Technical Report

Galaxy Morphological Classification using Deep Learning

Name: Vishvapriya Sangvikar

Module Code & Title (COM7003 - Artificial Intelligence)

University Name: Leeds Trinity University (City Campus)

Date of Submission: 21/03/2025

Table of Contents

Abstract

Galaxy morphological classification is a critical activity in astrophysics that aids in understanding galaxy formation, evolution, and the large-scale structure of the universe. Traditional classification methods rely on human judgments, which are time-consuming and prone to inconsistencies. Deep learning models have been made a possible solution for automatic galaxy classification with the advent of large-scale astronomical surveys. This report compares the performance of three architectures of deep learning—Custom Convolutional Neural Networks (CNNs), ResNet-34, and Xception—on the 'Resized and Reduced Galaxy Zoo 2 Images' dataset. The models were trained for galaxy classification into three morphological categories: Elliptical, Spiral, and Spiral Barred. The findings pinpoint the potential of deep learning for machine classification of galaxies and necessitate the provision of higher-resolution data and explainable AI techniques. The future will be focused on improving the accuracy of models and deploying the classification system in an interactive platform for large-scale astronomy research. This research contributes to the development of scalable, automated tools for galaxy classification, supporting advancements in astrophysics and large-scale astronomical surveys.

Introduction

A galaxy is a complex, giant system of stars, star remnants, interstellar gas, dust, clouds, and dark matter bound together by gravity. According to (Harvey et al. 2022), the universe contains an estimated 200 billion galaxies, each with varying size, structure, and content. Galaxies were initially revealed in early telescopic observations, whereby astronomers identified them as nebulae—extended, fuzzy objects in the night sky. Early observations by Christiaan Huygens during the mid-seventeenth century, and later by Edmond Halley and Nicolas-Louis de Lacaille in the eighteenth century, established the foundations of galaxies as separate stellar systems and not as part of the Milky Way (Cimatti et al., 2019).

Galaxies are formed through the gravitational collapse of massive gas and dust clouds, resulting in the development of stars and planetary systems after millions of years. They are formed and structured due to the interaction of dark matter, gravity, and various astrophysical processes. It was Edwin Hubble in 1926 who put forward the morphological galaxy classification system that was later advanced by de Vaucouleurs

(1959) in an attempt to put galaxies into classified groups such as elliptical, lenticular, irregular, and spiral (Zaritsky et al, 1995). Elliptical galaxies, for instance, are characterized as smooth ellipsoidal in shape and lacking conspicuous features such as arms or a disk. They are almost spherical to highly elongated and are typically found in dense regions of the universe, such as galaxy clusters. Spiral galaxies, however, consist of a flat, rotating disk with a central bulge and spiral arms that extend outward. The arms are star-forming regions, typically bright because of hot, young stars. Some spiral galaxies have a bar-shaped structure in the middle connecting the spiral arms, known as barred spirals. The Milky Way, for example, is a barred spiral galaxy.

Galaxy classification is a valuable tool for astronomers, providing a systematic way of studying their diverse structures and behaviors. Classifying galaxies based on their shapes, compositions, and evolutionary histories allows astronomers to identify patterns and relationships that offer hints about their formation and evolution. But with the introduction of big-astronomical surveys like the Sloan Digital Sky Survey (SDSS) and the future Vera C. Rubin Observatory's Legacy Survey of Space and Time (LSST), there has been an exponential rise in data. It is time-consuming and liable to human errors and biases to classify millions of galaxies manually.

To address these challenges, advanced computational tools like Convolutional Neural Networks (CNNs) have become indispensable. CNNs, a type of deep learning algorithm designed for image analysis, offer a powerful solution for automating galaxy classification. They can efficiently analyse vast datasets, extract intricate features from galaxy images, and provide consistent, objective, and highly accurate classifications. Their ability to handle complex, noisy, and multi-wavelength data makes them particularly well-suited for modern astronomy, enabling researchers to uncover subtle patterns and relationships that might otherwise go unnoticed.

For this project, the 'Resized and Reduced Galaxy Zoo 2 Images' dataset from Kaggle has been utilised. This dataset comprises approximately 240,000 images of galaxies, resized and reduced to a manageable resolution for computational analysis. Derived from the Sloan Digital Sky Survey (SDSS), the images have been meticulously labelled through the Galaxy Zoo 2 citizen science initiative, providing detailed morphological classifications such as the presence of spiral arms, bars, or elliptical shapes. This dataset serves as a valuable resource for training and evaluating machine learning models for galaxy classification.

The primary objective of this study is to compare the performance of Convolutional Neural Networks (CNNs), ResNet, and Xception models in the classification of galaxies.

By leveraging the 'Resized and Reduced Galaxy Zoo 2 Images' dataset, the study aims to evaluate the accuracy, efficiency, and robustness of these models in categorising galaxies into distinct morphological types, such as spiral, elliptical, lenticular, and irregular. The comparison will focus on their ability to handle complex and noisy astronomical data, as well as their effectiveness in extracting intricate features from galaxy images. Through this analysis, the study seeks to identify the most suitable model for automated galaxy classification, contributing to the development of scalable tools for large-scale astronomical surveys and enhancing our understanding of galaxy formation and evolution.

Fig 1. Convolutional Neural Network Architecture

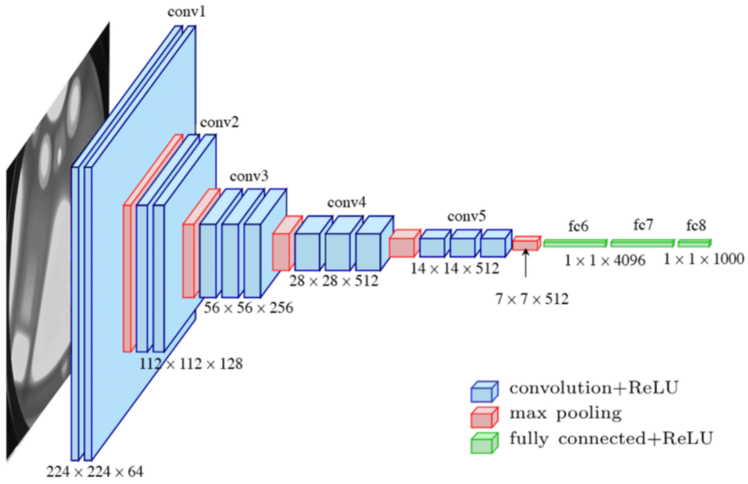
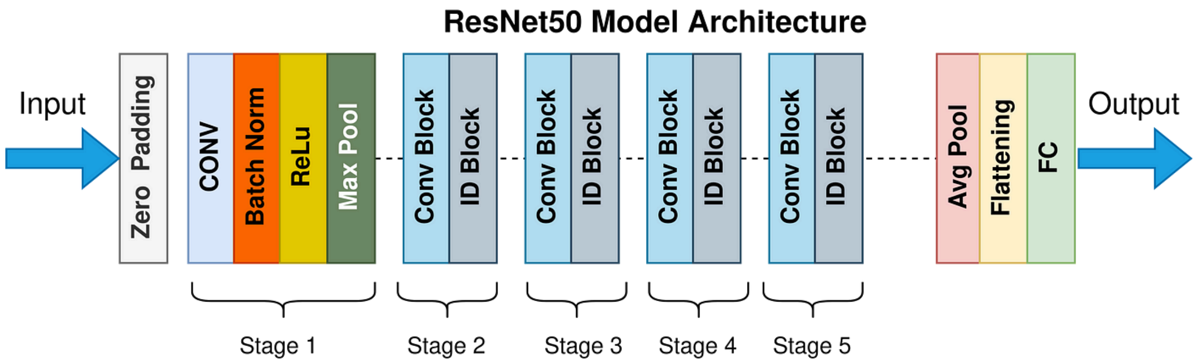


Fig 2. ResNet Architecture



Problem Definition and Business Context

The problem addressed in this project is the automated classification of galaxies into distinct morphological types, such as spiral, elliptical and spiral barred based on their visual features. This is achieved through a comparative study of advanced deep learning models, including Convolutional Neural Networks (CNNs), ResNet, and Xception. The goal is to evaluate and compare the performance of these models in accurately categorising galaxies, ensuring robustness and efficiency in handling complex and noisy astronomical data. By leveraging the 'Resized and Reduced Galaxy Zoo 2 Images' dataset, this project aims to identify the most effective model for automated galaxy classification, contributing to advancements in astrophysics and enabling scalable solutions for large-scale astronomical surveys.

Data Preparation

The dataset comprises images of galaxies in three distinct sizes: 69x69, 227x227, and 299x299 pixels, accompanied by a mapping file containing the metadata associated with each image. The images were organised into two separate directories: Test and Train, to facilitate the training and evaluation of the models. For this project, the dataset was categorised into three distinct classes: Elliptical, Spiral, and Spiral Barred, reflecting the primary morphological types of galaxies under study.

Following data collection from the source, the Kaggle Notebook service was employed as the integrated development environment (IDE) for conducting the analysis and implementing machine learning models. To enhance the training process, a Graphical Processing Unit (P100 GPU) was utilised, providing accelerated computational capabilities. The Kaggle platform offered a comprehensive and user-friendly environment for data preprocessing, model training, and evaluation, ensuring a streamlined and efficient workflow throughout the project. This approach enabled effective management of tasks, from initial data preparation to the final stages of model deployment, while maintaining a high standard of performance and accessibility.

After loading the model in my dataset with the help of Kaggle API, I performed the data preprocessing on the mapping file to understand the metadata of the images. The dataset consisted of agreement scores and total votes. These metrics reflect the level of consensus among volunteers regarding the classification of each galaxy. In this stage, I checked if there are any duplicate entries, incorrect entries, missing image metadata. I cleaned the dataset to make sure there are no inconsistencies.

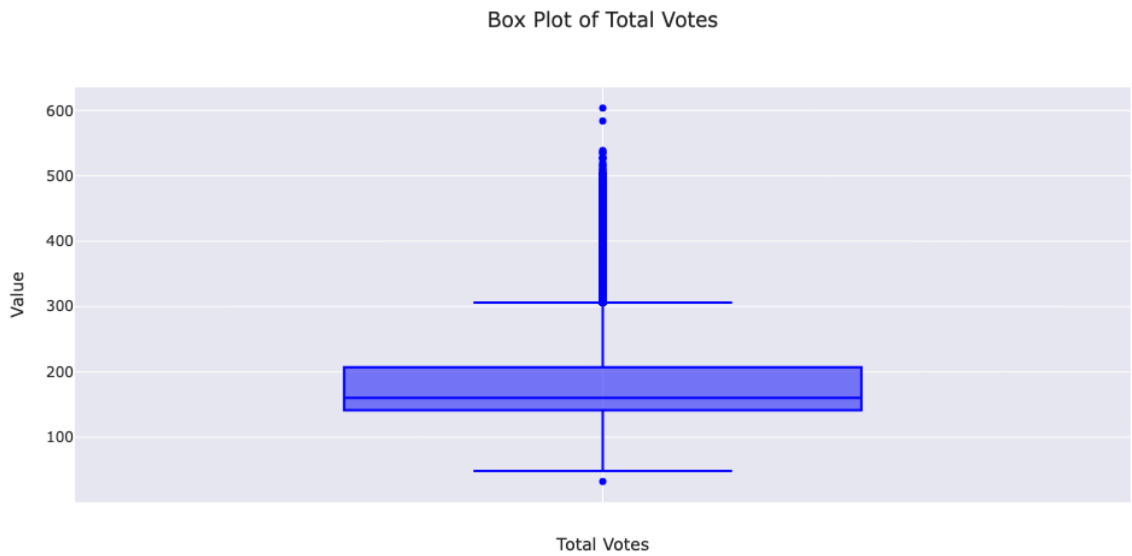
During the data processing, I identified that the classes are not balanced properly which I balanced by creating symlinks. Symlinks or Symbolic links are essentially shortcuts or references that point to the original files without duplicating them. By generating additional symlinks for underrepresented classes, I effectively balanced the dataset, ensuring that each class had an equal number of samples for training. This approach not only resolved the imbalance but also optimised storage usage, as it avoided the need to create redundant copies of the images.

Exploratory Data Analysis (EDA)

After analysing the data, I plotted graphs to visualise and understand the distribution of features within the dataset. The initial features selected for this analysis were the agreement scores and total votes. Upon examination, it was observed that there is no direct correlation between these two features, indicating that the level of consensus among volunteers (agreement scores) does not necessarily depend on the total number of votes cast for each galaxy. After visualising the features, I started to visualise the images which I did in the Image Exploration section in the code file.

After calculating the mean grey values for the images to analyse the pixel distribution across the dataset, I observed that the resolution of the images was relatively low and the images were quite dark. This limitation restricted the level of detail available for advanced processing techniques. Initially, I considered applying edge detection to enhance feature extraction, as it could have highlighted structural differences in the images. However, the low resolution made this approach impractical, prompting me to explore alternative preprocessing methods to effectively analyse the data within the constraints of the image quality. Subsequently, I plotted graphs for the different colour channels (e.g., red, green, and blue) to examine the distribution of pixel intensities across the three galaxy categories: Elliptical, Spiral, and Spiral Barred. This analysis provided insights into the variations in colour distribution among the categories, aiding in the understanding of the dataset's characteristics and guiding further preprocessing steps.

Fig 3. Box Plot of Total Votes in the dataset



Model Selection and Justification

At this point, I chose three different models to compare their performance in galaxy classification: a Custom Convolutional Neural Network (CNN), ResNet34, and Xception.

The Custom CNN was constructed from scratch, one that would be specially trained for this task with architecture flexibility and the ability to modify the hyperparameters. ResNet was employed due to its innovative application of residual connections to address the issue of vanishing gradients as well as allowing very deep training networks. Xception was chosen on the basis of efficiency and performance with depthwise separable convolutions for reducing complexity during computation with equal precision retained. Comparison of these models assisted me in selecting the most appropriate methodology to categorise galaxies into the predefined types: Elliptical, Spiral, and Spiral Barred. The comparison is a significant part of the project and emphasizes the strength and weaknesses of each model while processing astronomical image data.

For the Custom CNN, I applied data augmentation techniques, such as random rotation and flipping, to increase the diversity of the training data. This was a critical stage to prepare the model to generalise well to new images by mimicking different orientations and perspectives. I started by defining the model parameters, configuring safely the input shape, convolutional layer sizes, and activation functions. Custom CNN architecture was created with 7 layers. The architecture rendered the model efficient as well as robust, most suited for the task of galaxy classification

Next, I trained the custom CNN model on the 69x69-sized dataset of images because they were simpler to train. The resolution was chosen because it was in-between having sufficient detail for accurate classification and for keeping computational efficiency during training. Image reduction reduces the complexity and resource requirements of the model and, as a result, accelerates the training process and makes it more practical with large datasets. The images retained enough structural information to allow the model to acquire discriminative features for distinguishing the three galaxy classes even with the limited resolution.

After the initial training, I proceeded to create tailored Xception and ResNet34 models similar to the custom CNN. I chose Xception and ResNet34 because they are lighter, faster to train, and require less computational resources compared to their large counterparts. I build the models up to 18 and 34 layers respectively. This not only improved the training process efficiency but also provided me with greater flexibility to experiment with different layer configurations, filter sizes, and regularisation techniques. By calibrating these models to the specific profile of my data, I was able to achieve a balance between performance and resource usage so that I could have more iterative refinement and testing.

Fig 4- Custom CNN Model Training code snippet

```
reduce_lr = ReduceLROnPlateau(patience=1, factor=0.5, min_lr=1e-6)
timestp = TimestampCallback()
early = EarlyStopping(patience=10, restore_best_weights=False, verbose=1)

metrics = ['accuracy']
callbacks = [reduce_lr, timestp]
```

```
history_s = run_cnn('small', sizes=sizes, dropout=dropout, callbacks=callbacks)
```

```
Found 20946 files belonging to 3 classes.
Using 14663 files for training.
Using 6283 files for validation.
Epoch 1/30
459/459 ━━━━━━━━━━━━━━━━━━━━ 15s 20ms/step - accuracy: 0.4454 - loss: 1.0320 - val_a
Epoch 2/30
459/459 ━━━━━━━━━━━━━━━━━━━━ 2s 4ms/step - accuracy: 0.5611 - loss: 0.9149 - val_acc
Epoch 3/30
459/459 ━━━━━━━━━━━━━━━━━━━━ 2s 4ms/step - accuracy: 0.5808 - loss: 0.8831 - val_acc
Epoch 4/30
459/459 ━━━━━━━━━━━━━━━━━━━━ 2s 4ms/step - accuracy: 0.6020 - loss: 0.8515 - val_acc
Epoch 5/30
459/459 ━━━━━━━━━━━━━━━━━━━━ 2s 4ms/step - accuracy: 0.6212 - loss: 0.8234 - val_acc
Epoch 6/30
459/459 ━━━━━━━━━━━━━━━━━━━━ 2s 4ms/step - accuracy: 0.6310 - loss: 0.7895 - val_acc
Epoch 7/30
459/459 ━━━━━━━━━━━━━━━━━━━━ 2s 4ms/step - accuracy: 0.6449 - loss: 0.7669 - val_acc
Epoch 8/30
459/459 ━━━━━━━━━━━━━━━━━━━━ 2s 4ms/step - accuracy: 0.6694 - loss: 0.7332 - val_acc
Epoch 9/30
```

Model Optimisation and Testing

After completing the training of all the models I proceeded with hyperparameter tuning. For hyperparameter tuning and data augmentation in the custom models, I adopted a systematic approach to optimize performance. I tried with various hyperparameters, such as dropout rates, learning rates, and filter sizes, to identify the best configuration for each model. Data augmentation techniques, including random rotation, flips and zooms, were applied to enhance the diversity of the training data and improve the model's ability to generalise to unseen data.

Model Evaluation and Results

After training all the models, I plotted the training and validation metrics across epochs to analyse their convergence behavior. My primary objective was to assess whether the models were learning effectively while avoiding issues like overfitting, underfitting, divergence, slow convergence, or instability. By examining these plots, I could evaluate each model's learning dynamics and overall generalisation capability.

After extensive experimentation and performance analysis, I selected Xception with data augmentation as the final model for testing. It demonstrated the highest accuracy and generalisation on both the validation and test datasets. When tested on the final test set, Xception achieved an accuracy of **74%,** outperforming ResNet-34 **(71%)** and the Custom CNN **(69%).**

Fig 5- Final Model Accuracy Evaluation

Fig 6- Custom CNN Model Evaluation

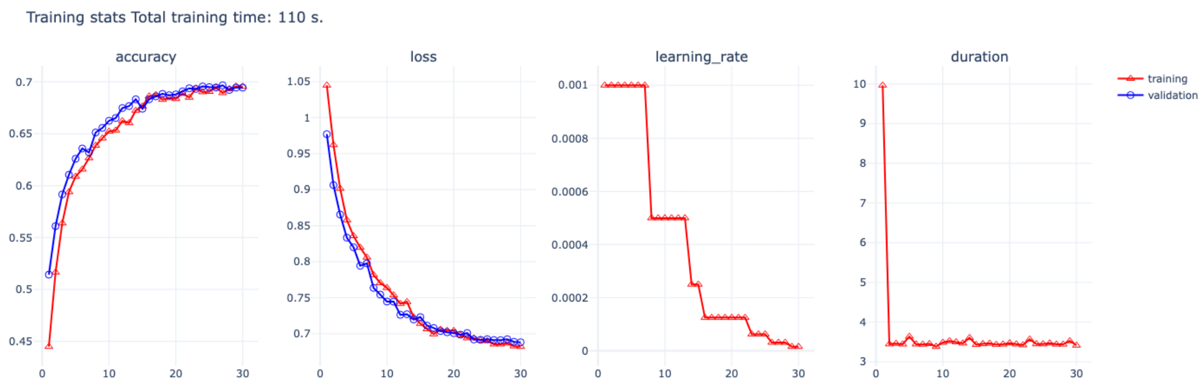


Fig 7- Custom Xception Model Evaluation

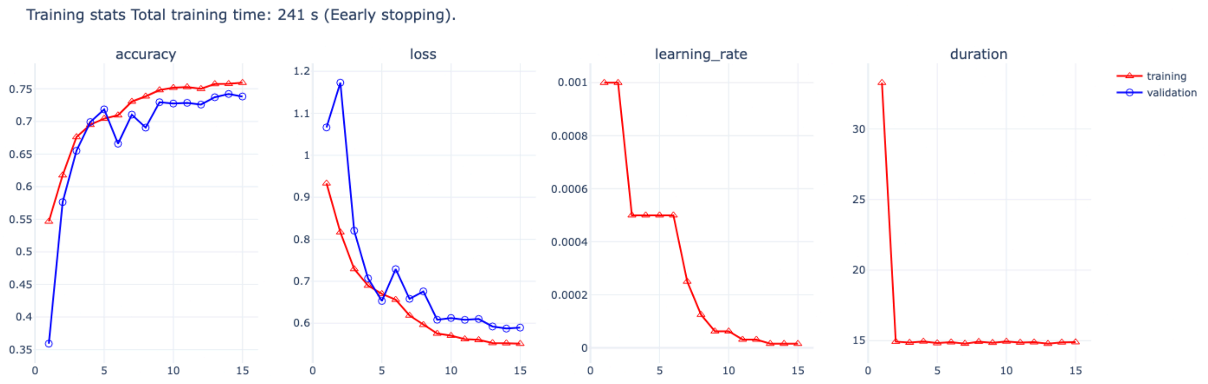


Fig 8- Custom ResNet34 Model Evaluation

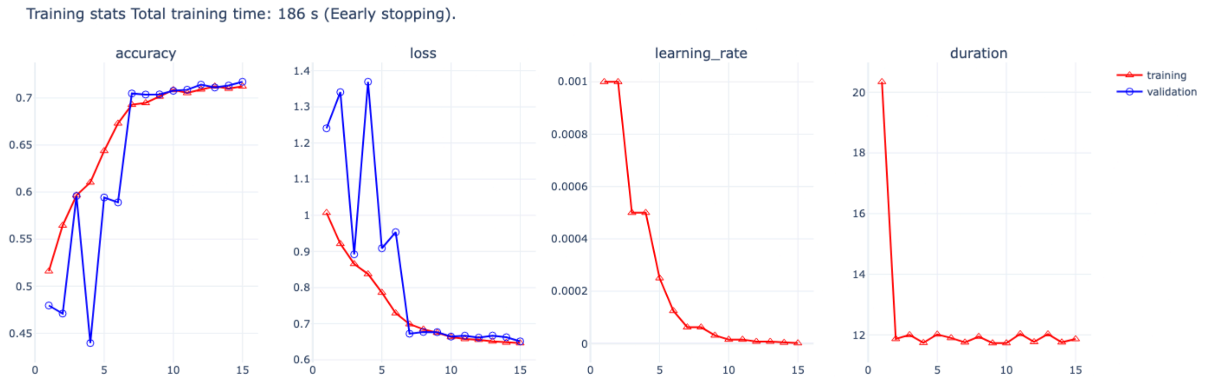
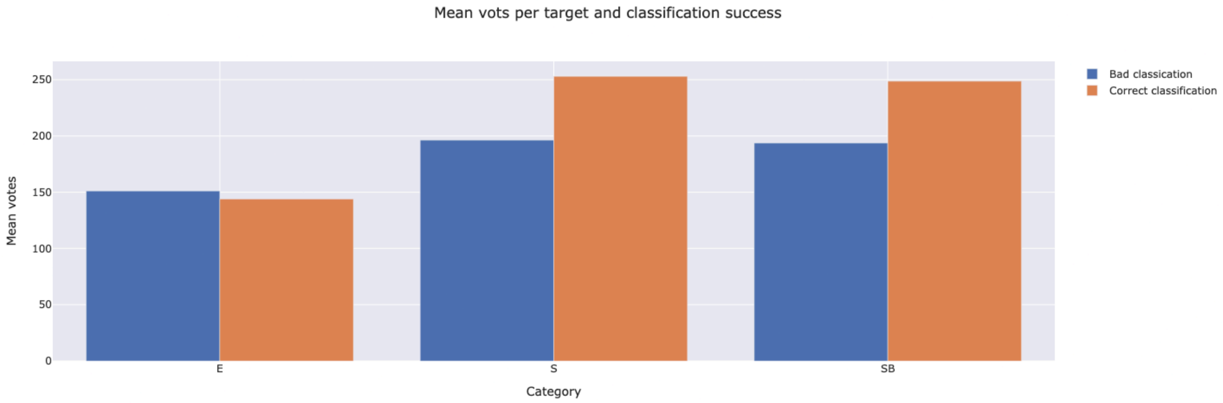


Fig 9- All Models Evaluation comparison

Model type	Input data	Augmentation	Epoch counts	Max accuracy (val)	Min loss (val)	Total duration (s)
CNN	small	False	30	0.6637	0.7771	64
CNN	small	True	30	0.6968	0.6882	110
Xception	small	False	15	0.714	0.6695	238
Xception	small	True	15	0.7422	0.5873	241
ResNet	small	False	15	0.7019	0.6744	189
ResNet	small	True	15	0.7172	0.6511	186

Fig 10- Classification success evaluation of the model



Challenges during the project

During this project, I encountered several challenges that required careful problem-solving and adaptation. One of the primary issues was dataset imbalance, as the number of images for the Elliptical galaxy class significantly outweighed the other classes, potentially leading to a biased model. To address this, I created symbolic links (symlinks) for the underrepresented classes, which helped balance the dataset and improve model fairness. Without this approach, the model exhibited bias towards the overrepresented class, highlighting the importance of symlinks in ensuring balanced training. Additionally, I faced technical issues during compilation which required debugging and adjustments to the codebase. Another challenge arose when I attempted to use edge detection to better understand the structural differences between galaxies. However, due to the low resolution of the images, this approach did not yield significant improvements in accuracy. Finally, I initially tried executing the project on AWS SageMaker but encountered challenges due to the high costs associated with GPU usage. As a result, I migrated to Kaggle Notebooks, which provided a more cost-effective and accessible environment for training and experimentation. Overcoming these challenges not only improved the robustness of the project but also enhanced my problem-solving and technical skills.

Ethical and Practical Considerations

In the development and deployment of machine learning models to classify galaxies, ethical as well as practical issues need to be addressed to ensure fairness, transparency, and accessibility. From an ethical standpoint, it is essential to deal with biases in the data, i.e., class imbalances, to prevent biased predictions that under-represent certain kinds of galaxies. Further, the black-box nature of deep learning models makes interpretability hard, which can undermine trust in AI-based classifications. Offering transparency through explainable AI techniques can help establish trust in the results.

A significant practical obstacle is that deep learning models are computationally costly to train, taking huge resources and energy consumption, which translates to a high carbon footprint (Hao, K., 2019). Furthermore, limited access to cutting-edge computational resources and high-quality astronomical data may disadvantage researchers from under-resourced institutions and developing nations, increasing inequalities in scientific research.

Socially, automating galaxy classification could reduce the tasks of citizen scientists and human astronomers, potentially affecting public interest and opportunities for hands-on astronomy education negatively (Kohler, S., 2016). Achieving a balance between automation and human skill and encouraging AI-supported collaborative research can enable us to retain scientific interest and engagement.

To alleviate these issues, it is crucial to promote open methodology, green computing principles, and equitable access to AI tools and datasets so that AI-enabled advancements in astronomy benefit the broader scientific community and minimize unintended consequences.

Conclusion and Future Work

In this study, I investigated the application of deep learning techniques for galaxy classification, employing Convolutional Neural Networks (CNNs), ResNet-34, and Xception to categorize galaxies into three morphological classes: Elliptical, Spiral, and Spiral Barred. The performance of these models was systematically evaluated based on accuracy, generalisation capability, and computational efficiency. While the deep learning models demonstrated promising results, the study also revealed several challenges. The high computational cost associated with training these models remains a significant concern, alongside the inherent difficulty in interpreting their decisions due to their "black-box" nature. Additionally, the low resolution and dark characteristics of the images limited the extraction of finer details, which could have further improved classification accuracy.

To address these limitations and enhance the model's performance and applicability, several future directions are proposed. Incorporating higher-resolution images or multi-wavelength astronomical data (e.g., infrared, ultraviolet) could provide more comprehensive features for classification, potentially improving accuracy. Furthermore, the integration of explainable AI (XAI) techniques, such as GRAD-CAM or SHAP values (Cogswell et al., 2019), could enhance the transparency and interpretability of the model's decision-making process. Once these improvements are implemented, the model could be integrated into a web-based or cloud-based interactive platform, enabling astronomers and citizen scientists to classify galaxies more efficiently.

Overall, this report intends to contribute to the advancement of AI-driven astronomy by demonstrating the potential of deep learning models in large-scale astronomical surveys. Future work will focus on improving model accuracy, efficiency, and interpretability, ensuring that AI-based classification methods are scalable, accessible, and transparent for the broader scientific community.

References

- Zaritsky , D. (1995, July 12). *SPECTRAL CLASSIFICATION OF GALAXIES ALONG THE HUBBLE SEQUENCE*. Arxiv. https://arxiv.org/pdf/astro-ph/9507044
- Cimatti, A., & Fraternali , F. (2019, December 3). Cambridge University Press. https://arxiv.org/pdf/1912.06216
- Sandage, A. (2004). Episodes in the Development of the Hubble Galaxy Classification. *Astrophysics and Space Science Library*, 39–56. https://doi.org/10.1007/978-1-4020-2862-5_3
- Raul Ioan Urechiatu, & Frincu, M. (2024). Improved Galaxy Morphology Classification with Convolutional Neural Networks. *SSRN*. https://doi.org/10.2139/ssrn.4699260
- *Vera Rubin Observatory's Legacy Survey of Space and Time | Kavli Institute for Particle Astrophysics and Cosmology (KIPAC)*. (2013). Stanford.edu. https://kipac.stanford.edu/research/projects/vera-rubin-observatorys-legacy-survey-space-and-time
- Hubble classification of galaxies using neural networks. (1994). *Vistas in Astronomy*, *38*, 273–280. https://doi.org/10.1016/0083-6656(94)90037-X
- Cavanagh, M. K., Bekki, K., & Groves, B. A. (2021). Morphological classification of galaxies with deep learning: comparing 3-way and 4-way CNNs. *Monthly Notices of the Royal Astronomical Society*, *506*(1), 659–676. https://doi.org/10.1093/mnras/stab1552
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2020). Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *International Journal of Computer Vision*, *128*(2), 336–359. https://doi.org/10.1007/s11263-019-01228-7
- Cheng, T.-Y., H. Domínguez Sánchez, Jesús Vega-Ferrero, Conselice, C. J., M. Siudek, Aragón-Salamanca, A., Bernardi, M., Cooke, R., Ferreira, L., M. Huertas-Company, J. Krywult, Palmese, A., A. Pieres, Plazas, A., A. Carnero Rosell, Gruen, D., Thomas, D., Bacon, D., Brooks, D., & James, D. J. (2022). Lessons learned from the two largest Galaxy morphological classification catalogues built by convolutional neural networks. *Monthly Notices of the Royal Astronomical Society*, *518*(2), 2794–2809. https://doi.org/10.1093/mnras/stac3228

- Howell, E. (2018, March 20). *How Many Galaxies Are There?* Space.com; Space.com. https://www.space.com/25303-how-many-galaxies-are-in-the-universe.html

Generative AI Use Statement

This assignment used generative AI in the following ways for the purposes of completing the assignment- research, generating sample code and feedback