**University of Dhaka**
**Department of Robotics and Mechatronics Engineering**

Master's Thesis Proposal

# Morphological and Spectral Classification and Stratification of Celestial Objects using Physics Informed Deep Learning

Submitted By:
**Md Fairuz Siddiquee**
Session: 2021-2022

Thesis Supervisor:
**Dr. Md Mehedi Hasan**
Assistant Professor
Department of Robotics and Mechatronics Engineering
University of Dhaka

______________________________

Signature of Supervisor

# Contents



# List of Figures

# 1 Introduction

The exploration and understanding of the cosmos has been an endeavour that has captivated human curiosity for centuries. As we gaze at the night sky, the myriad of celestial objects scattered across the vast expanse continues to evoke questions about their origins, compositions, and characteristics. In the pursuit of unravelling these mysteries, astronomers have amassed an unprecedented amount of data through various observational instruments and surveys.

Any naturally occurring physical entity, group, or structure that exists inside the observable universe is referred to as an astronomical object or celestial object [23], which includes galaxies, planetary systems, star clusters, nebulae, quasars, exoplanets, galaxy clusters, pulsars etc. With advancements in technology and data acquisition techniques, there is a growing need to develop innovative methods that can unveil hidden patterns and unprecedented insights from vast and complex astronomical datasets. One of the main tasks handled by machine learning (ML) algorithms is classification [13], [11]. A classifier is typically trained to recognize patterns in input data in order to predict the label that will be given to instances of data that haven't been observed before [4].

With the development of contemporary technology, astronomers' ability to analyse data has increased [22]. As a result, the field of astroinformatics [3] has emerged to assist in the analysis of the data given. Typically, this includes categorising enormous sets of astronomical photographs [1] – [18]. Our aim is to develop a comprehensive methodology for the morphological and spectral classification and stratification of celestial objects, with focus on the morphological and redshift classification of galaxies, utilising a combination of deep learning techniques and physics-informed neural networks.

The two primary groups (morphologies), elliptical and spiral, are intended to be distinguished by the morphological categorization of galaxy images [10]. Galaxies may additionally be divided into the following categories: irregular, elliptical, spiral, and lenticular. Using galaxy images (Merge Galaxies), we are able to identify galaxies that are colliding, which can shed light on how they evolved. In addition, we can classify galaxies based on their shape (Galaxy Morphology), which can help us comprehend how they form and develop. When sub-classes are added, the classification problem becomes more challenging [14]. The morphology is an essential clue for comprehending the physical processes and inner structure of the galaxy, as well as offering insight on the inception and growth of the cosmos [20].

This methodology aims to enhance the cataloguing of astronomical objects, expedite the exploration of exoplanets, and improve predictions regarding fundamental queries like the age of the universe and early evolution of the cosmos. Leveraging the immense volumes of data from projects like the Sloan Digital Sky Survey (SDSS) and the Galaxy Zoo dataset, this research aims to develop robust models that can classify galaxies based on their morphological features and redshift values. Convolutional neural networks (CNNs) [2] - [7], which are capable of using a raw image as input and executing an implicit feature extraction procedure along with the classification in a single step, provide satisfactory outcomes [19]. Modern state-of-the-art CNNs often include a relatively large number of layers [21], when dealing with challenging classification problems [9].

The study of morphological and spectral celestial stratification can help us to understand the formation and evolution of the universe. It will contribute to the field of discovering and identifying new astronomical objects and to study their properties. Our research work will accelerate the development of new techniques for observation and analysis.

# 2 Problem Statement

Astronomical data, such as pictures and spectra taken from telescopes and observatories, can include detailed characteristics and information that are not immediately detectable using typical analytic methods. This is because astronomical data are gathered from extremely far away. The problem comes in efficiently processing and analysing these datasets in order to extract useful knowledge about the morphology and spectral fingerprints of the celestial objects. This knowledge will give essential information on the objects' age, mass, composition, and evolutionary stage. The challenge that will be investigated in this study is how to make use of deep learning techniques in conjunction with physics-informed methodologies in order to categorise and organise celestial objects according to the morphological and spectral characteristics of such objects.

# 3 Research Questions

The following research aspects will be addressed through this research proposal:

- Finding the way to analyse and classify the structures of celestial objects by understanding the morphology and spectra of the celestial objects.
- Evaluation of the existing physics informed neural networks on the basis of the contribution to improve the accuracy of celestial object classification. Comparison of the deep learning algorithms to investigate efficient approaches.
- Incorporation of the Redshift (Z-Values) of the galaxies to perform stratification according to their respective relative distance.
- Exploring the fields where the results of this research be used to advance the cataloguing of astronomical objects, accelerate exploration of searching for exoplanets, and contribute to a more accurate prediction of the age of the universe and the early evolution of the universe.

# 4 Research Objectives

- Development of a deep learning model capable of accurately classifying celestial objects based on their morphological features.
- Incorporation of spectral data into the classification process using physics-informed neural networks to improve classification accuracy.
- Implementation of a Redshift-based stratification technique to enhance the precision of the classification process.
- Categorisation of the subdivision of the classification according to the orientation of the galaxies and the Redshift (z) value to extract information regarding the distance and age of the galaxies (Depth Perspective). Sorting of the 'Don't Know Category' into Stars, Quasars, Asteroids or Merging Galaxies. Processing of the CMB (Cosmic Microwave Background) Radiation, Anomalies, Artefacts or Noise.

- Validation of the proposed methodology using the Sloan Digital Sky Survey (SDSS) and Galaxy Zoo datasets and evaluation of the performance of the proposed models.
- Assessment of the impact of the proposed research on the advancement of the cataloguing of astronomical objects.

# 5 Literature Review

A growing body of research has focused on applying deep learning and physics-informed neural networks to astronomy and cosmology. Some of this research has focused on the use of deep learning for image classification and segmentation of astronomical objects. Other research has focused on the use of physics informed neural networks for modelling the dynamics of astronomical systems.

A number of studies have investigated the application of deep learning techniques in astronomy and have demonstrated the effectiveness of deep learning for image classification and segmentation of astronomical objects. Convolutional Neural Networks (CNNs) have been used to identify galaxies, stars, and other objects in images. Recurrent Neural Networks (RNNs) have been employed to analyse time-series data from astronomical sources. Additionally, the integration of physical knowledge into neural networks, known as physics-informed neural networks, has shown promise in various scientific domains. However, the combination of both deep learning and physics-informed techniques for celestial stratification remains relatively unexplored.

González et al. [6] presented a method for automatic detection and classification of galaxies using deep learning and data augmentation. The method is implemented in the AstroCV open source computer vision library. The Galaxy Zoo and the Sloan Digital Sky Survey (SDSS) are two examples of open datasets that were used to train the detection and classification algorithms.

The authors first trained a convolutional neural network (CNN) to detect galaxies. The CNN was trained on a dataset of images that had been labelled with the presence or absence of galaxies to identify the features of galaxies, such as their shape, brightness, and size. In this paper, researchers classified the galaxies into 5 categories: Elliptical, Spiral, Edge-on, DK (Don't Know), and Merge. The authors then used data augmentation to increase the size of the training dataset. Data augmentation is a technique that creates new training data by applying transformations to existing data. This helped to make the CNN more robust to variations in the input data.

The real-time object detection system YOLO and the deep learning framework DARKNET were used to train the detection and classification algorithms. They trained the YOLO model on a dataset of galaxy images that had been augmented using a variety of techniques, such as rotating, flipping, and cropping the images. This data augmentation contributed in making the model more resilient to changes in the manner in which galaxies appear. The trained algorithm was utilised by the authors to identify and categorise galaxies in new images. On a dataset of images from the Sloan Digital Sky Survey (SDSS), the authors assessed the effectiveness of their strategy.

The authors demonstrated that deep learning methods can learn to identify the features of galaxies, even if they are faint or distorted and through data augmentation the accuracy of the model can be increased. As the method was tested only on a limited number of galaxy types, there is an uncertainty of the performance of this method on rare or unusual galaxy types. The proposed method can be used to detect galaxies in real time. The accuracy was 80% for the T2 Dataset and 81% for the T4 Dataset, which is impressive given the nature of Galaxy Zoo categorization based on votes, in which certain objects have divided votes but are allocated to a single classification (e.g., dwarf spheroidal or S0 galaxies).

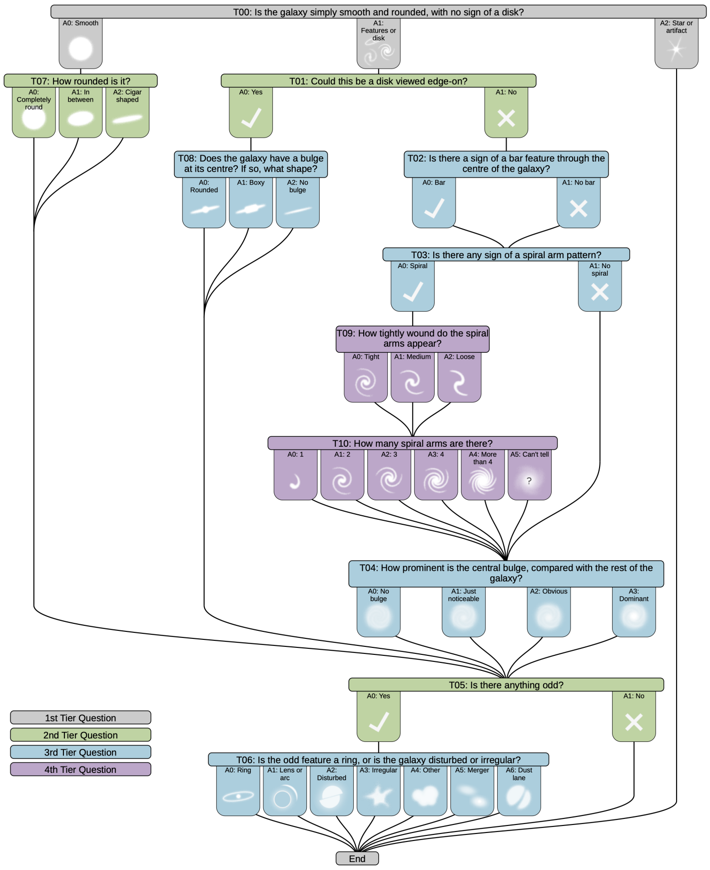


Figure 1: Galaxy Zoo Decision Tree

The research [17] used a deep convolutional neural network (CNN) to estimate photometric redshifts and related probability distribution functions (PDF) for galaxies in the Sloan Digital Sky Survey (SDSS) Main Galaxy Sample at z ¡ 0.4. The machine learning algorithm was trained to associate the observed brightnesses of galaxies with their redshifts. Researchers tested 4 sizes of training sets: 400k, 250k, 100k, and 10k galaxies (80%, 50%, 20% and 2% of the full database, respectively). In all but the final case researchers obtained a MAD dispersion $\sigma$MAD = 0.0091. This value is significantly lower than the best one published so far, obtained from another machine learning technique (KNN).

Jia et al. [12] proposed a classifier for the Z-wise and S-wise spirals. A Z-wise spiral is a spiral galaxy whose arms are oriented parallel to the line of sight, while an S-wise spiral is a spiral galaxy whose arms are oriented perpendicular to the line of sight. The proposed classifier, called the chirality equivariant residual network (CE-ResNet), is a convolutional neural network that is manifestly equivariant under a reflection of the input image. This means that the network will not be biassed towards classifying Z-wise spirals over S-wise spirals, or vice versa. The CE-ResNet was trained on images from the Sloan Digital Sky Survey (SDSS), with training labels provided by the Galaxy Zoo 1 (GZ1) project. The GZ1 project is a citizen science project that crowdsourced the classification of millions of galaxies.

On the test set, the CE-ResNet achieved an accuracy of 92%. This is significantly higher than the accuracy of human classifiers, which is typically around 80%. The network was able to correctly classify Z-wise spirals as Z-wise spirals and S-wise spirals as S-wise spirals, even when the galaxy was not perfectly aligned with the line of sight.

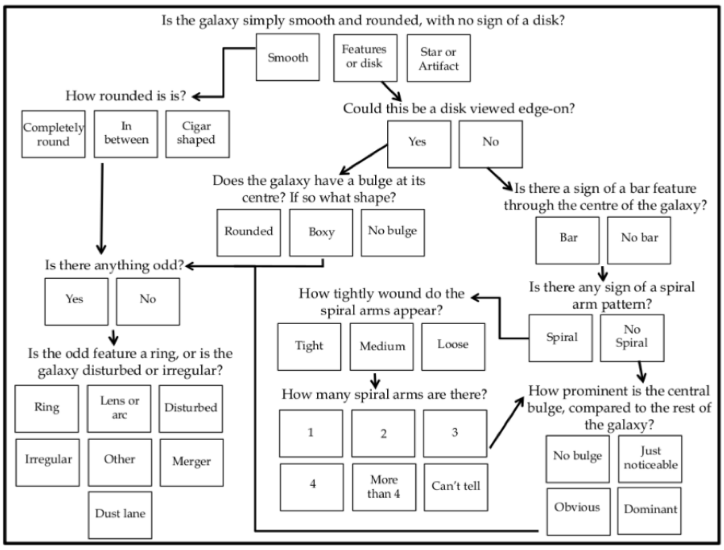


Figure 2: Schematic Diagram of Decision Tree of Galaxy Zoo 2 Classification [15]

Gharat et al. [5] suggested a deep learning method for categorising galaxies in the Sloan Digital Sky Survey (SDSS) in their study. To extract information from the images, the authors applied a convolutional neural network (CNN) and then classify them into one of 10 classes: elliptical, spiral, lenticular, barred spiral, unbarred spiral, irregular, compact, dwarf elliptical, dwarf spiral, and unknown.

The authors trained their CNN on a dataset of 304,122 SDSS images. They used a transfer learning approach, where they initialised the weights of the CNN with the weights of a CNN that had been trained on a large dataset of natural images. This helped the CNN to learn features that are relevant to galaxy classification. This study showed that deep learning methods have the potential to classify galaxies more accurately and efficiently than traditional methods. The authors used a large and diverse dataset of SDSS images to ensure that the CNN was able to learn features that are relevant to galaxy classification. Moreover, a transfer learning approach was used, which helped the CNN to learn features more quickly and efficiently. The authors evaluated their CNN on a test set of 10,000 SDSS images. The CNN achieved an accuracy of 84.73%, which is comparable to the accuracy of human classifiers.

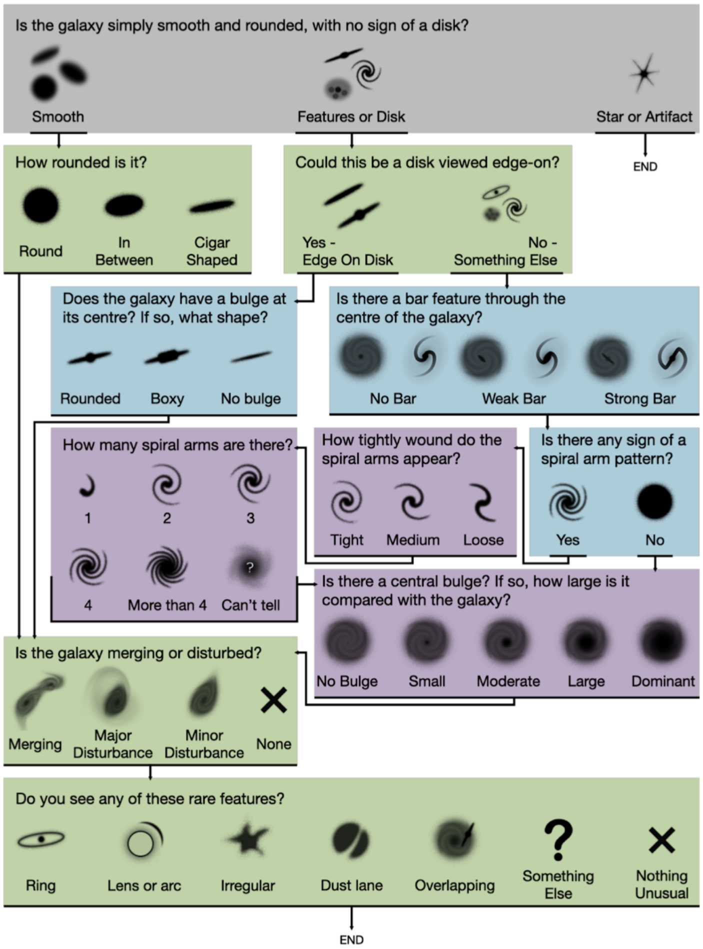


Figure 3: Flowchart of Galaxy Classification of Galaxy Zoo [24]

# 6 Research Methodology

The proposed research will use a combination of deep learning and physics informed neural networks to study morphological and spectral celestial stratification. The deep learning techniques will be used to learn the complex patterns in astronomical data, while the physics informed neural networks will be used to incorporate physical constraints into the learning process and to improve the accuracy of celestial object classification.

### 6.1 Data Collection and Preprocessing:

We will preprocess the data to normalise the intensities, remove noise, calibrate redshift, and perform data augmentation after gathering morphological data (images) and spectral data from a variety of astronomical datasets, including:

- Sloan Digital Sky Survey (SDSS): The Sloan Digital Sky Survey (SDSS) is one of the most ambitious and impactful astronomical surveys in history. The survey began collecting data in 2000; [8] the final imaging data release (DR9) covers over 35% of the sky, with photometric observations of approximately 1 billion objects, while the survey continues to acquire spectra, having collected spectra of over 4 million objects so far. The main galaxy sample has a median redshift of z = 0.1; there are redshifts for luminous red galaxies as far as z = 0.7, and for quasars as far as z = 5; and the imaging survey has been involved in the detection of quasars beyond a redshift z = 6.
- Galaxy Zoo JWST and Cosmic Evolution Early Research Science (CEERS) survey: JWST is being used by CEERS to capture extraordinarily deep photos of the sky covering 100 square arcmin.
- Galaxy Zoo Cosmic Dawn: The galaxies on Galaxy Zoo were photographed using the Subaru telescope's Hyper Suprime-Cam (HSC) instrument. Because these galaxies are normally further distant from earth compared to those in prior iterations of Galaxy Zoo, their colours and quality might look a little different, appearing redder and blurrier in general.
- Galaxy Zoo DECaLS: These galaxy images come from the Dark Energy Camera Legacy Survey (DECaLS). DECaLS is 10 times more sensitive to light than the Sloan Digital Sky Survey, which provided images for the first iteration of Galaxy Zoo because it utilises a larger telescope.
- The James Webb Space Telescope (JWST) is a space telescope intended primarily for infrared astronomy. Its high-resolution and high-sensitivity sensors allow it to see things that the Hubble Space Telescope cannot because they are too ancient, distant, or dim.[16]

### 6.2 Deep Learning Architecture

Deep neural networks, such as CNNs and RNNs, will be designed and trained to analyse the morphological and spectral features of celestial objects and to learn intricate features from images. The deep learning techniques will be used to learn the patterns in a dataset of astronomical images. Experimentation with different architectures and optimization techniques will help to achieve accurate classification and stratification.

## 6.3 Physics-Informed Integration

We will develop physics-informed neural networks that incorporate known physical principles and relationships into the learning process. Physics Informed Neural Network architecture has to handle both morphological and spectral data. The physics-informed neural networks will be utilised to model the celestial dynamics. The network structure should incorporate relevant physical equations and constraints, like the redshift-distance relation, to guide the learning process and to enhance the estimation accuracy. The integrated framework of the deep learning model and the physics-informed model will predict galaxy morphologies and redshifts with high accuracy and efficiency. This integration can enhance the accuracy of predictions and facilitate the interpretation of results.

## 6.4 Training and Validation

We will train the PINN using the acquired data, and optimise it to satisfy physical laws while minimising prediction errors. Validation will be performed using a separate dataset or cross-validation techniques.

## 6.5 Data Stratification and Interpretation

Implementing a stratification framework that combines morphological and spectral features to categorise celestial objects will be our next step. We will develop visualisation techniques to interpret the learned features and relationships within the network.

## 6.6 Evaluation and Validation

We will quantitatively evaluate the proposed methodology using appropriate metrics for classification accuracy, and stratification performance. Validation of the results of the integrated model will be done through comparisons with existing methods using both SDSS and Galaxy Zoo datasets. We will assess the model's performance using various evaluation metrics, considering the overall accuracy and its performance on different redshift bins.

### 6.7 Workflow Diagram

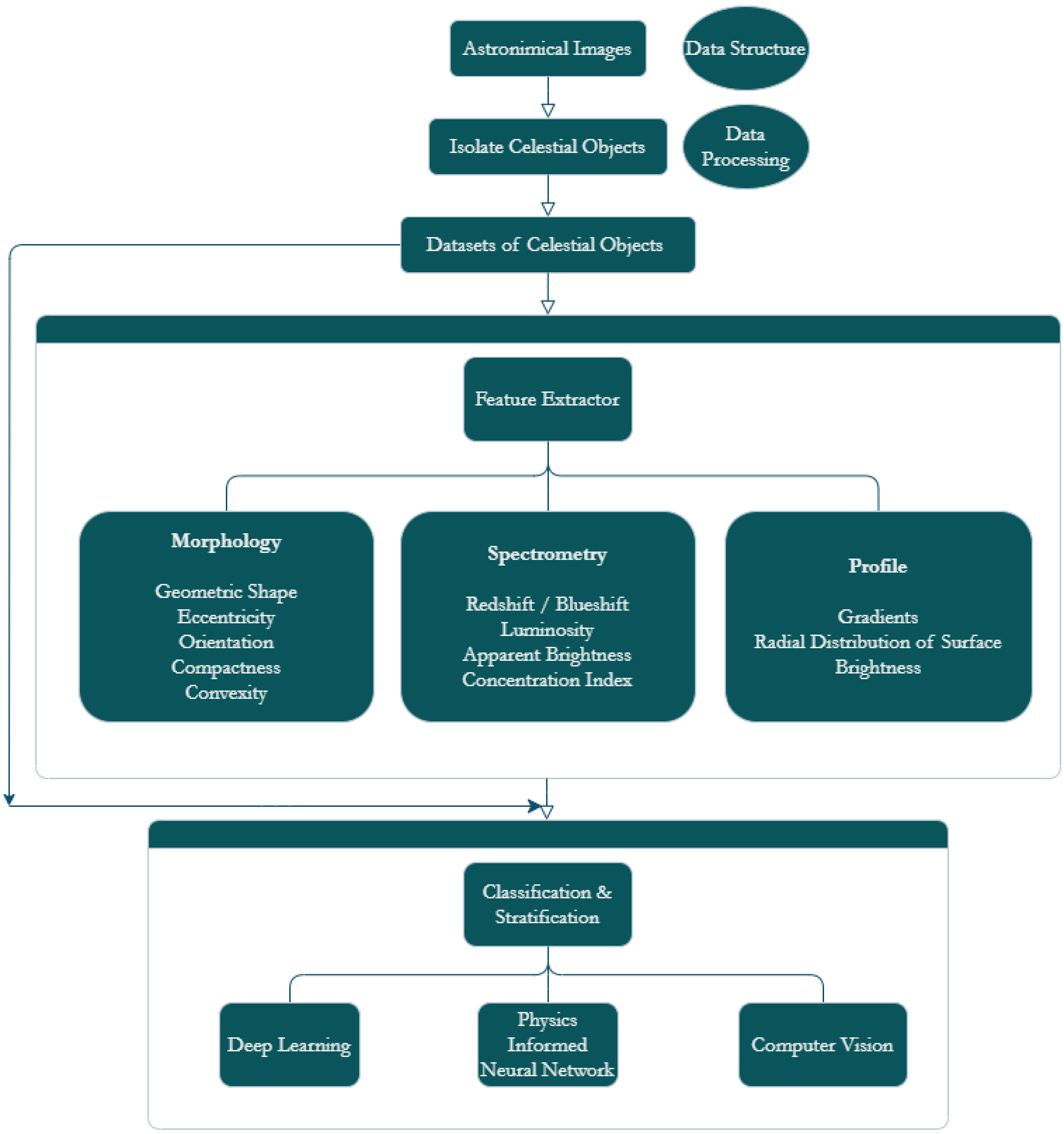


Figure 4: Workflow Diagram for the proposed methodology

# 7 Conclusion

This thesis proposes an approach that combines deep learning and physics-informed neural networks to address the challenge of morphological and spectral celestial stratification. By leveraging the power of these techniques, we aim to enhance our understanding of celestial bodies, their characteristics, and their underlying physical processes. Through the fusion of machine learning and domain-specific knowledge, this research can contribute significantly to our understanding of the universe's complexity and diversity. The combination of advanced computational techniques with astrophysical insights is expected to yield a comprehensive and accurate stratification of celestial objects. By automating and enhancing the analysis process, this approach has the potential to unlock new discoveries and deepen our understanding of the universe.

# References

[1] Sandro Ackermann, Kevin Schawinski, Ce Zhang, Anna K Weigel, and M Dennis Turp. Using transfer learning to detect galaxy mergers. *Monthly Notices of the Royal Astronomical Society*, 479(1):415–425, 2018.

[2] Yoshua Bengio et al. Learning deep architectures for ai. *Foundations and trends® in Machine Learning*, 2(1):1–127, 2009.

[3] Kirk D Borne. Astroinformatics: data-oriented astronomy research and education. *Earth Science Informatics*, 3:5–17, 2010.

[4] Richard O Duda, Peter E Hart, and David G Stork. Pattern classification second edition john wiley & sons. *New York*, 58:16, 2001.

[5] Sarvesh Gharat and Yogesh Dandawate. Galaxy classification: a deep learning approach for classifying sloan digital sky survey images. *Monthly Notices of the Royal Astronomical Society*, 511(4):5120–5124, 2022.

[6] Roberto E González, Roberto P Munoz, and Cristian A Hernández. Galaxy detection and identification using deep learning and data augmentation. *Astronomy and computing*, 25:103–109, 2018.

[7] Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, Gang Wang, Jianfei Cai, et al. Recent advances in convolutional neural networks. *Pattern recognition*, 77:354–377, 2018.

[8] James E Gunn, Walter A Siegmund, Edward J Mannery, Russell E Owen, Charles L Hull, R French Leger, Larry N Carey, Gillian R Knapp, Donald G York, William N Boroski, et al. The 2.5 m telescope of the sloan digital sky survey. *The Astronomical Journal*, 131(4):2332, 2006.

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[10] Edwin P Hubble. Extra-galactic nebulae. In *A Source Book in Astronomy and Astrophysics, 1900–1975*, pages 716–724. Harvard University Press, 1979.

[11] Eibe Frank Ian H. Witten. *Data Mining: Practical Machine Learning Tools and Techniques.* Morgan Kaufmann, 2016.

[12] He Jia, Hong-Ming Zhu, and Ue-Li Pen. Galaxy spin classification. i. z-wise versus s-wise spirals with the chirality equivariant residual network. *The Astrophysical Journal*, 943(1):32, 2023.

[13] Sotiris B. Kotsiantis. Supervised machine learning: A review of classification techniques. In *Informatica*, 2007.

[14] Ana Carolina Rodrigues Cavalcante Martinazzo, Mateus Espadoto, and Nina Sumiko Tomita Hirata. Deep learning for astronomical object classification: a case study. In *Proceedings*, 2020.

[15] Karen L Masters, Robert C Nichol, Ben Hoyle, Chris Lintott, Steven P Bamford, Edward M Edmondson, Lucy Fortson, William C Keel, Kevin Schawinski, Arfon M Smith, et al. Galaxy zoo: bars in disc galaxies. *Monthly Notices of the Royal Astronomical Society*, 411(3):2026–2034, 2011.

[16] Clara Moskowitz. How jwst is changing our view of the universe, Dec 2022.

[17] Johanna Pasquet, Emmanuel Bertin, Marie Treyer, Stéphane Arnouts, and Dominique Fouchez. Photometric redshifts from sdss images using a convolutional neural network. *Astronomy & Astrophysics*, 621:A26, 2019.

[18] CE Petrillo, CRESCENZO Tortora, S Chatterjee, G Vernardos, LVE Koopmans, G Verdoes Kleijn, NICOLA ROSARIO Napolitano, G Covone, LS Kelvin, and Andrew M Hopkins. Testing convolutional neural networks for finding strong gravitational lenses in kids. *Monthly Notices of the Royal Astronomical Society*, 482(1):807–820, 2019.

[19] Waseem Rawat and Zenghui Wang. Deep convolutional neural networks for image classification: A comprehensive review. *Neural computation*, 29(9):2352–2449, 2017.

[20] Allan Sandage. The classification of galaxies: Early history and ongoing developments. *Annu. Rev. Astron. Astrophys.*, 43:581–624, 2005.

[21] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[22] J Anthony Tyson. Optical synoptic telescopes: new science frontiers. In *Ground-based and Airborne Telescopes III*, volume 7733, pages 46–57. SPIE, 2010.

[23] International Astronomical Union. Naming of astronomical objects. `https://www.iau.org/public/themes/naming/`.

[24] Mike Walmsley, Chris Lintott, Tobias Géron, Sandor Kruk, Coleman Krawczyk, Kyle W Willett, Steven Bamford, Lee S Kelvin, Lucy Fortson, Yarin Gal, et al. Galaxy zoo decals: Detailed visual morphology measurements from volunteers and deep learning for 314 000 galaxies. *Monthly Notices of the Royal Astronomical Society*, 509(3):3966–3988, 2022.