

<https://doi.org/10.15407/knit2022.01.003>  
UDC 524.7+52-735

<sup>1,2</sup> **I. B. VAVILOVA**, Dr. Sci. Hab. in Phys.&Math., Prof.  
Head of the Department of the Extragalactic Astronomy and Astroinformatics  
E-mail: irivav@mao.kiev.ua  
<https://orcid.org/0000-0002-5343-1408>

<sup>3</sup> **V. KHRAMTSOV**, PhD student, Department of Astronomy & Space Informatics  
E-mail: vld.khramtsov@gmail.com  
<https://orcid.org/0000-0003-1744-7071>

<sup>1</sup> **D. V. DOBRYCHEVA**, PhD. in Phys.&Math.,  
Senior Scientist of the Department of the Extragalactic Astronomy and Astroinformatics  
E-mail: daria@mao.kiev.ua  
<https://orcid.org/0000-0001-5557-3453>

<sup>1,4</sup> **M. Yu. VASYLENKO**, PhD student,  
Junior Scientist of the Department of the Extragalactic Astronomy and Astroinformatics  
E-mail: vasmax@mao.kiev.ua  
<https://orcid.org/0000-0002-7714-0779>

<sup>1</sup> **A. A. ELYIV**, Ph.D. in Phys.&Math.,  
Senior Scientist of the Department of the Extragalactic Astronomy and Astroinformatics  
E-mail: andrii.elyiv@gmail.com  
<https://orcid.org/0000-0001-6215-1048>

<sup>1</sup> **O. V. MELNYK**, Ph.D. in Phys.&Math.,  
Senior Scientist of the Department of the Extragalactic Astronomy and Astroinformatics  
E-mail: melnykol@gmail.com

<sup>1</sup> Main Astronomical Observatory of the NAS of Ukraine  
27 Akademik Zabolotny Str., Kyiv 03143 Ukraine

<sup>2</sup> Astronomical Observatory of the I. I. Mechnikov National University of Odesa  
1v, Marazliyivska Str., Odesa 65014 Ukraine

<sup>3</sup> Institute of Astronomy, V.N. Karazin Kharkiv National University  
35 Sumska Str., Kharkiv 61022 Ukraine

<sup>4</sup> Institute of Physics, National Academy of Sciences of Ukraine  
Nauka av., 46, Kyiv 02000 Ukraine

## MACHINE LEARNING TECHNIQUE FOR MORPHOLOGICAL CLASSIFICATION OF GALAXIES FROM SDSS. II. THE IMAGE-BASED MORPHOLOGICAL CATALOGS OF GALAXIES AT $0.02 < Z < 0.1$

---

*We applied the image-based approach with a convolutional neural network model to the sample of low-redshift galaxies with  $-24^m < M_r < -19.4^m$  from the SDSS DR9. We divided it into two subsamples, SDSS DR9 galaxy dataset and Galaxy Zoo 2 (GZ2) dataset, considering them as the inference and training datasets, respectively. To determine the principal parameters of galaxy morphology*

Цитування: Vavilova I. B., Khramtsov V., Dobrycheva D. V., Vasylenko M. Yu., Elyiv A. A., Melnyk O. V. Machine learning technique for morphological classification of galaxies from SDSS. II. The image-based morphological catalogs of galaxies at  $0.02 < z < 0.1$ . *Space Science and Technology*. 2022. **28**, № 1 (134). С. 03–22. <https://doi.org/10.15407/knit2022.01.003>

defined within the GZ2 project, we classified the galaxies into five visual types (completely rounded, rounded in-between, smooth cigar-shaped, edge-on, and spiral). Using GZ2 galaxy morphology classification, we were able to define 34 morphological features of galaxies from the inference set of our SDSS DR9 sample, which do not match with the GZ2 training set. As a result, we created the morphological catalog of 315782 galaxies at  $0.02 < z < 0.1$ , where morphological five classes and 34 detailed features were first defined for 216148 galaxies by image-based CNN classifier. For the rest of galaxies, the initial morphological classification was re-assigned as in the GZ2 project.

Our method shows the promising performance of morphological classification attaining  $>93\%$  of accuracy for five classes morphology prediction except the cigar-shaped ( $\sim 75\%$ ) and completely rounded ( $\sim 83\%$ ) galaxies. Main results are presented in the catalog of 27378 completely rounded, 59194 rounded in-between, 18862 cigar-shaped, 7831 edge-on, 23119 spiral in the inference data set of the studied SDSS sample. As for the classification of galaxies by their detailed structural morphological features, our CNN model gives the accuracy in range 92–99% in depending on features, number of galaxies with the given feature in the inference dataset, and, of course, the galaxy image quality. As a result, for the first time we assigned 34 morphological detailed features (bar, rings, number of spiral arms, mergers, etc.) for more than 160000 low-redshift galaxies from the SDSS DR9. We demonstrate for the first time that implication of the CNN model with adversarial validation and adversarial image data augmentation improves classification of smaller and fainter SDSS galaxies with  $\eta < 17.7$ .

The proposed CNN model allows solving a bunch of galaxy classification problems, for example, such as a quick selection of galaxies with a bar, bulge, ring, and other morphological features for their subsequent analysis.

**Keywords.** Methods: data analysis, machine learning, convolutional neural networks; galaxies: general, morphological classification, galaxy catalogs, large-scale structure of the Universe.

## 1. INTRODUCTION

Since the beginning of extragalactic astronomy and the first catalogs of galaxies by Ch. Messier, F. W. Hershel and J. F. W. Hershel, J. L. E. Dreyer, the image-based morphological classifications of galaxies have played a vital role in reflecting the evolutionary history of various types of galaxies and the large-scale structure of the Universe as a whole (Davis et al. [26], Peebles [74], Barrow & Saich [8], Yang et al. [110], Bundy et al. [14], Peng et al. [75], Reid et al. [78], Leung et al. [60]).

Manual galaxy morphological classification as the most precise method requires extensive usage of human resources, either from highly skilled professionals or, in some cases, amateur astronomers and volunteers such as in the Galaxy Zoo (GZ) project [109]. Current and near-term galaxy observational surveys as the SDSS, LSST, DES, KiDS, SKA, the Euclid satellite, JWST, etc., are approaching the Exabyte scale multiwavelength databases of hundreds of millions of galaxies, which is impossible to classify manually. For instance, the Vera C. Rubin Observatory (LSST), which will be operated starting from 2022, is expected to generate about 30 TB of data per night, revealing  $\sim 20$  million galaxies over this time [49], more than the Sloan Digital Sky Survey (SDSS) over its lifetime [10]. It is also worth mentioning the Euclid survey, which aims to detect billions of galaxies over 15,000 square degrees of the celestial sky

[3], and other big data astronomical projects (see, for example, a recent review of multiwavelength surveys and catalogs by Vavilova et al. [99]). Moreover, the human mind is not able to comprehend complex correlations in the diverse space of parameters, and multidimensional mathematical analysis is the best tool for determining the various common features between different types of objects. All that exaggerates the interest to use the alternatives in the form of machine learning (ML) techniques, including deep learning (DL), for the classification, forecasting, and discovery of various properties of galaxies (see, for example, [13, 43, 98]).

In this context, we note several recent prospective applications of Convolutional Neural Networks (CNNs) to classify the galaxies by their different parameters.

Cabayo et al. [15] have demonstrated the CNN capability to avoid distorting effects when extract the galaxy photometry from astronomical images as Lumos architecture. Exploiting the PAU imaging survey, the authors combined a CNN and a Mixture Density Network that allowed them to measure the photometry of a blended galaxy with the high accuracy. Diego et al. [29], in their work with DL in classifying early- and late-type galaxies in the OTELO and COSMOS databases, have used optical and infrared photometry and available shape parameters (the Sérsic index or the concentration index). Regardless of

slight differences in the photometric bands used in each catalog, their neural network architecture operates well with missing data.

The distance moduli and photometric redshift estimates benefit from the ML utilization into the big data sets, which provide a wide number of galaxy features for learning. Pasquet-Itam & Pasquet [72] used DL for classifying, detecting, and predicting photometric redshifts of quasars in SDSS. In works by Kugler & Gianniotis [58], Speagle & Eisenstein [88], Disanto et al. [30], Salvato et al. [82], and Elyiv et al. [42] the machine learning methods were applied to assign and predict photometric redshifts within large-scale galaxy surveys with good accuracy. The GAN approach serves as a basis for restoring galaxy distribution in the Zone of Avoidance (Schawinski et al. [85], Vavilova et al. [97]) and generating dark matter structures in cosmological simulations (Diakogiannis et al. [28]).

Among the CNNs modeling in tasks of multi-wavelength sky surveys we note as follows: search for blazar candidates in the Fermi-LAT Clean Sample [51]; boosted decision tree for detecting the faint gamma-ray sources with future Cherenkov Telescope Array [57, 81]; infrared color selection of Wolf-Rayet star candidates in our Galaxy using the Spitzer GLIMPSE catalog [70]; cosmic string searches in 21-cm temperature CMB maps [24]; neural network-based Faranoff-Riley classifications of radio galaxies from the Very Large Array archive [4]; deep learning classification of compact and extended radio source from Radio Galaxy Zoo project [62]; CNN for morphological assignment to radio-detected galaxies with active nuclei [64]. Scaife et al. in recent work presented the first application of group-equivariant CNN to radio galaxy and their image translations, rotations, and reflections [84].

Deep learning is promising for generating various synthetic catalogs and mock images, which helps to interpret the observational data [53] and to discover new galaxies as, for example, high- $z$  «blue nuggets» from the CANDELS survey [48]; to reveal structural properties of dark matter halos to their assembly history and environment [21]; to establish a topology of the large-scale structure of the Universe in LCDM cosmological simulation [90]; to separate the radiation from active galactic nuclei and star-forming gal-

axies with recognition method based on Deep Neural Network [19].

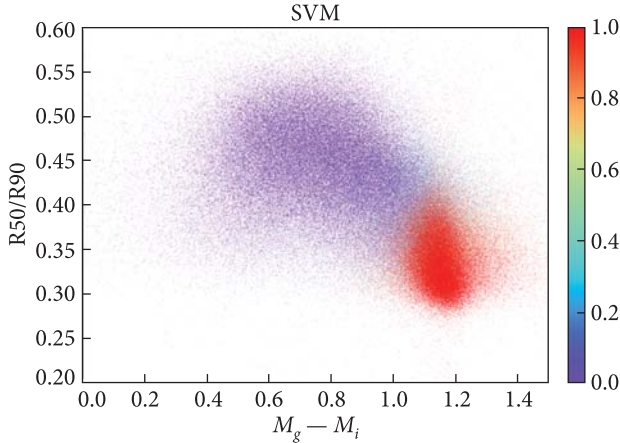
As for the discovery of new classes of celestial bodies with CNN, we highlight the works related to the gravitational lenses [50, 55] and the transient events and objects as supernovae, gamma-ray bursts, jets, etc. For example, the Catalina Real-Time Transient Survey serves as the platform for their detection and monitoring (see, for example, [31, 65–66]) as well as the Zwicky Transient Factory [9] as the LSST precursor and ML models implementation in synoptic sky surveys.

The CNN models have played a crucial role in analyzing data streams from the Advanced Laser Interferometer Gravitational-Wave Observatory (LIGO) detectors allowing to register gravitational wave signals from coalescing black hole binaries. Among such works are ones on the training an ML system for real-time glitch classification [111], on the classification of gravitational wave signals, events, and instrumental noise [45–46] as well as to test the theories on binary black hole mergers upon which the models are based [47].

So, CNNs reliably manage with tasks for a variety of image-based classification, regression, prediction, and discovery of galaxies and other celestial bodies (see, obstacle recent overviews by Baron [6], Fluke & Jacobs [43], Vavilova et al. [99]).

In our work, we used a deep CNN model for the image-based morphological classification of  $\sim 300000$  galaxies ( $0.02 < z < 0.1$ ) from SDSS DR9. To do this, we divided galaxies by their images [112] into five morphological classes (completely rounded, rounded in-between, cigar-shaped, edge-on, and spiral galaxies) as in the Galaxy Zoo 2 (GZ2) project. In our previous works [54, 93], we used binary classification but, as it turned out, it is difficult to correctly divide galaxies into two classes using the assigned label of galaxies from the GZ2. We investigate the problem of differences in these datasets and suggest ways to overcome adversarial validation. We also used our CNN model to predict 34 detailed structural morphological features (bar, ring, bulge, mergers, number of spiral arms, etc.) of these galaxies with were labeled in GZ2 project [107].

The structure of our paper is as follows. The sample of galaxies is described in Section 2. CNN model



**Fig. 1.** Diagram of color indices ( $g-i$ ) and inverse concentration indexes  $R50/R90$  of the studied low-redshift galaxies from the SDSS DR9 after applying the Support Vector Machine (SVM) method: red color — early E (from elliptical to lenticular) and blue color — late L (from S0a to irregular Im/BCG) morphological types. Color bar from 0 to 1 shows SVM probability to classify galaxy as the late to the early morphological type [98]

as the image morphological classifier, training and inference galaxy datasets are presented in Section 3. The created galaxy catalogs and results are given in Section 4 as well as discussion and conclusion in Section 5 and 6, respectively.

## 2. SAMPLE OF LOW-REDSHIFT GALAXIES FROM THE SDSS

We used a representative sample of the 316031 SDSS galaxies at  $0.02 < z < 0.1$  (with velocities correction on the velocity of Local Group,  $V_{LG} > 1500$  km/s). This sample was studied by us practically as “galaxy by galaxy” in previous works [18, 32–36, 41, 67, 77, 92–95, 98, 100, 105]. Our most previous research was to apply the ML photometry-based approach for binary morphological classification of these galaxies [98] and to create the catalog of their morphological types (early and late) obtained with the Support Vector Machine with an accuracy of 96.4 % [102].

The main stages in preparing this sample were as follows. A preliminary set of galaxies at  $z < 0.1$  with the absolute stellar magnitudes  $-24^m < M_r < -13^m$  from the SDSS DR9 contained of  $\sim 724000$  galaxies. Following the SDSS recommendation, we limited the sample to  $m_r < 17.7$  in  $r$ -band to avoid typical statis-

tical errors in spectroscopic flux. After excluding the images with stars and artifact objects as well as the duplicates of galaxy images, the final sample consisted of  $N = 315782$  galaxies. To clear the sample from segmented images of the same galaxy, we used our code based on the minimal angle distances between such SDSS objects. The absolute stellar magnitude of the galaxy was obtained by the formula

$$M_r = m_r - 5 \lg(D_L) - 25 - K_r(z) - ext_r,$$

where  $m_r$  is the visual stellar magnitude in  $r$ -band,  $D_L$  is the luminosity distance,  $ext_r$  is the Galactic absorption in  $r$ -band in accordance to [86],  $K_r(z)$  is the cosmological  $k$ -correction in  $r$ -band according to [22, 23]. The color indices were calculated as

$$M_g - M_i = (m_g - m_i) - (ext_g - ext_i) - (K_g(z) - K_i(z)),$$

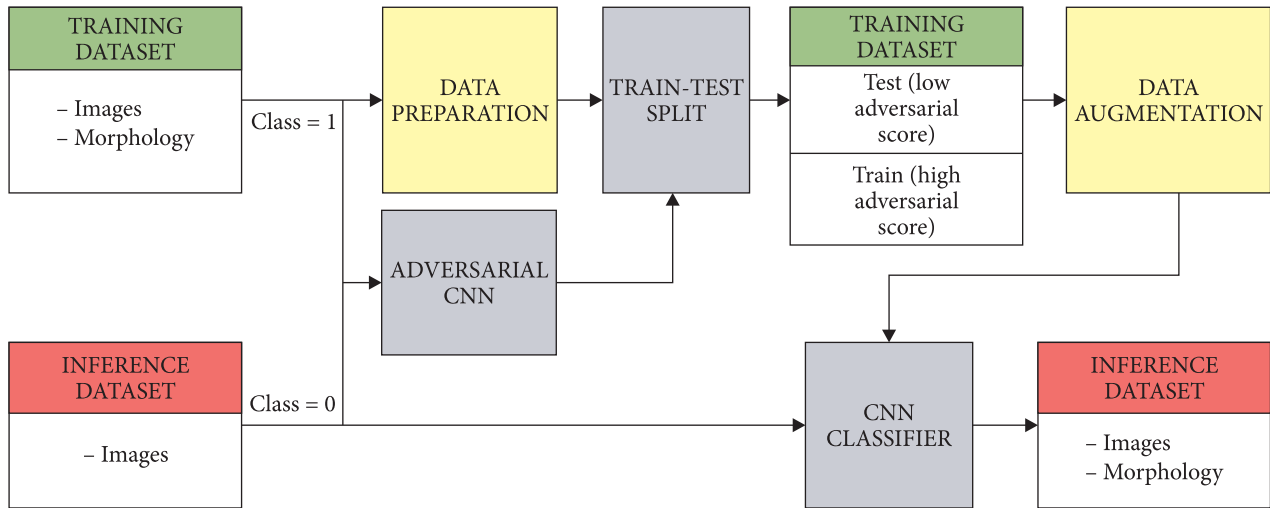
where  $m_g$  and  $m_i$  are visual stellar magnitude in  $g$ - and  $i$ -band;  $ext_g$  and  $ext_i$  are the Galactic absorption in  $g$ - and  $i$ -band;  $K_g(z)$  and  $K_i(z)$  are the  $k$ -correction in  $g$ - and  $i$ -band, respectively.

We found that Support Vector Machine gives the highest accuracy exploiting different galaxy classification techniques: human labeling, multi-photometry diagrams, and five supervised ML methods. Namely, 96.1 % early E and 96.9 % late L types of galaxies [98]. We verified dependencies between accuracy and redshifts, human labeling bias, the overlap of different morphological types for galaxies with the same color indices, edge-on and face-on galaxy shape to determine the ability of each method to predict the galaxy morphological type. Distribution of 315782 galaxies from the studied SDSS sample by their morphological type (early and late) is given in Fig. 1.

## 3. CNN MODELS FOR IMAGE-BASED MORPHOLOGICAL MULTI-LABEL CLASSIFICATION OF GALAXIES

The studied sample of 315782 galaxies from SDSS DR9 is tightly overlapped with the data from the Galaxy Zoo 2, GZ2 [109]. It allows us to divide it into two datasets: “*inference dataset*” of 143410 galaxies, which do not match with GZ2 dataset; “*training dataset*” of 172372 galaxies, which match the galaxies from our studied sample (Fig. 2).

For each galaxy from both datasets we have their SDSS images, but morphological classes are defined



**Fig. 2.** Block-scheme of the image-based classification of galaxies with CNN model for five morphological classes. Algorithm consists of the training/inference datasets, image data preparation, adversarial validation, special train-test split via adversarial scores, data augmentation, CNN classifier

only for galaxies from the training GZ2 dataset. The 315782 RGB images of galaxies were requested from the SDSS (<http://skyserver.sdss.org/dr15/en/help/docs/api.aspx##cutout>). They are composed of *gri*-bands [63] having color scaling, each of  $100 \times 100 \times 3$  pixels ( $39.6 \times 39.6$  arcsec) in each channel of the RGB image.

**3.1. Galaxy morphological image classification into five classes by shape.** It is important to note the principal difference between galaxy images in our inference dataset and the GZ2 training dataset. Galaxies from the inference dataset are much shallower than those from the GZ2 dataset. As we mentioned in Section 2, the galaxies from the studied sample were pre-selected via  $m_r < 17.7$  limit following the SDSS recommendation. This affects the value of the 90 % Petrosian flux parameter. Thereat, the galaxies, which do not match the GZ2 dataset, are fainter and smaller on average than galaxies from the training GZ2 dataset. To get around this problem, we used an adversarial CNN to compare these two datasets (training and inference). Namely, we trained it on all galaxy images of our sample, passing the class ‘0’ for the inference dataset and class ‘1’ for the training dataset (Fig. 2).

To develop the CNN model based on the images of galaxies, we used the GZ2 assigned labels for five morphological classes by shape: completely round-

ed, rounded in-between, cigar-shaped, edge-on, and spiral galaxies.

We provided additional data cleaning of 172372 galaxy images from the training sample and took into consideration only those galaxies for which GZ2’s volunteers gave the most votes for a more accurate result (Fig. 2). It turned out to be 72738 galaxies. The criteria for each image of the galaxy were defined in GZ2 project [109]; their description is available through web-site <https://data.galaxyzoo.org/>. The criteria with (\*\_count) prefix indicate the number of votes of volunteers; other criteria correspond to the debiased fraction of votes assigned in the GZ2 catalog as (\*\_debiased). So, we applied criteria for galaxies belonging to the five morphological classes by shape as follows

- completely rounded: smooth (number of votes  $>0.469$ ), completely\_round ( $>0.469$ ), smooth\_count ( $>25$ ), completely\_round\_count ( $>25$ );
- rounded in-between: smooth ( $>0.469$ ), rounded\_in\_between ( $>0.5$ ), smooth\_count ( $>25$ ), rounded\_in\_between\_count ( $>25$ );
- cigar-shaped: smooth ( $>0.469$ ), cigar\_shaped ( $>0.5$ ), smooth\_count ( $>25$ ), cigar\_shaped\_count ( $>25$ );
- edge-on: features\_or\_disk ( $>0.43$ ), edgeon\_yes ( $>0.602$ ), features\_or\_disk\_count ( $>25$ ), edgeon\_yes\_count ( $>25$ );

- spiral: features\_or\_disk (>0.43), edgeon\_no (>0.715), spiral (>0.619), features\_or\_disk\_count (>25), edgeon\_no\_count (>25), spiral\_count (>25).

These criteria with vote scores >0.4 and higher were found to be quite good for providing reliable image-based morphological classification of galaxies. The galaxy image data preparation and augmentation for our CNN model is described in detail in our paper [56].

The adversarial CNN resulted in the fact that the training dataset contains galaxies, which properties are not common with inference one. The useful parameter to solve this task is the adversarial score, which means the probability of the galaxy with some feature to be similar to the galaxy with the same feature from the GZ2 training set. So, we can test galaxies with a low adversarial score from the training dataset in a way to train them on galaxies with the high adversarial score from the training set (Fig. 2). On the step of this “train-test split”, we randomly selected ~9000 galaxies with an adversarial score less than 0.7 to test the CNN classifier (test-split training dataset), and the rest part of this dataset (~63000) to train the CNN classifier (train-split training dataset). Because this score is related to the presence of fainter and smaller galaxies in the inference galaxies, we have done the image data augmentation of galaxies from the training sample (we decreased the intensity of pixels and reduced the image size). Distribution of predicted labels vs. true labels of five morphologi-

cal classes for these ~9000 galaxies with the high adversarial score as the confusion matrix is shown in Table 1. One can see in Table 1 that our model for CNN classifier guarantees > 93 % of accuracy for rounded in-between, edge-on, and spiral morphological classes, 83 % for the completely rounded and 75 % for the cigar-shaped galaxies.

As a result of the CNN classifier, we got the morphological classes of 72738 galaxies from the training set as follows: 19468 completely rounded, 27321 rounded in-between, 3235 cigar-shaped, 4099 edge-on, and 18615 spiral galaxies.

Meanwhile, knowing the morphological class of galaxies from the training dataset, we are able to classify galaxies from the inference dataset with CNN. We compared a few CNN models for the five-class morphological classification. Following our previous works [54, 93] the best neural network for our task is DenseNet-201.

Our CNN model consists of two main parts. The first one is the convolutional part, where CNN performs the image processing with a gradually decreasing size. The highlight of CNN architectures is to use the fully connected layers at the tail. This tail corresponds to the neural network classifier, which transforms the output of the convolutional part into the dense layer, the number of neurons in which is equal to the number of classes (see, for example, <http://cs231n.stanford.edu/>). So, the second part of our model is the fully connected part, where the processed galaxy image comes through a few layers of connected neurons up to the last layer, consisting of five neurons, the output of which corresponds to the probability of a galaxy being one of five defined classes.

Our CNN model was completed by the two dense layers of neurons (with the number of neurons equal 128 and 5, respectively) and, after, by the global max-pooling. The activation functions at the tail of the CNN model were the same as in adversarial validation. As an optimizer, we used the ADAM with an initial learning rate of  $10^{-4}$ ; the optimizer minimized the categorical\_crossentropy loss function (see, in detail, [56]).

**3.2. Galaxy morphological multi-label classification by 34 features.** Besides classification into five morphological classes, the image galaxies from the training dataset attribute 37 parameters of the detailed morphology. They are estimated for each galaxy according

*Table 1. Distribution of predicted labels vs. assigned labels as the probabilities for galaxy to belong to one of five morphological classes (CNN classifier for test-split training galaxy dataset of 9000 galaxies). Each row represents the fraction of galaxies from a certain class (defined at the horizontal row) to be classified as galaxies from other classes*

True labels	Predicted labels				
	Completely rounded	Rounded in-between	Cigar-shaped	Edge-on	Spiral
Completely rounded	<b>0.83</b>	0.16	0	0.00038	0.012
Rounded in-between	0.054	<b>0.93</b>	0.0047	0.00025	0.015
Cigar-shaped	0	0.17	<b>0.75</b>	0.065	0.017
Edge-on	0	0.0076	0.049	<b>0.93</b>	0.0092
Spiral	0.0075	0.022	0.0017	0.0092	<b>0.96</b>

to volunteers' answers and form the decision tree [80, 109]. The principal restriction for classification with CNN is the presence of the only parameter, which characterizes this class of objects on an image [59]. We introduced the "similarity learning" approach: if two galaxies have similar images, then their morphological parameters are similar. In other words, we exploited the galaxy images from the training dataset, which are most similar to the galaxy images from the inference dataset by their 37 morphological detailed features. The algorithm is finding the nearest galaxies from the training dataset to the target galaxy of the inference dataset in the penultimate CNN layer of multi-parameter space and assigns the attributes of nearest neighboring galaxies to the target galaxy.

We also applied the adversarial validation to predict 37 detailed morphological features of galaxy images from the inference dataset with some adversarial score. The three very sparse features were removed from the consideration. So, we worked with the inference dataset of 160471 galaxies and with 34 galaxy morphological features (bar, ring, various number of spiral arms, disks, dust lane, merger etc.).

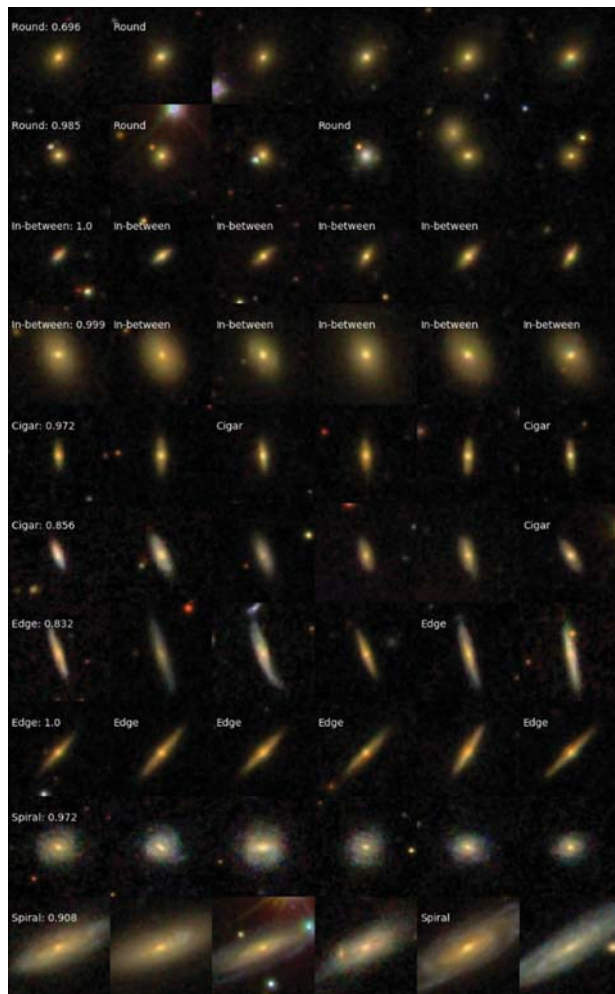
These morphological features are listed in Table 2. The names of features ("parameter") are given in the first column as they are labeled by the GZ2's volunteers. The numbers of galaxies in the inference dataset with the given feature are given in column 2. Columns 3–5 contain the ROC AUC<sup>1</sup> classification quality metric (ROCtest) for galaxies of GZ2 dataset. Two resulting accuracy scores, measured with ROC AUC classification quality metric, give the score for the model trained with adversarial augmentations (ROCtest AUG, column 4) and for the model, trained without these augmentations (ROCtest NOAUG, column 3). As the binary classification quality metric we used *Area Under Receiver Operating Characteristic Curve* (ROC AUC). To estimate the ROC AUC, one needs to plot the following curve: the fraction of true positives out of the positives (TPR = true positive rate) versus the fraction of false positives out of the negatives (FPR = false positive rate) at various threshold settings to estimate the area under this curve. For ideal classification, ROC AUC

equals 1; for random, it is 0.5. For more information about ROC AUC classification quality metric measuring, see the paper by Bradley [12]. Column 5 gives the difference between ROCtest AUG and ROCtest NOAUG values.

Table 2. Quality metrics of morphological detailed features of galaxies from the inference dataset

Feature's name [109]	Number of galaxies with the given feature in the inference dataset	ROCtest NOAUG, %	ROCtest AUG, %	ROCtest difference, %
1	2	3	4	5
smooth	624	89.25	88.59	-0.66
features_or_disk	19 770	92.54	91.88	-0.66
star_or_artifact	6	95.36	97.63	2.28
edgeon_yes	2 079	98.81	98.65	-0.16
edgeon_no	7 504	97.21	96.82	-0.39
bar	90	93.99	92.41	-1.57
no_bar	1 762	90.69	89.80	-0.90
spiral	1 199	93.40	92.88	-0.52
no_spiral	47	86.30	84.78	-1.52
no_bulge	63	98.36	98.35	-0.01
odd_yes	1 096	94.78	93.37	-1.41
odd_no	61 537	84.62	83.51	-1.11
completely_round	6 018	96.17	95.60	-0.58
rounded_in_between	20 107	92.31	91.46	-0.85
cigar_shaped	12 434	97.96	97.73	-0.23
ring	52	96.97	96.43	-0.54
irregular	41	96.74	96.94	0.20
other	4	95.93	89.20	-6.74
merger	8	91.79	88.89	-2.90
dust_lane	4	99.39	99.40	0.02
bulge_shape_rounded	18	96.73	96.27	-0.47
bulge_shape_no_bulge	664	98.65	98.52	-0.13
arms_winding_tight	3	89.45	88.60	-0.85
arms_winding_medium	2	75.33	77.59	2.26
arms_winding_loose	100	94.95	94.41	-0.54
arms_number_2	338	90.55	89.99	-0.56
arms_number_3	1	93.54	93.47	-0.07
arms_number_4	1	93.84	85.45	-8.39
arms_number_more_than_4	1	97.79	97.51	-0.27
arms_number_cant_tell	1	86.13	86.07	-0.06

<sup>1</sup> ROCAUC is the Receiver Operating Characteristics, which is determined with Area Under Curve quantitate value.



**Fig. 3.** A set of the inference galaxies (first column) with their five nearest neighbors from the GZ2 training dataset (the rest four columns). Each row represents the morphological class, which is intrinsic to the galaxy from the inference set. The number in the left upper corner of each image of the inference galaxies is a value of the corresponding probability of being this galaxy in a given class. Some classes of galaxies from the GZ2 training set are pointed out when possible (not all of the GZ2 galaxies fit our criteria for CNN classifier)

One can compare these scores and estimate the degree of influence of image data augmentations on the classification quality of a trained model. Scores are given for the dataset of 9000 galaxies, expanded with a fraction of galaxies, which do not pass the criteria of morphological classification.

There is a particular discrepancy in the numbers of galaxies with detailed morphological features from

the training dataset. It can be explained, among other things, by the fact that the GZ' volunteers did not notice certain morphological details while the CNN classifier found.

#### 4. IMAGE-BASED CATALOGS OF LOW-REDSHIFT SDSS GALAXIES CLASSIFIED BY FIVE MORPHOLOGICAL CLASSES AND 34 MORPHOLOGICAL FEATURES

Applying the CNN classifier to the inference galaxy dataset (low panel in Fig. 2), we took into account the following labels of galaxies: predictions of belonging to one of five morphological classes (Table 1) and 34 detailed morphological features (Table 2). The augmentation procedures for image data of galaxies («decrease in stellar magnitude and correction of sizes») from the training dataset are described in our work [56].

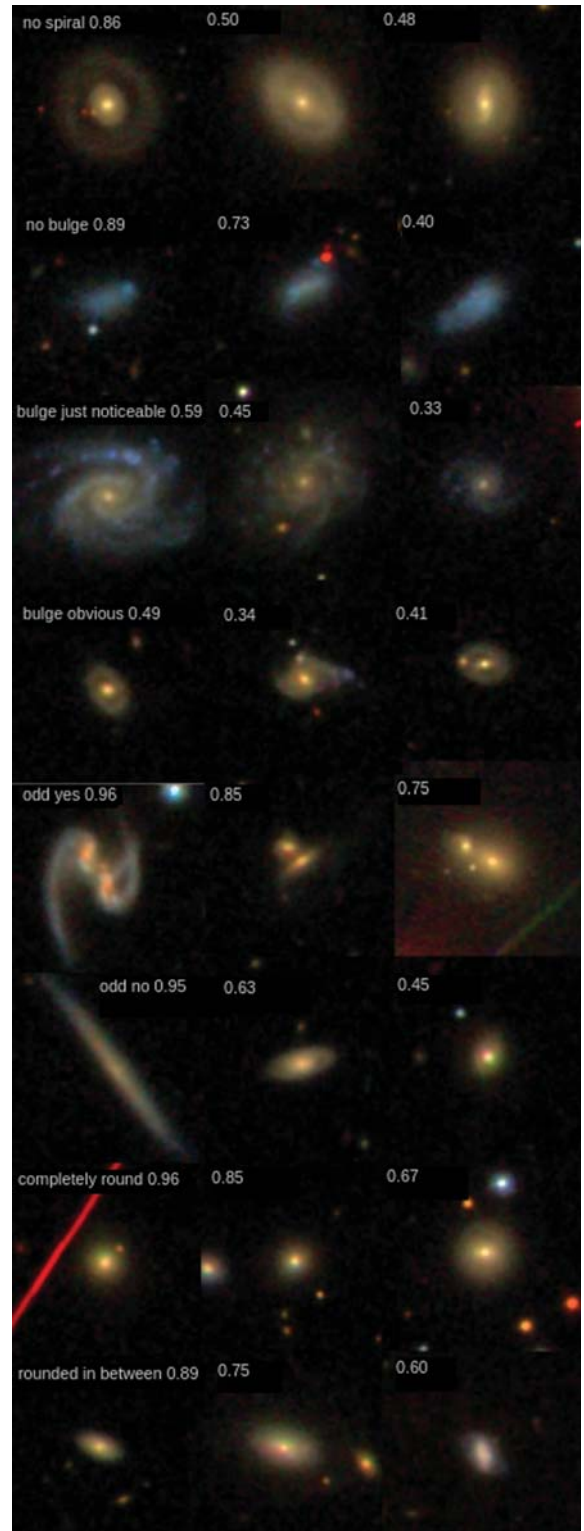
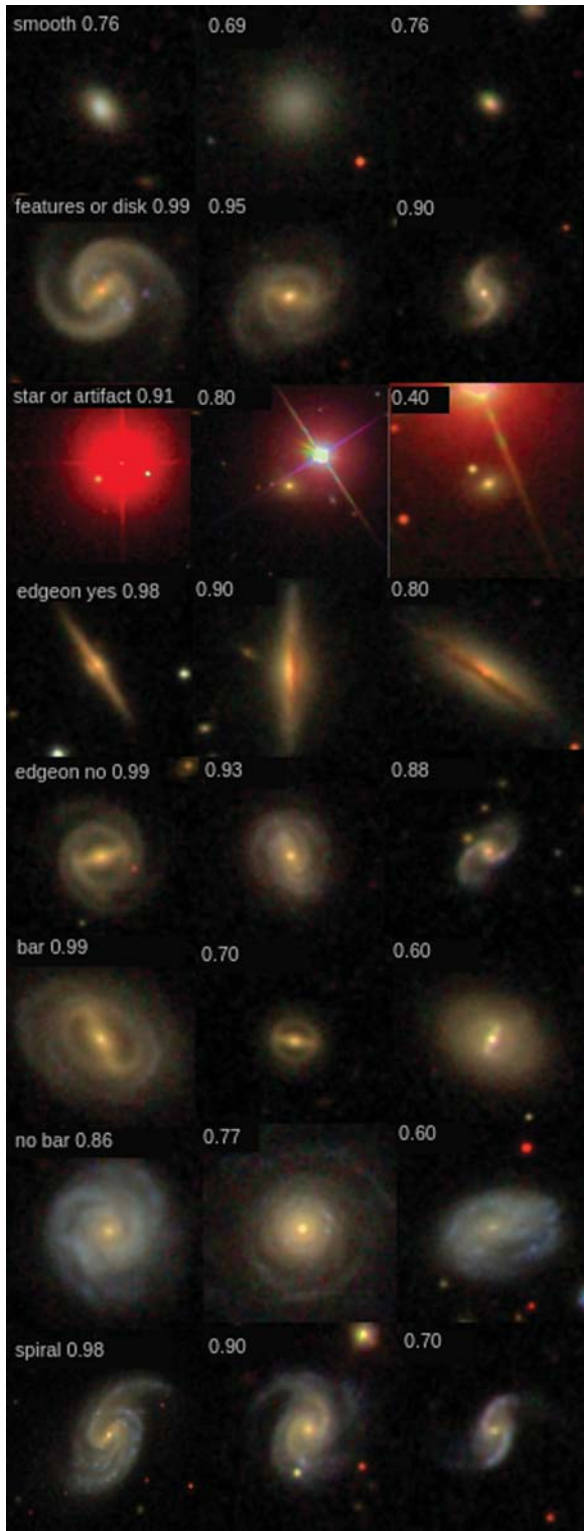
We have trained our CNN classifier having attained the overall accuracy of 89.3% on the test set of 9000 galaxies obtained after splitting the training galaxy set (see, the distribution of predicted labels vs. true labels as the probabilities for galaxy to belong to one of five morphological classes in Table 1). Assuming that a galaxy is in a certain morphological class if the probability is the highest one, we found that the inference dataset contains of 27378 completely rounded, 59194 rounded in-between, 18862 cigar-shaped, 7831 edge-on, and 23119 spiral galaxies.

So, a common classification of the studied sample of 315782 low-redshift SDSS galaxies with  $m_r < 17.7$  and  $V_{LG} > 1500$  km/s into five morphological classes consists of the following parts:

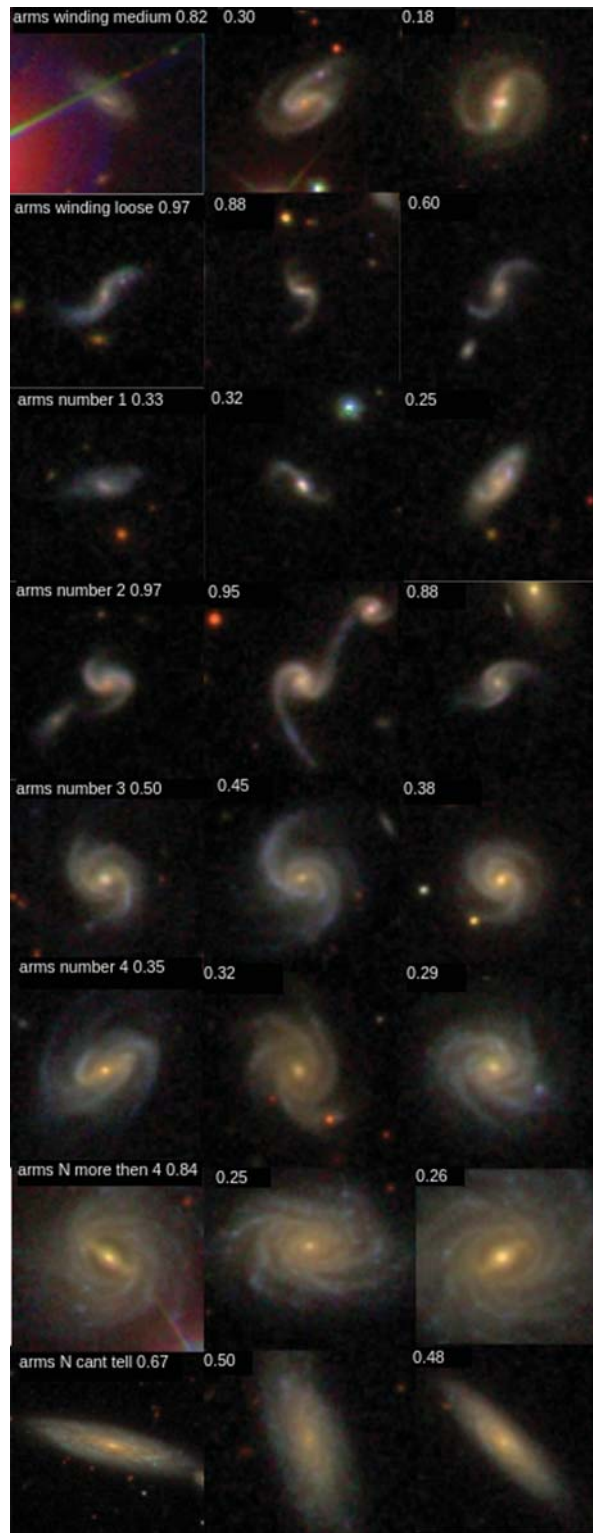
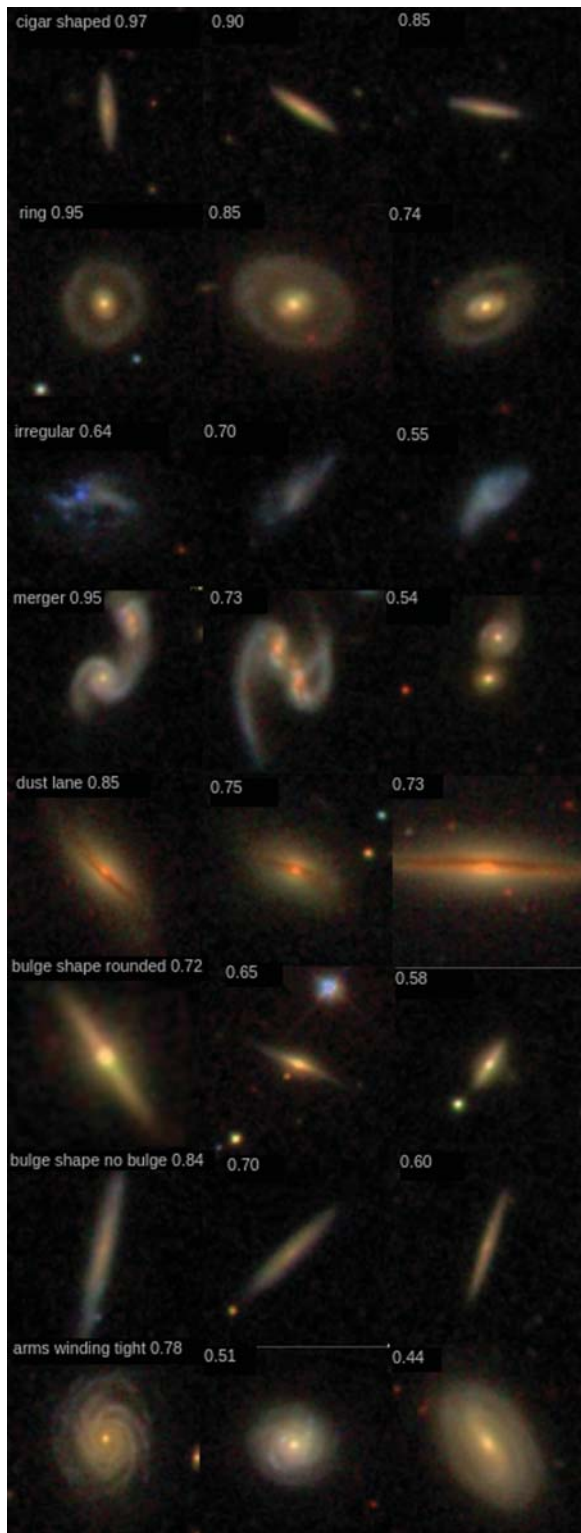
- 72738 galaxies from the training dataset and 143410 galaxies from the inference dataset, which have undergone CNN model with the high adversarial score and the accuracy pointed in Table 1. It turned out 46846 completely rounded, 86515 rounded in-between, 22097 cigar-shaped, 13930 edge-on, and 41738 spiral galaxies.
- 105560 galaxies from the studied sample were not classified with CNN model because of their low adversarial score (98534 galaxies) or technical reasons (7026 galaxies). We left the initial morphological classification for these galaxies into five classes as it was assigned in the GZ2 project.

Examples of the inference galaxies with their five nearest neighbors (in multi-label parametric space)





**Fig. 4** (parts 1–2). The examples of galaxies with some morphological features (smooth, features or disk, star or artifact, edge on yes, edge on no, bar, no bar, spiral, no spiral, no bulge, bulge just noticeable, bulge obvious, odd yes, odd no, completely rounded, rounded in between, see Table 2) from the inference SDSS dataset with their two nearest neighbors from the GZ2 training dataset



**Fig. 4** (parts 3—4). The examples of galaxies with some morphological features (cigar shaped, ring, irregular, merger, dust lane, bulge shape rounded, bulge shape no bulge, arms winding tight, arms winding medium, arms winding loose, arms number 1, arms number 2, arms number 3, arms number 4, arms N more then 4, arms N can't tell, see Table 2) from the inference SDSS dataset with their three nearest neighbors from the GZ2 training dataset

from the GZ2 training dataset classified onto five morphological classes with a given accuracy are shown in Fig. 3.

Also, we determined the number of galaxies that passed the 0.5 cut-off for the acceptance of the detailed morphological features. The number of such galaxies with the certain feature in the inference dataset is presented in Table 2. The examples of inference galaxies with some of these morphological features (ring, bar, merger, irregular, arms\_winding\_loose etc.) with two nearest neighbors (in multi-label parametric space) from the GZ2 training dataset are shown in Fig. 4. The mosaics of galaxies in Fig. 3 and

Fig. 4 illustrate well how our CNN classifier work in finding the similar morphological features of various galaxies, for instance, to find the edge-on galaxies turned to the observer under the same angles (see seventh and eighth rows in Fig. 3) or to find the galaxies with similar morphological features as ring, bar or bulge (Fig. 4, part 3–4).

We used additional morphological parameters such as a bar or ring to predict the presence of these features in galaxies from our inference set. Because these features are not mutually exclusive, we introduce a different approach to determine their types. Namely, we assumed that the penultimate layer of a

**Table 3. Examples of five galaxies from the Catalog of morphological classes of low-redshift galaxies from SDSS DR9, which have the highest probability to belong to the completely rounded, rounded in-between, cigar-shaped, edge-on, and spiral morphological classes**

Identifier	Coordinates		Probability				
	Ra	Dec	Completely rounded	Rounded in-between	Cigar-shaped	Edge-on	Spiral
SDSS ObjID 1237655468061294796 2MASS J15063732+0113452	226.655531	1.229153	<b>0.9976</b>	0.0023	0.0602	0.0554	0.0417
SDSS ObjID 1237660960793428158 2MASS J09283138+3507071	142.130736	35.118625	0.0001	<b>0.9998</b>	0.0424	0.0365	0.0263
SDSS ObjID 1237648722321473797 2MASS J14352175+0050296	218.840662	0.841566	0.0439	0.0004	<b>0.9995</b>	0.0257	0.0093
SDSS ObjID 1237648722290606136 2MASS J09532021+0041516	148.334281	0.697737	0.0338	0.1159	0.0004	<b>0.9997</b>	0.0053
SDSS ObjID 1237660240312795177 2MASS J03114746-0024108	47.947759	-0.402997	0.0059	0.0038	0.0414	0.0279	<b>0.9994</b>

**Table 4. Examples of five galaxies from the Catalog of morphological classes of low-redshift galaxies from SDSS DR9, which have the highest probability to have ring, bar, irregular shape, dust lane, two spiral arms**

Identifier	Coordinates		Probability				
	Ra	Dec	Bar	Ring	Irregular	Dust lane	Arms number 2
SDSS ObjID 1237662198283633002 2MASS J15105965+0829209	227.748594	8.489149	<b>0.9922</b>	0.0657	0.0000	0.0000	0.3236
SDSS ObjID 1237665101137641844 2MASS J14201619+3017044	215.067510	30.284630	0.0002	<b>0.9528</b>	0.0000	0.0000	0.0129
SDSS ObjID 1237668311087972614	229.630080	14.677338	0.0004	0.0000	<b>0.7009</b>	0.0000	0.0113
SDSS ObjID 1237662262714368418 2MASS J15344528+0549459	233.688611	5.829438	0.0003	0.0000	0.0000	0.8559	0.0061
SDSS ObjID 1237662196139163742 2MASS J12453820+4332122	191.409199	43.536667	0.0002	0.0000	0.0000	0.0000	<b>0.9778</b>

neural network consisting of 128 neurons must clearly characterize the galaxy. In its turn, the neighboring galaxies in the multidimensional parameter space must have the same characteristics. By determining the optimal number of nearest neighbors for the most accurate prediction and the optimal value of trimming the likelihood of signs, we test our hypothesis on a deferred GZ2 dataset and found a good confirmation.

We created the catalog of 315782 low-redshift galaxies from SDSS DR9, where morphological classes and detailed features were defined for the first time for 216148 galaxies by an image-based CNN classifier. For the rest of the galaxies (with the lower adversarial score) the initial morphological classification was re-assigned as in the GZ2 project. These new data will be added to our previous “Binary morphology SDSS galaxies catalog” [102]<sup>2</sup>, see also: <http://skyserver.sdss.org/dr9> (SDSS DR9 Home Page). Examples of five galaxies from this catalog, which have the highest probability to belong to the completely rounded, rounded in-between, cigar-shaped, edge-on, and spiral morphological classes, are listed in Table 3. Examples of five galaxies from this catalog, which have the highest probability to assign ring, bar, dust lane, and other morphological, structural features, are listed in Table 4.

## 5. DISCUSSION

Classification of morphological types and features of galaxies is one of the cornerstones for extragalactic astronomy and observational cosmology. Galaxies of different morphological types are distributed non-uniformly across the sky and along the cosmological scale. The early-type galaxies predominate in the central part of the cluster. They also have larger masses, less gas, higher velocity dispersion, and diverse stellar population than the late-type galaxies (see, for example, these works [5, 7, 17, 25, 32, 33, 52, 71, 76, 77, 79, 89, 91, 100, 106]).

Astronomical surveys have accumulated a large number of galaxy images and data that need deep scientific exploration. For this purpose, it is very convenient to use relevant methods for a reliable galaxy morphological automated classification. There are

many different options for sorting galaxies by type now, but each has its own drawbacks. For example, broad spectra of data are lost in spectroscopy classification because not all galaxies have spectra of good enough quality. Classifications based on the photometry give an error when trying to classify red spirals [92, 98], i.e., galaxies with a high content of old stars or interacting galaxies which affect the photometric characteristics of each other [11, 68, 73].

In favor of our choice of CNN as a basic model, we will mention several papers related to the image-based and photometry-based galaxy morphology classification with accentuating on the SDSS and Galaxy Zoo samples [61].

Cabrera et al. [16] explained how the human labeled biases in morphological photometry-based classification could be reduced through supervised ML. This coincides with our conclusion [101], where we discuss which factors and properties of galaxies exactly affect the accuracy of supervised methods. In that paper, we concluded that one could not get the accuracy significantly exceeding 76 % when using the GZ2 data as a training set for ML with the photometry-based approach. One of the reasons is the attribution of irregular galaxies in the GZ2, which have the redder color indices, to the elliptical (early) type, and vice versa the elliptical galaxies with the bluer color indices to the spirals. In any case, the morphology obtained with the ML trained photometric parameters demonstrates significantly less bias than morphology based on citizen-science classifiers. This conclusion is also important for galaxies with low surface brightness galaxies [40].

For example, Cheng et al. [20] used a set of ~2800 galaxies from Dark Energy Survey with visual classification from GZ1 to compare ML methods for galaxy classification: CNN, K-Nearest Neighbor, Logistic Regression, Support Vector Machine, Random Forest, and Neural Networks. They revealed that CNN was the most successful method in their study giving an accuracy of 99.4 % for the binary morphological classification of ellipticals and spirals. As for the combination of photometry and image galaxy SDSS data with Galaxy Zoo labels, we note the work by Hayat et al. [1], who applied self-supervised representation learning. Mittal et al. [69] introduced the data augmentation-based MORphological Clas-

<sup>2</sup> <https://cdsarc.cds.unistra.fr/viz-bin/cat/J/A+A/648/A122>

sifier Galaxy (daMCOG CNN) using convolutional neural network and obtained a testing accuracy of 98 %. Their datasets of 4614 images were collected from SDSS Image Gallery, Galaxy Zoo challenge, and Hubble Image Gallery.

Walmsley et al. [107] used Bayesian CNNs and a novel generative model of Galaxy Zoo volunteer responses to infer posteriors for the visual morphology of galaxies. They show that training Bayesian CNNs with active learning requires up to 35–60 % fewer labeled galaxies depending on the morphological feature being classified. They concluded that in the synergy of human and machine intelligence, the Galaxy Zoo would be able to classify surveys of any conceivable scale, providing massive and detailed morphology catalogs to support research into galaxy evolution. These authors in their next paper [108] used Galaxy Zoo data (SDSS DR8) and DECaLS data [27] to provide the detailed visual morphology measurements in *grb*-bands with Bayesian DL classifier for 314000 galaxies brighter than  $m_r = 17.77$  at  $z < 0.15$ . Applying RGB image construction and various methods for the data processing, these authors were able to improve decision trees for GZ2' volunteer classification of galaxy morphological features.

Several useful catalogs were developed with the GZ classification scheme. Willett et al. [109] issued a catalog of morphological types from the GZ2 in the synergy with the SDSS DR7, which contains more than 16 million morphological classifications of 304122 galaxies and their finer morphological features (bulges, bars, and the shapes of edge-on disks as well as parameters of the relative strengths of galactic bulges and spiral arms). Simmons et al. [87] cross-verified morphological features of ~48000 galaxies from the CANDELS survey and GZ project (clumpiness, bar instabilities, spiral structure, merging). It allowed them to create a list of galaxies with featureless discs at  $1 \leq z \leq 3$ , representing “a dynamically warmer progenitor population to the settled disc galaxies seen at later epochs”.

Dominguez-Sanchez et al. [37] presented a morphological catalog for ~670000 SDSS-galaxies in two options: T-type, related to the Hubble sequence, and GZ2 types. Their models with DL for the GZ2 type questions have the highest accuracy (>97 %), when applied to a test dataset with the same characteristics

as the one used for the training dataset. In the recent work [38], they presented the MaNGA Deep Learning Morphological Value Added Catalog as a part of the SDSS DR17. This catalog includes a number of morphological properties: e.g. a T-Type, a finer separation between elliptical and S0, as well as the identification of edge-on/barred galaxies and a separation between early/late types.

Vega-Ferrero et al. [103] presented morphological classifications of ~27 million galaxies from the Dark Energy Survey (Data Release 1). They provided CNN model to classify these galaxies by early and late types (accuracy ~87 %) as well as by face-on and edge-on galaxies (accuracy 73 %). In work by Domingo-Sanchez et al. related to the algorithm for preparing this largest automated morphological catalog up to date [39], the authors describe how their trained SDSS image data of galaxies were transferred on Dark Energy survey images. They also modeled fainter objects by simulating what the brighter objects with well-determined classifications would look like if they were at higher redshifts. This is the same as we applied in our approach.

The results mentioned above are quite comparable in accuracy with each other in determining the peculiarities of galaxies being in good agreement with our results. They evident that CNN models are effective enough for image-based classification of galaxy morphological features.

In general, our method shows a satisfactory level of morphological classification performance, attaining more than 90 % of accuracy for most morphological classes (Table 1). Such value of the accuracy is in good agreement with the accuracy obtained in the work by Walmsley et al. [107], who used Bayesian CNN to study Galaxy Zoo volunteer responses and achieved coverage errors of 11.8 % within a vote fraction deviation of 0.2. As well as with work by Gauthier et al. [44], who applied both supervised and unsupervised methods to study the Galaxy Zoo dataset of 61578 pre-classified galaxies (spiral, elliptical, round, disk). They found that the variation of galaxy images is correlated with brightness and eccentricity, and the accuracy for galaxies to be associated with each of these four classes is about 94 %.

As for the classification of galaxies by their detailed structural morphological features, our CNN

model gives the accuracy in the range of 92–99 % depending on features, a number of galaxies with the given feature in the inference dataset, and, of course, the galaxy image quality (Table 2). As a result, for the first time we assigned 34 morphological detailed features for more than 160000 low-redshift galaxies with  $m_r < 17.7$  from the SDSS DR9, which have the highest adversarial score by our CNN classifier. If we compare our result with a largest galaxy morphological catalog presented by Vega-Ferrero et al. [103], where the face-on and edge-on galaxies were classified by their images with an accuracy of 73 %, we constitute that our CNN model gives a more significant output.

## 6. CONCLUSION

We developed a CNN model for image-based galaxy morphological classification. The studied sample of 315782 galaxies with  $m_r < 17.7$  from SDSS DR9 at  $0.02 < z < 0.1$  is overlapped with the data from the Galaxy Zoo 2 (GZ2). It allowed us to divide it into two datasets: “inference dataset” of 143410 galaxies, which do not match with the GZ2 dataset; “training dataset” of 172372 galaxies, which match with galaxies from our studied sample.

To develop the CNN model based on the images of galaxies, we used the GZ2 assigned labels for five morphological classes by shape and for 34 detailed structural morphological features of galaxies. We revealed that adversarial validation is very helpful when the labeled datasets are biased in magnitude distribution for the training dataset, and such a difference could bias the final prediction of the classifier on the inference dataset. To avoid this problem, we applied the adversarial validation method for analyzing the homogeneity of these two datasets and for modeling fainter galaxies. As a result, the galaxies were selected from the training dataset with the highest adversarial score that are most closely coincided with the inference dataset, and the images were normalized to be similar. Our CNN classifier has demonstrated >93 % of accuracy for rounded in-between, edge-on, and spiral morphological classes, 83 % for the completely rounded, and 75 % for the cigar-shaped galaxies. Assuming that a galaxy is in a certain morphological class if the probability is the highest one, we found that the inference dataset contains 27378 completely rounded, 59194 rounded

in-between, 18862 cigar-shaped, 7831 edge-on, and 23119 spiral galaxies.

As for the detailed structural features of galaxies, we worked with the inference dataset of 160471 galaxies and with 34 galaxy morphological features (bar, ring, various number of spiral arms, disks, dust lane, merger, etc.) as they are labeled by the GZ2’s volunteers. We used ROC AUC (Area Under Receiver Operating Characteristic Curve) as the binary classification quality metric, which gives accuracy scores for the model trained with adversarial augmentations and for the model trained without these augmentations. Our CNN model provides the accuracy in the range of 92–99 % depending on the features, the number of galaxies with the given feature in the inference dataset, and, of course, the galaxy image quality. As a result, for the first time, we have also visually inspected the galaxy images. As a result, for the first time we assigned 34 morphological detailed features for more than 160000 low-redshift galaxies with  $m_r < 17.7$  from the SDSS DR9, which have the highest adversarial score by our CNN classifier.

In general, we created the catalogs of 315782 low-redshift galaxies from SDSS DR9, where morphological classes and detailed features were defined for the first time for 216148 galaxies by the image-based CNN classifier. For the rest of galaxies (with the lower adversarial score) the initial morphological classification was re-assigned as in the GZ2 project. These catalogs can be accessed through the VizieR CDS platform. A vector representation of the probability distribution of a galaxy having one or another feature (the penultimate layer of our CNN model) can be founded at the Ukrainian Virtual Observatory [96] web-site (<http://ukr-vo.org/catalogs>). This will be of interest to those who will study the similarities between galaxies in more detail. Our approach to the image data augmentation can be applied as the mathematical tools in tasks of positional and photometrical processing CCD frames, archive astroplates in various bands, transient objects, artifacts [31, 70, 81, 83, 104]. The proposed CNN model allows solving a bunch of galaxy classification problems, for example, such as a quick selection of galaxies with a bar, bulge, or ring for their subsequent analysis. Our approach consumes the time at the stage of preliminary preparation of the

studied galaxy dataset and can be useful for further studies of the morphology, image, photometry, and spectroscopic data of galaxies.

**Acknowledgements.** This work was done in the frame of the budgetary program «Support for the development of priority fields of scientific research» of the NAS of Ukraine (CPCEL 6541230). The use of

the SDSS [2, 10] and SAO/NASA Astrophysics Data System was extensively applicable. The authors thank the Galaxy Zoo team.

Vavilova I. B. thanks the Wolfgang Pauli Institute, Vienna, Austria, for the support in frame of «The Pauli Ukraine Project» (2022) under the «Models in plasmas, Earth and space science» program.

## REFERENCES

1. Abul Hayat Md., Stein G., Harrington P., et al. (2020). Self-Supervised Representation Learning for Astronomical Images. eprint arXiv:2012.13083.
2. Ahn C. P., Alexandroff R., Allende Prieto C., et al. (2012). The Ninth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the SDSS-III Baryon Oscillation Spectroscopic Survey. *Astrophys. J. Suppl. Ser.*, **203**, 2, 21. <https://doi.org/10.1088/0067-0049/203/2/21>
3. Amiaux J., Scaramella R., Mellier Y., et al. (2012). Euclid mission: building of a reference survey. *SPIE Proceedings*, **8442**, Space Telescopes and Instrumentation: Optical, Infrared, and Millimeter Wave; 84420Z. <https://doi.org/10.1117/12.926513>
4. Aniyani A. K., Thorat K. (2017). Classifying Radio Galaxies with the Convolutional Neural Network. *Astrophys. J. Suppl. Ser.*, **230**, 2, 20. <https://doi.org/10.3847/1538-4365/aa7333>
5. Babyk I., Vavilova I. (2014). The distant galaxy cluster XLSSJ022403.9-041328 on the  $L_X$ - $T_X$ -M scaling relations using Chandra and XMM-Newton observations. *Astrophys. and Space Sci.*, **353**, 2, 613–619. <https://doi.org/10.1007/s10509-014-2057-x>
6. Baron Dalya. (2019). Machine Learning in Astronomy: a practical overview. eprint arXiv:1904.07248 <https://arxiv.org/pdf/1904.07248.pdf>
7. Barchi P. H., de Carvalho R. R., Rosa R. R., et al. (2020). Machine and Deep Learning applied to galaxy morphology. A comparative study. *Astronomy and Computing*, **30**, 100334. <https://doi.org/10.1016/j.ascom.2019.100334>
8. Barrow J. D., Saich P. (1993). Growth of large-scale structure with a cosmological constant. *Mon. Notic. Roy. Astron. Soc.*, **262**, 3, 717–725. <https://doi.org/10.1093/mnras/262.3.717>
9. Bellm E. C., Kulkarni S. R., Graham M. J., et al. (2019). The Zwicky Transient Facility: System Overview, Performance, and First Results. *Publ. Astron. Soc. Pacif.*, **131**, 995, id. 018002. <https://doi.org/10.1088/1538-3873/aaecbe>
10. Blanton M. R., Bershady M. A., Abolfathi B. (2017). Sloan Digital Sky Survey IV: Mapping the Milky Way, Nearby Galaxies, and the Distant Universe. *Astron. J.*, **154**, 28. <https://doi.org/10.3847/1538-3881/aa7567>
11. Bottrell C., Hani M., Teimoorinia H., et al. (2019). Deep learning predictions of galaxy merger stage and the importance of observational realism. *Mon. Notic. Roy. Astron. Soc.*, **490**, 4, 5390–5413. <https://doi.org/10.1093/mnras/stz2934>
12. Bradley A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, **30**, No. 7, 1145–1159. [https://doi.org/10.1016/S0031-3203\(96\)00142-2](https://doi.org/10.1016/S0031-3203(96)00142-2).
13. Brugere I., Gallagher B., Berger-Wolf T. Y. (2016). Network Structure Inference, A Survey: Motivations, Methods, and Applications. eprint arXiv:1610.00782 <https://arxiv.org/pdf/1610.00782.pdf>
14. Bundy K., Scarlata C., Carollo C. M. (2010). The Rise and Fall of Passive Disk Galaxies: Morphological Evolution Along the Red Sequence Revealed by COSMOS. *Astrophys. J.*, **719**, 2, 1969–1983. <https://doi.org/10.1088/0004-637X/719/2/1969>
15. Cabayol L., Eriksen M., Amara A., et al. (2021). The PAU survey: Estimating galaxy photometry with deep learning. *Mon. Notic. Roy. Astron. Soc.*, **506**, 3, 4048–4069. <https://doi.org/10.1093/mnras/stab1909>
16. Cabrera-Vives G., Miller C. J., Schneider J. (2018). Systematic Labeling Bias in Galaxy Morphologies. *Astron. J.*, **156**, 6, 284. <https://doi.org/10.3847/1538-3881/aae9f4>
17. Cassata P., Giavalisco M., Guo Y., et al. (2011). The Relative Abundance of Compact and Normal Massive Early-type Galaxies and Its Evolution from Redshift  $z \sim 2$  to the Present. *Astrophys. J.*, **743**, 1, 96. <https://doi.org/10.1088/0004-637X/743/1/96>
18. Chesnok N. G., Sergeev S. G., Vavilova I. B. (2009). Optical and X-ray variability of Seyfert galaxies NGC 5548, NGC 7469, NGC 3227, NGC 4051, NGC 4151, Mrk 509, Mrk 79, and Akn 564 and quasar 1E 0754. *Kinematics and Physics of Celestial Bodies*, **25**, 2, 107–113. <https://doi.org/10.3103/S0884591309020068>
19. Chen Bo Han, Goto Tomotsugu, Kim Seong Jin. (2021). An active galactic nucleus recognition model based on deep neural network. *Mon. Notic. Roy. Astron. Soc.*, **501**, 3, 3951–3961. <https://doi.org/10.1093/mnras/staa3865>
20. Cheng Ting-Yun, Conselice C. J., Arag S. (2020). Optimizing automatic morphological classification of galaxies with machine learning and deep learning using Dark Energy Survey imaging. *Mon. Notic. Roy. Astron. Soc.*, **493**, 3, 4209–4228. <https://doi.org/10.1093/mnras/staa501>

21. Chen Yangyao, Mo H. J., Li Cheng. (2020). Relating the Structure of Dark Matter Halos to Their Assembly and Environment. *Astrophys. J.*, **899**, 1, 81. <https://doi.org/10.3847/1538-4357/aba597>
22. Chilingarian I. V., Melchior A.-L., Zolotukhin I. Y. (2010). Analytical approximations of K-corrections in optical and near-infrared bands. *Mon. Notic. Roy. Astron. Soc.*, **405**, 1409–1420. <https://doi.org/10.1111/j.1365-2966.2010.16506.x>
23. Chilingarian I. V., Zolotukhin I. Y. (2012). A universal ultraviolet-optical colour-colour-magnitude relation of galaxies. *Mon. Notic. Roy. Astron. Soc.*, **419**, 1727–1739. <https://doi.org/10.1111/j.1365-2966.2011.19837.x>
24. Ciuca R., Hernandez O. F. (2017). A Bayesian framework for cosmic string searches in CMB maps. *J. Cosmology and Astropart. Phys.*, **8**, 28. <https://doi.org/10.1088/1475-7516/2017/08/028>
25. Davies R. L., Efstathiou G., Fall S. M. (1983). The kinematic properties of faint elliptical galaxies. *Astrophys. J.*, **266**, 41–57. <https://doi.org/10.1086/160757>
26. Davis M., Efstathiou G., Frenk C. S., White S. D. M. (1985). The evolution of large-scale structure in a universe dominated by cold dark matter. *Astrophys. J.*, Part 1, **292**, 371–394. <https://doi.org/10.1086/163168>
27. Dey A., Schlegel D. J., Lang D., et al. (2019). Overview of the DESI Legacy Imaging Surveys. *Astron. J.*, **157**, 5, 168. <https://doi.org/10.3847/1538-3881/ab089d>
28. Diakogiannis F. I., Lewis G. F., Ibata R. A. (2019). Reliable mass calculation in spherical gravitating systems. *Mon. Notic. Roy. Astron. Soc.*, **482**, 3, 3356–3372. <https://doi.org/10.1093/mnras/sty2931>
29. de Diego J. A., Nadolny J., Bongiovanni A. (2020). Galaxy classification: deep learning on the OTELO and COSMOS databases. *Astron. and Astrophys.*, **638**, A134. <https://doi.org/10.1051/0004-6361/202037697>
30. D’Isanto A., Cavuoti S., Gieseke F. (2018). Return of the features. Efficient feature selection and interpretation for photometric redshifts. *Astron. and Astrophys.*, **616**, A97. <https://doi.org/10.1051/0004-6361/201833103>
31. Djorgovski S. G., Graham M. J., Donalek C. (2016). Real-Time Data Mining of Massive Data Streams from Synoptic Sky Surveys. eprint arXiv:1601.04385. <https://arxiv.org/ftp/arxiv/papers/1601/1601.04385.pdf>
32. Dobrycheva D. V., Melnyk O. V., Vavilova I. B., Elyiv A. A. (2014). Environmental Properties of Galaxies at  $z < 0.1$  from the SDSS via the Voronoi Tessellation. *Odessa Astron. Publ.*, **27**, 26.
33. Dobrycheva D. V., Melnyk O. V., Vavilova I. B. (2015). Environmental Density vs. Colour Indices of the Low Redshifts Galaxies. *Astrophys.*, **58**, 2, 168–180. <https://doi.org/10.1007/s10511-015-9373-x>
34. Dobrycheva D. V. (2013). The New Galaxy Sample from SDSS DR9 at  $0.003 < z < 0.1$ . *Odessa Astron. Publ.*, **26**, 187.
35. Dobrycheva D. V. (2017). Morphological content and color indices bimodality of a new galaxy sample at the redshifts  $z < 0.1$ . PhD Thesis in Phys.-Math. Sciences. Kyiv: Main Astronomical Observatory, NAS of Ukraine, 132 p.
36. Dobrycheva D. V., Vavilova I. B., Melnyk O. V. (2018). Morphological Type and Color Indices of the SDSS DR9 Galaxies at  $0.02 < z < 0.06$ . *Kinematics and Phys. Celestial Bodies*, **34**, 6, 290–301. <https://doi.org/10.3103/S0884591318060028>
37. Dominguez-Sanchez H., Huertas-Company M., Bernardi M., et al. (2018). Improving galaxy morphologies for SDSS with Deep Learning. *Mon. Notic. Roy. Astron. Soc.*, **476**, 3, 3661–3676. <https://doi.org/10.1093/mnras/sty338>
38. Dominguez Sánchez H., Margalef B., Bernardi M., Huertas-Company M. SDSS-IV DR17: Final release of MaNGA Py-Morph photometric and deep learning morphological catalogs. *Mon. Notic. Roy. Astron. Soc.*, Advance Access. <https://doi.org/10.1093/mnras/stab3089>
39. Dominguez Sanchez H., Vega-Ferrero J., Huertas-Company M., Bernardi M. (2021). Constructing the Largest Galaxy Morphological Catalogue with Supervised Deep Learning ... with No Training Sample. Amer. Astron. Soc. meeting #238, id. 119.01. *Bull. American Astron. Soc.*, **53**, No. 6. e-id 2021n6i119p01
40. Du Wei, Cheng Cheng, Wu Hong et al. (2019). Low Surface Brightness Galaxy catalogue selected from the a.40-SDSS DR7 Survey and Tully-Fisher relation. *Mon. Notic. Roy. Astron. Soc.*, **483**, 2, 1754–1795. <https://doi.org/10.1093/mnras/sty2976>
41. Elyiv A., Melnyk O., Vavilova I. (2009). High-order 3D Voronoi tessellation for identifying isolated galaxies, pairs and triplets. *Mon. Notic. Roy. Astron. Soc.*, **394**, 3, 1409–1418. <https://doi.org/10.1111/j.1365-2966.2008.14150.x>
42. Elyiv A. A., Melnyk O. V., Vavilova I. B. (2020). Machine-learning computation of distance modulus for local galaxies. *Astron. and Astrophys.*, **635**, A124. <https://doi.org/10.1051/0004-6361/201936883>
43. Fluke Christopher J., Jacobs Colin. (2020). Surveying the reach and maturity of machine learning and artificial intelligence in astronomy. *WIREs Data Mining and Knowledge Discovery*, 10, 2, article id. e134910. <https://doi.org/10.1002/widm.1349>
44. Gauthier A., Jain A., Noordeh E. (2016). *Galaxy Morphology Classification*. e-proceedings <http://cs229.stanford.edu/proj2016/report/GauthierJainNoordeh-GalaxyMorphology-report.pdf>, p. 1–6
45. George D., Huerta E. A. (2018). Deep neural networks to enable real-time multimessenger astrophysics. *Phys. Rev. D* **97**, 044039. <https://doi.org/10.1103/PhysRevD.97.044039>
46. George D., Huerta E. A. (2018). Deep Learning for real-time gravitational wave detection and parameter estimation: Results with Advanced LIGO data. *Phys. Lett. B*, **778**, 64–70. <https://doi.org/10.1016/j.physletb.2017.12.053>



47. Huerta E. A., Moore C. J., Kumar P. (2018). Eccentric, nonspinning, inspiral, Gaussian-process merger approximant for the detection and characterization of eccentric binary black hole mergers. *Phys. Rev. D*, **97**, 2, id. 024031. <https://doi.org/10.1103/PhysRevD.97.02403>
48. Huertas-Company M., Primack J. R., Dekel A. (2018). Deep Learning Identifies High-*z* Galaxies in a Central Blue Nugget Phase in a Characteristic Mass Range, *Astrophys. J.*, **858**, 2, 114. <https://doi.org/10.3847/1538-4357/aabfed>
49. Ivezić Z., Kahn S. M., Tyson J. A. (2019). LSST: From Science Drivers to Reference Design and Anticipated Data Products. *Astrophys. J.*, **873**, 2, 111. <https://doi.org/10.3847/1538-4357/ab042c>
50. Jacobs C., Collett T., Glazebrook K. (2019). Finding high-redshift strong lenses in DES using convolutional neural networks. *Mon. Notic. Roy. Astron. Soc.*, **484**, 4, 5330–5349. <https://doi.org/10.1093/mnras/stz272>
51. Kang Shi-Ju, Fan Jun-Hui, Mao Weiming, et al. (2019). Evaluating the Optical Classification of Fermi BCUs Using Machine Learning. *Astrophys. J.*, **872**, 2, 189. <https://doi.org/10.3847/1538-4357/ab0383>
52. Karachentseva V. E., Vavilova I. B. (1994). *Clustering of Low Surface Brightness Dwarf Galaxies in the Local Supercluster*. European Southern Observatory Conference and Workshop Proceedings, **49**, 91–100.
53. Khalifa Nour Eldeen M., Taha Mohamed Hamed N., Hassanien Aboul Ella. Deep Galaxy: Classification of Galaxies based on Deep Convolutional Neural Networks. Eprint arXiv:1709.02245 (2017).
54. Khramtsov V., Dobrycheva D. V., Vasylenko M. Yu. (2019). Deep Learning for Morphological Classification of Galaxies from SDSS. *Odessa Astron. Publ.*, **32**, 21. <https://doi.org/10.18524/1810-4215.2019.32.182092>
55. Khramtsov V., Sergeev A., Spiniello C. (2019). KiDS-SQuAD — II. Machine learning selection of bright extragalactic objects to search for new gravitationally lensed quasars. *Astron. and Astrophys.*, **632**, A56. <https://doi.org/10.1051/0004-6361/201936006>
56. Khramtsov V., Vavilova I. B., Vasylenko M. Yu., Dobrycheva D. V., Elyiv A. A., Akhmetov V. S., Dmytrenko A., Khlamov S. (2022). Machine learning technique for morphological classification of galaxies from SDSS. III. CNN Image-based inference of detailed morphology. *Astron. and Computing*. (submitted)
57. Krause M., Pueschel E., Maier G. (2017). Improved gamma hadron separation for the detection of faint gamma-ray sources using boosted decision trees. *Astropart. Phys.*, **89**, 1–9. <https://doi.org/10.1016/j.astropartphys.2017.01.004>
58. Kugler S. D., Gianniotis N. (2016). Modelling multimodal photometric redshift regression with noisy observation. eprint arXiv:1607.06059. <https://arxiv.org/pdf/1607.06059.pdf>
59. LeCun Yann, Chopra Sumit, Hadsell Raia, et al. (2006). *A tutorial on energy-based learning*. Predicting Structured Data, MIT Press. <http://yann.lecun.com/exdb/publis/pdf/lecun-06.pdf>
60. Leung E., Bekki K., While L. (2019). Automated Simulations of Galaxy Morphology Evolution using Deep Learning and Particle Swarm Optimisation. arXiv:1904.02906. <https://arxiv.org/ftp/arxiv/papers/1904/1904.02906.pdf>
61. Lintott C. J., Schawinski K., Slosar A., et al. (2008). Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. *Mon. Notic. Roy. Astron. Soc.*, **389**, 3, 1179–1189. <https://doi.org/10.1111/j.1365-2966.2008.13689.x>
62. Lukic V., Bruggen M., Banfield J. K. (2018). Radio Galaxy Zoo: compact and extended radio source classification with deep learning. *Mon. Notic. Roy. Astron. Soc.*, **476**, 1, 246–260. <https://doi.org/10.1093/mnras/sty163>
63. Lupton R., Blanton M. R., Fekete G. (2004). Preparing Red-Green-Blue Images from CCD Data. *Publs Astron. Soc. Pacif.*, **116**, 816, 133–137. <https://doi.org/10.1086/382245>
64. Ma Zhixian, Xu Haiguang, Zhu Jie, et al. (2019). A Machine Learning Based Morphological Classification of 14,245 Radio AGNs Selected from the Best-Heckman Sample. *Astrophys. J. Suppl. Ser.*, **240**, 2, 34. <https://doi.org/10.3847/1538-4365/aaf9a2>
65. Mahabal A. A., Djorgovski S. G., Drake A. J. (2011). Discovery, classification, and scientific exploration of transient events from the Catalina Real-time Transient Survey. *Bull. Astron. Soc. India*, **39**, 3, 387–408.
66. Mahabal Ashish, Rebbapragada Umaa, Walters R. (2019). Machine Learning for the Zwicky Transient Facility. *Publs Astron. Soc. Pacif.*, **131**, 997, id. 038002. <https://doi.org/10.1088/1538-3873/aaf3fa>
67. Melnyk O. V., Dobrycheva D. V., Vavilova I. B. (2012). Morphology and color indices of galaxies in Pairs: Criteria for the classification of galaxies. *Astrophys.*, **55**, 3, 293–305. <https://doi.org/10.1007/s10511-012-9236-7>
68. Mezcua M., Lobanov A. P., Mediavilla E., Karouzos M. (2014). Photometric Decomposition of Mergers in Disk Galaxies. *Astrophys. J.*, **784**, 1, 16. <https://doi.org/10.1088/0004-637X/784/1/16>
69. Mittal A., Soorya A., Nagrath P., Hemant D. J. (2020). Data augmentation based morphological classification of galaxies using deep convolutional neural network. *Earth Sci. Inform.*, **13**, 601–617. <https://doi.org/10.1007/s12145-019-00434-8>
70. Morello G., Morris P. W., Van Dyk S. D. (2018). Applications of machine-learning algorithms for infrared colour selection of Galactic Wolf-Rayet stars. *Mon. Notic. Roy. Astron. Soc.*, **473**, 2, 2565–2574. <https://doi.org/10.1093/mnras/stx2474>

71. Parikh T., Thomas D., Maraston C., et al. (2019). SDSS-IV MaNGA: local and global chemical abundance patterns in early-type galaxies. *Mon. Notic. Roy. Astron. Soc.*, **483**, 3, 3420–3436. <https://doi.org/10.1093/mnras/sty3339>
72. Pasquet-Itam J., Pasquet J. (2018). Deep learning approach for classifying, detecting and predicting photometric redshifts of quasars in the Sloan Digital Sky Survey stripe 82. *Astron. and Astrophys.*, **611**, A97. <https://doi.org/10.1051/0004-6361/201731106>
73. Pearson W. J., Wang L., Trayford J. W., et al. (2019). Identifying galaxy mergers in observations and simulations with deep learning. *Astron. and Astrophys.*, **626**, A49. <https://doi.org/10.1051/0004-6361/201935355>
74. Peebles P. E. (1993). *Principles of Physical Cosmology*. Princeton, New Jersey: Princeton Univ. Press, 718 p.
75. Peng Ying-jie, Lilly S. J., Kova K., et al. (2010). Mass and Environment as Drivers of Galaxy Evolution in SDSS and zCOSMOS and the Origin of the Schechter Function. *Astrophys. J.*, **721**, 1, 193–221. <https://doi.org/10.1088/0004-637X/721/1/193>
76. Rodriguez-Puebla A., Calette A. R., Avila-Reese V., et al. (2020). The bivariate gas-stellar mass distributions and the mass functions of early- and late-type galaxies at  $z \sim 0$ . *Publ. Astron. Soc. Austral.*, **37**, article id. e024. <https://doi.org/10.1017/pasa.2020.15>
77. Pulatova N. G., Vavilova I. B., Sawangwit U., Babyk Iu., Klimanov S. (2015). The 2MIG isolated AGNs. I. General and multiwavelength properties of AGNs and host galaxies in the northern sky. *Mon. Notic. Roy. Astron. Soc.*, **447**, 3, 2209–2223. <https://doi.org/10.1093/mnras/stu2556>
78. Reid B. A., Samushia L., White M., et al. (2012). The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: measurements of the growth of structure and expansion rate at  $z = 0.57$  from anisotropic clustering. *Mon. Notic. Roy. Astron. Soc.*, **426**, 4, 2719–2737. <https://doi.org/10.1111/j.1365-2966.2012.21779.x>
79. dos Reis S. N., Buitrago F., Papaderos P., et al. (2020). Structural analysis of massive galaxies using HST deep imaging at  $z < 0.5$ . *Astron. and Astrophys.*, **634**, A11. <https://doi.org/10.1051/0004-6361/201936276>
80. Richert Willi, Luis Pedro Coelho (2013). *Building Machine Learning Systems with Python*. URL: <http://gen.lib.rus.ec/book/index.php?md5=7a375749558682503761fa801a67d7ec> (Last access 07.03.2022).
81. Ruhe Tim. (2020). Application of machine learning algorithms in imaging Cherenkov and neutrino astronomy. *Int. J. Modern Phys. A*, **35**, 33, 2043004. <https://doi.org/10.1142/S0217751X20430046>
82. Salvato M., Ilbert O. (2019). Hoyle Ben, the many flavours of photometric redshifts. *Nature Astron.*, **3**, 212–222. <https://doi.org/10.1038/s41550-018-0478-0>
83. Savanevych V. E., Khlamov S. V., Vavilova I. B., et al. (2018). A method of immediate detection of objects with a near-zero apparent motion in series of CCD-frames. *Astron. and Astrophys.*, **609**, id. A54, 11 p. <https://doi.org/10.1051/0004-6361/201630323>
84. Scaife A. M. M., Porter F. (2021). Fanaroff-Riley classification of radio galaxies using group-equivariant convolutional neural networks. *Mon. Notic. Roy. Astron. Soc.*, **503**, 2, 2369–2379. <https://doi.org/10.1093/mnras/stab530>
85. Schawinski Kevin, Zhang Ce, Zhang Hantian (2017). Generative adversarial networks recover features in astrophysical images of galaxies beyond the deconvolution limit. *Mon. Notic. Roy. Astron. Soc. Lett.*, **467**, 1, L110–L114. <https://doi.org/10.1093/mnrasl/slx008>
86. Schlegel D. J., Finkbeiner D. P., Davis M. (1998). Maps of Dust Infrared Emission for Use in Estimation of Reddening and Cosmic Microwave Background Radiation Foregrounds. *Astrophys. J.*, **500**, 2, 525–553. <https://doi.org/10.1086/305772>
87. Simmons B. D., Lintott C., Willett K. W., et al. (2017). Galaxy Zoo: quantitative visual morphological classifications for 48 000 galaxies from CANDELS. *Mon. Notic. Roy. Astron. Soc.*, **464**, 4, 4420–4447. <https://doi.org/10.1093/mnras/stw2587>
88. Speagle J. S., Eisenstein D. J. (2017). Deriving photometric redshifts using fuzzy archetypes and self-organizing maps. II. Implementation. *Mon. Notic. Roy. Astron. Soc.*, **469**, 1, 1205–1224. <https://doi.org/10.1093/mnras/stx510>
89. Sybilska A., Lisker T., Kuntschner H., et al. (2017). The hELENA project. I. Stellar populations of early-type galaxies linked with local environment and galaxy mass. *Mon. Notic. Roy. Astron. Soc.*, **470**, 1, 815–838. <https://doi.org/10.1093/mnras/stx1138>
90. Tsizh M., Novosyadlyj B., Holovatch Yu. (2020). Large-scale structures in the Lambda-CDM Universe: network analysis and machine learning. *Mon. Notic. Roy. Astron. Soc.*, **495**, 1, 1311–1320. <https://doi.org/10.1093/mnras/staa1030>
91. Tamburri S., Saracco P., Longhetti M., et al. (2014). The population of early-type galaxies: how it evolves with time and how it differs from passive and late-type galaxies. *Astron. and Astrophys.*, **570**, A102. <https://doi.org/10.1051/0004-6361/201424040>
92. Vasylenko M. Yu., Dobrycheva D. V., Vavilova I. B. (2019). Verification of Machine Learning Methods for Binary Morphological Classification of Galaxies from SDSS. *Odessa Astron. Publ.*, **32**, 46. <https://doi.org/10.18524/1810-4215.2019.32.182538>
93. Vasylenko M., Dobrycheva, D., Khramtsov V. (2020). Deep Convolutional Neural Networks models for the binary morphological classification of SDSS-galaxies. *Commun. BAO*, **67**, 354. <https://doi.org/10.52526/25792776-2020.67.2-354>

94. Vavilova I. B., Karachentseva V. E., Makarov D. I., Melnyk O. V. (2005). Triplets of Galaxies in the Local Supercluster. I. Kinematic and Virial Parameters. *Kinematics and Phys. Celestial Bodies*, **21**, 1, 3–20.
95. Vavilova I. B., Melnyk O. V., Elyiv A. A. (2009). Morphological properties of isolated galaxies vs. isolation criteria. *Astron. Nachr.*, **330**, 1004. <https://doi.org/10.1002/asna.200911281>
96. Vavilova I. B., Pakuliak L. K., Protsyuk Yu. I., et al. (2012). UkrVO Joint digitized archive and scientific projects. *Baltic Astron.*, **21**, 356–365. <https://doi.org/10.1515/astro-2017-0394>
97. Vavilova I. B., Elyiv A. A., Vasylenko M. Yu. (2018). Behind the Zone of Avoidance of the Milky Way: what can we Restore by Direct and Indirect Methods? *Radio Phys., Radio Astron.*, **23**, 4, 244–257. <https://doi.org/10.15407/rpra23.04.244>
98. Vavilova I., Dobrycheva D., Vasylenko M. (2020). *Multiwavelength Extragalactic Surveys: Examples of Data Mining*. Knowledge Discovery in Big Data from Astronomy and Earth Observation, 1st Edition. Ed. by Petr Skoda and Fathallahman Adam. Elsevier, 307–323 <https://doi.org/10.1016/B978-0-12-819154-5.00028-X>
99. Vavilova I., Pakuliak L., Babyk Iu. (2020). *Surveys, Catalogues, Databases, and Archives of Astronomical Data*. Knowledge Discovery in Big Data from Astronomy and Earth Observation, 1st Edition. Ed. by Petr Skoda and Fathallahman Adam. Elsevier, 57–102. <https://doi.org/10.1016/B978-0-12-819154-5.00015-1>
100. Vavilova I., Elyiv A., Dobrycheva D., Melnyk O. (2021). *The Voronoi tessellation method in astronomy*. Intelligent Astrophysics. Ed. by I. Zelinka, M. Brescia and D. Baron. Emergence, Complexity and Computation, **39**. Springer, Cham, 57–79. [https://doi.org/10.1007/978-3-030-65867-0\\_3](https://doi.org/10.1007/978-3-030-65867-0_3)
101. Vavilova I. B., Dobrycheva D. V., Vasylenko M. Yu. (2021). Machine learning technique for morphological classification of galaxies from the SDSS. I. Photometry-based approach, **648**, A122. <https://doi.org/10.1051/0004-6361/202038981>
102. Vavilova I. B., Dobrycheva D. V., Vasylenko M. Yu. (2021). *VizieR Online Data Catalog: SDSS galaxies morphological classification* (Vavilova+, J/A+A/648/A122, 2021). <https://doi.org/10.26093/cds/vizier.36480122>
103. Vega-Ferrero J., Dominguez Sanchez H., Bernardi M., et al. (2021). Pushing automated morphological classifications to their limits with the Dark Energy Survey. *Mon. Notic. Roy. Astron. Soc.*, **506**, 2, 1927–1943. <https://doi.org/10.1093/mnras/stab594>
104. Villarroel B., Soodla J., Comerón S., et al. (2020). The Vanishing and Appearing Sources during a Century of Observations Project. I. USNO Objects Missing in Modern Sky Surveys and Follow-up Observations of a “Missing Star”. *Astron. J.*, **159**, 1, article id. 8, 19 p. <https://doi.org/10.3847/1538-3881/ab570f>
105. Vol’vach A. E., Vol’vach L. N., Kut’kin A. M., et al. (2011). Multi-frequency studies of the non-stationary radiation of the blazar 3C 454.3. *Astron. Reports*, **55**, 7, 608–615. <https://doi.org/10.1134/S1063772911070092>
106. Vulcani B., Poggianti B. M., Aragon-Salamanca A., et al. (2011). Galaxy stellar mass functions of different morphological types in clusters, and their evolution between  $z = 0.8$  and 0. *Mon. Notic. Roy. Astron. Soc.*, **412**, 1, 246–268. <https://doi.org/10.1111/j.1365-2966.2010.17904.x>
107. Walmsley M., Smith L., Lintott C. (2020). Galaxy Zoo: probabilistic morphology through Bayesian CNNs and active learning. *Mon. Notic. Roy. Astron. Soc.*, **491**, 2, 1554–1574. <https://doi.org/10.1093/mnras/stz2816>
108. Walmsley M., Lintott C., Geron T., et al. (2021). Galaxy Zoo DECaLS: Detailed Visual Morphology Measurements from Volunteers and Deep Learning for 314,000 Galaxies. *Mon. Notic. Roy. Astron. Soc.*, **509**, 3, 3966–3988. <https://doi.org/10.1093/mnras/stab2093>
109. Willett K. W., Lintott C. J., Bamford S. P., et al. (2013). Galaxy Zoo 2: detailed morphological classifications for 304 122 galaxies from the Sloan Digital Sky Survey. *Mon. Notic. Roy. Astron. Soc.*, **435**, 4, 2835–2860. <https://doi.org/10.1093/mnras/stt1458>
110. Yang Xiaohu, Mo H. J., van den Bosch Frank C. (2003). Constraining galaxy formation and cosmology with the conditional luminosity function of galaxies. *Mon. Notic. Roy. Astron. Soc.*, **339**, 4, 1057–1080. <https://doi.org/10.1046/j.1365-8711.2003.06254.x>
111. Zevin M., Coughlin S., Bahaadini S., et al. (2017). Gravity Spy: integrating advanced LIGO detector characterization, machine learning, and citizen science. *Classical and Quantum Gravity*, **34**, 6, id. 064003. <https://doi.org/10.1088/1361-6382/aa5cea>
112. Zhu Xiao-Pan, Dai Jia-Ming, BianChun-Jiang, et al. (2019). Galaxy morphology classification with deep convolutional neural networks. *Astrophys. and Space Sci.*, **364**, 4, 55. <https://doi.org/10.1007/s10509-019-3540-1>

Стаття надійшла до редакції 21.09.2020

Після доопрацювання 01.08.2021

Прийнято до друку 07.03.2022

Received 21.09.2020

Revised 01.08.2021

Accepted 07.03.2022

<sup>1</sup> I. B. Vavilova, зав. відділу, д-р фіз.-мат. наук, проф.

<sup>2</sup> старш. наук. співроб.

E-mail: irivav@mao.kiev.ua

<https://orcid.org/0000-0002-5343-1408>

<sup>3</sup> V. Khramtsov, аспірант

E-mail: vld.khramtsov@gmail.com

<https://orcid.org/0000-0003-1744-7071>

<sup>1</sup> D. V. Dobrycheva, старш. наук. співроб., канд. фіз.-мат. наук,

E-mail: daria@mao.kiev.ua

<https://orcid.org/0000-0001-5557-3453>

<sup>1</sup> M. Yu. Vasilenko, мол. наук. співроб.

<sup>4</sup> аспірант

E-mail: vasmah@mao.kiev.ua

<https://orcid.org/0000-0002-7714-0779>

<sup>1</sup> A. A. Elyiv, старш. наук. співроб., канд. фіз.-мат. наук

E-mail: andrii.elyiv@gmail.com

<https://orcid.org/0000-0001-6215-1048>

<sup>1</sup> O. V. Melnyk, старш. наук. співроб., канд. фіз.-мат. наук

E-mail: melnykol@gmail.com

<sup>1</sup> Головна астрономічна обсерваторія НАН України  
вул. Академіка Заболотного 27, Київ, Україна, 03143

<sup>2</sup> НДІ «Астрономічна обсерваторія» Одеського національного університету ім. І. І. Мечникова  
вул. Маразліївська 1 в, Одеса, Україна, 65014

<sup>3</sup> Інститут астрономії, Харківський національний університет ім. В. Н. Каразіна  
вул. Сумська 35, Харків, Україна, 61022

<sup>4</sup> Інститут фізики НАН України  
пр. Науки 46, Київ, Україна, 02000

## МАШИННЕ НАВЧАННЯ ДЛЯ МОРФОЛОГІЧНОЇ КЛАСИФІКАЦІЇ ГАЛАКТИК ІЗ ОГЛЯДУ SDSS.

### II. МОРФОЛОГІЧНІ КАТАЛОГИ ЗОБРАЖЕНЬ ГАЛАКТИК НА $0.02 < Z < 0.1$

Ми застосували згорткову нейронну мережу (CNN) до вибірки зображень галактик на малих червоних зміщеннях із  $-24^m < M_r < -19.4^m$  огляду неба SDSS DR9. Ми розділили її на дві підвибірки галактик SDSS DR9 і галактик Galaxy Zoo 2 (GZ2), розглядаючи їх як цільову з невідомими параметрами (inference) та навчальну (training), відповідно. Щоб визначити основні морфологічні параметри галактик, визначені в рамках проекту GZ2, ми класифікували галактики на п'ять візуальних класів (повністю заокруглені, майже заокруглені, гладкі сигароподібні, видимі з ребра, спіральні). Використовуючи класифікацію морфології галактик GZ2, ми також визначили 34 морфологічні характеристики галактик із вибірки SDSS DR9, які не збігаються з навчальною підвибіркою галактик GZ2. У результаті ми створили морфологічний каталог зображень 315782 галактик на  $0.02 < z < 0.1$ , де морфологічні п'ять класів і 34 детальні характеристики були вперше визначені для 216148 галактик із застосуванням CNN класифікатора. Для решти галактик початкову морфологічну класифікацію було перевизначено, як у проєкті GZ2.

Наш метод демонструє багатообіцяючу ефективність морфологічної класифікації, що досягає понад 93 % точності для прогнозування морфології п'яти класів, за винятком сигароподібних (~75 %) та повністю округлених (~83 %) галактик. В результаті були отримані каталоги 27378 повністю заокруглених, 59194 майже заокруглених, 18862 сигароподібних, 7831 видимих з ребра, 23119 спіральних галактик (inference) досліджуваної вибірки SDSS. Що стосується класифікації галактик за їхніми детальними структурними морфологічними особливостями, то наша модель CNN дає точність 92–99 % залежно від морфологічної ознаки та якості зображення галактики. Створено каталоги, де вперше 34 детальні морфологічні особливості (бар, кільця, кількість спіральних рукавів, злиття тощо) визначено для понад 160000 галактик цільової підвибірки SDSS DR9. Ми вперше показуємо, що застосування моделі CNN зі змагальною валідацією та математичними перетвореннями зображень галактик по-краще класифікацію менших за розмірами та слабкіших  $m_r < 17.7$  галактик SDSS.

Запропонована модель CNN дозволяє вирішити різні проблеми класифікації галактик, таких як швидкий відбір галактик із баром, балджем, кільцем та іншими морфологічними особливостями для їх подальшого аналізу.

**Ключові слова.** Методи: аналіз даних, машинне навчання, згорткові нейронні мережі; галактики: морфологічна класифікація, зображення галактик, каталоги галактик, великомасштабна структура Всесвіту.