AUT Journal of Electrical Engineering
10.22060/eej.2024.23499.5620

Classification of galaxy images using vision transformers

Ahmadreza Yeganehmehr [1], Hossein Ebrahimnezhad[2*]

[1]Master of Science, Faculty of Electrical Engineering, Sahand University of Technology, Tabriz, Iran
[2]Full Professor, Faculty of Electrical Engineering, Sahand University of Technology, Tabriz, Iran

Abstract:

The deep gaze of humans into the night sky, aimed at uncovering the mysteries of the cosmos using advanced telescopes, has generated an immense volume of data. The classification of stars and galaxies present in these images, due to the vast amount of data, was a highly time-consuming process for astronomers. As a result, the "Galaxy Zoo" citizen science project, in which volunteers participated in the classification of this data, was introduced by researchers, significantly reducing the classification time. In recent decades, the introduction of machine learning and deep learning models has accelerated the classification of galaxies, leading to the replacement of manual classification methods with automated machine-based approaches. Recently, Vision Transformers (ViTs) have emerged as a significant innovation in machine learning, demonstrating substantial potential in various research fields. These models have particularly garnered attention in the analysis, detection, and classification of images and computer vision, due to their ability to process large datasets and learn complex patterns. The need to develop advanced methods for the automatic analysis of galaxy images to increase detection and classification accuracy in the shortest possible time motivated the current research to classify galaxy images from the Galaxy10 DECaLS dataset into 10 classes with an accuracy of 99.85% using the ViT model. The results obtained have been promising in comparison with other competitors.

Keywords: Vision Transformer (ViT), Galaxy Classification Algorithms, Galaxy Morphology, Deep Learning, CNN.

* Corresponding Author: Prof. Hossein Ebrahimnezhad
Address: Computer Vision Res. Lab., Faculty of Electrical Engineering, Sahand University of Technology, Tabriz, Iran. Email: ebrahimnezhad@sut.ac.ir.

10.22060/eej.2024.23499.5620

1. Introduction

Human advancement in constructing telescopes with varying optical wavelengths over the past few decades has revealed astonishing images of the cosmos. However, due to the vastness of space, only a tiny fraction of it has been imaged so far. The immense volume of images captured by telescopes, aimed at identifying celestial objects in the background of the universe, has made the task of recognizing and classifying them increasingly difficult and time-consuming. In 1926, individuals like Edwin Hubble attempted to identify luminous objects similar to stars in images under the name of galaxies. Hubble realized that galaxies are abundant in space, each hosting billions of stars and planets. His research on galaxies showed that, based on their morphology, each galaxy exhibits a distinct appearance, leading him to classify galaxies into three categories: elliptical, spiral, and irregular. He also identified another group of galaxies with unusual and unconventional shapes and appearances. Building on Hubble's efforts to classify galaxies, other researchers and scientists have since defined six classes for observable galaxies (de Vaucouleurs, 1959 [1]), which continue to be used in current research and the classification of cosmic observations.

Galaxy morphology has become a significant area of focus for researchers in the recognition and classification of galaxies, leading to the introduction of various techniques such as visual crowd-sourced classification (Lintott [2]), automated computational methods including machine learning (Cancesis [3]; Lutz [4]; Freeman [5]), deep learning (Vega-Ferrero [6]; Walmsley [7]; Gupta [8]; Farias [9]; Barchi [10]; Domínguez [11]; Banerji [12]), and feature extraction (Ferrari [13]; Abdelaziz [14]; Shamir [15]; Marin [16]; Abd el Aziz [17]; González [18]; York [19]; Gardner [20]; Grogin [21]; M. J. Baumstark [22]). Manual classification of a large number of galaxies poses challenges such as low accuracy and speed. Over the past decade, machine learning and deep learning-based classifiers have achieved high accuracy in classifying galaxies. In some studies, these methods have even led to the identification of new galaxy classes (Kyle W. Willett [23]; Paulo Henrique Barchi [24]; Soroush Gharaat [25]; Koketso Mohale [26]; Xinrui Tan [27]). However, the implementation of these models is often complex and challenging, requiring significant computational resources that can be costly (Dieleman [28]; Nour Eldin M. Khalifa [29]; Jia Ming Dai [30]; Joshua Yao-Yu Lin [31]; Ji Cao[32]). Recently, Google developed a new architecture called Vision Transformer (ViT) for image classification.

In this research, we utilized the ViT model for galaxy morphology classification. Our results indicate that ViT can exhibit competitive performance compared to CNN and other methods, especially showing significant capability in classifying smaller and fainter galaxies. Given the promising initial results, we believe that the ViT architecture could serve as a powerful tool for morphological classification of galaxies in next-generation surveys.

2. Dataset

This project utilizes the AstroNN[2] dataset, which is derived from the DESI Legacy Imaging Survey, with labels sourced from the Galaxy Zoo project. The Galaxy10 DECaLS dataset includes 17,736 color images of galaxies, each with dimensions of 256×256 pixels, captured in the g, r, and z bands. These images are categorized into 10 different classes, obtained from the DESI Legacy Imaging Surveys. The dataset file includes columns for images with dimensions (256, 256, 3), along with additional metadata including the angle (ans), right ascension (ra), declination (dec), redshift, and pixel scale (pxscale, measured in arcseconds per pixel). The dataset is categorized as follows:

- Class 0 (1,081 images): Disturbed Galaxies
- Class 1 (1,853 images): Merging Galaxies
- Class 2 (2,645 images): Round Smooth Galaxies
- Class 3 (2,027 images): In-between Round Smooth Galaxies
- Class 4 (334 images): Cigar Shaped Smooth Galaxies
- Class 5 (2,043 images): Barred Spiral Galaxies
- Class 6 (1,829 images): Unbarred Tight Spiral Galaxies

[2] https://astronn.readthedocs.io/en/stable/galaxy10.html

10.22060/eej.2024.23499.5620

- Class 7 (2,628 images): Unbarred Loose Spiral Galaxies
- Class 8 (1,423 images): Edge-on Galaxies without Bulge
- Class 9 (1,873 images): Edge-on Galaxies with Bulge

These diverse classifications enable us to rigorously test the ViT model for accurate morphological classification of galaxies based on real and complex data. The dataset can also be manually compiled using sources such as the Hubble image set and other similar digital surveys. After preprocessing, these data can be used to train various models. An example of the dataset images is shown in Figure (1).

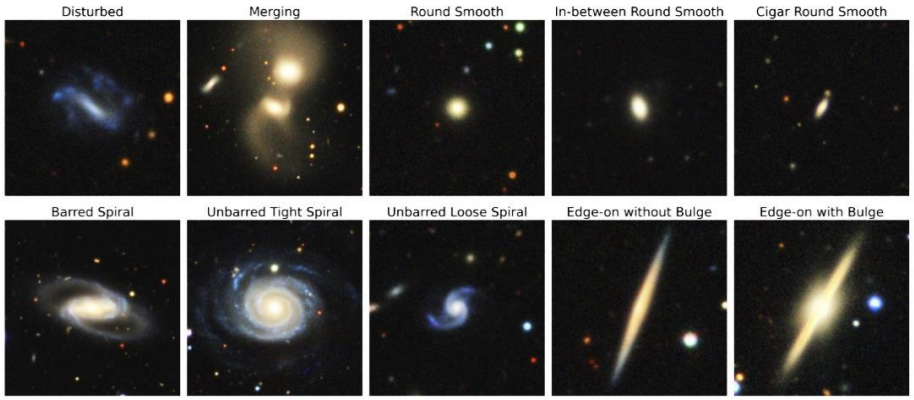

Figure (1): Sample images of each class from Galaxy10 DECaLS Dataset

3. Research Methodology and Structure

In this section, we will first introduce the Vision Transformer (ViT) model to provide an understanding of its architecture, allowing us to explain the proposed method in detail. In the methodology section, using the ViT code available on Kaggle[3], we will perform the classification of a galaxy dataset by modifying the code's parameters. The selected model will be evaluated by solving the complex and challenging problem of classifying 10 galaxy classes. In Section 4, the results will be compared with other models and techniques for galaxy image classification.

3-1 Introduction to Vision Transformer (ViT)

Transformers, first introduced by Vaswani et al. in 2017 [33], have revolutionized natural language processing (NLP). These models utilize neural network architectures that are not dependent on convolution and can identify long-term dependencies and complex relationships in sequential data. One of the key innovations of transformers is the attention mechanism, which allows them to assign different weights to the relative importance of words in a sentence. Transformers have not only made a significant impact in NLP but have also achieved success in computer vision.

Vision Transformers (ViTs) use a self-attention mechanism to better identify global dependencies in images, leading to improvements in tasks such as image classification and object detection. Additionally, these technologies enable cross-modal learning and can be applied to tasks that involve the combination of text and images, such as generating captions for images and visual question answering.

The success of transformers in NLP has significantly influenced the computer vision research community, leading to the development of transformer-based models for vision tasks. Models such as

[3] https://www.kaggle.com/code/ahmadalijamali/ultrasound-vision-transformer-classification/notebook

DETR[4], ViT[5], DeiT[6], and Swin[7] have rapidly grown in prominence and are recognized for their advancements in object detection, image classification, and improved image understanding. These models have achieved significant improvements in applying transformers to computer vision tasks. ViT has demonstrated competitive and sometimes superior performance compared to traditional CNN[8] models, especially when large training datasets are available.

3-2 Methodology

Applying innovative techniques and models for the simultaneous classification of galaxy types with high accuracy and speed, compared to previous methods where classification did not reach 100%, has been challenging, and there is a strong need to address these shortcomings. In this study, we aim to classify 10 types of galaxies, which include Disturbed, Merging, Round Smooth, In-between Round Smooth, Cigar-shaped Smooth, Barred Spiral, Unbarred Tight Spiral, Unbarred Loose Spiral, Edge-on without Bulge, and Edge-on with Bulge galaxies, using the ViT model. Our research into image processing models and algorithms has revealed the successful performance of the ViT model in terms of image quality, noise, and background lighting, making ViT a reliable and efficient method for this experiment. Figure (2) illustrates the architecture of the ViT model. We will further elaborate on the enhanced sections.

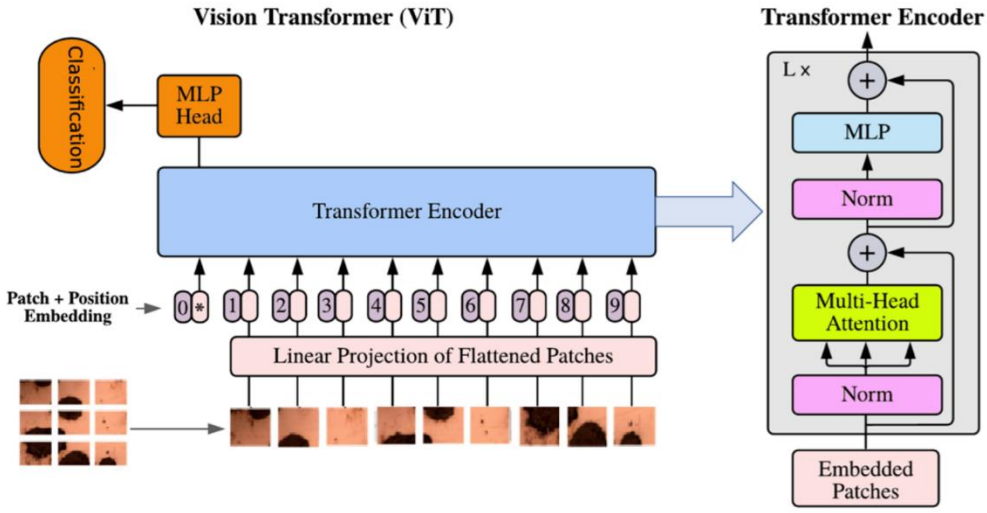

Figure (2): Vision Transformer (ViT) model architecture [37]

3-2-1 Architecture Description

The ViT (Vision Transformer) model, as shown in Figure 2, differs from models like DeTr, Swin, and BeTr[9]. ViT employs a method similar to conventional transformers, where the image is first divided into smaller pieces called "patches," and these patches are then fed as input to the transformer network. This model works directly on visual data without utilizing structures like CNNs. In contrast, the DeTr model uses a combination of CNN and transformers for object detection. The Swin Transformer is an optimized variant of the transformer, where the input patches are processed in a "swin" manner to

[4] Detection Transformer*
[5] Vision Transformer
[6] Data-efficient Image Transformers
[7] Shifted Window Transformer
[8] Convolutional Neural Network
[9] Better Transformer

accommodate different scales. The BeTr model introduces improvements over previous transformer models and is designed for various applications. Due to the complexity of the structure and texture of galaxies in images, their detection and classification are challenging. Given the results obtained from datasets across different fields by ViT, employing this model is a smart choice for classifying galaxies based on morphology.

3-2-2 Preprocessing Stage

Before beginning the classification process, the dataset images require preprocessing. Preprocessing is a fundamental step in data processing and preparation for machine learning and deep learning. Without proper preprocessing, the data may contain unnecessary information, noise, and excessive dimensions, which can reduce model performance. Preprocessing helps the model receive data in a more suitable and higher-quality format, allowing it to achieve greater efficiency, ultimately leading to improved performance, faster training, and higher model accuracy. During the preprocessing stage, all images were resized to a standard dimension of 256 x 256 pixels. This optimal size was chosen to preserve image details while facilitating processing by the model. Pixel values were normalized within a standard range to enable the model to recognize patterns more quickly and improve accuracy.

To enhance model accuracy in identifying and classifying galaxy classes, a galaxy dataset comprising 17,736 galaxy images across 10 classes was utilized. We used 80% of the dataset for model training, and the remaining 20% was reserved for model testing. This division helps the model generalize better by exposing it to a set of unseen data. For morphological feature extraction, this model uses visual features to recognize different types of galaxies. The data is labeled through the Galaxy Zoo project and includes information such as angle (ANS), right ascension (RA), redshift, and pixel scale. To adapt to ViT structure, images are divided into small blocks or "patches." These patches are sent as separate inputs to the model, facilitating the detection of local and global features across various scales.

To improve performance, different parameters such as patch size, the number of transformer layers, attention heads, and learning rate were fine-tuned. These settings were applied to increase accuracy and reduce computational time. After completing the preprocessing and configuration stages, the data was prepared for training the ViT model, and its performance was evaluated in the testing phase. Accuracy and loss charts over training and validation epochs assist in analyzing model efficiency. These preprocessing steps have helped optimize and enhance the model's performance, enabling it to classify galaxies with high accuracy.

3-2-3 Modified Parameters

To accurately identify the morphology of each galaxy, some model parameters required rewriting and fine-tuning. Table 1 presents the modifications made to the model's parameters.
These adjustments were crucial for adapting the model to the task of galaxy classification, ensuring that it could effectively handle the distinctive features of astronomical images. The modifications included changes in learning rate, batch size, patch size, and other parameters to optimize performance for this specific dataset.

Table (1): Reset parameters of ViT model

Changes applied	Before	After
Increasing the size of patches **(**patch_size**)**	4	16
Increasing the number of transformer layers **(**n_transformers**)**	1	2
Increasing the number of attention heads **(**n_heads Attention**)**	2	4
Increasing the number of MLP units **(**mlp_units**)**	[2048 ,1024]	[4096 ,2048]
Decreasing the learning rate	0.001	0.0001
Increasing the number of epochs	10	50
Reducing the input size of images	(256,256,3)	(120,120,3)

10.22060/eej.2024.23499.5620

Each of these changes has been designed to enhance the model's ability to extract and process complex galaxy features. These adjustments reflect a strategic approach aimed at improving the model's learning capacity and complexity, ultimately applied to increase its performance. Below is an analysis of each modification:

- **Increasing Patch Size (from 4 to 16):** By increasing patch size, the model can focus on identifying larger and more general patterns in images rather than small details. These patterns can represent the overall structure of galaxies, such as classifying them into spiral, elliptical, and other types. This enables the model to learn on a more holistic level, enhancing the likelihood of high accuracy. Additionally, increasing patch size reduces the number of patches the model must process, which can reduce computational load.
- **Number of Transformer Layers (from 1 to 2):** Adding another layer increases the model's depth, allowing it to identify more complex and hierarchical relationships between patches. This depth increase can enhance the model's ability to generalize and recognize more intricate patterns, which is useful in distinguishing subtle differences between galaxy types.
- **Increasing the Number of Attention Heads (from 2 to 4):** Increasing the number of attention heads allows the model to simultaneously attend to various aspects of the input, improving its ability to identify diverse and nuanced patterns within each image. This is particularly effective for images where multiple galaxy features may require attention.
- **Increasing MLP Layer Units (from [2048, 1024] to [4096, 2048]):** This adjustment boosts the model's capacity to learn complex, nonlinear transformations. This enhancement is particularly valuable for classification tasks that require a rich, complex feature representation, such as distinguishing different galaxy types. These layers act as a classifier booster, maximizing accuracy.
- **Decreasing Learning Rate (from 0.001 to 0.0001):** Lowering the learning rate results in more stable convergence and reduces the risk of large jumps in optimization. This is balanced by the model's increased complexity and helps guide it toward a more optimal point.
- **Increasing the Number of Epochs (from 10 to 50):** Extending the number of epochs allows the model more time to learn from the data, potentially improving accuracy. However, the risk of overfitting increases, so monitoring validation error during training is essential.
- **Reducing Input Image Size (from (256,256,3) to (120,120,3)):** Reducing input size may seem counter-intuitive at first but can lead to faster training and reduced overfitting, as it allows the model to focus more on key, general galaxy features rather than becoming caught up in finer details. This can be especially beneficial if the target galaxy structures are large enough to be recognizable at lower resolutions. It can also prevent unnecessary learning and improve the model's speed and learning accuracy.

The adjustments made to the ViT model's parameters also contributed to a reduction in computation time. The program was executed on a laptop with the following hardware specifications: an Intel(R) Core(TM) i7-1065G7 processor, 8 GB of RAM, and an Nvidia GeForce MX330 2GB GPU, within the Anaconda Navigator 2.4.2 environment. The operation was completed in 2,107 seconds. With more powerful processors, the computation time could be expected to decrease significantly. Due to its complexity and high computational demands, this model takes significant time on standard hardware. Access to powerful GPUs, such as the V100-SXM2-32GB, can provide better processing performance for this model. In some ViT-based models, like EfficientViT, computational complexity is reduced using techniques like Linformer, which provides better speeds even on moderate hardware compared to the classic ViT. Although this model does not reach the processing speeds of MobileViT, it requires less processing time than ViT and performs well on average GPUs.

MobileViT, with its lightweight structure and high processing speed, is designed specifically for use in resource-limited devices that require fast processing. On servers like Nvidia V100-SXM2, MobileViT needs less than 45 milliseconds per step, making it ideal for rapid applications and lightweight environments. In contrast, CvT, by combining CNNs with transformers, achieves even faster processing than classic ViT on powerful servers like the V100. This model ranks well in both speed and accuracy

categories, especially on high-performance computing servers capable of processing larger, more complex models.
ViT achieves the best accuracy on powerful GPUs, but its processing speed is considerably slower on lightweight hardware. Despite the limitations of our tools and resources, we were able to achieve promising and acceptable results, which will be discussed and analyzed in Section 4.

4. Results

This section presents the results obtained from galaxy classification using the ViT model. The analytical and processing capabilities of the ViT model were notably impressive and satisfactory, demonstrating acceptable results and outperforming previous methods and models in terms of accuracy and speed in astronomical contexts. To evaluate the performance of the ViT model, we utilized a dataset comprising 17,736 galaxy images. From this, 14,189 images were allocated for training, and 3,547 images were reserved for testing. Figure (3) illustrates the distribution of images across the 10 galaxy classes used for testing the model. The results show that the ViT model provided substantial improvements in both accuracy and processing speed compared to traditional methods. The model's performance across different galaxy classes was evaluated based on various metrics, and the findings underscore the effectiveness of the ViT model in handling and classifying complex astronomical data. In the following sections, we will delve into a detailed analysis of these results and compare them with other classification techniques and models in the field of astronomy.

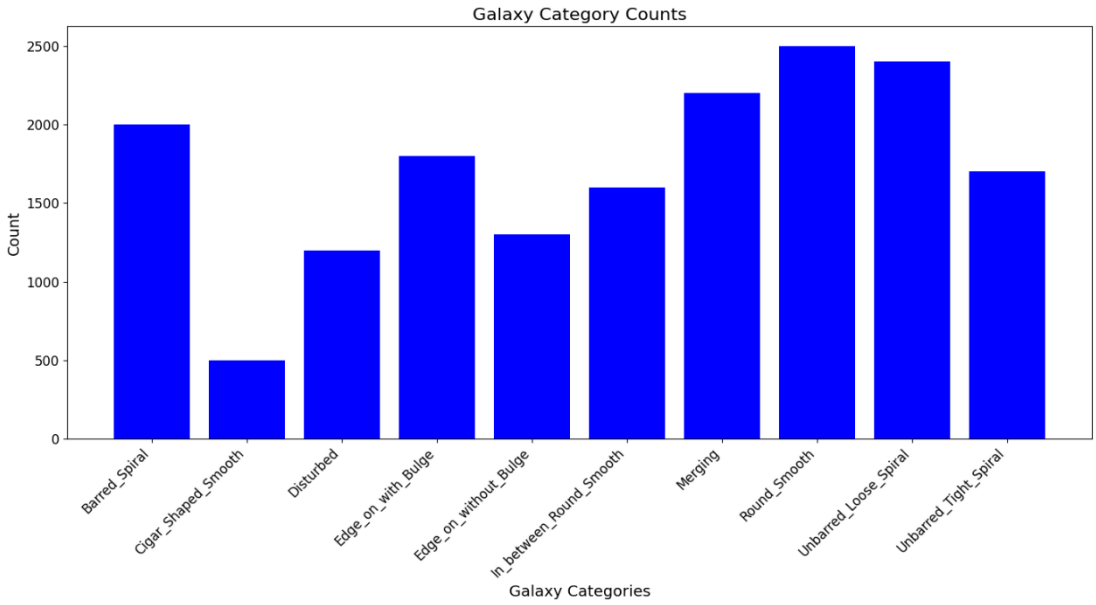

Figure (3) Frequency diagram of dataset of galaxy classes images

In Figure (4), the results indicate a perfect classification accuracy of 100% for six galaxy classes: edge-on without bulge, barred spiral, Disturbed, Unbarred Tight Spiral, Cigar Shaped Smooth, and In-between Round Smooth Galaxies. These results demonstrate that the ViT model achieved an overall success rate of 99.85%. Out of the 3,547 test images, the model accurately classified 3,542 galaxies, with only 5 images being misclassified, as shown in Figure (5).

10.22060/eej.2024.23499.5620

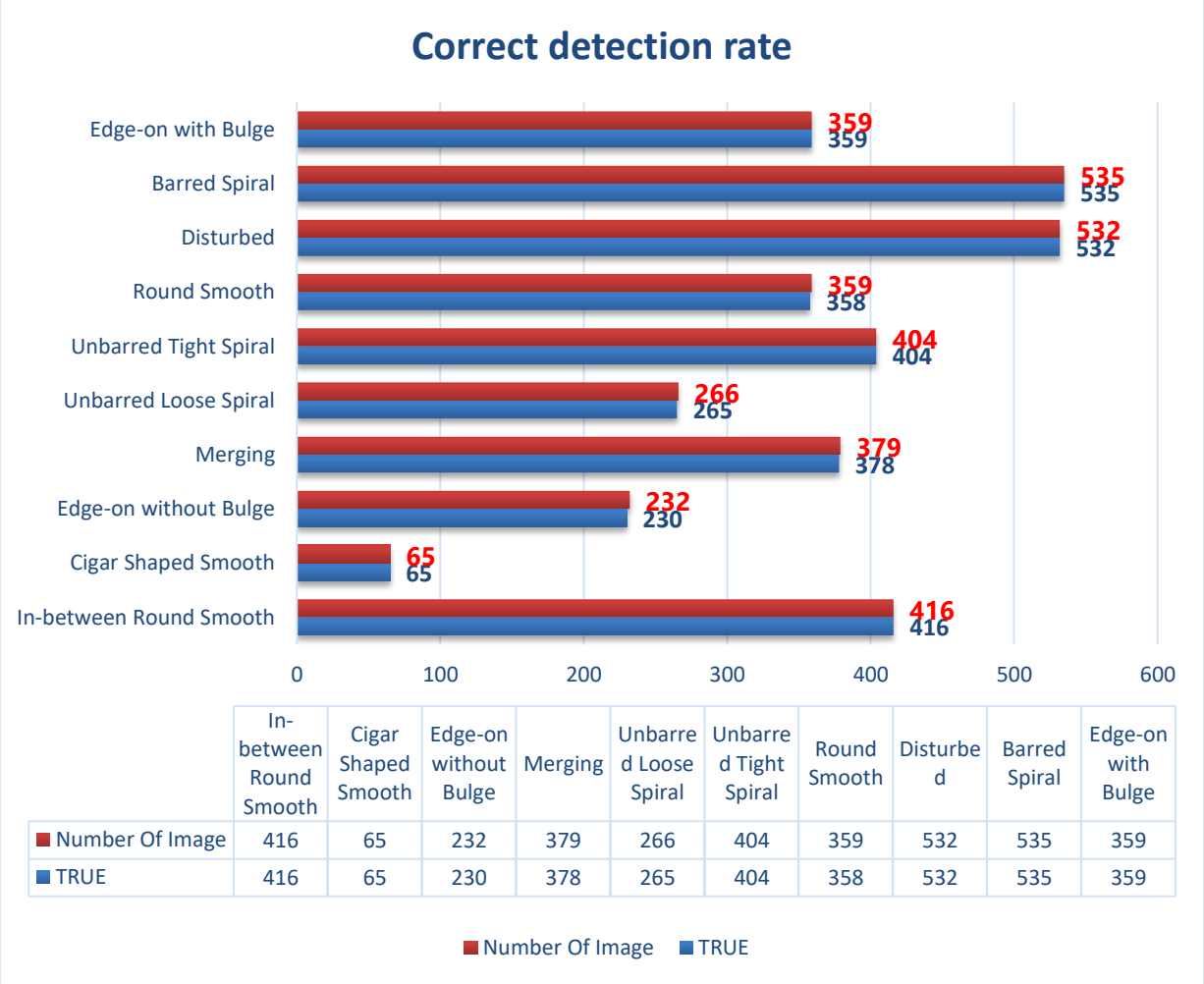

Figure (4) The number of correct classification of each class, red color (number of images of each class) - blue color (number of correct recognition)

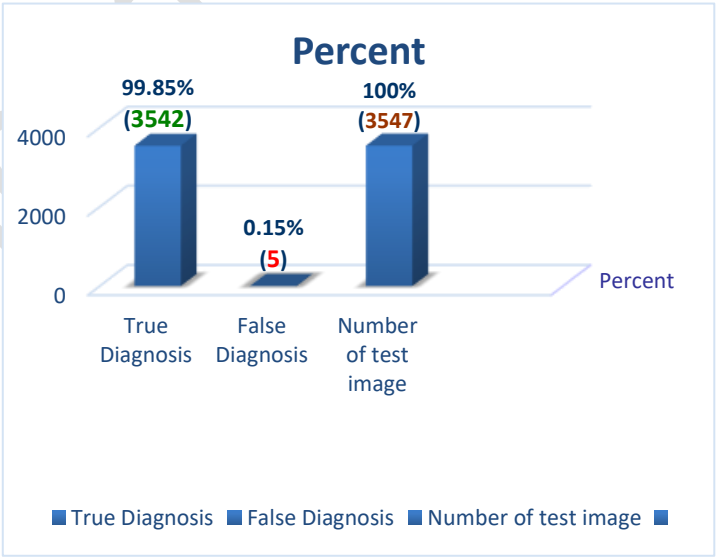

Figure (5) The success rate of the ViT model for correct and incorrect classification of galaxies from a total of 3547 test images.

The images of galaxies within each class can be challenging to classify due to stellar scattering around the arms or edges and the presence of dispersed stars in the background. For example, the classification

of irregular and merging galaxies, as well as edge-on galaxies with bulges, presents a significant challenge due to their diverse shapes and structures. Training a model to accurately classify these types of galaxies is notably more complex compared to other classes. This complexity arises from the unique and often overlapping features that complicate the detection and differentiation of these galaxy types.

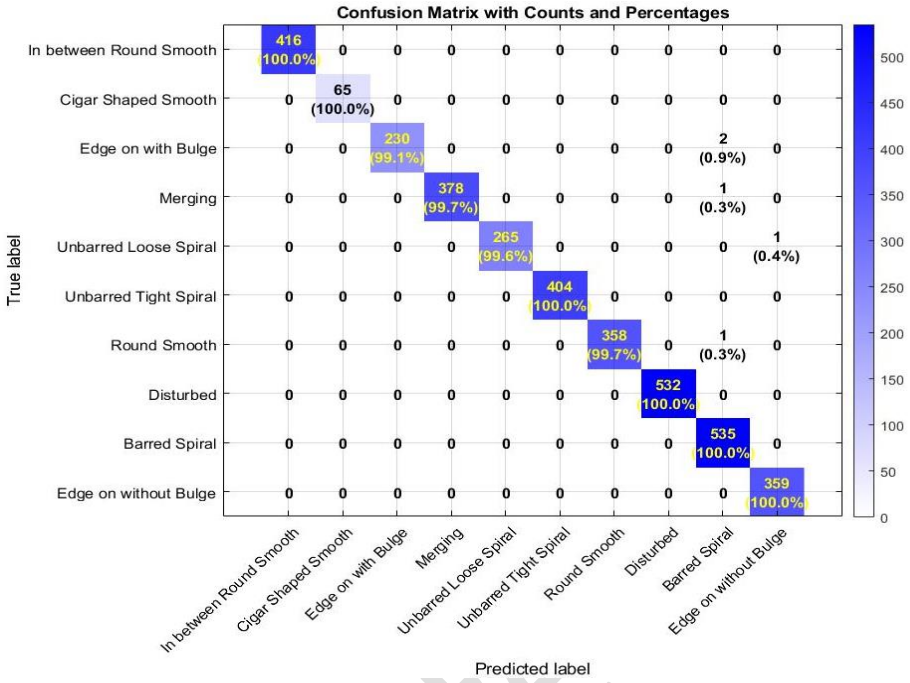

Figure (6) confusion matrix for classification of 10 galaxy class

As shown in the confusion matrix in Figure (6), the ViT model correctly classified all images in six galaxy classes. However, in four other classes, it made errors in categorizing five images. These errors include:

- Two instances of misclassifying edge-on with bulges galaxy as barred spirals.

- One instance of misclassifying a merging galaxy as a barred spiral.

- One instance of misclassifying an unbarred Loose spiral galaxy as an edge-on without bulge.

- One instance of misclassifying a Round Smooth galaxy as a barred spiral.

Given the complexity of galaxy morphologies and the number of test images, the classification performed by the ViT model has been highly successful, demonstrating the model's strong capability in handling complex image classification tasks. It is also possible that these five misclassifications could be due to the model's higher precision compared to human classification, where these images might have been misclassified by humans. Some factors can lead to misclassifications in the model. For instance, certain galaxy classes are structurally and visually similar. Spiral and irregular galaxies, for example, can appear very similar in certain images, especially when taken from different angles. This similarity may cause the ViT model to misclassify these classes. Variations in galaxy brightness due to changes in instrumentation or observational conditions can also prevent the ViT model from accurately identifying key features, leading to incorrect classification of some classes.

The ViT model requires a large number of images for each class for effective learning. If some classes are underrepresented, the model may struggle to recognize them accurately. In the Galaxy10 DECaLS dataset, certain classes may be underrepresented or display physical differences across samples. In many galaxy images, the scale and viewing angle can vary, presenting another challenge for the ViT

model. Galaxies viewed from oblique or edge-on angles are generally more complex than those viewed head-on, making them prone to misclassification. Finally, ViT models are primarily sensitive to local patterns and short-range information, while some galaxy features require a broader view and an understanding of structural characteristics across the whole image. This may be another factor contributing to errors in classes with complex, composite structures. In Table 2, the performance of the model for 10 classes is presented. As seen in the table, classes 1, 2, 6, and 8 have achieved a value of 100% across all metrics (Accuracy, F1-Score, Recall, and Precision). This indicates that the model has successfully identified these classes without any errors. This excellent performance may be attributed to the distinct and prominent features of these classes, which make them easy for the model to recognize. Class 3, with an accuracy of 99.944% and an F1-Score of 99.567%, shows that this class generally performs well; however, since Recall (99.138%) is lower than Precision (100%), it means that the model has misclassified some samples into other classes (false positives). Class 9 also exhibits high performance with an accuracy of 99.887% and an F1-Score of 99.628%, similar to Class 3, but with a slight drop in Precision (99.258%), which may indicate the presence of false positives.

Classes 4, 5, and 10 have accuracies close to 100%, with Recall at the highest level and slightly lower Precision. For example, Class 10 has a Recall of 100% and a Precision of 99.722%. This indicates that the model is proficient at identifying positive samples, but there are still false positives, likely due to visual similarities with other classes. In some classes, such as Classes 3 and 9, Precision is lower than Recall, suggesting the existence of false positives, possibly due to visual similarities between classes. For instance, irregular galaxies may be easily mistaken for other types of galaxies. Given that most classes have high accuracy and only a few are affected by false positives, it can be concluded that the ViT model performs exceptionally well in identifying galaxies.

Table 2. Classification Performance of ViT Model.

Class	Precision	Recall	F1_Score	Accuracy
1	100	100	100	100
2	100	100	100	100
3	100	99.138	99.567	99.944
4	100	99.736	99.868	99.972
5	100	99.624	99.812	99.972
6	100	100	100	100
7	100	99.721	99.861	99.972
8	100	100	100	100
9	99.258	100	99.628	99.887
10	99.722	100	99.861	99.972

Figure 7 illustrates the changes in the model's training accuracy and validation accuracy over 50 training epochs. The horizontal axis represents the number of epochs, and the vertical axis shows accuracy. The graph demonstrates that, in the initial epochs, both training and validation accuracies increase rapidly, indicating the model’s swift learning from patterns in the data. After these initial epochs, the model's accuracy gradually approaches a value close to 1.0 (or 100%), with the model reaching near-maximum accuracy around epoch 15. This suggests that the model is well-trained and almost saturated. Both validation accuracy (orange) and training accuracy (blue) remain at a stable level with minimal gap between them. This indicates that the model has not overfitted and has successfully retained its generalization ability. The final loss value is low (0.0093), indicating minimal error in predicting the correct classes.

10.22060/eej.2024.23499.5620

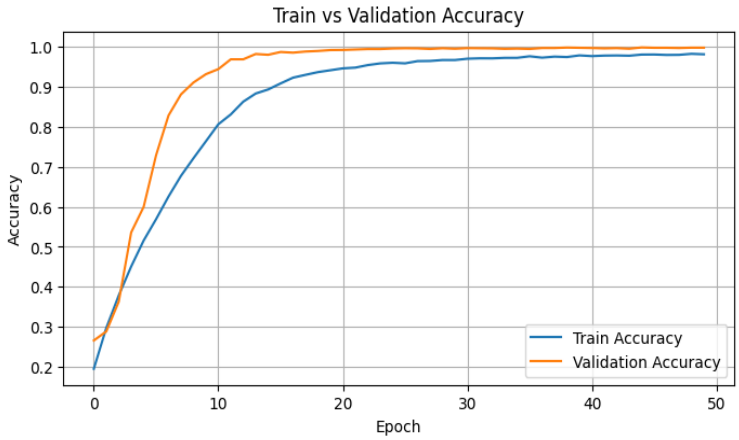

Figure 7: Trend of training accuracy and validation accuracy over training epochs.

The graph shows that, in many epochs, validation accuracy (orange) exceeds training accuracy (blue). This may be due to the use of data augmentation techniques during training, such as rotation, cropping, color changes, or flipping. These augmentations add variations to the training set, making predictions more challenging for the model. Consequently, training accuracy is slightly lower than validation accuracy, as the validation set is generally presented without such augmentations. Additionally, if the model employs optimization techniques like mini-batch SGD, the results of each epoch might vary due to the randomness in batch selection. These variations can cause small fluctuations in recorded accuracies and, in some cases, validation accuracy to exceed training accuracy.

Occasionally, minor differences between training and validation samples may lead to this outcome. For instance, if the validation data distribution is simpler than that of the training data, the model might perform better on the validation set. This scenario is often observed in complex models that utilize various regularization and augmentation techniques and generally does not indicate a problem. The slight discrepancy observed here reflects the model's good generalization ability, showing that the model has not become overly reliant on the training data. This graph highlights that the model performs exceptionally well in classifying galaxy classes, achieving high accuracy. Additionally, the balanced training and validation accuracy indicates that the model is likely to perform well in real-world scenarios.

Figure (8) shows the validation loss curve for the model. This chart is typically generated during the training of deep learning models and reflects the model's performance on the validation set throughout the training process. The validation set is a subset of data used to evaluate the model during training to prevent overfitting and estimate the model's overall performance.

The x-axis of Figure (8) represents the number of training epochs, while the y-axis shows the mean validation loss. Initially, the model learns rapidly, with the loss decreasing significantly. This indicates that the model is effectively capturing the primary patterns in the data. After several epochs, the loss stabilizes near a constant value, suggesting that the model's performance has converged and further training does not lead to substantial improvements.

If the validation loss were to increase, it would signal overfitting, meaning that the model has become too tailored to the training data and is losing its ability to generalize to new, unseen data. The provided chart demonstrates good model performance, as the validation loss has reached a minimum and accuracy is maximized. The absence of a significant increase in validation loss and the stabilization of accuracy indicate that the model has not overfitted and has reached a converged state. Continuing training beyond this point is unlikely to yield significant performance improvements.

10.22060/eej.2024.23499.5620

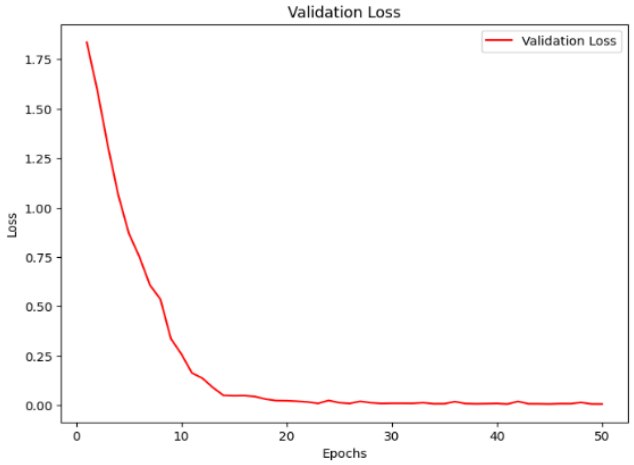

Figure (8) loss diagram of model validation data

A review of previous literature indicates that automated methods for galaxy classification using machine learning and neural networks have been extensively utilized due to their ability to analyze large volumes of data and extract complex features. In recent years, several successful models and methods for high-accuracy galaxy classification have been introduced. Table 3 presents a comparison of some of these models along with their performance. While the accuracy of the models listed in the table is acceptable considering the number of classes and test images, our model demonstrates superior performance compared to other competitors. This highlights the effectiveness of Vision Transformers (ViTs) as a reliable and successful approach for classifying galaxy images, particularly those of small, faint galaxies with high noise levels.

Table (3) Comparison of the success rates of successful models and methods in galaxy classification (in percentage)

Paper Title	Year	Number of galaxy class	Dataset	Accuracy (%)
Deep Galaxy: Classification of Galaxies based on Deep Convolutional Neural Networks **[29]**	2017	3	EFIGI	97.27
Galaxy detection and identication using deep learning and data augmentation **[18]**	2018	5	Galaxy Zoo	81
Galaxy Morphology Classification with Deep Convolutional Neural Networks **[30]**	2018	5	Galaxy Zoo	95.20
Galaxy Classification: A deep learning approach for classifying Sloan Digital Sky Survey images **[25]**	2021	10	Galaxy Zoo	84.73
Galaxy Morphological Classification with Efficient Vision Transformer **[31]**	2021	8	GZ 2	80.55
Automated detection and classification of galaxies based on their brightness patterns **[35]**	2022	3	EFIGI	97.2
Galaxy Morphological Classification with Deformable Attention Transformer **[36]**	2022	6	AMIGA & EFIGI	94
Spiral-Elliptical automated galaxy morphology classification from telescope images **[22]**	2023	2	SDSS	95.5
Accurate and efficient galaxy classification based on mobile vision transformer **[27]**	2024	10	Galaxy10 DECals	87
Galaxy morphology classification based on Convolutional vision Transformer (CvT) **[32]**	2024	5	GZ 2 & CANDLES	98
Proposed method	-	10	Galaxy10 DECals	**99.85**

As shown in Table 3, our improved ViT model outperforms other ViT-based models such as CvT, MobileViT, and EfficientViT in classifying galaxy images. This advantage is primarily due to the specialized ViT architecture and fine-tuning of its parameters.

The model's architecture utilizes multi-head attention and direct processing of image patches, which facilitates the identification of both global and distributed features within galaxy images. Additionally, the use of more transformer layers and MLP units allows for learning more complex galaxy features. Careful model tuning, including a lower learning rate and a higher number of epochs, enables the model to train with more stable convergence and achieve higher accuracy. Consequently, this model is ideal for projects that require high accuracy and significant computational resources, effectively detecting complex and distinct galaxy features.

In contrast, the EfficientViT model employs the Linformer method and low-rank matrices to reduce computational complexity, leading to faster processing and lower resource consumption, although it may affect the accuracy in learning certain finer details. This model is trained with a higher learning rate and more epochs, making it well-suited for processing large and imbalanced datasets like Galaxy Zoo 2. Thus, EfficientViT is more applicable for large-scale projects or systems with limited resources.

The MobileViT model combines MobileNet and ViT, utilizing lightweight MobileNetV2 blocks and transfer learning from ImageNet. With an accuracy of 87.12% on the Galaxy10 DECaLS dataset and an inference speed of 45 milliseconds per step, it is a suitable choice for low-power devices and applications requiring quick processing. Due to its unique architecture, MobileViT is better suited for applications requiring speed and efficient resource usage. Overall, our ViT model, thanks to its specialized architecture and fine-tuning, is an ideal choice for projects requiring high accuracy and substantial computational power. On the other hand, EfficientViT and MobileViT models are optimized respectively for large-scale processing and fast deployment on low-power devices.

5. Discussion

As shown in Table 3, the methods employed, considering the number of galaxy classes, have been able to provide satisfactory results in testing the detection and classification power. The use of neural network models in galaxy classification offers several advantages. Deep neural networks, due to their multi-layered structure and ability to learn complex features from data, perform highly accurately in galaxy classification. These models can identify various features of galaxies, such as shape, brightness, and color, which may not be perceptible to the human eye. Furthermore, they can learn hidden patterns and relationships from the training data provided to them. This capability allows them to perform well even on new and unseen data.

Neural networks are capable of processing large volumes of image data in parallel, which is crucial for applications like classifying millions of galaxy images. Galaxies exhibit complex features that might not be immediately apparent. Deep neural networks can detect highly intricate features in data, resulting in more accurate classification compared to traditional methods. They can be employed for various tasks such as detecting and classifying types of galaxies (e.g., elliptical, spiral, and irregular galaxies) and even predicting other physical properties. Neural networks also perform well with incomplete or noisy data due to their ability to generalize and extract useful patterns from imperfect data. Neural networks are adaptable and modifiable, meaning that models can be optimized for improved performance in specific applications, such as classifying particular types of galaxies.

Overall, the use of neural networks, especially Vision Transformers (ViTs), in galaxy classification offers numerous advantages due to their ability to learn complex patterns and generalize to new data. This results in improved accuracy and efficiency in this field. The use of ViTs in galaxy classification has several key advantages that make this model a strong choice compared to traditional methods like Convolutional Neural Networks (CNNs). For example, ViT relies on the Transformer architecture rather than local filters like CNNs. This allows it to effectively model long-range dependencies between different pixels in an image. This feature is highly beneficial for classifying galaxies, which often have

complex patterns and large-scale structures. ViTs are also easily scalable and can be adapted to analyze images at various resolutions, which is particularly useful for high-resolution galaxy images.

ViT can leverage pre-trained models on large datasets like ImageNet, enabling the model to perform well without requiring an extensive amount of training data. This is particularly valuable for galaxy classification, where labeled data might be scarce. Unlike CNNs, which require complex filter designs and adjustments, ViT automatically identifies and models important image features using the Attention Mechanism. This simplifies the model development process and reduces the need for specialized expertise in designing neural networks. ViT generally performs better in classifying complex data like galaxy images, which contain extensive details, due to its robust pattern modeling capabilities.ViT is inherently more resilient to changes in scale and orientation of images because its architecture relies on the entire image and the relationships between tokens, rather than local filters. This feature is especially useful for classifying galaxies, which can vary in size and orientation. Overall, utilizing ViT in galaxy classification can lead to significant improvements in accuracy and efficiency due to its superior modeling power, scalability, and reduced data requirements.

6. Conclusion

This research represents a significant contribution to the field of galaxy morphology classification, particularly by leveraging the power of machine learning, deep learning, and Vision Transformers (ViTs). Below is a summary of the key points and analytical insights:

- **Objective**:

The aim was to enhance galaxy morphology classification in large datasets by emulating human visual classification. The ViT model, originally designed for general image classification, has proven to be effective for classifying galaxies as well. In most cases, ViT achieves high accuracy in galaxy classification, especially when ample and diverse data are available. Despite the complexity of the ViT model, efficient processing can be achieved with suitable hardware (such as GPUs[10] or TPUs[11]). ViT often outperforms traditional models like CNNs in certain scenarios, particularly in the classification of more complex data.

- **Challenges**:

- Need for Large Datasets: ViT requires a substantial amount of data to perform optimally. If galaxy data is limited, techniques such as data augmentation or additional preprocessing may be necessary.

- High Computational Resources: Running large models like ViT demands significant computational resources.

- **Future Directions:**

- Transfer Learning: By using pre-trained ViT models on large datasets, accurate galaxy classification models can be developed even with limited data.

- Combination with Other Models: Integrating ViT with other models, such as CNNs, could improve classification accuracy.

[10] Graphics Processing Unit
[11] Cloud Tensor Processing

Using ViT for galaxy classification is a modern and powerful approach that can offer more precise results compared to traditional methods, though it does require substantial data and computational resources. Our research extends the boundaries of automated galaxy classification and provides tools and datasets that are likely to influence future studies in this domain. The high levels of accuracy, especially with deep learning models, highlight the potential of artificial intelligence in deciphering the complexities of large-scale cosmic structures.

References:

[1] G. De Vaucouleurs, ‘‘Classification and morphology of external galaxies”. In:Astrophysik iv: ternsysteme/astrophysics iv: Stellar systems. Springer, Vol.53, pp.275–310, July. 1959

[2] C. J. Lintott, K. Schawinski, A. Slosar, 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, No.3, pp.1179–1189, Sep. 2008

[3] C. J. Conselice, ‘‘The relationship between stellar light distributions of galaxies and their formation histories”. The Astrophysical Journal Supplement Series, Vol.147, No.1, July. 2003

[4] J. M. Lotz, J. Primack, P. Madau, ‘‘A new nonparametric approach to galaxy morphological classification”. The Astronomical Journal, Vol.128, No.1, pp.163–182, July. 2004

[5] P. Freeman, R. Izbicki, et al, ‘‘New image statistics for detecting disturbed galaxy morphologies at high redshift”. Monthly Notices of the Royal Astronomical Society, Vol.434, No.1, pp.282–295, Jun. 2013

[6] J. Vega-Ferrero, H. Domínguez Sánchez, M. Bernardi, et al, ‘‘Pushing automated morphological classifications to their limits with the dark energy survey”. Monthly Notices of the Royal Astronomical Society, Vol.506, No.2, pp.1927–1943, Sep. 2021

[7] M. Walmsley, C. Lintott, T. G´eron, S. Kruk, et al,‘‘Galaxy zoo decals: Detailed visual morphology measurements from volunteers and deep learning for 314000 galaxies”. Monthly Notices of the Royal Astronomical Society, Vol.509, No.3, pp.3966–3988, Jan. 2022

[8] R. Gupta, P. Srijith, S. Desai, ‘‘Galaxy morphology classification using neural ordinary differential equations”. Astronomy and Computing, Vol.38, No.5533, pp.100543, Jan. 2022

[9] H. Farias, D. Ortiz, G. Damke, M. J. Arancibia, M. Solar, ‘‘Mask galaxy: Morphological segmentation of galaxies.Astronomy and Computing”,Vol.33, No.100420, Aug. 2020

[10] P. H. Barchi, R. R. de Carvalho, R. R. Rosa, R. Sautter, et al, ‘‘Machine and deep learning applied to galaxy morphology-a comparative study”. Astronomy and Computing, Vol.30, No.100334, Jan. 2020

[11] H. Domínguez Sánchez's, M. Huertas-Company, M. Bernardi, D. Tuccillo, J. L. Fischer, ‘‘Improving galaxy morphologies for sdss with deep learning. Monthly Notices of the Royal Astronomical Society, Vol.476, No.3, pp.3661–3676, May. 2018

[12] M. Banerji, O. Lahav, C. J. Lintott, et al, ‘‘Galaxy Zoo: reproducing galaxy morphologies via machine learning. Monthly Notices of the Royal Astronomical Society, Vol.406, No.1, pp.342–353, July. 2010

[13] F. Ferrari, R.R. de Carvalho, M. Trevisan, ‘‘Morfometryka—a new way of establishing morphological classification of galaxies”, The Astrophysical Journal, Vol. 814, No.1, pp.55, Nov. 2015

[14] M. Abd el aziz, K. M. Hosny, I.M. selim, ‘‘Galaxies imageclassification using artificial bee colony basedonorthogonal Gegenbauer moments”, springer, Soft Comput, Vol.23, No.19, pp.9573–9583, Oct. 2019.

[15] L. Shamir, ‘‘Automatic morphological classification of galaxy images”, Monthly Notices of the Royal Astronomical Society, Vol.399, No.3, pp.1367–1372, Nov. 2009

[16] M. Marin, L.E. Sucar, J.A. Gonzalez, R. Diaz, ‘‘A Hierarchical Model for Morphological Galaxy Classification”, in: Proceedings of the Twenty-Sixth International Florida Artificial Intelligence Research Society Conference, Jan. 2013

[17] M. E. Abd el aziz, I. M. selim., and X. Shengwu., ‘‘Automatic Detection of Galaxy Type From Datasets of Galaxies Image Based on Image Retrieval Approach”, Scientific Reports (www.nature.com/scientificreports) , Jun. 2017

[18] R. E. González, R. P. Muñoz, C. A. Hernández, ‘‘Galaxy detection and identi_cation using deep learning and data augmentation”, Astronomy and Computing, Vol.25, pp.103-109, Oct. 2018

[19] D. G. York, J. Adelman, J. E. Jr. Anderson, et al, ‘‘The sloan digital sky survey:Technical summary”, The Astronomical Journal, Vol.120, No.3, pp.1579-1587, Sep. 2000

[20] J. P. Gardner, J. C. Mather, M. Clampin, R. Doyon, M. A. Greenhouse, H. B. Hammel, J. B. Hutchings, P. Jakobsen, S. J. Lilly, K. S. Long, et al, ‘‘The james webb space telescope”. Space Science Reviews, Vol.123, No.4, pp.485–606, Apr. 2006

[21] N. A. Grogin, D. D. Kocevski, S. M. Faber, et al, ‘‘Candels: the cosmic assembly near-infrared deep extragalactic legacy survey”. The Astrophysical Journal Supplement Series, Vol.197, No.2, pp.39, Dec. 2011

[22] M. J. Baumstark, G. Vinci, ‘‘Spiral-Elliptical automated galaxy morphology classification from telescope images”. Astronomy and Computing, Vol. 46, No.100770, Oct. 2023

[23] K.W. Willett, C. J. Lintott, S. P. Bamford, 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, Vol. 435, Issue 4, pp 2835–2860, Nov. 2013

[24] P. H. Barchi, R. R. de Carvalho, R. R. Rosa, et al, ‘‘Machine and Deep Learning Applied to Galaxy Morphology - A Comparative Study”. Astronomy and Computing, Vol. 30, N. 100334, Nov. 2019

[25] S. Gharat, Y. Dandawate, ‘‘Galaxy Classification: A deep learning approach for classifying Sloan Digital Sky Survey images”, Monthly Notices of the Royal Astronomical Society, Vol.511, No.4, pp.5120–5124, Apr. 2022

[26] K. Mohale, M. Lochner, ‘‘Enabling Unsupervised Discovery in Astronomical Images through Self-Supervised Representations”. MNRAS, Vol. 530, Issue 1, pp. 1274--1295, May 2024

[27] X.Tan, ‘‘Accurate and efficient galaxy classification based on mobile vision transformer”. Applied and Computational Engineering, Vol.33(1), pp.118-125, Jan. 2024

[28] S. Dieleman, K. W. Willett, J. Dambre, ‘‘Rotation-invariant convolutional neural networks for galaxy morphology prediction”, Monthly Notices of the Royal Astronomical Society, Vol450, No.2, pp.1441–1459, Jun. 2015

[29] N. E. M. Khalifa, M. H. N. Taha, A. E. Hassanien, I. M. Selim, ‘‘Deep Galaxy: Classification of Galaxies based on Deep Convolutional Neural Networks”, Computer Vision and Pattern Recognition (cs.CV), Sep. 2017

[30] J. M. Dai, J. Tong, ‘‘Galaxy Morphology Classification with Deep Convolutional Neural Networks”.Astrophysics and Space Science, Vol. 364(4), Jul. 2018

[31] J. Y. Y. Lin, S. M. Liao, H. J. Huang, W. T. Kuo, O. H. Min Ou, ‘‘Galaxy Morphological Classification with Efficient Vision Transformer”. accepted by the NeurIPS Machine Learning and the Physical Sciences workshop, Oct. 2021

[32] J. Cao, T. Xu, Y. Deng, et al, ‘‘Galaxy morphology classification based on Convolutional vision Transformer (CvT)”. A&A,Vol. 683, pp.11, A42, Mar. 2024

[33] A. Vaswani, N. Shazeer, N. Parmar, et al, ‘‘Attention Is All You Need”. Advances in Neural Information Processing Systems 30 (NIPS 2017), Jun. 2017

[34] M. Vafaeezadeh, H. Behnam, P. Gifani, ‘‘Ultrasound Image Analysis with Vision Transformers—Review”. Diagnostics, Vol. 14, No.5, pp. 542, Mar. 2024

[35] M. Eassa, I. M. Selim, W. Dabour, P. Elkafrawy, ‘‘Automated detection and classification of galaxies based on their brightness patterns”, Alexandria Engineering Journal, Vol.61, No.2, pp.1145-1158, Feb. 2022

[36] S. Kang, M. S. Shin, T. Kim, ‘‘Galaxy Morphological Classification with Deformable Attention Transformer”, Machine Learning and the Physical Sciences workshop, NeurIPS. 2022

[37] K. Alrfou, A. Kordijazi, T. Zhao, ‘‘Computer Vision Methods for the Microstructural Analysis of Materials: The State-of-the-art and Future Perspectives”, Materials Science, Computer Science, Engineering, Jul. 2022