CS401 Machine Learning Project Report - Classifying Galaxies

By Cai Wood, Jonathan Fay and Alexandria Dempsey

Abstract

Classifications of galactic images can be a long and laborious task to do by hand. It is both repetitive and never ending. The sheer volume of data being produced by astronomers nightly is astronomical. This is where well-trained machine learning algorithms can come in real handy. The aim of this project is to educate both ourselves and our models about galactic classification by using the labels created by the volunteer astronomers who took part in the Galaxy Zoo project. Over a hundred thousand internet users between 2007 and 2009 were invited to classify images of galaxies based on their morphologies (the fancy astronomical word for shape). Over the course of the first project (of which there are currently four) almost 900,000 galaxies were classified. These labels have been useful for astronomers to attempt to train machine learning models to classify galaxies for them. This project will attempt to do the same using the labels from Galaxy Zoo 1 (Linott, 2011) to train a small variety of models (CNNs and MLPs) using different types of samples based on debiased voting statistics from the data produced by volunteer astronomers, as well as a curated set of images and labels from the follow-up project Galaxy Zoo 2 (Willet, 2013).

Introduction

“What is a galaxy?” Good question! A galaxy is a gravitationally bound system consisting of planets, stars and all manner of nebulous material such as dust and gas! The scale of these objects is astronomical, ranging in size from a few thousand to several hundred thousand light years in diameter (A. B. Bhattacharya, 2008). We (yes, you!) live in the sea of planets and solar systems that is the Milky Way which is home to one hundred thousand million stars. The morphology of the Milky Way is not unique, it is a barred spiral galaxy, meaning that it has an elongated central bulge (like a bar), with spiral arms at each end of the central bar. There are different morphologies possible for galaxies, the broad categories being spiral, elliptical and irregular. Irregular galaxies have no common traits describing them, they do not have the structure observed in spiral and elliptical galaxies. It is for this reason that the focus of this project is on elliptical and spiral galaxies, which together make up approximately 80% of galaxies.

Classification of galaxies can be achieved using either a system developed by Edwin Hubble in 1936 or by general classification using general properties (A. B. Bhattacharya, 2008).

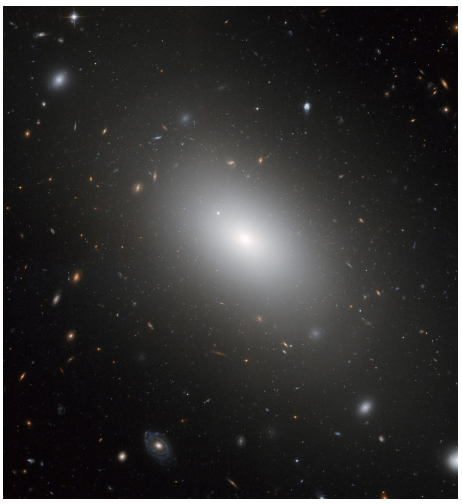

Figure 1: The elliptical galaxy NGC 1132 (Hubble Heritage Team, 2008)

Elliptical galaxies are elliptical masses of stars, usually surrounded by globular clusters. They are more three-dimensional in structure compared to spiral galaxies and the stars are in random orbit around the centre. Elliptical galaxies appear to be less common than spiral galaxies, however this is not the case. Light emitted from galaxies is produced by star formation, however there is little to no star formation in elliptical galaxies and so they appear dim and red.

Spiral galaxies are far more visually interesting. They are flat disk galaxies with a central bulge, off of which are spiral arms. The spiral arms form because the speed of rotation is dependent on the radial distance. The speed differential is thought to create density waves sweeping through the galactic disk which creates and gives stability to the spiral structure (Holliday, 1999). Spiral galaxies can form due to asymmetrical central mass distribution or the gravitational effect of a neighbouring galaxy.

Figure 2: The spiral galaxy NGC 1232 (ESO, 1998)

It is thought that spiral galaxies evolve into elliptical galaxies under the process of secular evolution - a process in which “disk material is slowly rearranged through the collective interaction of instabilities, such as bars, ovals, spirals, and triaxial dark matter haloes” (Kormendy & Kennicutt, 2004). Edge on spiral galaxies are harder to distinguish as they have the central bulge, making them similar to elliptical galaxies as their spiral arms are not easily noticeable.

In astronomy, there is far too much data to process by hand. As of this year, the rate at which data is being produced is 90 TB per night (Kremer et al., 2017). In the last 10 years, machine learning has become widely used in astronomy, being used in classification of stars based on their spectra, detection of exoplanets and supernovae and tracking important changes in light curves from stars or activity from our own sun. This project plans to use neural networks to classify galaxies into two categories, spiral and elliptical.

The Data

The data used in this project is from the Sloan Digital Sky Survey (SDSS) - Data Release 7 with labels provided by the first Galaxy Zoo project (GZ1), where volunteer astronomers were asked to classify galaxies into categories on the Galaxy Zoo website. There were six categories in total; “clockwise (CW) spiral”, “anti-clockwise (ACW) spiral”, “edge-on (other) spiral”, “elliptical”, “merger” and “star/don’t know”. The users would first have 15 galaxies to classify and if 11 out of the 15 were correct then they would proceed to the main part of the site. Typically, any galaxy that had >80% votes on one category would be classed as that category in the dataset, however, after inspecting the data it was seen that this was not always the case, this was because of the debiasing used on the data (Linott, 2011). The debiasing method used was based on the known physical characteristics of the galaxies, eg. redshift, luminosity, size and distance. It was done to prevent galaxies that were small or faint from being classified as ellipticals since the images would not clearly show the spiral’s arms. The data used in the project were images of galaxies based on spiral and elliptical galaxies as well as those that were uncertain, which were downloaded from DESI Legacy Imaging Surveys using their “jpeg-cutout” feature.

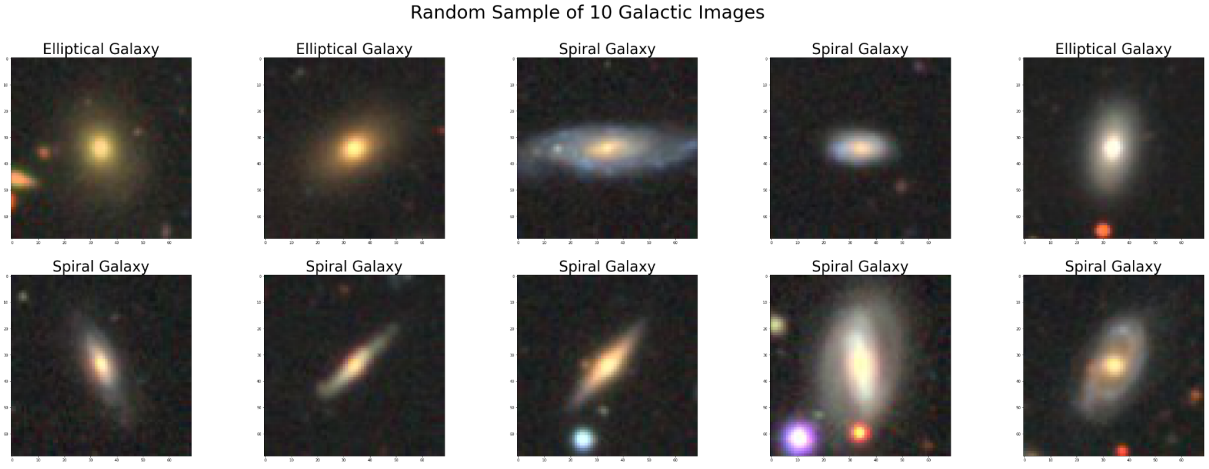

Figure 3: Random sample of 10 download images showing elliptical galaxies and a variety of spiral galaxies.

Table 1: First 16 rows of GZ1 Table 2 from which the labels were drawn. Each row pertains to a single galactic object in the SDSS sample. 'SPIRAL', 'ELLIPTICAL' and 'UNCERTAIN' contain the label information of each image which were calculated based on the debiased statistics with a debiasing method discussed here (source needed gz1). 'RA' and 'DEC' contain positional information for the galactic objects which were used to download the images

	OBJID	RA	DEC	NVOTE	P_EL	P_CW	P_ACW	P_EDGE	P_DK	P_MG	P_CS	P_EL_DEBIASED	P_CS_DEBIASED	SPIRAL	ELLIPTICAL	UNCERTAIN
0	587727178986356823	00:00:00.41	-10:22:25.7	59	0.610	0.034	0.000	0.153	0.153	0.051	0.186	0.610	0.186	0	0	1
1	587727227300741210	00:00:00.74	-09:13:20.2	18	0.611	0.000	0.167	0.222	0.000	0.000	0.389	0.203	0.797	1	0	0
2	587727225153257596	00:00:01.03	-10:56:48.0	68	0.735	0.029	0.000	0.147	0.074	0.015	0.176	0.432	0.428	0	0	1
3	587730774962536596	00:00:01.38	+15:30:35.3	52	0.885	0.019	0.000	0.058	0.019	0.019	0.077	0.885	0.077	0	1	0
4	587731186203885750	00:00:01.55	-00:05:33.3	59	0.712	0.000	0.000	0.220	0.068	0.000	0.220	0.640	0.290	0	0	1
5	587727180060098638	00:00:01.57	-09:29:40.3	28	0.857	0.000	0.036	0.000	0.107	0.000	0.036	0.830	0.060	0	0	1
6	587731187277627676	00:00:01.86	+00:43:09.3	38	0.500	0.000	0.053	0.289	0.105	0.053	0.342	0.351	0.473	0	0	1
7	587727223024189605	00:00:02.00	+15:41:49.8	26	0.423	0.000	0.000	0.577	0.000	0.000	0.577	0.143	0.857	1	0	0
8	587730775499407375	00:00:02.10	+15:52:54.2	62	0.355	0.016	0.210	0.323	0.000	0.097	0.548	0.355	0.548	0	0	1
9	587727221950382424	00:00:02.41	+14:49:19.0	31	0.484	0.129	0.065	0.258	0.065	0.000	0.452	0.109	0.789	1	0	0
10	587730774425665704	00:00:02.58	+15:02:28.3	24	0.583	0.042	0.125	0.167	0.083	0.000	0.333	0.147	0.701	0	0	1
11	587730773888794751	00:00:02.82	+14:42:55.9	26	0.654	0.077	0.000	0.077	0.192	0.000	0.154	0.621	0.185	0	0	1
12	588015507658768464	00:00:03.24	-01:06:46.8	57	0.474	0.088	0.000	0.263	0.175	0.000	0.351	0.324	0.480	0	0	1
13	587727178449485858	00:00:03.33	-10:43:16.0	24	0.125	0.000	0.000	0.875	0.000	0.000	0.875	0.024	0.976	1	0	0
14	587730773351858407	00:00:03.46	+14:11:53.6	64	0.625	0.016	0.016	0.250	0.078	0.016	0.281	0.245	0.597	0	0	1
15	587731187277693069	00:00:04.12	+00:45:07.9	30	0.933	0.000	0.033	0.000	0.033	0.000	0.033	0.913	0.054	0	1	0

The second galaxy zoo project (GZ2) expanded upon GZ1 by extending the classifications from the brightest and biggest galaxies available in SDSS data release 7, thus, the images present were more detailed. The set of labels were downloaded from the Galaxy Zoo website. The image set from GZ2 used in the project was provided by the astroNN python package. This curated set of images was created for educational purposes as a dummy set for training models, meaning that the data may not be reliable for training machines with the intent of applying to real world classification tasks.

Three types of samples of the labels were taken from the GZ1 data: a random sample, a curated sample with 60% of the sample being the galaxies with a debiased statistic of over 95% for one category, 30% of the sample was >90% debiased statistic for one galaxy type and the remaining 10% had over 80% for the debiased statistic for one category, and an uncertain sample which consisted of the 60-80% debiased statistic. Each sample had ~10,000 elliptical galaxies and ~10,000 spiral galaxies included.

The right ascension (RA) and declination (DEC) of the sampled labels from table 2 were converted to degrees and were passed into the DESI Legacy Imaging Survey and the images were downloaded using the jpeg-cutout feature. A scaling of 0.55 was chosen to centre the images and ensure the galaxies chosen were of a suitable size. To match existing images in the GZ2 curated sample, 69x69 pixels were chosen for the rgb images, meaning our inputs would be of the same size (69x69x3) and it would allow our models to be evaluated on the different data sets more easily.

Methods

This project explored two types of supervised learning models: Multilayer Perceptrons (MLP) and Convolutional Neural Networks (CNN). All models were built using the keras interface for Google's Tensorflow python package. The original aim of the project was to explore how the performance of these models changed with depth (number of layers), however due to unforeseen problems with data acquisition, a single model of each type was implemented: CNNs with two convolution layers, and MLPs with one hidden layer. Each model performed a binary classification task on spiral and elliptical galaxies. They were trained on balanced samples from the SDSS7 and labels using the Galaxy Zoo 1 dataset, and one pair of models was trained on the curated sample of GZ2 labels of images also from SDSS7.

The Multilayer Perceptron's leading purposes are for pattern classification, prediction, recognition and approximation. The MLP is a supplement of feed-forward neural network. It consists of three types of layers; the input layer, the output layer and the hidden layer, of which there can be multiple. MLPs employ a procedure known as backpropagation for training. Backpropagation is an algorithm that, after each forward pass through the network, a backward pass is performed while altering the weights and biases in the model. The input layer accepts the input that is to be processed. The output layer predicts and classifies the signal.

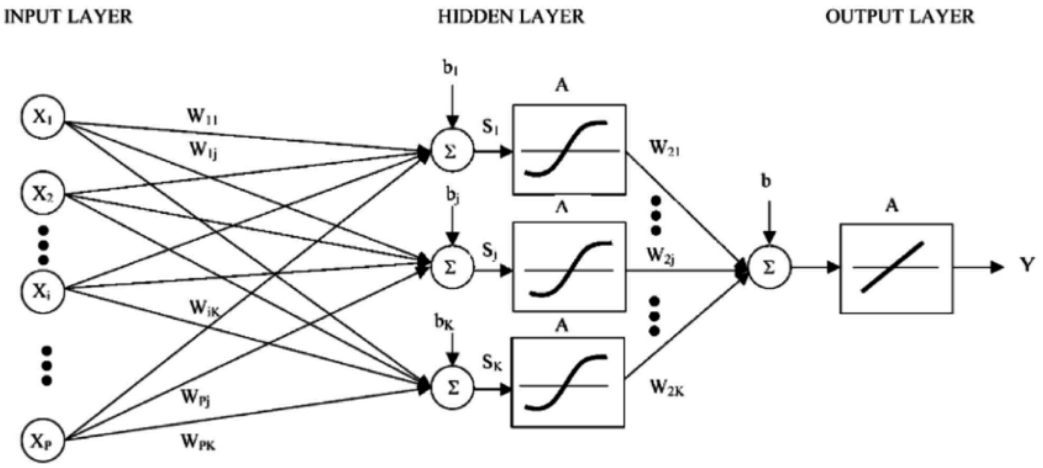

Figure 4: Outline of MLP procedure with one hidden layer, using a sigmoid activation function (Vuckovic, 2002)

The MLP's structure used consisted of an input layer which receives the flattened image, a single hidden layer with a thousand nodes, and an output layer of two nodes to give a binary classification. The latter two layers used a rectified linear unit and sigmoid activation function respectively. Each model was trained over twenty iterations over the data.

Convolutional Neural Networks are well suited to learning from images. Each filter passes over the image inputted into the model, generating a numeric value to represent each region and repeats this as the filter passes over the image. The output convolution layer is then compared to neighbouring values and condensed by a process known as pooling. These steps are repeated for each filter and compared at the end. Each filter aims to detect some

feature of the image for further processing. For example, consider a machine designed to detect the presence of koalas in an image. The first convolution step takes in the raw image and passes filters of the image looking for features such as eyes, nose, ears, feet, and hands. This information is then pooled together. Then a second set of filters is passed over the detected features in the second convolution layer. A filter may, for example, detect that eyes and noses in close proximity in the image which would imply the existence of a koala's head, while another filter notes that feet and hands are present in some proximity to each other and infer the existence of the koala's body. This information would then be passed into a smaller neural network to determine that there is a koala in the image. However, this example assumes one knows what the filters are looking for. In reality, all one can do when programming such a model is tell the model how many filters one wants to have in a given layer. The machine then takes the data and learns for itself what those filters should be in order to recognise the koala! This to me is the most intriguing property of CNNs!

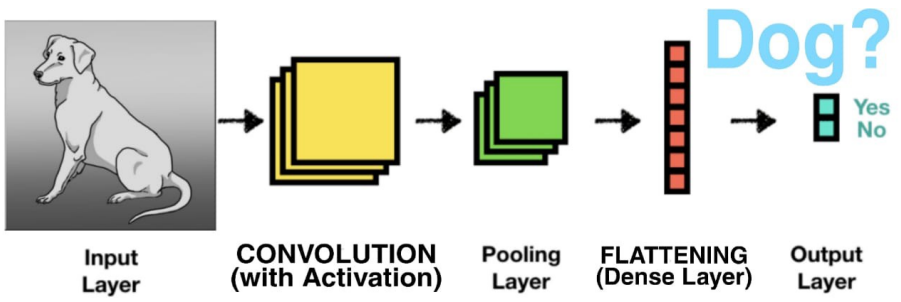

Figure 5: Overview of the CNN process with one convolution layer for a machine detecting dogs in images instead of koalas (Patel, 2020)

The CNN motif used consisted of two convolution layers, each with 3x3 filters with a stride of one. The first of these layers used ten filters, while the second used 20 filters. After each convolution step, the outputs were pooled in 2x2 cells. The resultant image was then flattened and passed into a dense hidden layer with 64 nodes with a rectified linear unit activation function. The hidden layer was then outputted into the output layer of two nodes with a sigmoid activation function which produces the prediction of the image's class. Each model was trained over twenty epochs for each data type.

As mentioned above three sets of 20,000 images were created from the labels sourced in the first Galaxy Zoo project. The random set of images was a completely balanced random sample of galactic images. The curated set was balanced between types of galaxies but favoured images that received a higher proportion of votes for the assigned label after debiasing, indicating a more confident classification. The uncertain data, as the name suggests, was a balanced sample taken from images which received between 60% and 80% of their votes such that the label would be the most likely of the two but not confident enough for these images to be properly classified. This dataset was created with the purpose of seeing how well our trained models agreed with the uncertain classifications of the volunteer astronomers in the Galaxy Zoo project as a metric of our learning. Each model trained on one of these three datasets was trained on a sample of ~16,000 images and evaluated on the remaining ~4,000 images.

The pair of machines trained on the GZ2 data were trained using the same model set ups (number of layers, filters, etc). These models were then evaluated on the three above data sets and vice versa in order to attempt to validate how each model fared with a different source of (what was thought to be) comparable data. Both sets of images were downloaded from SDSS7, however this did not yield the desired results.

Discussion of Results

Four pairs of models were trained on the random, curated, uncertain and GZ2 datasets mentioned above. The training accuracy vs training cycle (or epoch) was tracked for each model and is shown in Figure 6.

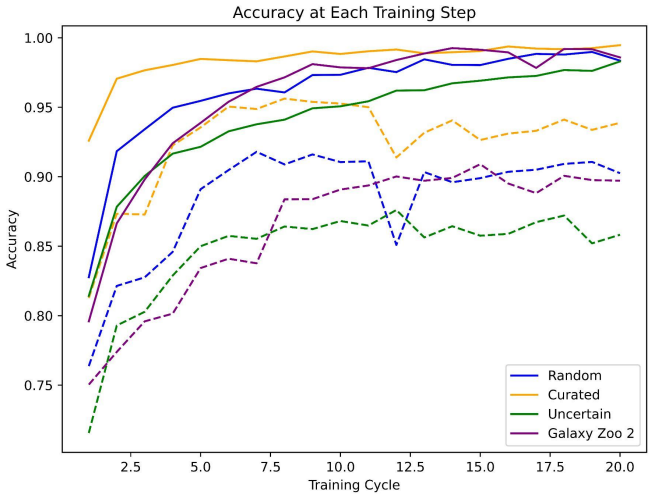

Figure 6: Graph of training accuracy at each training step (20 in total) for each model. Solid lines represent CNNs while the dashed lines denote MLPs. The lines are colour coded to the data they were trained on (Blue: Random, Orange: Curated, Green: Uncertain, Purple: GZ2)

The metric that was used for validation was accuracy. The accuracy is defined as:

$$Accuracy = \frac{tp + tn}{tp + fp + tn + fn}$$

where tp and tn are the true positive and negative values, and fp and fn are the false positive and false negative values respectively.

Figure 6 above depicts a graph of the accuracy for models vs each training step. It is clear from the graph that CNNs outperforms MLPs in both learning rate and accuracy. The CNN models have a similar learning rate for each set of data, with each CNN converging to approximately 98% accuracy exhibiting a logarithmic behaviour. Note that the MLP's are subject to random increases and decreases in their accuracy, whereas the CNN's accuracy increases steadily with each epoch. This randomness is exemplified in the accuracy after 20

epochs, with an 8% difference between the MLP's compared to a 2% difference in CNN accuracy. It is also evident that CNNs perform better from initial training, with average initial training accuracy being approximately 82%, compared to the average initial accuracy of 75% exhibited by the majority of the MLP's. The accuracy metric is valid to use in this scenario as the models were trained on data which is balanced with respect to galactic objects.

Table 2 shows the achieved training accuracy for each model, and accuracy of each model's predictions on test sets of each data type.

Table 2: Model name denotes "model_type training_data_type". Training accuracy is the final accuracy attained in the final training cycle. The right-hand side of the table shows the accuracy of predictions on different test samples (~20% of available data reserved for this purpose)

Model		Evaluated On			
Name	Training Accuracy	Random	Curated	Uncertain	GZ2
CNN Random	0.988	0.948	0.981	0.865	0.585
MLP1 Random	0.903	0.916	0.941	0.859	0.712
CNN Curated	0.995	0.945	0.984	0.884	0.648
MLP1 Curated	0.939	0.890	0.944	0.836	0.751
CNN Uncertain	0.983	0.929	0.955	0.896	0.436
MLP1 Uncertain	0.852	0.744	0.733	0.689	0.759
CNN GZ2	0.986	0.666	0.696	0.594	0.9573
MLP1 GZ2	0.897	0.517	0.512	0.502	0.827

As seen in Table 2 above, the training accuracy is relatively high overall with an average of 0.943, the lowest accuracy being 0.852 for the "MLP1 Uncertain" model and reaching as high as 0.995 with the "CNN Curated" model, therefore, a range of only 0.143. The CNN models are always higher than the MLP1 models in training.

The prediction accuracy on the different test samples is generally higher when applying the CNN models to the data also, however, this is not the case when the random, curated and uncertain models are applied to the GZ2 test data (seen in the right-most column of the table) with as big of a difference as 0.323 between the "CNN Uncertain" model and the "MLP1 Uncertain" model in favour of the MLP1 model, more than double the range of the training accuracy. The pattern goes back to CNN being more accurate after the GZ2 models are both applied to the GZ2 test set, with "CNN GZ2" going up to 0.9573 and "MLP1 GZ2" as 0.827, which is still higher than the previous MLP1 models that were applied to the GZ2 test data.

The curated test set performed better than the random set in terms of accuracy which is noticed by looking at Table 2. There are six instances where the curated set had higher accuracy though there are only slight differences, for example, "CNN Random" when applied to both sets resulted in 0.948 (random) and 0.981 (curated). Conversely, there are only two times that occur where the random set has a higher accuracy than the curated set, such as

the “MLP1 Uncertain” model prompting 0.744 (random) and 0.733 (curated). This is likely because the images used in the curated set were chosen according to their superior debiased statistics, with more than half of the images being over 95%. Thus, the curated set was more likely to have clear images that would be easier to classify for the models.

It is clear from Table 2 that the uncertain test set had higher accuracy than the GZ2 test sample. There is a much more noticeable distinction between the values in this case, although there are five occurrences rather than the six seen in the comparison between the curated and random set. Once again, the “MLP1 Uncertain” model causes one of the discrepancies, allowing the uncertain set to have 0.689 and the GZ2 set to have 0.759 accuracy. The other models that cause the uncertain set to be lower than the GZ2 set being “CNN GZ2” and “MLP1 GZ2”.

Overall, CNN worked better than MLP models except for when they are applied to the GZ2 test data. They produce better results because CNNs work well on images due to the fact that CNNs can understand spatial relation between pixels of images unlike MLPs. Thus, they are better able to identify characteristics of a complicated image.

Receiver operating characteristic (ROC) curves are used to compare accuracies of different models. ROC curves plot the true positive rate (TPR) versus the false positive rate (FPR) at various thresholds for the model. Since our models are probabilistic (meaning the output given is in a sense, the probability of the input being the suggested label) the TPR and FPR values are calculated for different thresholds on the outputs of the model for a set of inputs.

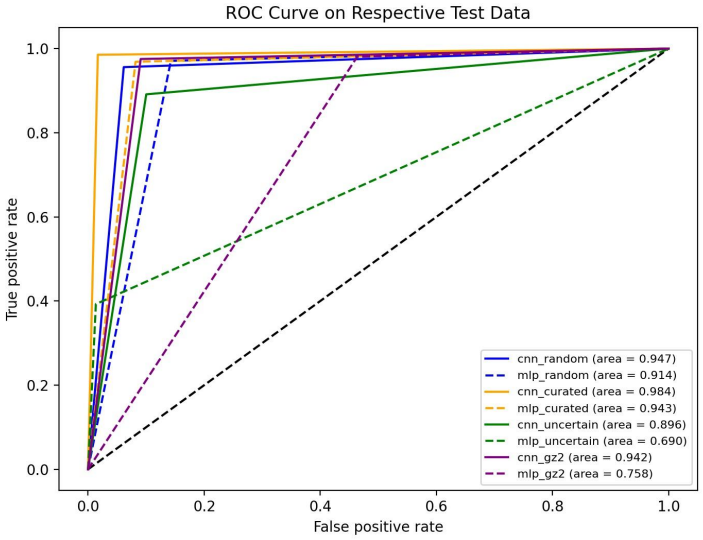

Figure 7: Graph of ROC curves on the respective test data. Solid lines represent CNNs, dashed lines depict MLPs. The line colours constitute the data they are applied to (Blue: Random, Orange: Curated, Green, Uncertain and Purple: GZ2). The legend denotes the area under the curve (AUC) for each model.

The dashed line for TPR = FPR represents the expected output of a random classifier (i.e. completely random guessing). The closer to this line a curve is the worse the classifier is

said to be. Figure 7 shows the ROC curves for the models when evaluated on a unique test sample of the same dataset they were trained on. Thus, one can see that in general (combined with observations from Table 2) each model is performing well at the task they were trained for. Note that each CNN outperforms their MLP counterpart. In particular the CNN trained on the curated data set performs best, which is reflected in its training accuracy being the highest in Table 2. The MLPs for the uncertain data and the GZ2 data don't appear to be performing too well, the uncertain MLP being the closest to the TPR=FPR line. The AUC can be used as a measure of accuracy for a model. It is important to note that there can be situations where a model with a higher AUC performs worse than a model with a lower AUC, but the AUC is regarded as a good general measure of predictive accuracy. As such, this plot suggests that the CNN trained on curated data will perform best at the trained task.

The uncertain data set was created to attempt to quantify how well our models would agree with the volunteer astronomers from the GZ1 project with respect to images that didn't meet the threshold for labelling. The quality of how well our models fared using data from GZ2 was evaluated. Figure 8 below depicts two ROC curves, the first being for when each model was evaluated on the Uncertain Test data (left) and the second when evaluating on the GZ2 test data (right)

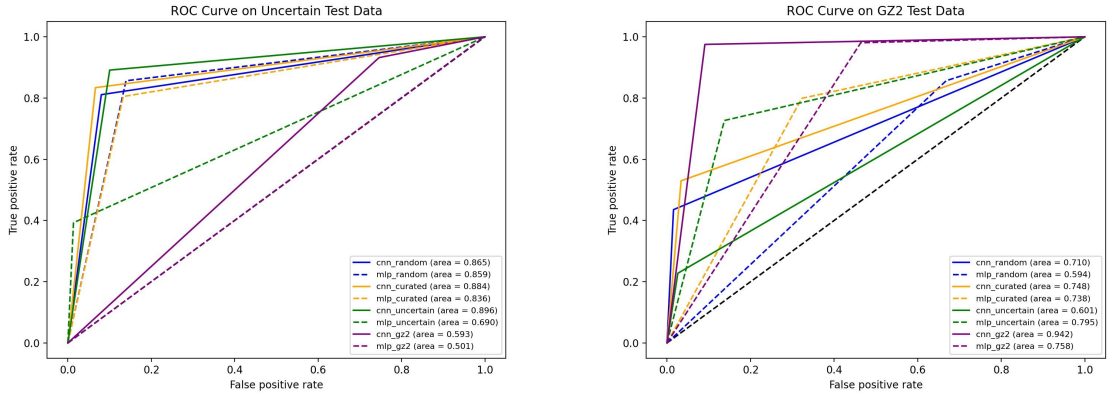

Figure 8: (left) ROC curve for each CNN (solid lines) and MLP (dashed line) trained on different datasets evaluated on the Uncertain Test dataset. (right) ROC curve for each CNN and MLP evaluated on the GZ2 test dataset. The legends contain the AUC value for each model in each situation.

For the Uncertain test data, in general the CNNs outperform their corresponding MLPs again, which is also reflected in their AUC values. The opposite is true for the GZ2 test data, with exception of the GZ2 trained models, the MLPs outperform the CNNs for the models trained on the uncertain data and the random data, while the CNN and MLP for the curated data have comparative performance based on their AUC values. For the Uncertain test data, the GZ1 data trained models vastly outperform the GZ2 models. This highlights that there must be an unknown difference between the two image sets. Both originated from the SDSS release 7. It is possible that there was some sort of pre-processing done to the images in the GZ2 set, which may explain the issues in performance for this task, but this is purely speculative. This same discrepancy is mirrored in the GZ2 ROC curve. The GZ2 trained

models outperform the GZ1 trained models. However, the GZ1 trained models still perform better at this task than the GZ2 models for the uncertain classification task. Another observation one can draw from these plots is that the Curated trained models performed best on the uncertain data task (second to the uncertain trained model of course). Surprisingly, the uncertain trained models outperformed the non GZ2 models in the GZ2 classification task. It was expected that the curated models would perform better at these tasks as they had more access to “higher quality data” to learn from.

Conclusion

The initial aim of this project was to explore the use of CNNs and MLPs in the binary classification of galaxies based on the primary morphological types. It was shown that for image classification of galaxies, CNNs are the preferred model (over MLPs) for a reason. In each scenario a model was trained for, the CNNs outperformed the MLPs. CNNs received a higher training accuracy than their MLP counterparts in the same number of training steps as shown in Figure 6 and is reflected in Table 2. This is also backed up by the ROC curves in Figure 7. In the case of evaluating on uncertain data, the CNNs performed better once again. However, the Curated and Random trained models performed worse than the uncertain trained CNN model which surprised us. Initially, we expected the models trained on “higher quality” to outperform the uncertain trained models on the Uncertain evaluation task. Neither of the curated or random trained models beat the uncertain CNN. Combining this with the results from the GZ2 evaluation tasks, it seems that a model trained on the target data, will perform better than a model trained on similar but different data. One thing that one could try differently is to take a more balanced data set with respect to quality (the percentage of votes for a given category) of data. Instead of taking a curated set of data as we did focusing on the “best images”, instead take a random sample of the data containing some of the uncertain data. It would be interesting to see how such a model would perform at these tasks. Another avenue of exploration could be scaling the data. One possible reason the GZ2 data behaved so unexpectedly is that the machine could have been relying on the colour of images in GZ1 to make its predictions, and since we do not know how the GZ2 data was processed. The colours may have some unexpected differences despite both image sets coming from the SDSS. The hope is that scaling the data would reduce the relative difference between colours in the images. A third avenue of exploration is to attempt to change this binary classification task into a tertiary classification task by attempting to classify according to all three labels from the GZ1 labelset. We have but scratched the tip of the iceberg when it comes to machine learning applications in astronomy. There is certain merit to training models to classify galactic objects for us, which will not only relieve astronomers of the boring and laborious task of classifying these objects by hand, but also be able to keep up with sheer volume of data being collected nightly.

References

Bhattacharya, A., Joardar, S. and Bhattacharya, R., 2008. Astronomy and astrophysics. Sudbury, Mass.: Jones and Bartlett Publishers.

Holliday, K. (1999). Introductory Astronomy. John Wiley & Sons. ISBN 0471983314 1st edition

Kormendy, John; Kennicutt, Robert C. (2004). Secular Evolution and the Formation of Pseudobulges in Disk Galaxies. Annual Review of Astronomy and Astrophysics, 42(1), 603–683. doi:10.1146/annurev.astro.42.053102.134024

Kremer, J., Stensbo-Smidt, K., Gieseke, F., Pedersen, K. and Igel, C., 2017. Big Universe, Big Data: Machine Learning and Image Analysis for Astronomy. IEEE Intelligent Systems, 32(2), pp.16-22.

C. Lintott et al., "Galaxy Zoo 1: data release of morphological classifications for nearly 900 000 galaxies," in Monthly Notices of the Royal Astronomical Society, vol. 410, no. 1, pp. 166-178, Jan. 2011, doi: 10.1111/j.1365-2966.2010.17432.x.

Schneider, P., 2016. Extragalactic astronomy and cosmology. Springer-Verlag Berlin An.

K.W. Willett et al., "Galaxy Zoo 2: detailed morphological classifications for 304,122 galaxies from the Sloan Digital Sky Survey, Monthly Notices of the Royal Astronomical Society", Volume 435, Issue 4, 11 November 2013, Pages 2835–2860, https://doi.org/10.1093/mnras/stt1458

Images
Spiral Galaxy: https://www.eso.org/public/images/eso9845d/
Elliptical Galaxy: Hubble Heritage Team (STScI/AURA)-ESA/Hubble Collaboration available from https://esahubble.org/images/heic0804a/

CNN:
Deep Patel, 2020
https://towardsai.net/p/deep-learning/convolutional-neural-network-cnn-in-the-easiest-way

MLP diagram:
Vuckovic, Aleksandra & Popović, Dejan & Radivojevic, Vlada. (2002). Artificial neural network for detecting drowsiness from EEG recordings. 155 - 158. 10.1109/NEUREL.2002.1057990.