



Galaxy Classification Using Deep Learning

Premanand Ghadekar, Kunal Chanda, Sakshi Manmode^(✉), Sanika Rawate, Shivam Chaudhary, and Resham Suryawanshi

Department of Information Technology, Vishwakarma Institute of Technology, Pune, India
sakshi.manmode19@vit.edu

Abstract. In this paper, the framework of Deep CNN (Deep Convolutional Neural Network) is basically used for classifying the galaxies. It is shown that galaxies can basically be classified with the help of distinct features into the different categories namely Disturbed galaxies, merging galaxies, Round Smooth galaxies, In-between Round Smooth galaxies, Cigar Shaped Spiral galaxies, Barred Tight Spiral galaxies, Unbarred Loose Spiral galaxies, Edge-on galaxies without bulge and Edge-on galaxy with the bulge. The model that we are proposing is a ConvNet galaxy architecture consists of one input layer having 16 filters, followed by 4 hidden layers, 1 penultimate dense layer, along with an Output Softmax layer. we also included data augmentation such as shear, zoom, rotation, rescaling, and flip. we used the activation function.

The dataset which is used in the proposed research is Galaxy 10 DECals, which has taken its images from DESI Legacy Imaging Surveys and got labeled by Galaxy Zoo. The dataset used contains 256×256 pixel-colored Galaxy images (g, r, and z band). The proposed model and framework is training over 17736 images and accomplished above 84.04% in testing accuracy. When a comparison is made between the results and the testing accuracy was compared with other existing models.

Keywords: Galaxy Classification · Deep convolutional neural networks · Feature extraction

1 Introduction

Galaxy is a spiral combination of stars, dark matter, dust, and glasses. Galaxies are of several types and kinds with different features. There are innumerable galaxies in the universe. The study of these galaxies, their types as well as their properties is really crucial [1]. The reason behind this is that galaxies are an important indication of the origin of the universe and its development [2]. Galaxy Classification plays an important role in evaluating the universe in order to understand galaxy formation [3]. Basically, Galaxy Classification is nothing but the system which dissects the galaxies morphologically on the basis of their appearance and impression. There are some strategies that are used to classify the galaxies morphologically in order to get accurate results. The challenges

faced by astrophysicists are solved by these classification strategies. Operations are performed on large databases containing information about galaxies so that astrophysicists could test the propositions and get new solutions [4]. Using this, they can elaborate on the formation of stars, the evolution of this universe, and the science behind the processes which administer the galaxies.

So, classifying galaxies is studying, operating, and graphically inspecting galaxy images in 2-D (two-dimensional) form and then differentiating them as per their look [5]. This process is becoming time-consuming day by day because of the evolution of cameras with charged coupled devices and the increasing size of telescopes which is resulting in an increase in the size of databases containing astronomical data, for example, the dataset which is used by the proposed model [6]. The size of the database makes it impossible to analyze the data manually. This model training and testing are most important based on the images to be classified. Classification of galaxies has become more challenging because galaxies are complicated in nature and image quality [7].

The most helpful part of galaxy classification is it will help the astronomers in differentiating the galaxies accurately. In recent years, the evolution of computational tools and training algorithms have advanced the research and analysis of galaxies and their appearance [8].

In the field of Visual Detection as well as Recognition, deep learning plays an important role as it has brought a great scope for improvement. Unprocessed data is fed as an input to Deep Learning Techniques for optimizing the segmentation parameters or feature designs which do not need any prior expertise in the field.

2 Literature Review

1. In recent years, data on galaxies is generated in a large amount due to some sky-surveys like Sloan Digital Sky Survey i.e., SDSS. That's why an efficient classification was a need in this speedy Deep Learning Era. So, Algorithms were intended to differentiate the galaxies into multiple classes. Earlier galaxies were classified into six classes. But the proposed model is classifying the galaxies into eight classes [1]. The proposed model is having a framework that extracts the important features making it reliable. The model got 84.73 percent of testing accuracy after considering important features.
2. As the neural network algorithms are broadly focusing on the survey information, the Radio Galaxy Classification has become a researcher's need. Whether the trained algorithms are having the cross-survey identification potential or not was the main query before researchers. The authors thought of building a Transfer Learning Model to solve this problem. It was used in different radio surveys. The model was also trained for Random Initialization and got good accuracy. First pre-trained images with inherited model weights were trained in such a way that they became performance boosters. The model was trained with NVSS data [10].
3. The paper presented the model with Deep Convolutional Network in order to classify Galaxies. According to features observed morphologically, classification was

done with three main classes namely Elliptical, Spiral and Regular. The framework which was proposed in the research had broadly 8 layers, along with one main layer consisting of 96 filters for extracting main features. The model was training near about 1356 images which give a good testing accuracy. The proposed model was performing well with the classification queries efficiently [12].

4. Morphological Classification of Galaxy is becoming trending in recent years. As far as we are discovering the universe in a broader vision, there is a need to build a robust and reliable model to classify the galaxies from their images. Researchers proposed two fundamental approaches, one was based upon traditional machine learning techniques with non-parametric morphology and the other was based on Deep Learning. Researchers built a system called CyMorph with a non-parametric approach. The dataset used was Sloan Digital Sky Survey i.e., SDSS with release 7 (DR7). In this work, three classes were considered. Overall Accuracy obtained was great because Decision Tree Models were used to approach the quality of the classification morphologically [13].

3 Methodology

3.1 Dataset Collection

The dataset which is used in our proposed research is the Galaxy10 DECals Dataset which is basically an improved version of the Galaxy10 Dataset. The dataset of Galaxy10 was

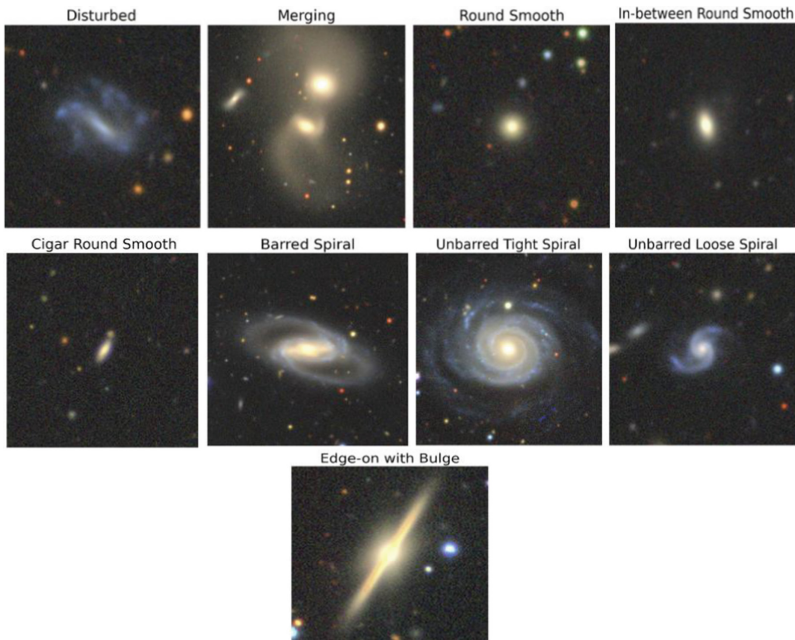


Fig. 1. Sample Images from each Class of Galaxy10 DECals

originally created by Galaxy Zoo (GZ) Data Release 2 where they were classifying 270 thousand SDSS galaxy images among which 22 thousand of those images were selected in the grouping of 10 broader classes. It was done on the basis of votes given to every image. Later, the images from DECals is DESI Legacy Imaging Surveys were utilized by Galaxy Zoo in order to get images perfectly resolved and greater image quality. The Galaxy10 DECals which merged all the three (GZ i.e., Galaxy Zoo DR2 with DECals images rather than SDSS images and DECals campaign ab, c) which then results in 441 thousand unique galaxies covered by DECals. After that 18 thousand of those images were selected in 10 broader classes. This was obtained after the continuous filtering with the votes obtained for each image. There was one more class named Edge-on Disk with Boxy Bulge which was later abandoned due to less number of images which was just 17 images. This was done to make the dataset of Galaxy10 DECals more distinct and precise in the formed 10 broader classes (Figs. 1, 2 and 3).

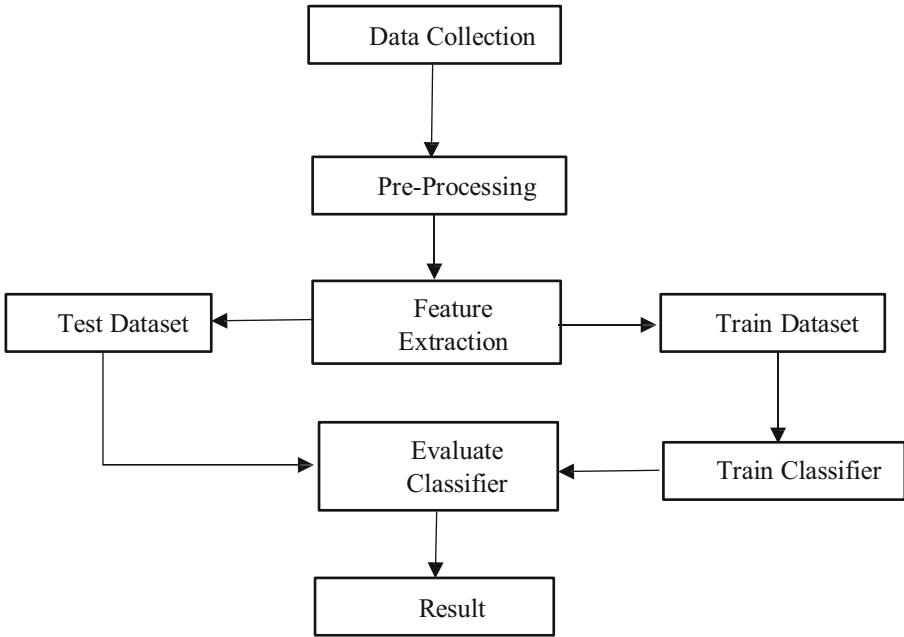


Fig. 2. Flow diagram of the proposed model

The dataset used contains 256×256 pixel-colored Galaxy images (g, r, and z band). It contains 17736 images which are broadly classified into 10 major classes. Images of the dataset come from DESI Legacy Imaging Surveys and labeling was done by Galaxy Zoo.

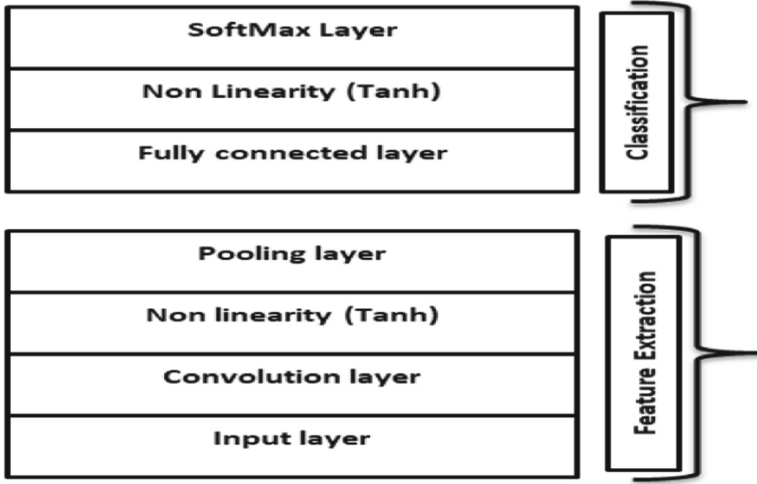


Fig. 3. Proposed model architecture

Table 1. Type of galaxies with the number of images

Name of class	Type of galaxy	Total pictures
Class-0	Disturbed Galaxies	1081
Class-1	Merging Galaxies	1853
Class-2	Round Smooth Galaxies	2645
Class-3	In-between Round Smooth Galaxies	2027
Class-4	Cigar Shaped Smooth Galaxies	334
Class-5	Barred Spiral Galaxies	2043
Class-6	Unbarred Tight Spiral Galaxies	1829
Class-7	Unbarred Loose Spiral Galaxies	2628
Class-8	Edge-on Galaxies without Bulge	1423
Class-9	Edge-on Galaxies with Bulge	1873

3.2 Proposed Deep Galaxies CNN Model

This project is based on morphological classification using Deep Convolution Neural Network. It shows how computational cosmology could help to make hard classification easy. The galaxy can be classified in various ways, they can be in three classes namely - Elliptical, Spiral, Irregular or they could be classified deeper as Disturbed, Merging, round smooth, barred spiral, bulge, etc.

This project uses astroNN dataset, this dataset is generated from [DESI Legacy Imaging Survey], and images got themselves labeled by [Galaxy Zoo].

The important advantage of the proposed model will be classifying the results into more classes (class 0 to class 9, as explained in Table 1). This is obligating to Astrophysicist and Astronomers.

3.3 Overview of Algorithms

3.3.1 Machine Learning

That works on building software applications being accurate in predicting the results and outcomes which will not need any external programming help. Machine Learning can help industries and companies to understand as well as examine their customers and consumers subordinately [4]. In addition to this, this predictive learning method can help in the formation of operating systems (OS) of self-driving cars, for example, Tesla. Data of the consumers is collected and correlated with their behaviors all the time. Machine Learning is used as an important as well as the primary driver for the business models of the companies such as Uber.

3.3.2 Deep Learning

In Deep Learning, a machine that imitates the neuron network in a human's brain can be created. The basis of it is on Deep Networks in which a task is divided and then distributed to machine learning algorithms. The algorithm includes many connected layers. A set of hidden layers is in between the input layer (1st layer) and the output layer (last layer), which gives the word Deep, meaning that the networks that join neurons are in more than 2 layers [8]. The interconnected neurons propagate the input signal and after that, it goes through the process [14].

3.3.3 Convolutional Neural Network

Convolutinnal Neural Networks or CNNs are types of artificial neural network that is ANN, which is broadly used for image classification and object recognition. They are types of discriminative models that are initially developed to work on images that are not pre-processed. Now, these models can also work with text and sound recognition and classification. CNN architecture was framed and designed for recognition models which can read the zip codes and digits, which was constructed in the era of the 90s (1990) which initiated the deep artificial neural networks or ANNs [6].

3.3.4 Dropout Layer

Dropout layer in CNN architecture acts as a filtering layer that nullifies the efforts of some neurons when passing to the next layer. A dropout layer is added to the input vector which normalizes some features. These layers play an important role in preventing the overfitting problem while training the data.

3.3.5 Optimizers

Optimizers are techniques or algorithms that are used to change the features of neural networks which includes learning rate and weights. These are used to minimize losses and

errors. Also, as the functions are minimized, optimizers are used in the Optimization Problems wherever needed. The following update equation is used to initialize some policies with renovated Epoch.

$$W_{New} = W_{old} - lr * (\nabla WL)W_{old} \quad (1)$$

Equation 1 is used to Renovate Epoch.

3.3.6 Pooling Layer

Pooling Layer is used to minimize the size of the image which is obtained from the reduction of dimensions of the feature maps. The number of learning parameters is reduced as well as computational efforts are minimized by using this Pooling Technique [11]. It basically recollects all the important features present in the convolution layer. There are two types of Pooling which are as follows:

- a. Max Pooling: Selection of the maximum values from the matrix of specified size (basically, the default size is 2 X 2) takes place. Due to this Feature Extraction of more important parameters takes place.
- b. Average Pooling: The average of all the pixel values of the matrix (basically, the default size is 2 X 2) takes place in the pooling layer.

3.3.7 Dense Layer

After the Pooling procedure, the output is then passed on to the Dense Layer. The constraint of the Dense layer is that the input should be in a One-Dimensional or 1-D shape (the array should be 1-D) [14]. The image classification takes place in this layer on the basis of the output of previous convolutional layers. The input received is known as Dense.

3.3.8 Flatten Method

Flatten method is used in the conversion of a multi-dimensional matrix to a single-dimensional matrix. This method avoids the overfitting problem thus data is easier to interpret. It includes Preprocessing of Data and Data Augmentation (Table 2).

4 Comparative Results

Table 2. Comparative analysis of existing and proposed model

Related work	Year	Paper title	Total accuracy
1	2022	Galaxy classification: a deep learning approach for classifying Sloan Digital Sky Survey images	84.73%
2	2019	Machine and Deep Learning applied to galaxy morphology—A comparative study	82%
3	2020	Optimizing automatic classification of galaxies with ML and DL using Dark Energy Survey imaging	92.2%
Proposed model	2022	Galaxy Classification using Deep Learning	84.04%

4.1 Model Accuracy

A validation method used in the classification problems in the Machine Learning Model is known as Accuracy. The overall accuracy of the proposed model is 84.04% (Figs. 4, 5, 6 and 7).

(1)	<u>Actual:</u> Edge-on Galaxies without bul	<u>Predicted:</u> Disturbed
(2)	<u>Actual:</u> Merging	<u>Predicted:</u> Merging
(3)	<u>Actual:</u> Cigar Shaped Smooth	<u>Predicted:</u> Cigar Shaped smooth
(4)	<u>Actual:</u> Round Smooth	<u>Predicted:</u> Round smooth
(5)	<u>Actual:</u> Disturbed	<u>Predicted:</u> Edge-on Galaxies with bulge
(6)	<u>Actual:</u> Unbarred loose spiral	<u>Predicted:</u> Edge-on Galaxies without bulge
(7)	<u>Actual:</u> Merging	<u>Predicted:</u> Merging
(8)	<u>Actual:</u> Distributed	<u>Predicted:</u> Merging
(9)	<u>Actual:</u> Distributed	<u>Predicted:</u> Unbarred
(10)	<u>Actual:</u> Unbarred loose	<u>Predicted:</u> Edge-on Galaxies
(11)	<u>Actual:</u> Cigar Shaped Smooth	<u>Predicted:</u> Cigar Shaped
(12)	<u>Actual:</u> Unbarred tight spiral	<u>Predicted:</u> Unbarred tight spiral
(13)	<u>Actual:</u> Merging	<u>Predicted:</u> Merging
(14)	<u>Actual:</u> Merging	<u>Predicted:</u> Merging
(15)	<u>Actual:</u> Round Smooth	<u>Predicted:</u> Round Smooth
(16)	<u>Actual:</u> Distributed	<u>Predicted:</u> Distributed
(17)	<u>Actual:</u> In-between Round Smooth	<u>Predicted:</u> In-between Round Smooth
(18)	<u>Actual:</u> Round Smooth	<u>Predicted:</u> Round Smooth
(19)	<u>Actual:</u> Merging	<u>Predicted:</u> Merging
(20)	<u>Actual:</u> Distributed	<u>Predicted:</u> Distributed
(21)	<u>Actual:</u> Merging	<u>Predicted:</u> Merging

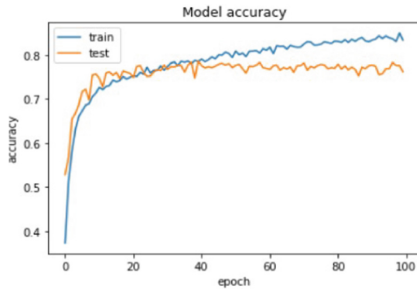


Fig. 4. Graph of model accuracy

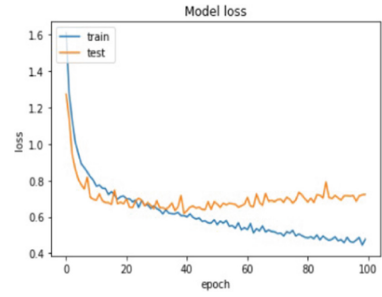


Fig. 5. Graph of model loss

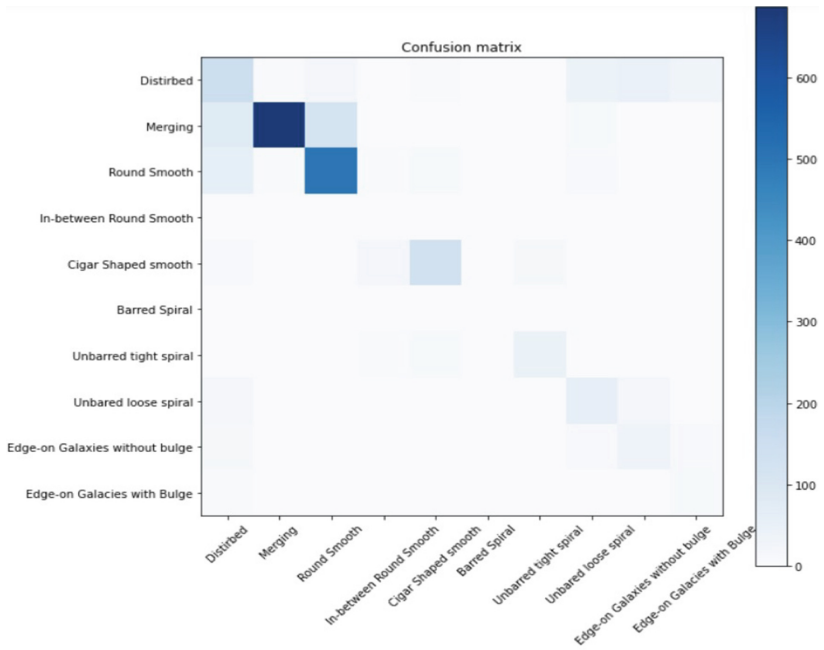


Fig. 6. Confusion matrix of result

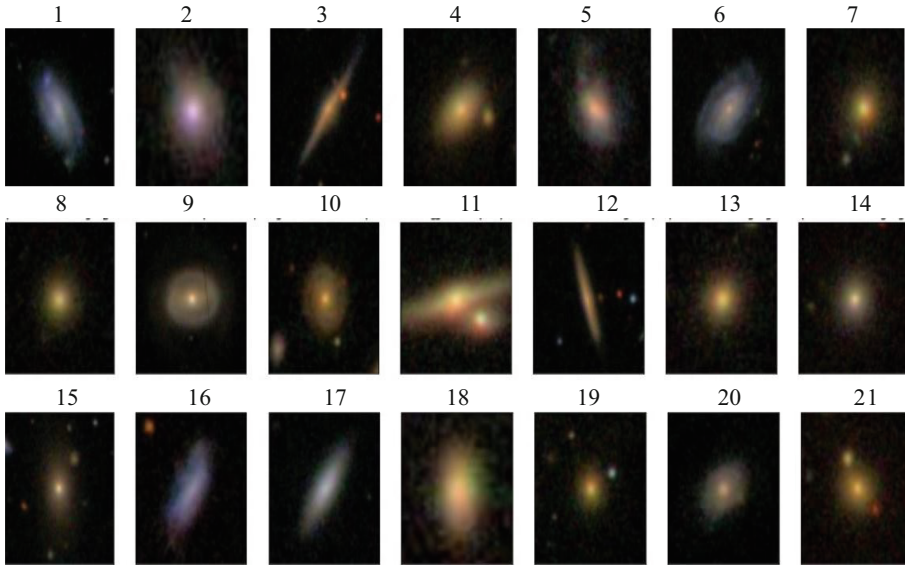


Fig. 7. Final result of classified galaxies

5 Conclusion and Future Scope

Galaxy Classification is a traditional yet interesting topic. It has helped researchers, astronomers, and astrophysicists in dissecting the galaxies morphologically on the basis of their appearance and impression. This paper was based on research on Deep Convolutional Neural Network Framework on the Galaxy Classification issue. The proposed model and framework have trained over 17736 images and accomplished 84.04% testing accuracy.

In future work, the focus will be on galaxy morphology classification should be done to a greater extent with more accuracy. The next step of the research will be training the model with larger and higher quality databases and the dataset. Advancement of the Algorithms for Galaxy Morphological Classification will be taken to a greater extent with the help of advanced Deep Learning Techniques.

References

1. Gharat, S., Dandawate, Y.: Galaxy classification: A deep learning approach for classifying Sloan Digital Sky Survey images. *Month. Notic. R. Astron. Soc.* **511**(4), 5120–5124 (2022)
2. Abd Elfattah, M., Elbendary, N., Elminir, H.K., Abu El-Soud, M.A., Hassanien, A.E.: Galaxies image classification using empirical mode decomposition and machine learning techniques. In: 2014 International Conference on Engineering and Technology (ICET), p. 15 (2014)
3. Nadolny, J., Bongiovanni, Á., Cepa, J., Cerviño, M., María Pérez García, A., Pović, M., Pérez Martínez, R., Sánchez-Portal, M., de Diego, J.A., Pintos-Castro, I., Alfaro, E., Castañeda, H.O., Gallego, J., Jesús González, J., Ignacio González-Serrano, J., Lara-López, M.A., Padilla Torres, C.P.: The OTELO survey is a morphological probe. Last ten Gyr of galaxy evolution. *Astron. Astrophys.* **647**, A89 (2021)

4. Dominguez, A.: A history of the convolution operation. *IEEE Pulse* **6**(1), 3849 (2015)
5. Kim, E.J., Brunner, R.J.: Star-galaxy classification using deep convolutional neural networks. *Month. Notic. R. Astrono. Soc.* **464**(4), 4463–4475 (2017)
6. Srinivas, S., Sarvadevabhatla, R.K., Mopuri, K.R., Prabhu, N., Kruthiventi, S.S.S., Babu, R.V.: A Taxonomy of deep convolutional neural nets for computer vision. *Front. Robot. AI* **2**, 36 (2016)
7. Jiang, H., Learned-Miller, E.: Face detection with the faster R-CNN. In: 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), p. 650657 (2017)
8. Barchi, P.H., et al.: Machine and deep learning applied to galaxy morphology: a comparative study. *Astron. Comput.* **30**, 100334 (2020)
9. Marin, M., Sucar, L.E., Gonzalez, J.A., Diaz, R.: A hierarchical model for morphological galaxy classification. In: Proceedings of the Twenty-Sixth International Florida Artificial Intelligence Research Society Conference, 438443 (2013)
10. Tang, H., Scaife, A.M.M., Wong, O.I., Shabala, S.S.: Radio galaxy zoo: Giant radio galaxy classification using multidomain deep learning. *Month. Notic. R. Astrono. Soc.* **510**(3), 4504–4524 (2022)
11. Selim, I.M., Abd El Aziz, M.: Automated morphological classification of galaxies based on a projection gradient nonnegative matrix factorization algorithm. *Exp. Astron.* **43**(2), 131144 (2017)
12. Cheng, T.-Y., et al.: Optimizing automatic morphological classification of galaxies with machine learning and deep learning using Dark Energy Survey imaging. *Month. Notic. R. Astrono. Soc.* **493**(3), 4209–4228 (2020)
13. Tang, H., Scaife, A.M.M., Leahy, J.P.: Transfer learning for radio galaxy classification. *Month. Notic. R. Astrono. Soc.* **488**(3), 3358–3375 (2019)
14. Mohan, D., Anna, M., Scaife, M., Porter, F., Walmsley, M., Bowles, M.: Quantifying uncertainty in deep learning approaches to radio galaxy classification. *Month. Notic. R. Astrono. Soc.* **511**(3), 3722–3740 (2022)