

# Galaxy Morphological Classification

**Robert Mifsud**

Supervisor: Mr Joseph Bonello

October, 2022

*Submitted in partial fulfilment of the requirements for the degree of B.Sc.  
Information Technology (Software Development).*



**L-Università ta' Malta**  
Faculty of Information &  
Communication Technology



**L-Università  
ta' Malta**

**FACULTY/INSTITUTE/CENTRE/SCHOOL** \_\_\_\_\_

**DECLARATIONS BY UNDERGRADUATE STUDENTS**

Student's I.D. /Code \_\_\_\_\_

Student's Name & Surname \_\_\_\_\_

Course \_\_\_\_\_

Title of Long Essay/Dissertation

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

Word Count \_\_\_\_\_

**(a) Authenticity of Long Essay/Dissertation**

I hereby declare that I am the legitimate author of this Long Essay/Dissertation and that it is my original work.

No portion of this work has been submitted in support of an application for another degree or qualification of this or any other university or institution of higher education.

I hold the University of Malta harmless against any third party claims with regard to copyright violation, breach of confidentiality, defamation and any other third party right infringement.

**(b) Research Code of Practice and Ethics Review Procedures**

I declare that I have abided by the University's Research Ethics Review Procedures.

\_\_\_\_\_  
Signature of Student

\_\_\_\_\_  
Name of Student (in Caps)

\_\_\_\_\_  
Date

# Contents

List of Figures	iv
List of Tables	v
<b>1 Background</b>	<b>1</b>
1.1 Astronomy . . . . .	1
1.1.1 Galaxy Morphologies . . . . .	1
1.2 Astronomical Data . . . . .	1
1.2.1 Photometric and Spectroscopic Data . . . . .	1
1.2.2 Sloan Digital Sky Survey . . . . .	2
1.2.3 Visually Classified Data - Galaxy Zoo . . . . .	2
1.3 Data Science . . . . .	3
1.3.1 Classification . . . . .	3
1.3.2 Convolutional Neural Network . . . . .	3
1.3.3 Key Performance Indicators . . . . .	3
1.4 Pre-processing techniques . . . . .	3
1.5 Data Augmentation . . . . .	3
1.5.1 Generative Adversarial Networks . . . . .	4
<b>2 Literature Review</b>	<b>5</b>
2.1 Classification in Astronomy . . . . .	5
2.2 Establishing classes . . . . .	5
2.3 Data Augmentation Techniques . . . . .	5
2.4 Pre-processing to improve performance . . . . .	5
<b>References</b>	<b>6</b>
<b>Appendix A Galaxy Zoo 2 Images</b>	<b>8</b>

# List of Figures

1.1	Hubble tuning fork diagram . . . . .	2
1.2	Classification scatter plot diagram . . . . .	4

# List of Tables

A.1 Galaxy Zoo 2 sample images . . . . .	8
--	---

# 1 Background

## 1.1 Astronomy

### 1.1.1 Galaxy Morphologies

Galaxy morphology refers to the physical shape of a galaxy. Two main types of morphologies are elliptical (early-type) and spiral (late-type) galaxies. In 1926, Edwin Hubble formulated developed an early classification scheme known as the Hubble tuning fork diagram (Figure 1.1). The diagram developed notation to refer to different traits found in morphologies.

## 1.2 Astronomical Data

Modern professional astronomy has evolved to rely on solely digital data. Telescopes use CCD (Charge-coupled device) cameras to produce digital images and spectrographs to measure spectra [1]. Spectrographs are prism-like devices which measure magnitude at different wavelengths [2]. Apart from looking extensively at one celestial object, nowadays telescopes conduct full-blown sky surveys which document whole sections of the observable universe. These surveys collect data into catalogues which combine other data points such as positional data and unique object identifiers [1].

### 1.2.1 Photometric and Spectroscopic Data

*Photometric* and *spectroscopic* are terms commonly used to distinguish between different types of astronomical data. Photometric data is collected from the images themselves and is a measure for the brightness in a portion of an image, more commonly referred to as magnitude. Data from the Sloan Digital Sky Survey (SDSS) includes multiple measures of photometry for a single image as an image is taken through five different coloured filters simultaneously [3]. Spectroscopic data measures light at each wavelength. This allows

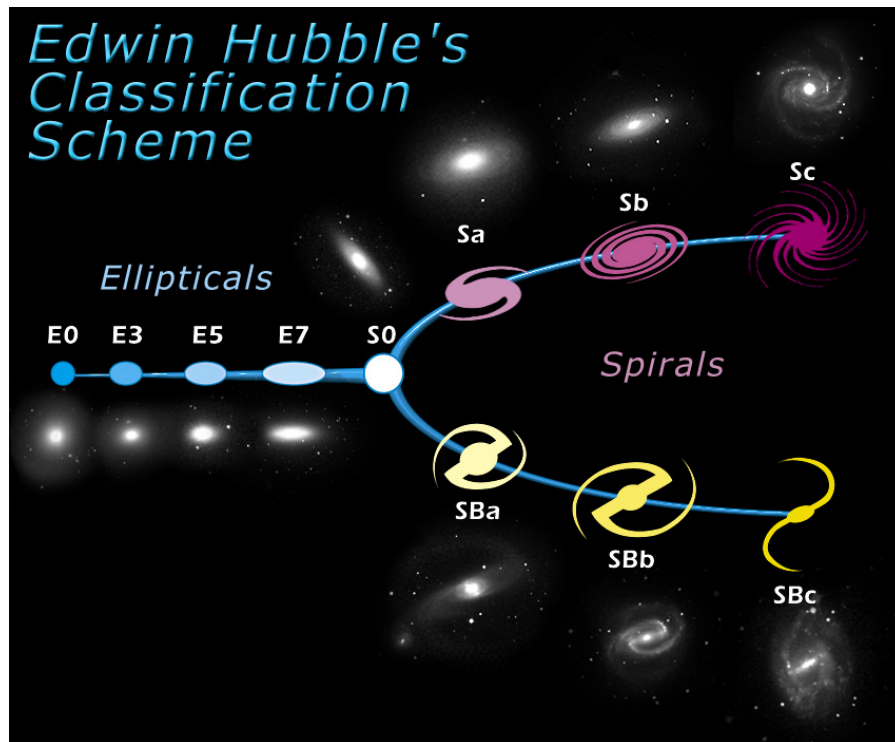


Figure 1.1: Edwin Hubble's tuning fork diagram via ESA/Hubble

the investigation of the chemical composition and physical properties of distant celestial objects [4].

### 1.2.2 Sloan Digital Sky Survey

The SDSS is a major survey founded with the intention of creating a detailed map of a large sector of the visible universe [5]. The corresponding telescope found at Apache Point Observatory saw first light in 1998 and is considered to have achieved its goals since. The SDSS comprises various sub surveys, each with different functions. One example is the Baryon Oscillation Spectroscopic Survey (BOSS), focused on creating the largest three-dimensional map of galaxies to date.

### 1.2.3 Visually Classified Data - Galaxy Zoo

Visually classified datasets comprise data which has been manually analysed by experienced astronomers. This process may involve assigning objects to their respective class, for example Stars, Galaxies and Quasars. Another example would be to assign galaxies to their corresponding morphology class.

This leads us into the Galaxy Zoo project. Galaxy Zoo is a large-scale crowdsourced project which uses an online tool to allow amateur astronomers to visually classify galaxies into morphological classes. The two main phases of the project both involved classifying images from SDSS data releases, with the second phase using a subset of the data from phase one, but the class count was significantly increased [6][7]. Refer to appendix A.1 to view sample Galaxy Zoo 2 images.

## 1.3 Data Science

Modern Data Science refers to the extraction of information from data [8]. In the era of *Big Data*, Data Science is crucial for gaining new and important knowledge. Good data not only helps us explain the past but also provides predictive power, which is where Data Science comes into play. One instance of this is Classification.

### 1.3.1 Classification

Classification problems involve taking labelled data and using it to train classifiers (Figure 1.2). The classifiers are then used to classify data which is not labelled during the testing phase. Classification makes for data which is structured, sortable, and easily analysed and understood [9]. In a context such as galaxy classification, this process can be applied to help understand the composition of the universe.

### 1.3.2 Convolutional Neural Network

### 1.3.3 Key Performance Indicators

## 1.4 Pre-processing techniques

## 1.5 Data Augmentation

Data Augmentation is crucial when working with imbalanced datasets to prevent overfitting. Overfitting occurs when training a classifier with data that is of low quality and quantity. This leads to classifiers which are biased to the training dataset and won't perform well with other samples. Data augmentation prevents overfitting by creating copies of the existing sample to enhance the dataset [10].



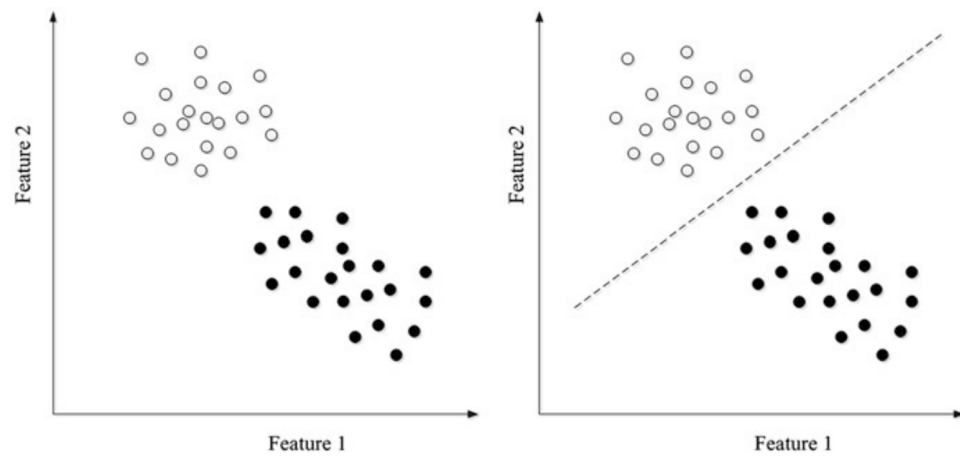


Figure 1.2: Classification is defined [9]

### 1.5.1 Generative Adversarial Networks

## **2 Literature Review**

### **2.1 Classification in Astronomy**

Applying the concepts of Machine Learning in the field of Astronomy is by no means a new endeavour. Ever since the advent of telescopes and surveys generating valuable data, research which exploits this data quickly followed suit. Promising attempts to classify galaxies using machine learning date back to 1992 [11]. Other applications include quasar [12] and multi class [13][14] classifications.

### **2.2 Establishing classes**

### **2.3 Data Augmentation Techniques**

### **2.4 Pre-processing to improve performance**

- cropping, rescaling (no noticeable effect?)[15]

# References

- [1] M. Pössel, "A Beginner's Guide to Working with Astronomical Data," *The Open Journal of Astrophysics*, vol. 3, no. 1, p. 10.21105/astro.1905.13189, Jan. 2020, arXiv:1905.13189 [astro-ph, physics:physics]. [Online]. Available: <http://arxiv.org/abs/1905.13189>
- [2] "SDSS Instruments." [Online]. Available: <http://skyserver.sdss.org/dr1/en/sdss/instruments/instruments.asp>
- [3] "Voyages | Types of Data." [Online]. Available: <https://voyages.sdss.org/preflight/capturing-recording-light/types-of-data/>
- [4] I Appenzeller (Immo) author, *Introduction to astronomical spectroscopy*, ser. Cambridge observing handbooks for research astronomers. Cambridge: University Press, 2013, no. 9.
- [5] M. R. Blanton *et al.*, "Sloan Digital Sky Survey IV: Mapping the Milky Way, Nearby Galaxies, and the Distant Universe," *The Astronomical Journal*, vol. 154, p. 28, Jul. 2017, aDS Bibcode: 2017AJ....154...28B. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2017AJ....154...28B>
- [6] K. W. Willett *et al.*, "Galaxy Zoo 2: detailed morphological classifications for 304,122 galaxies from the Sloan Digital Sky Survey," Aug. 2013, arXiv:1308.3496 [astro-ph]. [Online]. Available: <http://arxiv.org/abs/1308.3496>
- [7] C. J. Lintott *et al.*, "Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey," *Monthly Notices of the Royal Astronomical Society*, vol. 389, pp. 1179–1189, Sep. 2008, aDS Bibcode: 2008MNRAS.389.1179L. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2008MNRAS.389.1179L>
- [8] V. Dhar, "Data Science and Prediction." [Online]. Available: <https://cacm.acm.org/magazines/2013/12/169933-data-science-and-prediction/abstract>
- [9] S. Suthaharan, *Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning*, 1st ed., ser. Integrated Series in Information Systems. New York, NY: Springer, 2015, vol. 36, iSSN: 1571-0270.
- [10] J. Lemley, S. Bazrafkan, and P. Corcoran, "Smart Augmentation Learning an Optimal Data Augmentation Strategy," *IEEE Access*, vol. 5, pp. 5858–5869, 2017, conference Name: IEEE Access.
- [11] M. C. Storrie-Lombardi, O. Lahav, L. Sodré, Jr, and L. J. Storrie-Lombardi, "Morphological Classification of galaxies by Artificial Neural Networks," *Monthly Notices of the Royal Astronomical Society*, vol. 259, no. 1, pp. 8P–12P, Nov. 1992. [Online]. Available: <https://doi.org/10.1093/mnras/259.1.8P>
- [12] G. T. Richards *et al.*, "Bayesian High-redshift Quasar Classification from Optical and Mid-IR Photometry," *The Astrophysical Journal Supplement Series*, vol. 219, p. 39, Aug. 2015, aDS Bibcode: 2015ApJS..219...39R. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2015ApJS..219...39R>
- [13] C. Wang *et al.*, "J-PLUS: Support vector machine applied to STAR-GALAXY-QSO classification," *Astronomy & Astrophysics, Volume 659, id.A144, <NUMPAGES>23</NUMPAGES> pp.*, vol. 659, p. A144, Mar. 2022. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2022A%26A...659A.144W/abstract>
- [14] P. O. Baqui *et al.*, "The miniJPAS survey: star-galaxy classification using machine learning," *Astronomy &*

## REFERENCES

- Astrophysics*, Volume 645, id.A87, <NUMPAGES>19</NUMPAGES> pp., vol. 645, p. A87, Jan. 2021. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2021A%26A...645A..87B/abstract>
- [15] S. Dieleman, K. W. Willett, and J. Dambre, "Rotation-invariant convolutional neural networks for galaxy morphology prediction," *Monthly Notices of the Royal Astronomical Society*, vol. 450, pp. 1441–1459, Jun. 2015, aDS Bibcode: 2015MNRAS.450.1441D. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2015MNRAS.450.1441D>

# A Galaxy Zoo 2 Images





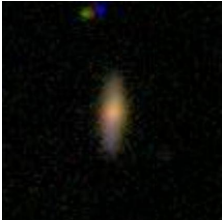

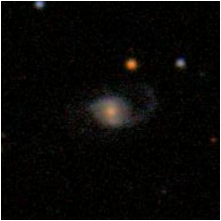





Class	Ex.1	Ex.2	Ex.3	Ex.4
Ei				
Er				
Sb				

Table A.1: Examples of Galaxy Zoo 2 images from different morphological classes