





Automated Detection of Galactic Rings from Sloan Digital Sky Survey Images

Linn Abraham^{1,2}, Sheelu Abraham^{1,3} , Ajit K. Kembhavi¹, N. S. Philip^{1,4}, A. K. Aniyani⁵, Sudhanshu Barway⁶ , and Harish Kumar²

¹ Inter-University Centre for Astronomy and Astrophysics, Pune 411007, India; linn.official@gmail.com, sheeluabraham@mtcc.ac.in, akk@iucaa.in, ninansajeethphilip@airis4d.com

² Puducherry Technological University, Puducherry, 605014, India; harishkumarholla@ptuniv.edu.in

³ Marthoma College, Chungathara, Nilambur, Kerala, 679334, India

⁴ Artificial Intelligence Research and Intelligent Systems, Kerala, 689544, India

⁵ DeepAlert Ltd, Bromley, BR1 1QE, London, UK; arun@deepalert.ai

⁶ Indian Institute of Astrophysics, Bengaluru, 560034, India; sudhanshu.barway@iiap.res.in

Received 2024 February 25; revised 2024 September 23; accepted 2024 September 28; published 2025 January 6

Abstract

Morphological features in galaxies—like spiral arms, bars, rings, and tidal tails, etc.—carry information about their structure, origin, and evolution. It is therefore important to catalog and study such features and to correlate them with other basic galaxy properties, the environments in which the galaxies are located, and their interactions with other galaxies. The volume of present and future data on galaxies is so large that traditional methods, which involve expert astronomers identifying morphological features through visual inspection, are no longer sufficient. It is therefore necessary to use AI-based techniques like machine learning and deep learning to find morphological structures quickly and efficiently. We report in this study the application of deep learning for finding ring-like structures in galaxy images from the Sloan Digital Sky Survey (SDSS) DR18. We use a catalog by R. J. Buta of ringed galaxies from SDSS to train the network, reaching good accuracy and recall, and generate a catalog of 29,420 galaxies, of which 4855 have ring-like structures with prediction confidence exceeding 90%. Using a catalog of barred galaxy images identified by S. Abraham et al. with deep-learning techniques, we identify a set of 2087 galaxies with bars as well as rings. The catalog should be very useful in understanding the origin of these important morphological structures. As an example of the usefulness of the catalog, we explore the environments and star formation characteristics of the ring galaxies in our sample.

Unified Astronomy Thesaurus concepts: [Astronomy data analysis \(1858\)](#); [Astronomy image processing \(2306\)](#); [Catalogs \(205\)](#); [Galaxies \(573\)](#)

Materials only available in the [online version of record](#): machine-readable table

1. Introduction

Galaxies display a wide variety of forms as a consequence of the differences in their intrinsic structure, interaction with other galaxies, as well as observational biases. Efforts to make sense of galaxy morphology with its multitude of forms have a long history, dating back to Edwin Hubble and others. The Hubble classification system (E. P. Hubble 1926) arranged galaxies into a sequence, ranging from ellipticals to lenticulars to spirals to irregular-shaped objects. Galaxies other than ellipticals were further classified into those with bar-like structures and those without such structures, resulting in a tuning-fork-like diagram. Other more detailed classification systems have been developed (see, e.g., S. Van den Bergh 1998), taking into account further morphological features.

In recent decades, the availability of many millions of higher-resolution multiwavelength images from large deep-field surveys with modern instrumentation have made possible detailed morphological studies (R. J. Buta 2011). Galaxy morphology plays a key role in bettering our understanding of the secular processes that underlie galaxy evolution. Several important questions concerning the formation and evolution of

galaxies can be addressed from the observed changes in morphology with epoch.

A fairly common feature of disk galaxies is the presence of a ring-shaped pattern in their light distribution (R. Buta & F. Combes 1996). These rings in galaxies have come to be considered an integral part of galaxy morphology (J. Kormendy 1979). Recognizing their importance, the de Vaucouleurs revised Hubble–Sandage classification system (VRHS; G. De Vaucouleurs 1959) added rings as another dimension to the two-dimensional Hubble tuning fork, turning it into a classification volume. Rings in galaxies can fundamentally be divided into normal rings, also called “resonance rings,” and “catastrophic rings,” which are a result of galaxy collisions (R. Buta & F. Combes 1996). Although different subclasses exist within the catastrophic rings, such as polar, accretion, and collisional rings, they constitute a very small minority of all observed rings. B. F. Madore et al. (2009) have estimated the abundance of collisional rings to be only about one in 1000. Several theories have been put forward to explain the different ring structures that are observed, with the resonance interpretation being the most popular. The “Manifold Theory” that has been proposed E. Athanassoula et al. (2009) has also had considerable success in explaining several aspects of ring morphology.

Several catalogs that contain the visual morphology classification of galaxies are available today. Notable among them are P. B. Nair & R. G. Abraham (2010) and M. Fukugita et al. (2007),



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

with the former containing the detailed classifications of 14,034 galaxies. Crowd-sourcing efforts such as the Galaxy Zoo citizen science project (C. Lintott et al. 2011) have managed to ramp up the number of galaxy morphology classifications to nearly 900,000, by involving volunteers. This vast database has motivated others to pick specific subtypes of galaxies for detailed classifications. For example, R. J. Buta (2017) has exhaustively classified 3962 ringed galaxies taken from this set. However, all these efforts involve a human in the loop for classifying galaxy morphologies.

With an explosion in the number of galaxy images being produced as part of imaging surveys like the Sloan Digital Sky Survey (SDSS; D. G. York et al. 2000), Panoramic Survey Telescope and Rapid Response System (N. Kaiser et al. 2002), and Dark Energy Survey (Dark Energy Survey Collaboration 2016), etc., automated methods of morphology classification have become very important. Although traditional image-processing techniques, many of which are histogram-based, have had some success (see, e.g., L. Shamir 2020), these are inappropriate for complex tasks, such as shape recognition from images. However, the advent of AI techniques like convolutional neural networks (CNNs; Y. LeCun & Y. Bengio et al. 1995) has significantly improved the automated efforts for galaxy morphology classification. One of the initial efforts in this area has been made by M. Banerji et al. (2010). S. Dieleman et al. (2015) have used CNNs to create a model that can match the combined consensus of the Galaxy Zoo volunteers on all of the questions regarding galaxy morphology that are part of the survey. H. Dominguez Sanchez et al. (2018) and S. Abraham et al. (2018) have used CNNs for the automated classification of galaxy morphological features such as bars, bulges, and edge-on morphology, etc. More recently, R. Shimakawa et al. (2024) have used a deep-learning model to obtain ring classification for a very large sample of galaxies taken from the Hyper Suprime-Cam Subaru Strategic Program.

In the present study, we investigate the efficacy of an automated method based on CNNs for the detection of rings in galaxies. We use galaxy images from SDSS (D. G. York et al. 2000) for the training and testing of our network. We treat the problem as a binary classification one, treating all subclasses of rings as a single `ring` type and distinguishing them from galaxies that do not contain any rings, the `non-ring` type. The trained network can be used to identify ringed galaxies from SDSS. We used the network’s predictions to generate a sample of 4855 ringed galaxies with a classification threshold of 0.90. Using this sample, we explored the connection between ring formation and star formation activity considering the main-sequence, green valley, and quenched galaxy populations. We also analyzed star formation by examining their distribution across different environments, categorized by their local surface densities.

The rest of the paper is organized as follows. In Section 2, we describe the data used in the analysis, and we introduce the neural network architecture in Section 3. We describe the data augmentation techniques and training procedure for the neural network in Sections 4 and 5, respectively, and in Section 6, we analyze the results on the trained network. In Section 7, we present the catalog of ringed galaxies generated using our trained network, and in Section 8, we present a subset of the ringed galaxies that have also been identified to have bars. In Section 9, we provide a discussion regarding the environments

and star formation rates (SFRs) for the ringed galaxies in our catalog. Finally, in Section 10, we give the concluding remarks.

2. Data

The performance of a supervised machine learning classifier depends to a large extent on the quality of the labeled data that are made available to it. For deep-learning models, the quantity becomes important as well, because of the large number of trainable or free parameters in the model. Many of the standard data sets available for galaxy morphology classification do not contain rings as a class (see, e.g., DECALS10).⁷ Even with the Galaxy Zoo 2 (GZ2) data set (K. W. Willett et al. 2013), ringed galaxies for which at least 50% of zoo volunteers agreed on a response are relatively small in number.

2.1. Rings and Nonrings

We used the R. J. Buta (2017) catalog as our primary source for identifying galaxies with rings. The author has provided detailed visual morphological classifications of 3962 such galaxies. The classifications are done within the framework of the comprehensive de Vaucouleurs revised Hubble–Sandage system (CVRHS; R. J. Buta et al. 2007, 2015). The catalog also provides the author’s comments, which are useful for creating a good-quality training set by identifying rare cases of rings, poorly resolved galaxies, etc. The galaxies used for classification by the author were picked from GZ2 (K. W. Willett et al. 2013).

Along with a good-quality training sample for rings, it is equally important to obtain a good sample of galaxies that do not have any presence of rings. Since the R. J. Buta (2017) catalog does not contain such galaxies, we used the P. B. Nair & R. G. Abraham (2010) catalog, which is one of the largest catalogs for the visual morphological classification of galaxies. They used galaxy images from SDSS DR4 (C. Stoughton et al. 2002; J. K. Adelman-McCarthy et al. 2006). Galaxies with spectroscopic redshifts in the range $0.01 < z < 0.1$ and extinction-corrected g' -band magnitude < 16 were selected from the spectroscopic main sample (M. A. Strauss et al. 2002). This led to their final sample size of 14,034 galaxies.

The classification scheme used in P. B. Nair & R. G. Abraham (2010) is not the CVRHS. Rather, it is primarily based on the Carnegie Atlas of Galaxies (A. Sandage & J. Bedke 1994), used in consultation with the Third Reference Catalog of Bright Galaxies (G. de Vaucouleurs et al. 1991). In order to use only galaxies without rings, we selected from the catalog galaxies that have the `ring` type and `ring flag` column set to zero.

2.2. Selection Criteria

The images used for the GZ2 survey come from SDSS DR7 (K. N. Abazajian et al. 2009). GZ2 includes galaxies with extinction-corrected Petrosian half-light magnitude in the r band < 17.0 , Petrosian radius $\text{petroR90}_r > 3''$, and spectroscopic redshift in the range $0.0005 < z < 0.25$, when it is known. Finally, galaxies that have SDSS flags that are either `SATURATED`, `BRIGHT`, or `BLENDED` without an accompanying `NODEBLEND` were removed. We further selected galaxies that have an extinction-corrected g' -band magnitude < 16 based on the selection criteria used for the nonring

⁷ https://astro.utoronto.ca/~hleung/shared/Galaxy10/Galaxy10_DECals.h5

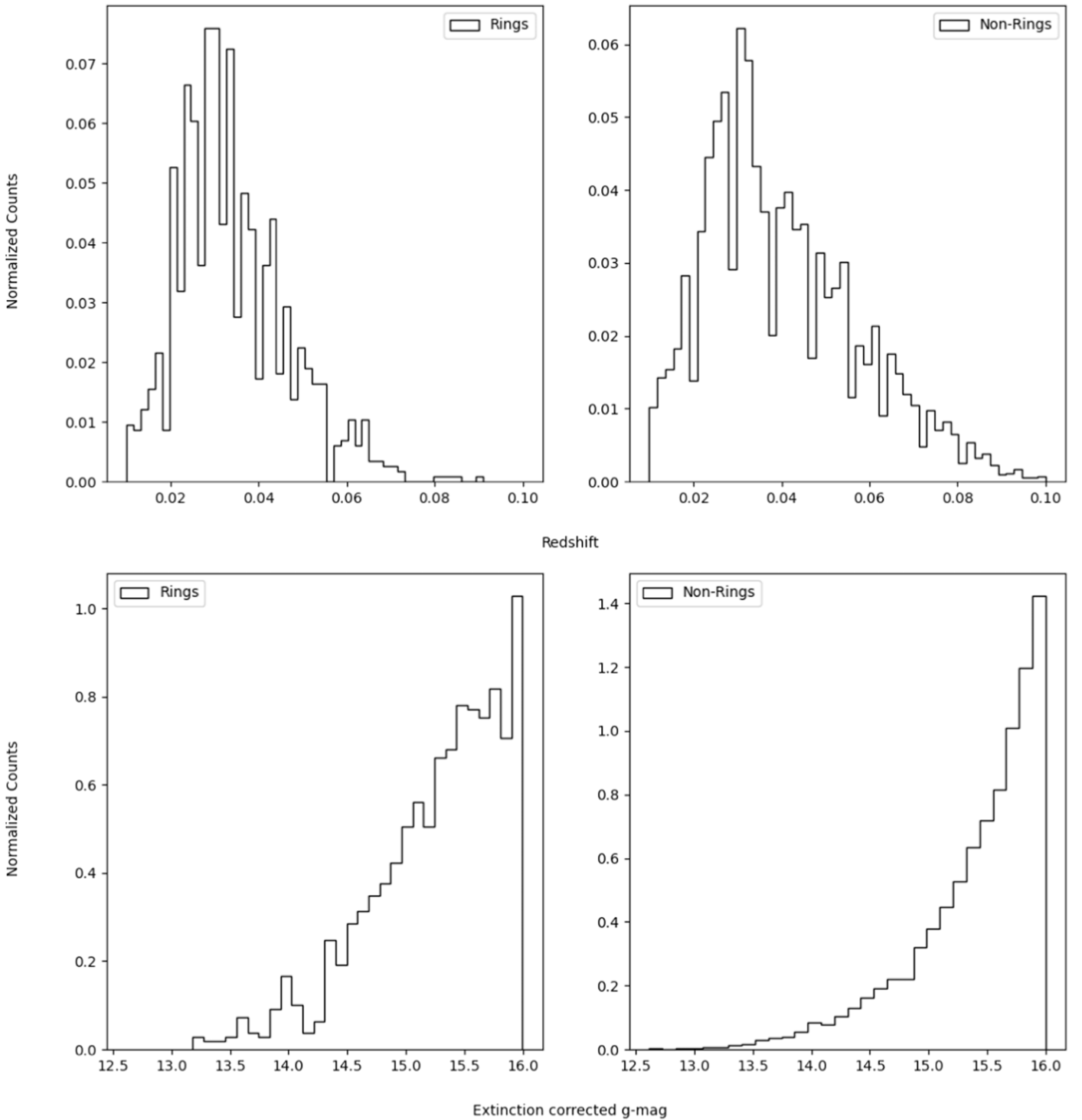


Figure 1. The normalized distribution of redshifts and extinction-corrected g -magnitude for the ringed galaxies (left) and nonringed galaxies (right).

galaxies. The spectroscopic redshift was also limited to be in the same range as that of the nonring galaxies, specifically $0.01 < z < 0.1$. By visual inspection, we removed 37 ringed and 797 nonringed galaxies that contained artifacts and those that were not clearly distinguishable. After the selection criteria were applied and following the manual removal, the number of ringed galaxies obtained was 1122 and the number of galaxies without rings came out to be 10,639. Figure 1 shows the distribution of redshift and extinction-corrected g -magnitude for the ringed and nonringed galaxies in our training set.

2.3. SDSS Image Cutouts

We modified a Python library⁸ that exists in the public domain for downloading the images we needed for training as well as catalog preparation. We incorporated multithreaded parallel downloading using the `Concurrent`⁹ Python module in the script to speed up the downloading process. Figure 2

⁸ <https://pypi.org/project/panstamps/>

⁹ <https://docs.python.org/3/library/concurrent.futures.html>

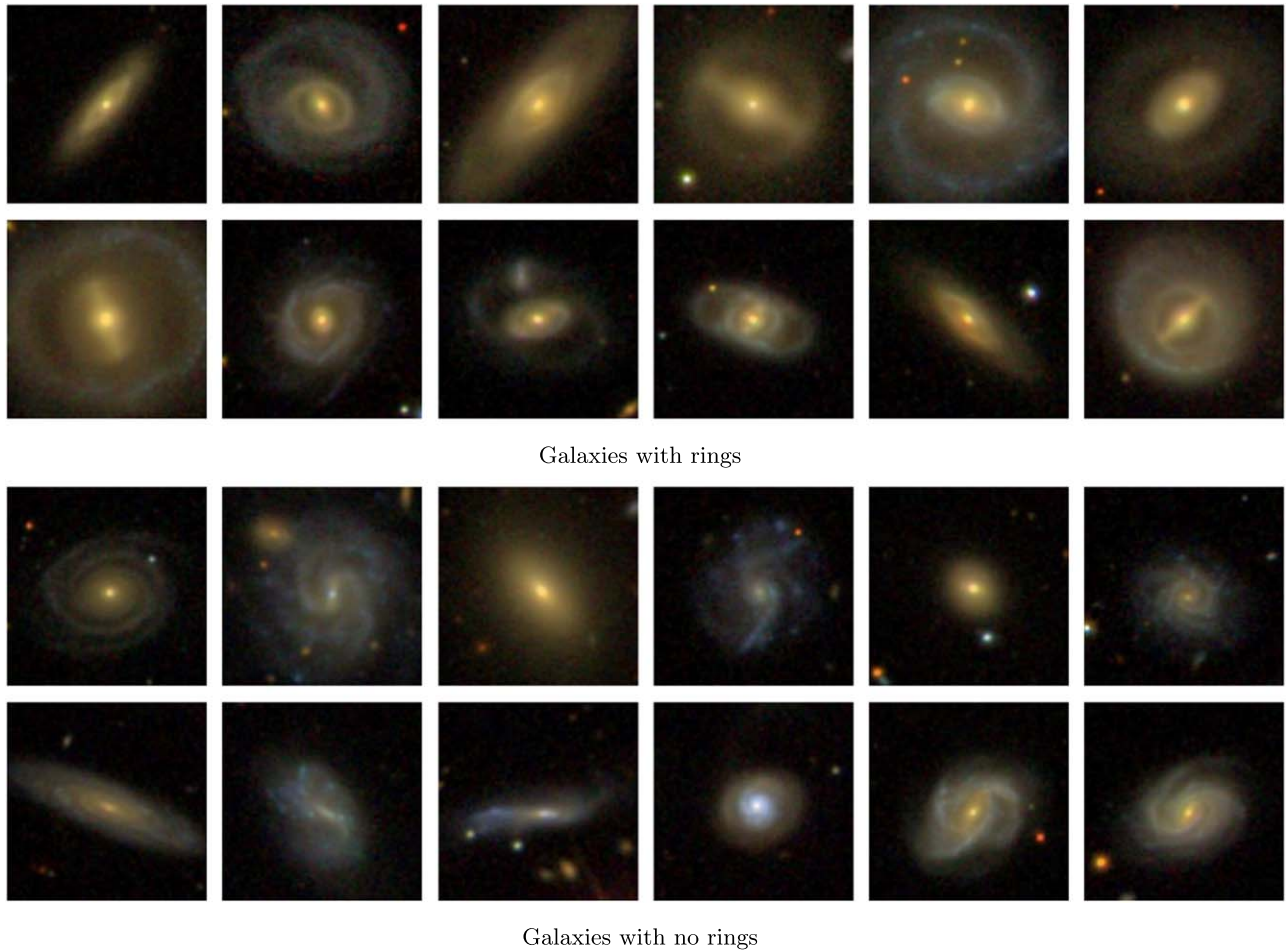


Figure 2. A sample of galaxies with and without rings from our training set.

shows a sample of the galaxies with and without rings in our training set.

3. Network Design

AlexNet (A. Krizhevsky et al. 2012) is one of the earliest CNN architectures that was designed and yet one of the most reliable. Having a smaller size than most modern architectures results in a lesser training time and also the requirement for a lesser amount of training data for this network. The original AlexNet architecture was designed for the ImageNet challenge (J. Deng et al. 2009). The network has eight layers, with the first five being convolutional layers and the remaining three being fully connected. The first convolutional layer filters the $224 \times 224 \times 3$ input image with 96 kernels of size $11 \times 11 \times 3$ with a stride of four pixels. All of the convolution and fully connected layers are immediately followed by a block consisting of a Rectified Linear Unit activation (V. Nair & G. E. Hinton 2010) and a batch normalization (NORM) layer. The first, second, and final convolution layers are each followed by a max-pooling layer, which leads to 50% downsampling of the convolution feature maps. These layers also contribute to the rotational invariance of the learnt model weights (Y.-L. Boureau et al. 2010). The max-pooling layers of the first two convolutions are followed by dropout layers (N. Srivastava et al. 2014), having a dropout fraction of 0.25. These prevent overfitting of the model by randomly dropping 25% of the weights during each training iteration.

We left the input layer size unchanged, since it was suitable for our purposes. However, galactic rings are features that are harder to distinguish in comparison to the classes in the ImageNet challenge. We therefore used a smaller size of 5 for the kernels in the first convolutional layer. This seemed to have a considerable impact on the accuracy of the network. Thus, in our network, the first convolution layer has 96 kernels of size $5 \times 5 \times 3$ and the output is fed to the second layer of convolution with 256 kernels each of size $5 \times 5 \times 48$. The third, fourth, and fifth convolutional layers have respectively 384, 384, and 256 kernels of size $3 \times 3 \times 256$, $3 \times 3 \times 192$, and $3 \times 3 \times 192$, respectively. These learn more complex features and their final output is fed into another max-pooling layer. The output of the final pooling layer is then fed into a series of fully connected layers of size 4096. These layers are connected by dropout layers, which have 0.50 dropout fraction. The final layer is the sigmoid activation function layer, which outputs a value that can be interpreted as the probability of the input belonging to the positive class. The architecture of our network is shown as a table and schematic diagram in Figure 3.

3.1. Selection of Hyperparameters

Hyperparameters are those parameters that are not learnt or updated during the learning process and hence need to be manually set by the programmer. These can include the network architecture itself and several other parameters, like the batch size used for the stochastic gradient descent or the

	Name	Type	Shape
0	conv2d	Conv2D	(None, 100, 100, 96)
1	activation	Activation	(None, 100, 100, 96)
2	batch_normalization	BatchNormalization	(None, 100, 100, 96)
3	max_pooling2d	MaxPooling2D	(None, 49, 49, 96)
4	dropout	Dropout	(None, 49, 49, 96)
5	conv2d_1	Conv2D	(None, 49, 49, 256)
6	activation_1	Activation	(None, 49, 49, 256)
7	batch_normalization_1	BatchNormalization	(None, 49, 49, 256)
8	max_pooling2d_1	MaxPooling2D	(None, 24, 24, 256)
9	dropout_1	Dropout	(None, 24, 24, 256)
10	conv2d_2	Conv2D	(None, 24, 24, 384)
11	activation_2	Activation	(None, 24, 24, 384)
12	batch_normalization_2	BatchNormalization	(None, 24, 24, 384)
13	conv2d_3	Conv2D	(None, 24, 24, 384)
14	activation_3	Activation	(None, 24, 24, 384)
15	batch_normalization_3	BatchNormalization	(None, 24, 24, 384)
16	conv2d_4	Conv2D	(None, 24, 24, 256)
17	activation_4	Activation	(None, 24, 24, 256)
18	batch_normalization_4	BatchNormalization	(None, 24, 24, 256)
19	max_pooling2d_2	MaxPooling2D	(None, 11, 11, 256)
20	dropout_2	Dropout	(None, 11, 11, 256)
21	flatten	Flatten	(None, 30976)
22	dense	Dense	(None, 4096)
23	activation_5	Activation	(None, 4096)
24	batch_normalization_5	BatchNormalization	(None, 4096)
25	dropout_3	Dropout	(None, 4096)
26	dense_1	Dense	(None, 4096)
27	activation_6	Activation	(None, 4096)
28	batch_normalization_6	BatchNormalization	(None, 4096)
29	dropout_4	Dropout	(None, 4096)
30	dense_2	Dense	(None, 2)
31	activation_7	Activation	(None, 2)

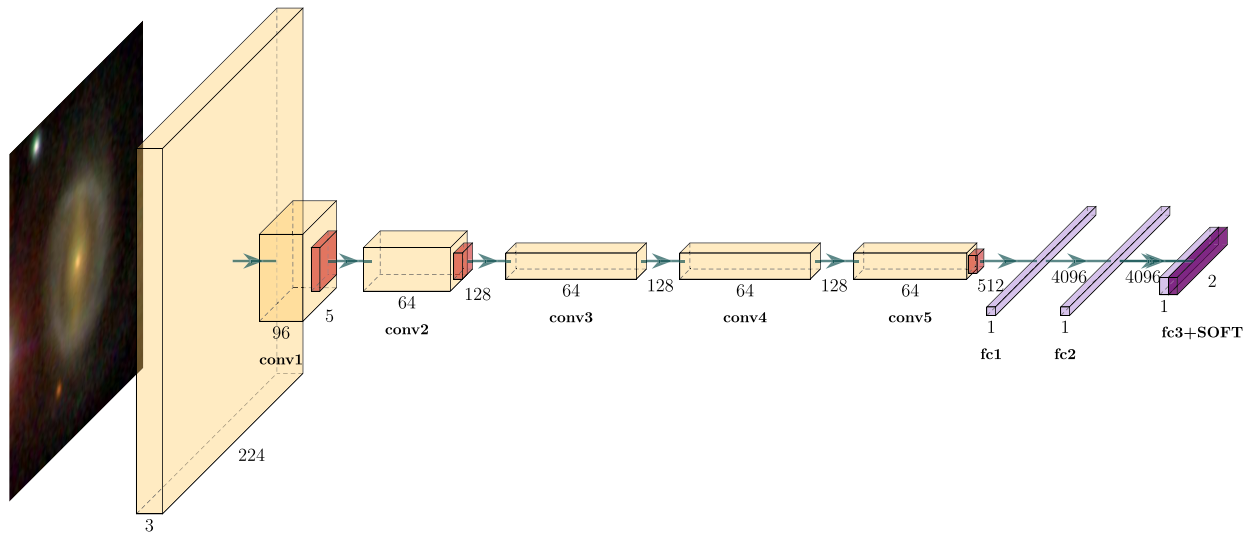


Figure 3. A tabular description of the AlexNet model architecture that we have used, generated with the Keras package (above) and a visual representation of the same (below). The code for generating the figure was adapted from H. Iqbal (2018).

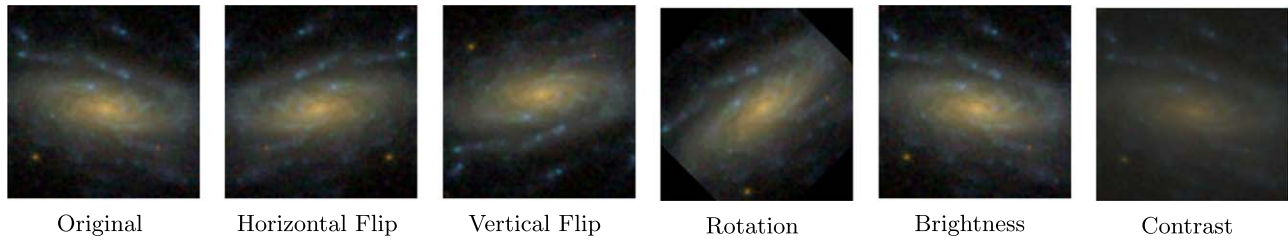


Figure 4. The different types of image augmentations that are done to each image in our training set.

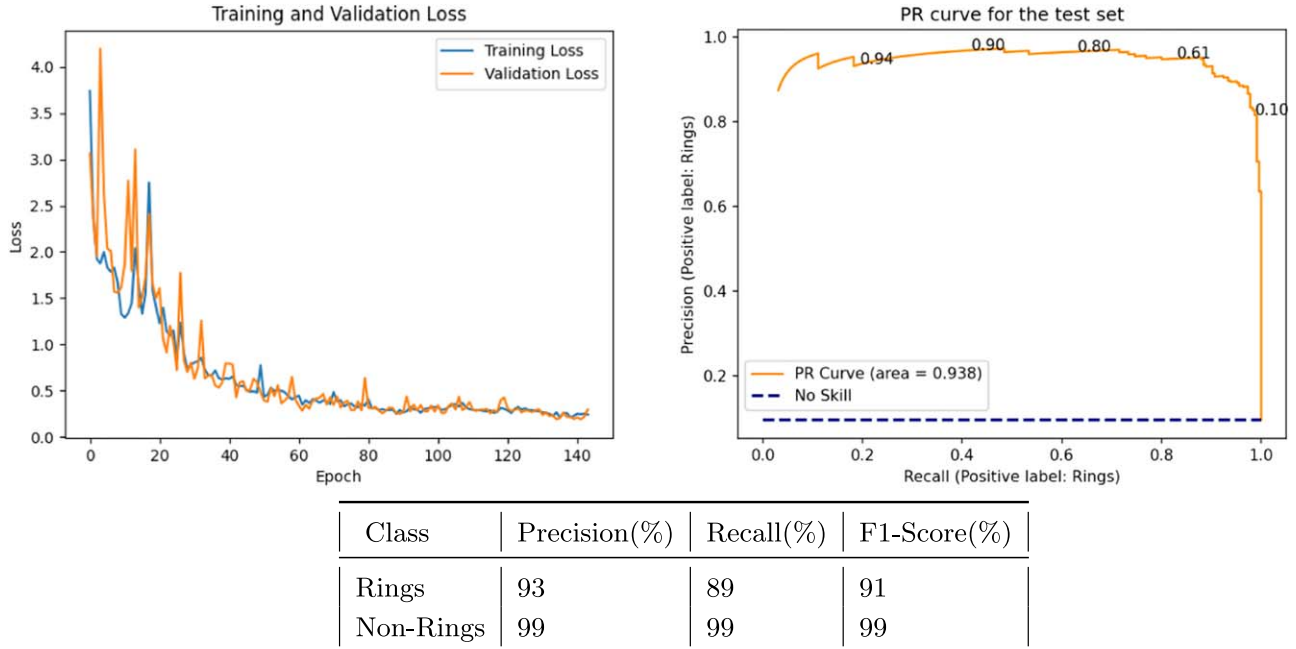


Figure 5. Top left: the training graph, which shows a plot of the training and validation loss against the number of epochs. Top right: the precision–recall curve, which shows the plot of precision against recall for ringed galaxies in our test set computed at various thresholds. The dashed line shows the curve expected from a no-skill classifier. A few thresholds taken at random are annotated on the curve. Bottom: the classification report showing the precision and recall for both rings and nonrings in our test set.

number of epochs of training, etc. Hyperparameter tuning is an important part of a deep-learning project, since these parameters have a direct bearing on the performance and accuracy of the model.

The loss function, which specifies the error and how much weightage should be given to individual errors, is an important hyperparameter to tune. We have used `categorical_crossentropy`,¹⁰ which is a standard loss function suitable for binary classification problems. The optimizer function has an effect on how the network weights are updated in order to minimize this loss. We used the Adam optimizer with a learning rate of 0.003. Regularization techniques are a set of safeguard measures that prevent the model from overfitting on the training data and thus losing its ability to generalize. We have used L2 weight regularization with a factor of 0.0002 to penalize the weights from growing too much during training. The batch size controls how many input samples are considered together for evaluating the loss function. This leads to a trade-off between accuracy and speed. A batch size of 16 was seen to be sufficient in our case. The number of epochs (or `num_epochs`) decides how many times the network gets to see

the complete training data. We trained the network for `num_epochs` without early stopping.

4. Image Augmentation

Neural networks are prone to overfitting if the data available for training are insufficient. In our case, the possibility of increasing the training data is limited by the availability of catalogs of ringed galaxies. It is therefore necessary to increase the training sample size, through data augmentation, by transforming the available images. We have used operations such as horizontal and vertical flipping, rotation through arbitrary angles, and brightness and contrast adjustments for augmenting our training set. Using rotations also makes the network invariant to rotations, which is needed because of the rotational symmetry of the ringed galaxies. The augmentations that we have used are illustrated in Figure 4 using a random galaxy.

We implemented image augmentations using the `on-the-fly` mode provided by the `keras-tensorflow` (F. Chollet et al. 2015) framework. Here, a predefined sequence of transformations is applied to each image from the training set, using parameters that are randomized, so that at each epoch, the network sees a different set of augmented images. In

¹⁰ https://www.tensorflow.org/api_docs/python/tf/keras/metrics/categorical_crossentropy

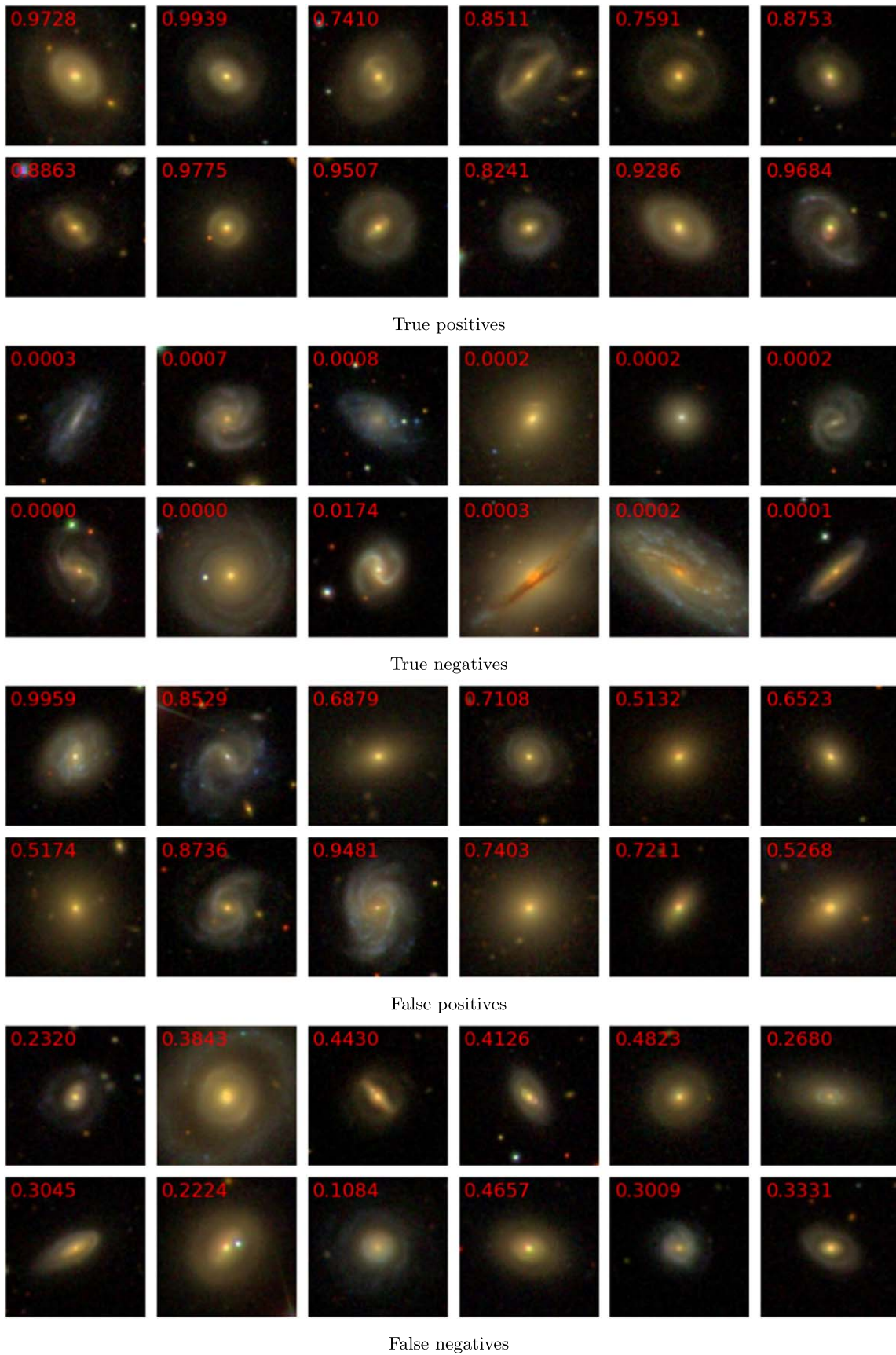


Figure 6. Examples of images from our test set that are both correctly and incorrectly classified.



Figure 7. Cutouts of a single galaxy downloaded at various zoom levels from the SDSS cutouts server and predicted on using the trained network. The prediction scores obtained for each are shown in red.

this way, we avoid generating all the augmented images before the start of the training and also save disk space.

5. Training

The galaxy image cutouts used for training are three-channel RGB 256×256 pixel images, which are dynamically resized by the Keras `image_data_set_from_directory`¹¹ function into the required input shape of 240×240 . The pixel intensities in the range (0–255) are rescaled to the range (0, 1) for easier computation. The input size and other parameters for the network, like `num_epochs`, `batch_size`, are controlled through a config file that is read at run time. The images are first resized to the desired input size and then loaded into memory in `n_batches` using the Keras `image_data_set_from_directory` function. A generator avoids the need for all of the images to be loaded into memory at once.

The data are initially split into a training and testing set using a ratio of 80:20. During training, the training set is further split internally into a training and validation set, using the same ratio. Before augmentations are applied, the number of ringed galaxies in our training set (including the validation split) is 897 and the number in the testing set is 225. Similarly, the number of galaxies without rings in our training set is 8511 and the number in the testing set is 2128. The network was trained for a total of 140 epochs. The training graph in Figure 5 shows the variation of the `categorical_crossentropy` loss function with increasing epochs for both the training and validation data. If there is overfitting, the validation loss of the network increases, whereas the training loss keeps decreasing. In our case, the training loss and validation loss are seen to decrease together with each epoch, and therefore there is no significant overfitting that occurs. There is further scope for training the network, albeit with minimal returns.

The training was done on the Amazon Web Services platform on an Intel Xeon workstation with four virtual CPUs and 16 GB RAM, making use of a Tesla (g4dn.xlarge) GPU accelerator

with 16 GB graphics memory. A community Amazon Machine Image running Arch Linux was modified to function as our base operating system. The data and model are version-controlled using Data Version Control (R. Kupriev et al. 2022) and stored on Amazon S3 cloud storage. The code itself is version-controlled and made available on github (L. Abraham 2024). The entire training session takes 2.5 hr to run with our hardware specifications.

6. Results

6.1. Evaluation Metrics for Unbalanced Data Sets

For evaluating the performance of the network in discriminating between galaxies with rings and those without rings, we use the three metrics accuracy, precision, and recall. Accuracy is the number of correct classifications as a fraction of the total number of classifications made:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}},$$

where TP denotes the number of true positives, FP the number of false positives, and FN the number of false negatives. Precision and recall are defined as

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}},$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}.$$

Precision is an indicator of the purity of the classification—that is, the extent to which a collection of galaxies identified as ringed are indeed ring galaxies. Recall, on the other hand, is a measure of the extent of completeness in identifying galaxies from a sample containing galaxies with and without rings. Precision and recall are therefore important for astronomers.

Our trained model obtained a classification accuracy of 98% on the testing set. The threshold used for converting prediction scores to class labels is 0.50. The classification report shown in Figure 5 summarizes the precision and recall we have obtained

¹¹ https://www.tensorflow.org/api_docs/python/tf/keras/utils/image_dataset_from_directory

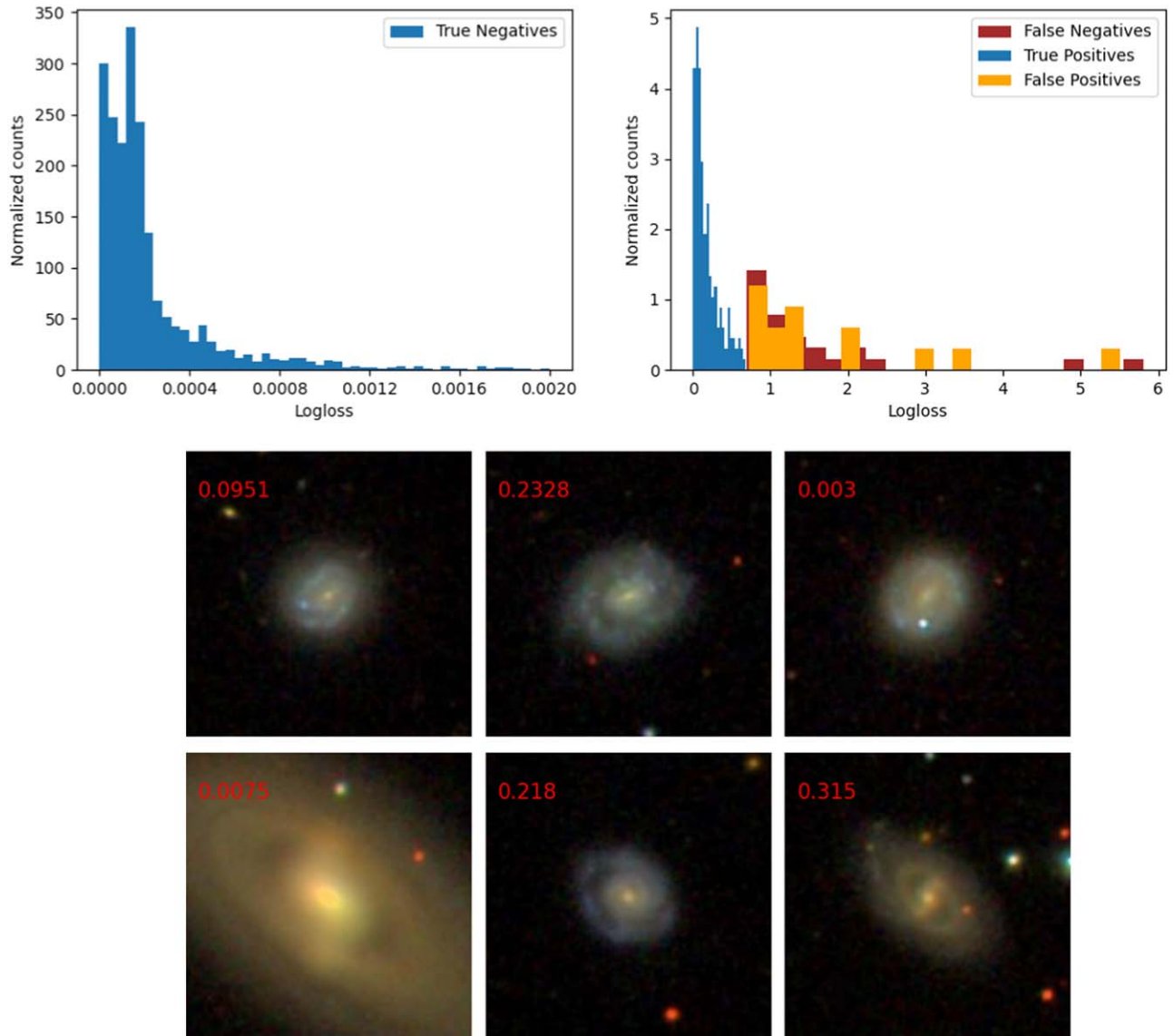


Figure 8. Top: the normalized distribution of log-loss in our test set. The true negatives are shown separately because of the difference in bin range. Bottom: a sample of galaxies from the false negatives having the highest log-loss values. The annotations shown in red denote the prediction scores.

for both the classes. The precision obtained for the rings was 93%. Our model also has a good recall value of 89%. Using the predicted probabilities of our network, a catalog with higher purity can be obtained by appropriately selecting the classification threshold. It is to be kept in mind that higher purity comes at a cost of reduced completeness. The nature of the trade-off can be understood using the precision–recall curve shown in Figure 5. Figure 6 shows galaxies in the test set that are correctly classified (True Positives and True Negatives) and failures (False Positives and False Negatives) obtained using our trained network.

6.2. Predictions at Different Zoom Levels

To get a better insight into what the model has learned, we downloaded a single galaxy at different zoom levels using the SDSS image cutout service. This is done by varying the scale factor and downloading the different images. The trained network was then used to predict on these images. Figure 7 shows different versions of a particular ringed galaxy along

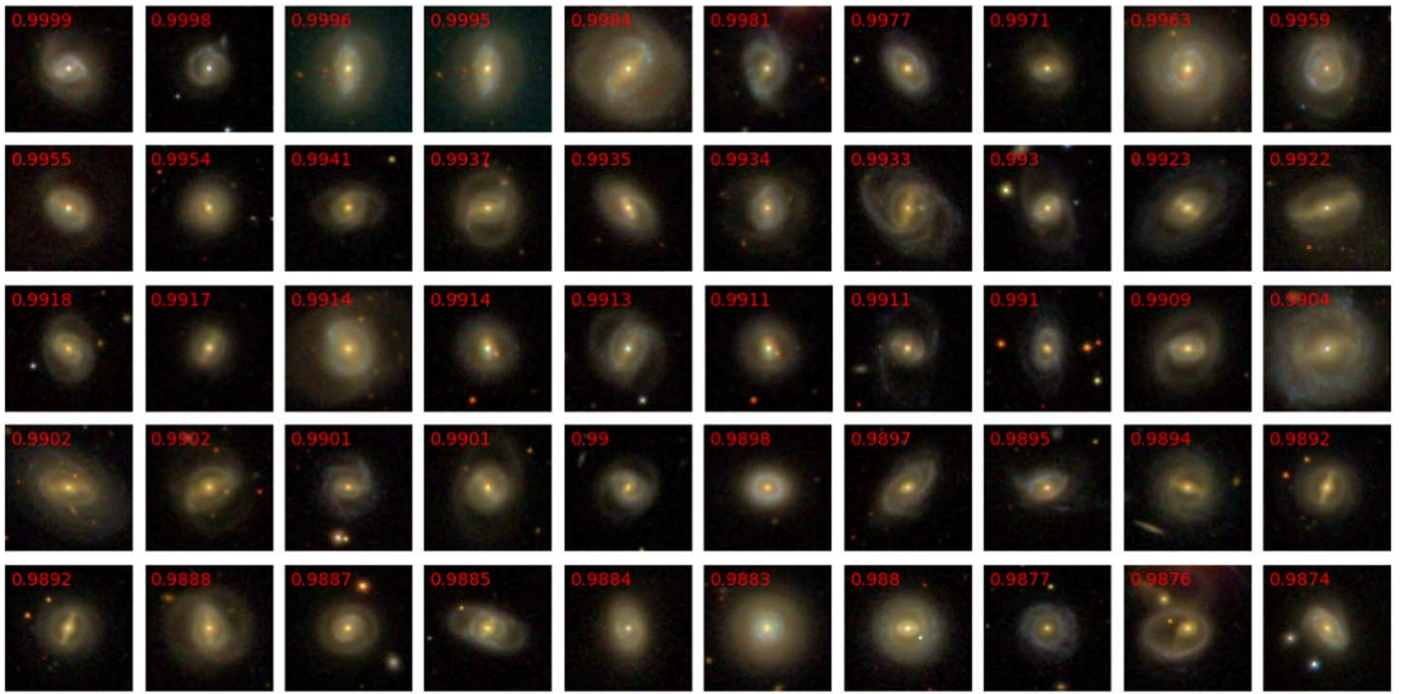
with the probability score returned by the network. It can be seen that the probability score goes to zero both when the images are too zoomed out or too zoomed in, such that the ring feature is not visible. For other zoom levels, the probability scores given by the network are comparable to what a human observer would assign.

6.3. Analysis of Failed Predictions

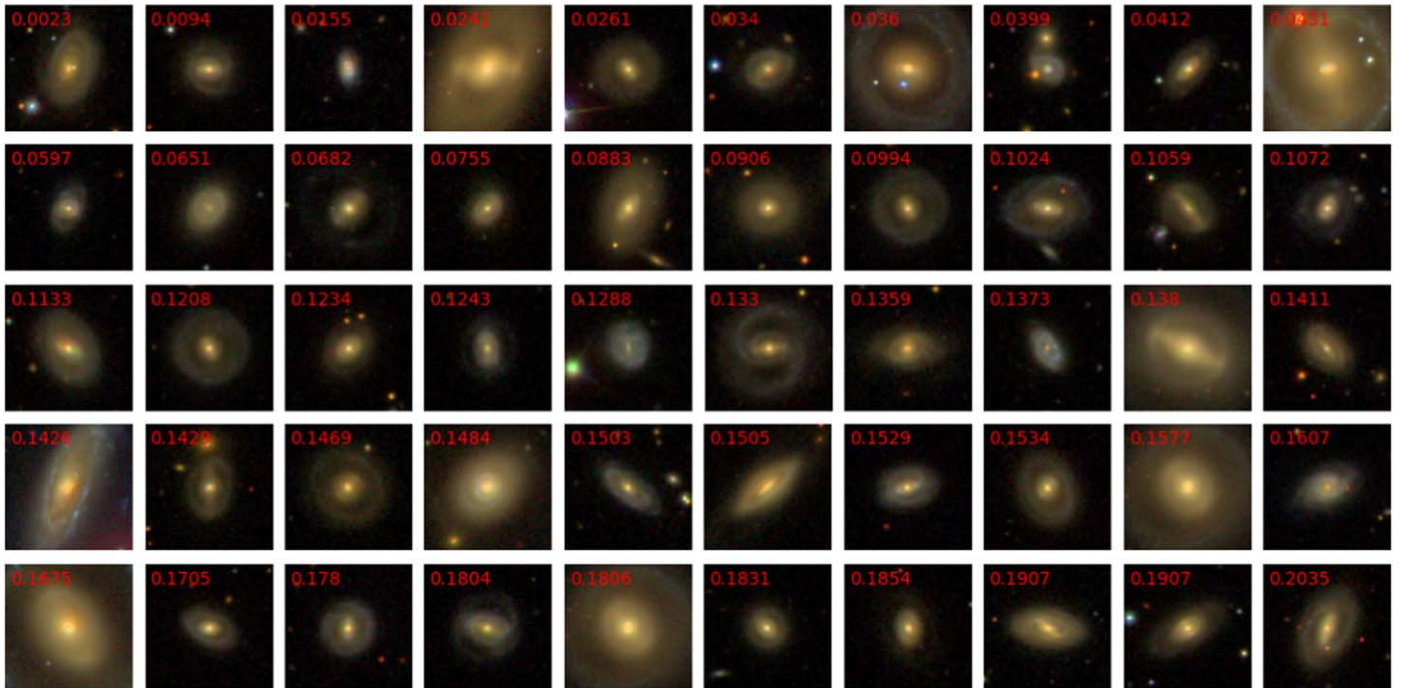
To get a better idea of the performance of our classifier, we analyzed the failed predictions. A quantitative way of doing this is to compute the log-loss for each sample in our test set. The log-loss is defined by

$$\text{logloss}_i = -[y_i \ln p_i + (1 - y_i) \ln (1 - p_i)],$$

where y_i is the true value, which is either 0 or 1, and p_i is the predicted probability score for observation i . In our binary classification problem, with a sigmoid activation for the last layer, the predicted value is a real-valued number between 0 and 1, which is interpreted as a probability of the sample actually



(a)



(b)

Figure 9. Ringed galaxies from our training sample predicted using the trained model and with a low log-loss (above) and high log-loss (below). The annotations shown in red denote the prediction scores.

belonging to the positive class. For the negative class, a low predicted probability implies a high confidence value. Hence, in this case, we need to take the difference of the value from one in order to get the real confidence. When analyzing the failures in our testing set, we can use a loss function that is weighted using the confidence with which a prediction has been made.

The histogram of these losses can then be used to get an idea about the overall performance of the model. Figure 8 shows the

histogram of log-loss obtained for our test set. The peak of the histograms is toward the lower end, which means that there are a lesser number of wrong predictions made with a high confidence value and a greater number of correct predictions made with a high confidence value. That is, it can be seen that the confidence of the wrong predictions is much lower than that of the correct predictions. The figure also shows representative images from our test set corresponding to the false negatives

Table 1
A Preview of the Catalog of Ringed Galaxies (Above) and of Barred Rings (Below), with the First 20 Rows from Each Catalog Shown

R.A.	Decl.	Objid	gmag	deVRad_r	deVAB_g	Redshift	Prediction	Label
40.285690	-0.714957	1237645941824356443	15.641980	9.327478	0.339572	0.040287	0.908651	Rings
57.025337	0.208845	1237645942905438473	15.497970	10.815760	0.871453	0.025475	0.598131	NonRings
56.781387	1.000343	1237645943979114582	15.879010	2.804902	0.321389	0.039847	0.250911	NonRings
56.847420	0.875488	1237645943979114622	15.182350	13.953940	0.724238	0.039371	0.865321	NonRings
57.248385	0.925979	1237645943979311221	15.657840	6.806947	0.535690	0.035788	0.818097	NonRings
57.674720	1.040755	1237645943979507954	15.344130	3.541592	0.822157	0.037123	0.516564	NonRings
243.708876	-0.915653	1237648672921485632	15.293520	8.426456	0.640077	0.030767	0.836999	NonRings
246.015172	-0.902869	1237648672922468973	15.979040	15.339940	0.550030	0.046526	0.494671	NonRings
245.367353	-0.457074	1237648673459077169	15.901310	8.323227	0.500005	0.059092	0.534696	NonRings
246.782081	-0.492432	1237648673459667234	15.630960	8.579305	0.437323	0.046192	0.457613	NonRings
189.522249	-0.027031	1237648673971437623	14.639720	18.461840	0.442339	0.012513	0.802063	NonRings
237.945144	-0.105170	1237648673992671592	15.529860	5.215971	0.949471	0.054426	0.630027	NonRings
243.236782	-0.096251	1237648673994965546	14.331720	27.210810	0.723929	0.030867	0.879053	NonRings
243.583196	-0.031564	1237648673995162093	15.782320	5.114430	0.559130	0.030811	0.536166	NonRings
245.381888	-0.072364	1237648673995948107	15.426320	3.416426	0.888818	0.027287	0.904141	Rings
248.064172	-0.049932	1237648673997127724	15.503360	5.104637	0.811834	0.044017	0.924116	Rings
195.644361	0.348565	1237648674510995594	14.989240	7.022532	0.659088	0.067754	0.922850	Rings
242.672630	0.276520	1237648674531639653	15.742830	5.721440	0.605837	0.062059	0.449378	NonRings
243.154609	0.379972	1237648674531836029	15.808500	3.519223	0.956060	0.043145	0.945323	Rings
249.820694	0.410283	1237648674534719840	15.602230	4.794272	0.273508	0.024328	0.069281	NonRings

R.A.	Decl.	SDSS_Objid	Prediction	Label
0.013400	-1.113000	1237663275780276407	0.001809	Barred-NonRings
0.019800	0.781700	1237657191978959126	0.650422	Barred-Rings
0.032300	-0.723700	1237663783123681369	0.039325	Barred-NonRings
0.056000	-1.213600	1237663275780276438	0.060768	Barred-NonRings
0.114500	14.962500	1237656495650570466	0.208050	Barred-NonRings
0.117000	14.381200	1237652942638481637	0.183115	Barred-NonRings
0.128700	-1.213000	1237663275780341888	0.532812	Barred-Rings
0.158900	14.623600	1237656495113699423	0.024679	Barred-NonRings
0.166500	-1.191200	1237663275780341936	0.036194	Barred-NonRings
0.228100	15.218700	1237652943712288939	0.000647	Barred-NonRings
0.241000	14.190400	1237656494576894171	0.162884	Barred-NonRings
0.241700	14.864400	1237652943175418025	0.215693	Barred-NonRings
0.261300	-0.257100	1237663783660617841	0.001189	Barred-NonRings
0.297800	15.985900	1237656496724443439	0.000518	Barred-NonRings
0.300100	15.351400	1237652943712288996	0.173043	Barred-NonRings
0.306000	-0.992900	1237657189831606381	0.008202	Barred-NonRings
0.368800	-0.224400	1237663783660683434	0.321357	Barred-NonRings
0.380900	14.407500	1237652942638612625	0.211441	Barred-NonRings
0.388600	-0.730500	1237663783123812523	0.000156	Barred-NonRings
0.398900	-0.395200	1237663783660683462	0.303564	Barred-NonRings

(This table is available in its entirety in machine-readable form in the [online article](#).)

and having high log-loss values. The rings in these galaxies are not very distinct, and one of the galaxies has a high inclination angle that might have confused the network.

6.4. Model Interpretability

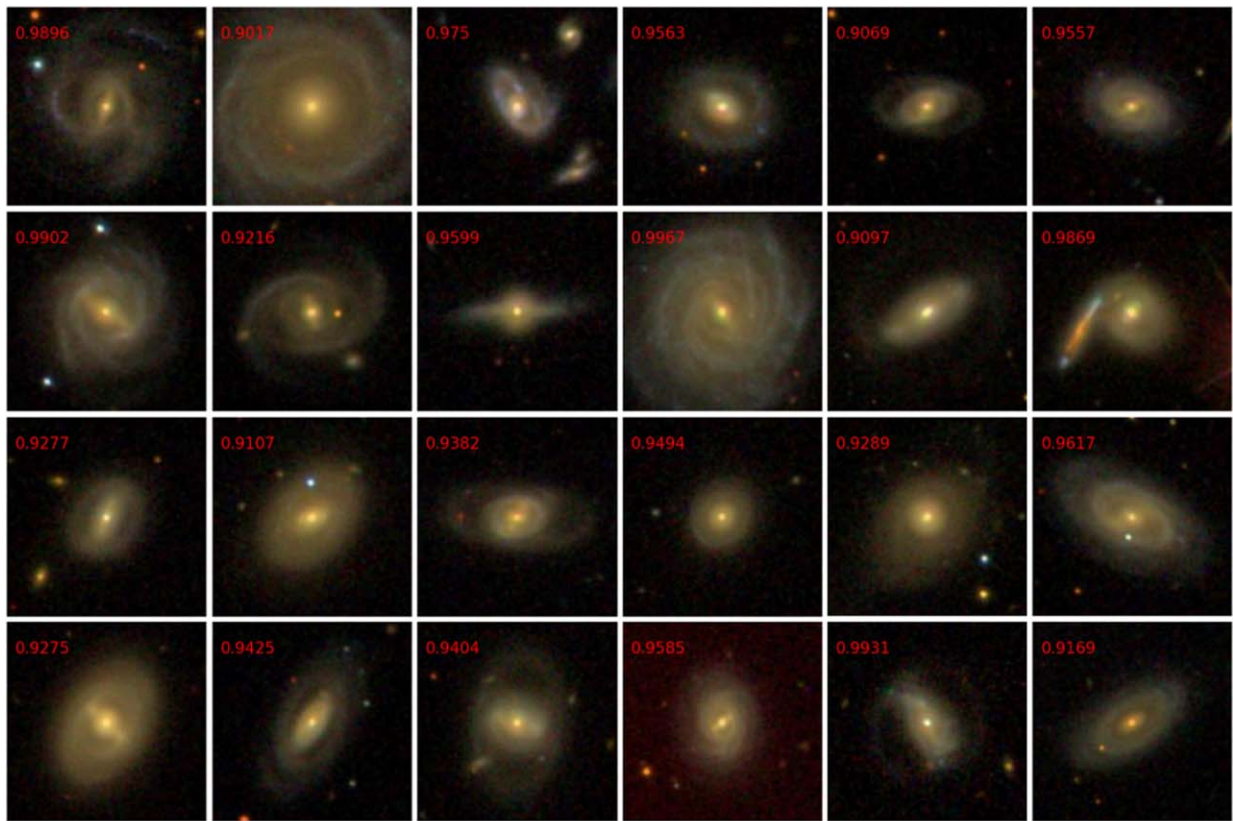
Once the model has been trained, if we use it to make predictions on the training data themselves, we will be able to get some idea of what the model has learned, which is known as the model interpretability. We used the trained model to predict the probability of each ringed galaxy in the sample being identified as ringed. The difference ($1 -$ the predicted probability) can be taken as a measure of the error. This error can be used for finding subsets of ringed galaxies in our training sample that are either easy for the network to learn or difficult to learn. The results for a sample are shown in Figure 9. It can be seen that among the galaxy images that the

network had trouble in learning, there are many with image artifacts or the presence of background stars or galaxies. Many of the galaxies have a zoomed-in view, as well. The galaxy images that were easily identified by the network had a better contrast with the background and the galaxies were at a normal zoom level.

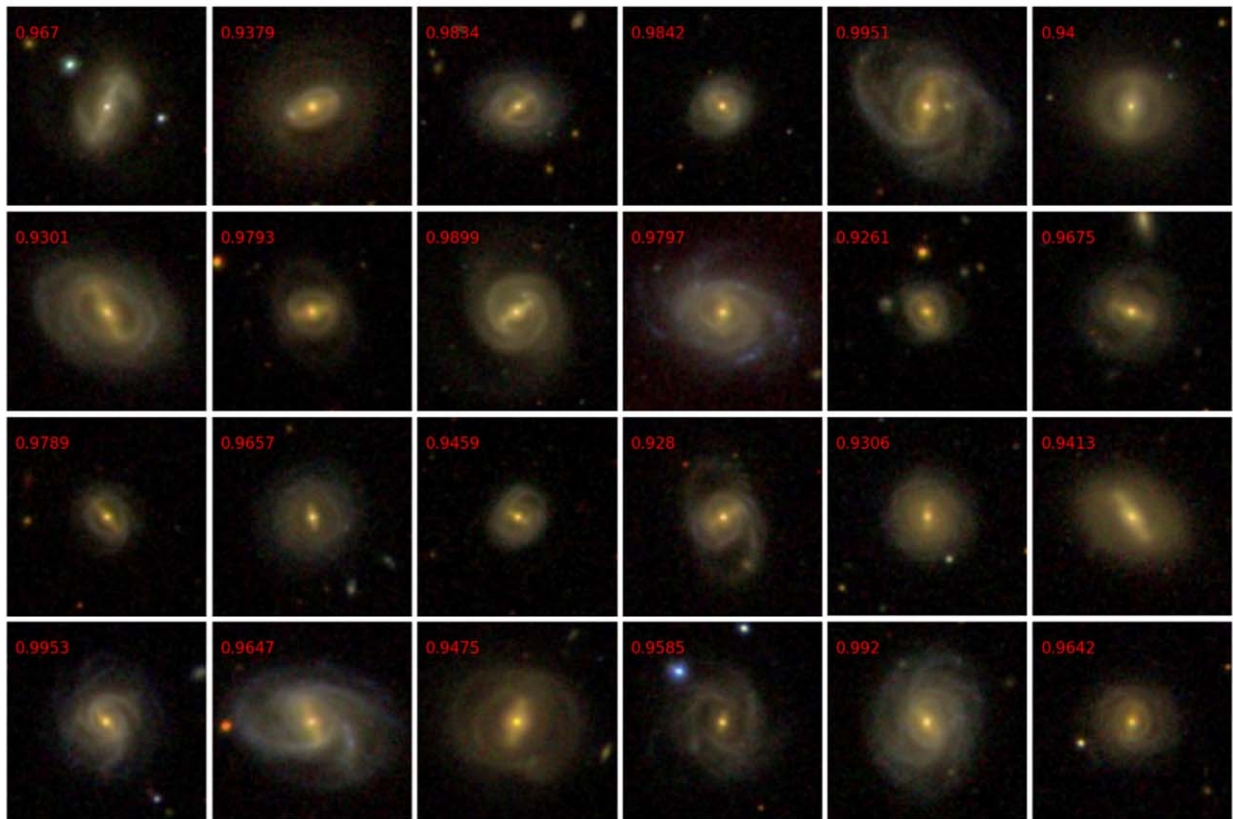
7. A Catalog of Ringed Galaxies

To prepare a catalog of ringed galaxies from SDSS, we used the same selection criteria that were used for the training data preparation. The SDSS data can be queried using the CASJOBS¹² interface. The complete query is given in the Appendix. The source list obtained from the CASJOBS server was downloaded using our bulk download script. The trained

¹² <https://skyserver.sdss.org/casjobs/default.aspx>



(a)



(b)

Figure 10. Some good examples of ringed galaxies (above) and barred ring galaxies (below) from our catalog. The annotations shown in red denote the prediction scores.

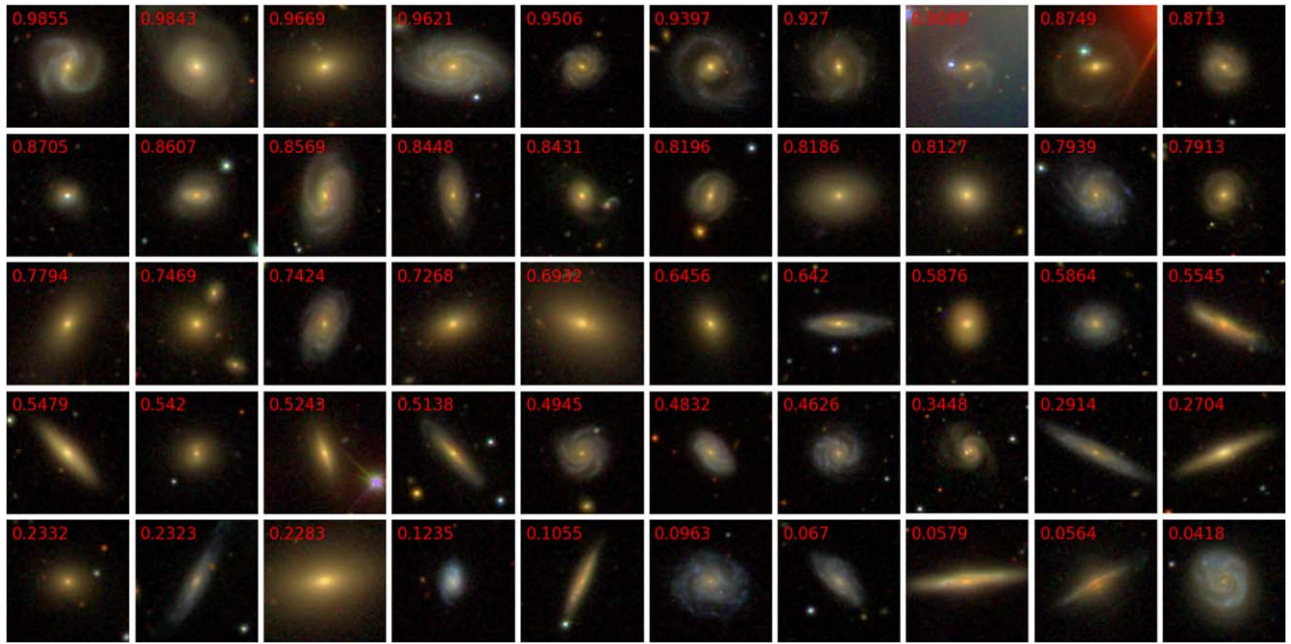


Figure 11. A mosaic of 50 randomly selected images from the prediction set sorted using the predicted probability of having a ring. Visual inspection shows that the rings have accumulated toward the bottom and the nonrings toward the top.

network was then used to predict the ringed or nonringed nature of these galaxies. A preview of the catalog is included in Table 1. Figure 10 shows a selection of galaxies with rings taken from our catalog.

In order to conduct a more thorough testing of our network on real world data, we show in Figure 11 an image mosaic created from 50 randomly selected galaxies from our prediction set, along with the machine-generated probability scores. The galaxies are sorted in descending order of the prediction score. Based on the natural abundance of ringed galaxies, we may expect less than 10 of these to be ringed. It can be seen on visual inspection that most of the galaxies with true rings have high prediction scores in comparison to the nonrings, and hence they are clustered toward the top of the image mosaic. We found 4855 galaxies that are predicted to have rings, using a classification threshold of 0.90. This corresponds to a ring galaxy fraction of 16.5%. At a slightly lower threshold of 0.89, we recover the same ring galaxy fraction of 18.2 as observed by P. B. Nair & R. G. Abraham (2010). Figure 12 shows how this fraction varies for all thresholds between 0 and 1. Further, in Figure 13, we show how the distribution of the extinction-corrected g -magnitude of the predictions compares with that of the training sample.

8. A Catalog of Ringed Galaxies with Bars

Bars in galaxies are much more common than rings. Approximately half of the observed galaxies are seen to have bars (J. H. Knapen 1999). However, galaxies that contain both bars and rings are certainly a smaller fraction. Dedicated catalogs for such galaxies are also quite rare. Here, we use the S. Abraham et al. (2018) catalog, which is an automated catalog of galaxies with bars in SDSS, to find galaxies with both bars and rings. The S. Abraham et al. (2018) catalog contains 111,838 galaxies that have been predicted as either barred or unbarred. We selected the barred galaxies, which came to be 25,781 in number, and used our trained network to predict which of these galaxies also have rings. Figure 10

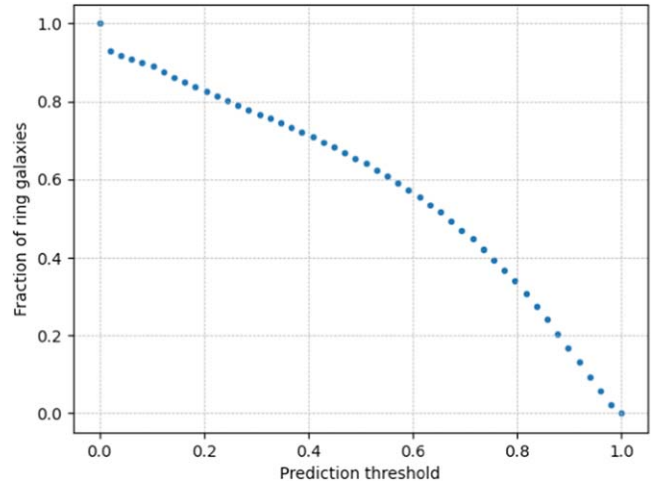


Figure 12. Plot showing the fraction of ring galaxies predicted at various thresholds.

shows a random selection of the barred galaxies that have been predicted to have rings with a probability score of greater than 90%. We found 2087 galaxies that contained both bars and rings using a classification threshold of 0.5. A preview of the catalog containing barred galaxies with predictions for the presence or absence of a ring is shown in Table 1.

9. Discussion

As a demonstration of the usefulness of the catalogs generated in this paper, in this section, we will explore the environments and star formation characteristics of the ring galaxies in our sample. Our goal is to analyze star formation within the framework of the main sequence of star-forming galaxies, the transitioning population of green-valley galaxies, and the quenched galaxies. Additionally, we will examine star formation in relation to the environmental variations from low

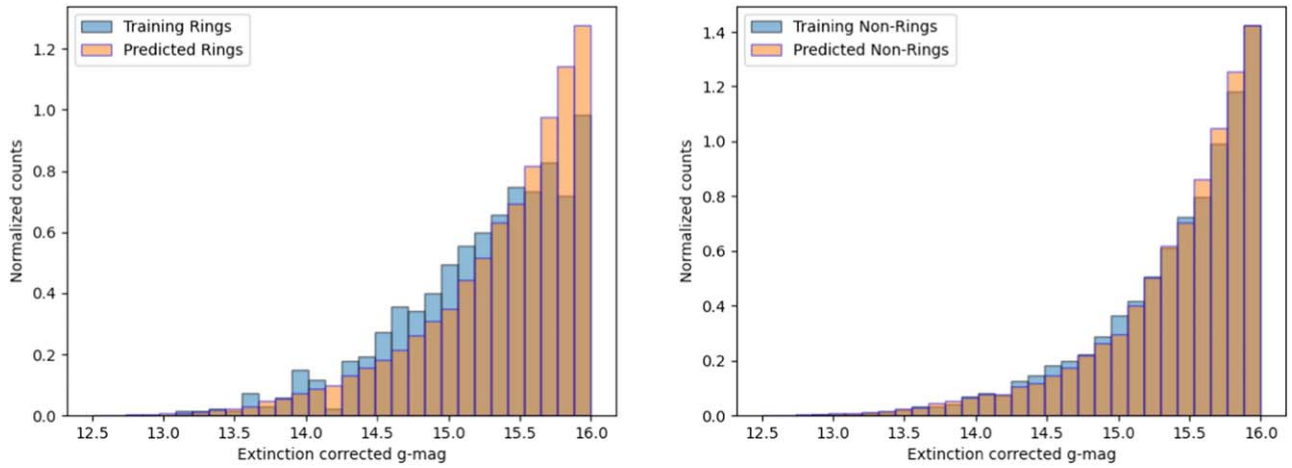


Figure 13. A comparison of the extinction-corrected g -magnitude distribution between the training and the prediction samples shown for both the ringed galaxies (left) and galaxies without rings (right).

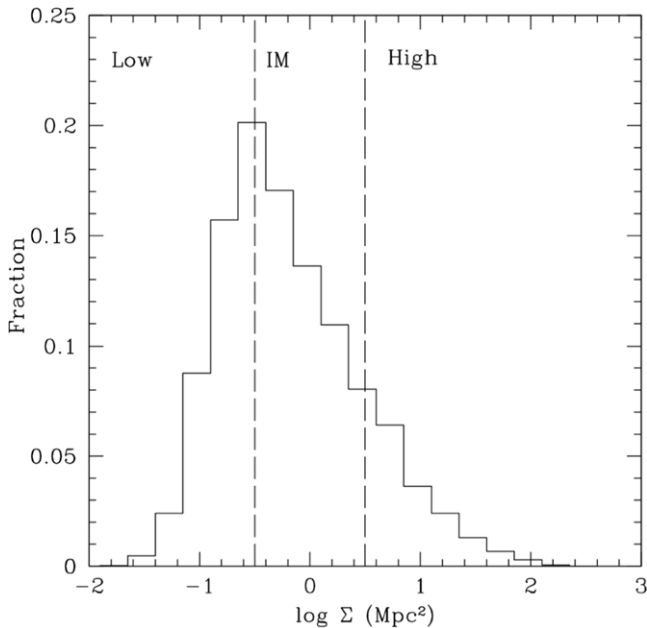


Figure 14. Histogram of local environmental density ($\log \Sigma$ (Mpc^{-2})) for our final sample of galaxies. The dashed lines separate the sample into low ($\log \Sigma$ (Mpc^{-2}) < -0.5), intermediate ($-0.5 < \log \Sigma$ (Mpc^{-2}) < 0.5), and high densities ($\log \Sigma$ (Mpc^{-2}) > 0.5).

to high densities. For this analysis, we use a sample of 4855 galaxies, with a classification threshold of 0.90, that are predicted to have rings.

We crossmatched this sample with I. K. Baldry et al. (2006) to find the local surface density as a measure of the environment. It is calculated using the relation $\Sigma_N = N/\pi d_N^2$, where d_N represents the distance to the N th nearest neighbor. These neighbors fall within the redshift range $\pm \Delta z_c = 1000 \text{ km s}^{-1}$ for galaxies with spectroscopic redshifts or within the 95% confidence interval for galaxies with only photometric redshifts. We use the best estimate, Σ , which is the average of Σ_N for the fourth- and fifth-nearest neighbors and categorize our sample into low-density ($\log \Sigma$ (Mpc^{-2}) < -0.5), intermediate-density ($-0.5 < \log \Sigma$ (Mpc^{-2}) < 0.5), and high-density ($\log \Sigma$ (Mpc^{-2}) > 0.5) groups.

Figure 14 presents the histogram of $\log \Sigma$ for the sample of 4471 ring galaxies with environment information. We find that

ring galaxies are primarily found in intermediate-density environments, i.e., in galaxy groups (49.3%) and in low-density environments, i.e., as isolated galaxies (35.1%). We see relatively few ring galaxies (15.6%) in high-density environments, in agreement with D. M. Elmegreen et al. (1992). Numerical simulations, such as ROMULUS (M. Tremmel et al. 2017), suggest that galaxies within group environments often undergo interactions or multiple mergers with gas-rich satellites, evolving into elliptical galaxies. Subsequently, these galaxies start to accrete gas and redevelop their disks, which then become unstable, triggering star formation in ring-like patterns. When galaxies display rejuvenated star-forming rings, two pathways for gas inflow are identified in numerical simulations: (a) diffuse cooling gas that feeds the galaxy, aiding in developing the disk/ring structure; and (b) the acquisition of ram-pressure-stripped gas streams from gas-rich satellites (S. L. Jung et al. 2022).

To study the star formation in our sample ring galaxies with environment information, we utilized the GALEX–SDSS–WISE Legacy Catalog (GSWLC-X2; S. Salim et al. 2016, 2018), and the crossmatch gave us 4022 ring galaxies with a stellar mass ($\log M_*$) range of 8.8–11.68. To investigate the influence of the environment on the star formation process in ring galaxies, we constructed a control sample of nonring galaxies with a classification threshold of 0.40 for being a ring galaxy, i.e., a threshold of 0.60 for being a nonring galaxy. This control sample has a similar r -band magnitude and redshift, and a crossmatch with the GSWLC-X2 catalog gave ~ 5000 nonring galaxies. The GSWLC-X2 catalog integrates ultraviolet data from GALEX, optical data from SDSS, and infrared data from the WISE all-sky survey to model the galaxy spectral energy distribution (SED). During the SED modeling, the catalog applies flexible dust attenuation laws and emission-line corrections to ensure accurate measurements of the SFR and stellar mass (M_*), as they might lead to the wrong identification of galaxies.

We calculated the specific SFR (sSFR) and, following S. Salim (2014), classified our sample galaxies into three subclasses as a function of environment, as shown in Figure 15. The star-forming region is identified with $\log \text{sSFR} \geq -10.8$; the green valley, which is a transition zone between the star-forming and quenched regions, is defined as $-11.8 < \log \text{sSFR} < -10.8$; and the quenched region is defined by $\log \text{sSFR} \leq -11.8$. In

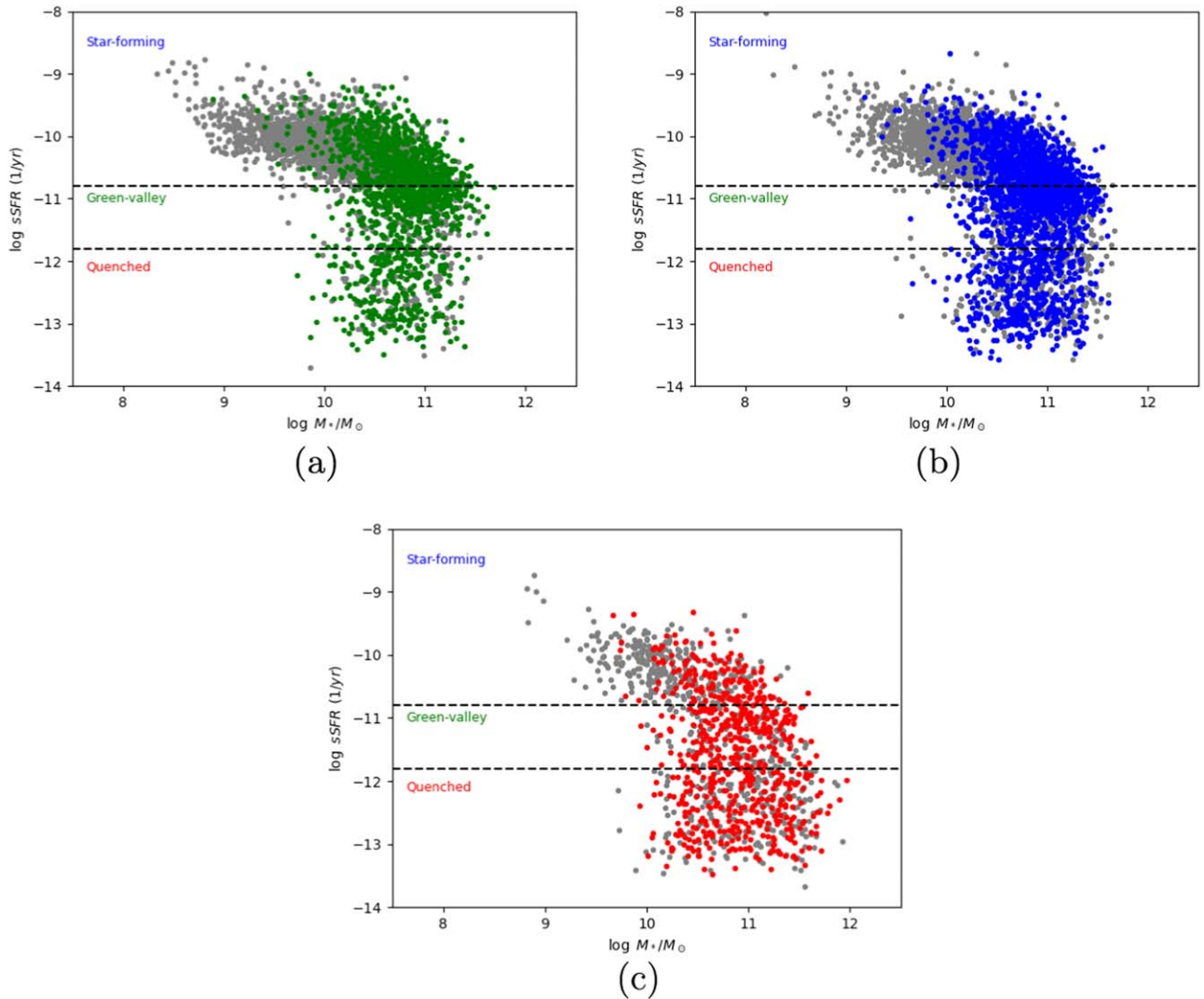


Figure 15. Dependence of sSFR on stellar mass (sSFR–mass plane) for our sample of ring galaxies and control sample of nonring galaxies (gray dots) as a function of (a) low-density ($\log \Sigma \text{ (Mpc}^{-2}) < -0.5$), (b) intermediate-density ($-0.5 < \log \Sigma \text{ (Mpc}^{-2}) < 0.5$), and (c) high-density ($\log \Sigma \text{ (Mpc}^{-2}) > 0.5$) environments.

high-density environments, most of the ring galaxies are in the quenched (42.3%) or green-valley region (33.6%); however, the control sample of nonring galaxies is almost equally distributed in the quenched (33.8%) and star-forming regions (47.8%), with fewer galaxies in the green valley (18.9%). In contrast, low- and intermediate-density environments are dominated by star-forming ring galaxies (52.4% and 43.2%, respectively), as are the nonring galaxies from the control sample (86.9% and 76.3%, respectively). However, the number of ring galaxies in green valleys increases for intermediate-density environments, though the number of nonring galaxies increases for high-density environments. It is apparent that most of the ring galaxies are located in the star-forming region for low- and intermediate-density environments, validating the scenario in which the rejuvenation of star formation happens by accreting gas and triggering star formation in ring-like features. Our results are somewhat different from those of J. Fernandez et al. (2024), who suggest that ringed galaxies exhibit lower star formation activity in intermediate-density environments, though it is much lower in high-density environments.

Many ring galaxies from our sample are also part of the MaNGA survey, and the study of spatially resolved stellar

populations of rings using such integral field units surveys will provide useful insights into the origin of star-forming rings, and this can be the subject of future work.

10. Conclusion

We have shown that a relatively simple deep neural network based on the AlexNet architecture can be used to detect galactic rings from color-composite galaxy images. The training required only a modest sample size of 1122 original ringed galaxy images and about 10,639 original images of galaxies without rings. To overcome the challenges of training with a small data set of labeled rings, we used a smaller network and also increased the total number of images available for training through realistic image transformations. We also used evaluation metrics suited for the class imbalance that was present in our data. We have prepared a catalog of candidate ring and nonring galaxies using the predictions of our trained network. The probability score provided in the catalog can be used as a confidence measure of the presence of rings, to obtain smaller catalogs with the required level of confidence for follow-up analysis. Additionally, we used the network to generate a catalog of galaxies with rings as well as bars, which may be of

special interest to galaxy morphologists. Finally, we explored the connection between the environments and star formation characteristics of the 4855 ringed galaxies in our catalog. We found that ring galaxies are star-forming and are primarily found in low- and intermediate-density environments.

Appendix CASJOBS Query

We provide the CASJOBS query used to create the catalog of ringed and nonringed galaxies. The query is run in the DR18 context.

```
SELECT
  p.objid,p.ra,p.dec,
  (p.u-p.extinction_u) as umag,
  (p.g-p.extinction_g) as gmag,
  (p.r-p.extinction_r) as rmag,
  (p.i-p.extinction_i) as imag,
  (p.z-p.extinction_z) as zmag,
  (p.deVRad_g) as g_dev, p.deVRad_r, p.deVAB_g,
  s.specobjid, s.class,
  s.z as redshift into mydb.MyTable
from PhotoObjAll AS p
JOIN SpecObj AS s ON s.bestobjid=p.objid
WHERE
  s.class='GALAXY'
  AND s.z BETWEEN 0.01 and 0.1
  AND (p.g-extinction_g) < 16
```

ORCID iDs

Sheelu Abraham  <https://orcid.org/0000-0001-9524-2739>
Sudhanshu Barway  <https://orcid.org/0000-0002-3927-5402>

References

- Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, *ApJS*, **182**, 543
- Abraham, L. 2024, Automated Detection of Galactic Rings from SDSS Images, v3, Zenodo, doi:10.5281/zenodo.13371749
- Abraham, S., Aniyani, A. K., Kembhavi, A. K., Philip, N. S., & Vaghmare, K. 2018, *MNRAS*, **477**, 894
- Adelman-McCarthy, J. K., Agüeros, M. A., Allam, S. S., et al. 2006, *ApJS*, **162**, 38
- Athanassoula, E., Romero-Gomez, M., & Masdemont, J. J. 2009, *MNRAS*, **394**, 67
- Baldry, I. K., Balogh, M. L., Bower, R. G., et al. 2006, *MNRAS*, **373**, 469
- Banerji, M., Lahav, O., Lintott, C. J., et al. 2010, *MNRAS*, **406**, 342
- Boureau, Y.-L., Ponce, J., & LeCun, Y. 2010, Proc. 27th Int. Conf. on Machine Learning, ICML'10 (Madison, WI: Omnipress), 111
- Buta, R., & Combes, F. 1996, *FCPh*, **17**, 95
- Buta, R. J. 2011, arXiv:1102.0550
- Buta, R. J. 2017, *MNRAS*, **471**, 4027
- Buta, R. J., Corwin, H. G., & Odewahn, S. C. 2007, The De Vaucouleurs Atlas of Galaxies (Cambridge: Cambridge Univ. Press)
- Buta, R. J., Sheth, K., Athanassoula, E., et al. 2015, *ApJS*, **217**, 32
- Chollet, F., et al. 2015, Keras, GitHub, <https://github.com/fchollet/keras>
- Dark Energy Survey Collaboration, Abbott, T., Abdalla, F. B., et al. 2016, *MNRAS*, **460**, 1270
- De Vaucouleurs, G. 1959, in *Astrophysik IV: Sternsysteme/Astrophysics IV: Stellar*, ed. S. Flügge & S. Flügge (Berlin: Springer), 275
- de Vaucouleurs, G., de Vaucouleurs, A., Corwin, G. H., Jr., et al. 1991, in *Third Reference Catalogue of Bright Galaxies*, ed. R. J. Buta, H. G. Corwin, & S. C. Odewahn (Springer: New York)
- Deng, J., Dong, W., Socher, R., et al. 2009, IEEE Conf. on Computer Vision and Pattern Recognition (Miami, FL: IEEE), 248
- Dieleman, S., Willett, K. W., & Dambre, J. 2015, *MNRAS*, **450**, 1441
- Dominguez Sanchez, H., Huertas-Company, M., Bernardi, M., Tuccillo, D., & Fischer, J. L. 2018, *MNRAS*, **476**, 3661
- Elmegreen, D. M., Elmegreen, B. G., Combes, F., & Bellin, A. D. 1992, *A&A*, **257**, 17
- Fernandez, J., Alonso, S., Mesa, V., & Duplancic, F. 2024, *A&A*, **683**, A32
- Fukugita, M., Nakamura, O., Okamura, S., et al. 2007, *AJ*, **134**, 579
- Hubble, E. P. 1926, *ApJ*, **64**, 321
- Iqbal, H. 2018, HarisIqbal88/PlotNeuralNet v1.0.0, Zenodo, doi:10.5281/zenodo.2526396
- Jung, S. L., Rennehan, D., Saeedzadeh, V., et al. 2022, *MNRAS*, **515**, 22
- Kaiser, N., Aussel, H., Burke, B. E., et al. 2002, *Proc. SPIE*, **4836**, 154
- Knapen, J. H. 1999, in *ASP Conf. Ser. 187, The Evolution of Galaxies on Cosmological Timescales*, ed. J. E. Beckman & T. J. Mahoney (San Francisco, CA: ASP), 72
- Kormendy, J. 1979, *ApJ*, **227**, 714
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. 2012, *Advances in Neural Information Processing Systems*, Vol. 25 (Red Hook, NY: Curran Associates Inc.)
- Kupriev, R., skshetry, Petrov, D., et al. 2022, DVC: Data Version Control—Git for Data & Models, v2.33.2, Zenodo, doi:10.5281/zenodo.7275755
- LeCun, Y., & Bengio, Y. 1995, in *The Handbook of Brain Theory and Neural Networks*, ed. M. A. Arbib (Cambridge, MA: MIT Press), 3361
- Lintott, C., Schawinski, K., Bamford, S., et al. 2011, *MNRAS*, **410**, 166
- Madore, B. F., Nelson, E., & Petriello, K. 2009, *yCat, J/ApJS/181/572*
- Nair, P. B., & Abraham, R. G. 2010, *ApJS*, **186**, 427
- Nair, V., & Hinton, G. E. 2010, Proc. 27th Int. Conf. on Machine Learning, ICML'10 (Madison, WI: Omnipress), 807
- Salim, S. 2014, *SerAJ*, **189**, 1
- Salim, S., Boquien, M., & Lee, J. C. 2018, *ApJ*, **859**, 11
- Salim, S., Lee, J. C., Janowiecki, S., et al. 2016, *ApJS*, **227**, 2
- Sandage, A., & Bedke, J. 1994, *The Carnegie Atlas of Galaxies*, Vol. 638 (Washington, DC: Carnegie Institution of Washington)
- Shamir, L. 2020, *MNRAS*, **491**, 3767
- Shimakawa, R., Tanaka, M., Ito, K., & Ando, M. 2024, *PASJ*, **76**, 191
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. 2014, *JMLR*, **15**, 1929, <http://jmlr.org/papers/v15/srivastava14a.htm>
- Stoughton, C., Lupton, R. H., Bernardi, M., et al. 2002, *AJ*, **123**, 485
- Strauss, M. A., Weinberg, D. H., Lupton, R. H., et al. 2002, *AJ*, **124**, 1810
- Tremmel, M., Karcher, M., Governato, F., et al. 2017, *MNRAS*, **470**, 1121
- Van den Bergh, S. 1998, *Galaxy Morphology and Classification* (Cambridge: Cambridge Univ. Press)
- Willett, K. W., Lintott, C. J., Bamford, S. P., et al. 2013, *MNRAS*, **435**, 2835
- York, D. G., Adelman, J., Anderson, J. E., Jr., et al. 2000, *AJ*, **120**, 1579