



Explainable and Robust Deep Learning Framework for Galaxy Morphology Classification

Prepared By

Name: Ahmad Yousef Alkahlout

ID No.: 253000753

Name: Mohmmmed Taleb Mohammed

ID No.: 252000097

Submitted in fulfillment of the requirements for the degree Master of
Science in Artificial Intelligence

At

Collage of Informatics

Midocean University

UAE

Supervisor Name: Dr. Mohamed Essa

Collage of Informatics, Midocean University, Comoros

Email: 253000753@st.midocean.edu.km /
252000097@st.midocean.edu.km

Acknowledgements

I would like to thank my supervisor, Dr. Mohamed Essa, for his continuous guidance and valuable feedback throughout this research. His expertise in both artificial intelligence and galaxy studies has greatly influenced the direction of this work. His comments challenged me to think more critically and approach the research from a scientific perspective rather than a purely technical one.

I am also thankful to the faculty members at the College of Informatics, Midocean University, for providing a supportive academic environment that inspired independent thinking and research development.

Finally, I would like to express my appreciation to my family for their patience, encouragement, and support during this journey.

Abstract

Galaxy morphology classification is an important task in astrophysics, as the structural appearance of galaxies is closely related to their shape and evolution. With the increasing volume of galaxy images generated by modern astronomical surveys, manual classification has become unworkable. As a result, automated approaches based on deep learning have gained significant attention.

Although convolutional neural networks have shown strong performance in galaxy image classification, most existing approaches focus primarily on accuracy. In scientific research, however, predictive performance alone is not sufficient. Researchers also require interpretability, stability under varying imaging conditions, and an understanding of model confidence.

This thesis proposes a deep learning framework for galaxy morphology classification that integrates robustness evaluation, uncertainty estimation, and explainability within a single methodology. The Galaxy10 dataset is used as a benchmark dataset due to its structured labels and predefined splits, which allow controlled experimentation.

The framework evaluates model stability under controlled worries, applies Monte Carlo Dropout for uncertainty estimation, and uses Grad-CAM to analyze whether model attention aligns with meaningful morphological structures. The aim of this work is to develop a classification approach that is not only accurate, but also reliable and scientifically interpretable.

Table of Contents

Acknowledgements	iii
Abstract.....	iv
Table of Contents	v
List of Abbreviations	vi
List of Figures.....	vii
List of Tables	ii
Chapter 1: Introduction	1
1.1 Overview	1
1.2 Problem Definition	2
1.3 Thesis Objectives	3
1.4 Thesis Contribution	3
1.5 Thesis Organization	4
Chapter 2 Background and Related Work.....	5
Chapter 3: Proposed Approach (Methodology)	6
Chapter 4: Experimental Results and Discussion	7
Chapter 5: Conclusion and Future Work	8
References.....	9

List of Abbreviations

- AI – Artificial Intelligence
- CNN – Convolutional Neural Network
- DL – Deep Learning
- Grad-CAM – Gradient-weighted Class Activation Mapping
- MC Dropout – Monte Carlo Dropout

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Overview

Galaxy morphology classification has always been one of the core problems in observational space science. The way a galaxy looks (whether it is spiral, elliptical, or irregular) tells us a great deal about its formation history, internal dynamics, and star formation activity. Traditionally, astronomers relied on visual inspection to perform this classification. While this approach was practical when datasets were relatively small, it has become increasingly unrealistic with the expansion of modern sky surveys.

Today, astronomical surveys generate enormous volumes of galaxy images. Manually reviewing and labeling such data is not only time-consuming but also subject to human bias and inconsistency. As a result, automated classification methods are no longer optional; they are necessary.

Deep learning models, particularly convolutional neural networks (CNNs), have demonstrated strong performance in image-based tasks. Their ability to learn features directly from raw pixel data makes them well suited for galaxy image analysis. However, high accuracy alone does not guarantee scientific usefulness. In astrophysics, researchers need to understand why a model makes a certain decision. They also need to know how stable that decision is under variations in imaging conditions, and how confident the model is in its prediction.

For these reasons, this thesis does not focus solely on classification performance. Instead, it aims to develop a framework that combines predictive accuracy with interpretability, robustness evaluation, and uncertainty estimation.

1.2 Problem Definition

Although many studies have applied deep learning to galaxy morphology classification, most of them emphasize performance metrics such as accuracy or F1-score. These metrics are important, but they do not fully address scientific reliability.

First, deep learning models are often treated as black boxes. In scientific research, this is problematic. If a model predicts that a galaxy is spiral, researchers should be able to verify that the decision is based on meaningful structures, such as spiral arms, rather than irrelevant background features.

Second, galaxy images are not always captured under identical conditions. Variations in brightness, noise, or small distortions may occur. A reliable classification system should remain stable under such conditions.

Third, galaxy morphology is not always clearly defined. Some galaxies exhibit transitional or ambiguous features. In these cases, it is important not only to assign a label, but also to estimate how confident the model is in that label.

Based on these considerations, the main research question of this thesis is:

How can we design a deep learning-based galaxy morphology classification framework that achieves strong performance while also ensuring interpretability, robustness, and uncertainty awareness suitable for scientific applications?

Finally, In this study, the Galaxy10 DECals dataset is used as the primary benchmark dataset for galaxy morphology classification. The dataset is derived from the Galaxy Zoo project and provides labelled galaxy images suitable for supervised deep learning experiments.

1.3 Thesis Objectives

The objectives of this research are:

1. To develop baseline convolutional neural network models for galaxy morphology classification.
2. To apply transfer learning techniques to improve performance and generalization.
3. To evaluate classification performance using standard quantitative metrics.
4. To analyze model robustness under controlled image perturbations.
5. To incorporate uncertainty estimation mechanisms into the prediction process.
6. To apply explainable AI techniques to interpret model decisions in a scientifically meaningful way.

1.4 Thesis Contribution

This work contributes in several ways.

First, it provides a structured evaluation of deep learning architectures for galaxy morphology classification.

Second, it introduces a robustness evaluation procedure that measures how performance changes under simulated observational variability.

Third, it integrates uncertainty estimation into the classification pipeline, which is particularly important in cases of ambiguous morphology.

Finally, it applies explainable AI techniques not only for visualization, but also for assessing whether model attention aligns with known morphological structures.

1.5 Thesis Organization

The remainder of this thesis is organized as follows.

Chapter 2 discusses the theoretical background and relevant research areas.

Chapter 3 presents the proposed methodology.

Chapter 4 analyzes the experimental results.

Chapter 5 concludes the study and suggests directions for future work.

Chapter 2 Background and Related Work

Chapter 3: Proposed Approach (Methodology)

Chapter 4: Experimental Results and Discussion

Chapter 5: Conclusion and Future Work

References