All the classifications are trained using k-folding (with $k = 15$ for the T-Type and $k = 10$ for the rest of the models), and the value reported in the catalogue is the average of the k models.

(iv) We report the standard deviation of the outputs of the k models, which can be used as a proxy for their uncertainties (see Figs 9 and 10).

(v) A visual classification (VC = 1 for E, VC = 2 for S0, and VC = 3 for S/Irr) and a visual flag (VF = 0 for reliable classification and VF = 1 for uncertain classifications) are also included.

By combining the different classification models, we find the following:

(i) Galaxies having inconsistent T-Type and $P_{\rm LTG}$ classifications tend to be faint and small galaxies with small central velocity dispersion, i.e. they are difficult to classify. In general, they share some properties with LTGs, but they have no obvious spiral features.

(ii) The larger discrepancy between the visual classification and the one provided by the combination of the T-Type and the $P_{\rm S0}$ (as defined in Section 3.3.2) is for galaxies classified as S0 by the former and S by the latter (see Table 2). This fraction is reduced from 10 per cent to 1 per cent when only galaxies with VF = 0 are considered.

(iii) The larger discrepancy between the visual classification and $P_{\rm LTG}$ is for galaxies classified as S by the former and ETG by the latter (see Table 2). This fraction is reduced from 16 per cent to 5 per cent when only galaxies with VF = 0 are considered.

By combining the two catalogues MPP-VAC-DR17 and MDLM-VAC-DR17, and despite the changes to the morphological classification, we find similar results to those found in F19:

(i) There is a strong correlation between the morphological classification and the values of n, $n_{\rm bulge}$, and B/T (see Fig. 11).

(ii) E galaxies tend to be bright (more negative M_r values), have large central velocity dispersion σ_0, and small ellipticity ϵ, while the trend is the opposite for the S galaxies, especially S2 (see Figs 12–14).

(iii) Separating galaxies according to the FLAG_FIT, we observe that Es and S2s tend to be dominated by galaxies better described by a one-component fit (FLAG_FIT = 1). On the other hand, the S0s and S1s tend to have more similar numbers of objects described by a one-

or two-component fit (FLAG_FIT = 1 and 2), with one-component objects tending to be less luminous and having smaller σ_0.

(iv) Since the PyMorph fits played no role in the morphological classification, the correspondence between FLAG_FIT and morphology is remarkable: FLAG_FIT should be used in scientific analyses of our photometric catalogue.

(v) We find a significant fraction of S galaxies with $P_{\rm smooth} > 0.6$ from GZ2 (Fig. 15). Most of these galaxies have have $n \sim 1$ and low B/T (Fig. 16), consistent with being disc galaxies. Therefore, as F19 noted previously, conclusions about Es that are based on GZ2 $P_{\rm smooth}$ should be treated with caution.

ACKNOWLEDGEMENTS

The authors thank the referee for useful comments, which helped to improve the quality of the paper. This work was supported in part by NSF AST-1816330. HDS acknowledges support from PIE2018-50E099 project: Cross-field research in space sciences. The authors gratefully acknowledge the computer resources at Artemisa, funded by the European Union ERDF and Comunitat Valenciana as well as the technical support provided by the Instituto de Física Corpuscular, IFIC (CSIC-UV).

Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions. SDSS-IV acknowledges support and resources from the Center for High Performance Computing at the University of Utah. The SDSS website is www.sdss.org.

SDSS-IV is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration, including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, Harvard–Smithsonian Center for Astrophysics, the Chilean Participation Group, the French Participation Group, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU)/University of Tokyo, the Korean Participation Group, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatário Nacional/MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

DATA AVAILABILITY

The catalogues described in this paper are part of the final data release of the MaNGA survey and will be released as part of the SDSS DR17. The catalogues are available at https://www.sdss.org/dr17/data_access/value-added-catalogs. The code used for the deep-learning algorithm may be shared upon request.

REFERENCES

Aguado D. S. et al., 2019, ApJS, 240, 23
Allen J. T. et al., 2015, MNRAS, 446, 1567
Bernardi M., Meert A., Vikram V., Huertas-Company M., Mei S., Shankar F., Sheth R. K., 2014, MNRAS, 443, 874
Bertin E., Arnouts S., 1996, A&AS, 117, 393
Blanton M. R. et al., 2017, AJ, 154, 28
Bundy K. et al., 2015, ApJ, 798, 7
Cappellari M. et al., 2011, MNRAS, 413, 813
Domínguez Sánchez H., Huertas-Company M., Bernardi M., Tuccillo D., Fischer J. L., 2018, MNRAS, 476, 3661 (DS18)
Domínguez Sánchez H., Bernardi M., Brownstein J. R., Drory N., Sheth R. K., 2019, MNRAS, 489, 5612
Drory N. et al., 2015, AJ, 149, 77
Fischer J.-L., Bernardi M., Meert A., 2017, MNRAS, 467, 490
Fischer J. L., Domínguez Sánchez H., Bernardi M., 2019, MNRAS, 483, 2057 (F19)
Graham M. T. et al., 2018, MNRAS, 477, 4711
Greene J. E. et al., 2017, ApJ, 851, L33
Gunn J. E. et al., 2006, AJ, 131, 2332
Lakshminarayanan B., Pritzel A., Blundell C., 2016, preprint (arXiv:1612.01474)
Law D. R. et al., 2015, AJ, 150, 19
Law D. R. et al., 2016, AJ, 152, 83
Meert A., Vikram V., Bernardi M., 2013, MNRAS, 433, 1344
Meert A., Vikram V., Bernardi M., 2015, MNRAS, 446, 3943 (DR7)
Meert A., Vikram V., Bernardi M., 2016, MNRAS, 455, 2440
Nair P. B., Abraham R. G., 2010, ApJS, 186, 427 (N10)
Peng C. Y., Ho L. C., Impey C. D., Rix H.-W., 2002, AJ, 124, 266
Sánchez S. F. et al., 2012, A&A, 538, A8
Simard L., Mendel J. T., Patton D. R., Ellison S. L., McConnachie A. W., 2011, ApJS, 196, 11 (S11)
Smee S. A. et al., 2013, AJ, 146, 32
Vega-Ferrero J. et al., 2021, MNRAS, 506, 1927
Vikram V., Wadadekar Y., Kembhavi A. K., Vijayagovindan G. V., 2010, MNRAS, 409, 1379
Wake D. A. et al., 2017, AJ, 154, 86
Willett K. W. et al., 2013, MNRAS, 435, 2835
Yan R. et al., 2016a, AJ, 151, 8
Yan R. et al., 2016b, AJ, 152, 197

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