MATEC Web of Conferences **392**, 01123 (2024) https://doi.org/10.1051/matecconf/202439201123
ICMED 2024

Discovery of astronomical objects in galaxies by means of deep learning

Ponguvala Haindavi[1,*], *Sunil* Kumar[2], Ganesh[3], *R*. Anil Kumar[4], *Gaurav* Pushkarna[5]

[1]Department of CSE – Data Science, K G Reddy College of Engineering and Technology,Hyderabadi, Telengana, India.
[2]Department of CSE,Mohan Babu University (Erstwhile Sree Vidyanikethan Engineering College(Autonomous), Tirupati, Andhra Pradesh , India.
[3]Department of Computer Science and Engineering, Stanley College of Engineering and Technology for Women, Abids, Hyderabad, Telangana, India
[4]Department of EEE, GRIET, Hderabad, Telangana, India.
[5]Lovely Professional University, Phagwara, Punjab, India,

Abstract. The study of space exploration has been studied for a very long time, and as technology has developed, so too have the methods and techniques employed, along with the quantity and type of data acquired. We now receive so much astronomical data, and so much brand new data is being generated every day, that it is physically impracticable to examine it all only by human work. In our study, we look at a number of astronomers face while working with this massive amount of data, and they use deep learning techniques to discover the best data for each objective. [1]previously SVM, KNN, the random forest approach, decision trees, and other multi-class classification algorithms are all used in the methodology. Galaxies' propensity to belong to specific classes is forecasted using regression. even if the findings from the random forest method were the best it was unable to effectively divide galaxies into the five groups. This approach does not explain real-time categorization and does not take outliers into consideration. The model's adaptability is constrained. This categorization scheme is unable to account for the modelling of galaxies as well as their evolution. Here, we suggest using Inception v3 for categorization and VGG-19 for image analysis. Segmentation is a method for discovering and classifying galaxies. These techniques greatly advance certain fields of study where there are enormous volumes of duplicate data that must be deleted in accordance with the demands of the study, thanks to Python's high performance in the investigation of image processing and computer vision.. As a result, there is less of a need for researchers to carefully sort through all of the data that has been collected from satellite telescopes, sky surveys, etc. [2]

* Corresponding author: haindavi284@gmail.com

Keywords: Deep Learning, Analytical Models, Instrument, Time Series Analysis, Predictive Models, Complexity Theory, VGG-19, CNN, SVM

1 Introduction

Anomaly detection is critical in many domains, including astronomy, where recognising rare and unexpected events or objects can reveal important insights into the universe. The detection of aberrant objects in galaxies has enormous potential in astronomy for discovering new events, understanding cosmic processes, and expanding our understanding of the cosmos. Astronomers have traditionally depended on manual inspection and expert analysis to detect anomalies in astronomical images and data. However, when the number and complexity of observational data increases, manual procedures become impractical and time-consuming[1]. This has led to the investigation of automated anomaly detection techniques, particularly ones that make use of deep learning's power.A branch of machine learning known as deep learning has achieved outstanding results across a variety of disciplines, including computer vision and natural language processing. It is a desirable option for anomaly identification in astronomical objects due to its capacity to learn intricate patterns and representations from unstructured data.The fundamental goal of this project is to create efficient and effective anomaly detection models for recognising unusual and rare objects in galaxies using deep learning techniques. Supernovae, quasars, unique galaxy shapes, and other surprising occurrences are examples of anomalies[2]. Astronomers can efficiently analyse vast datasets and focus their attention on the most intriguing and scientifically significant occurrences by automating the process of anomaly detection[3].The proposed research aims to leverage the power of deep neural networks to learn and extract meaningful features from astronomical images and data. By training the models on a large dataset of labeled normal objects, the models can learn to distinguish between regular and anomalous objects. The models will be optimized to achieve high detection accuracy while being computationally efficient to handle the vast amount of astronomical data available[4].In addition to the development of anomaly detection models, this research will also focus on interpretability. Understanding the features and characteristics that contribute to the detection of anomalies is essential for astronomers to learn more about the underlying physical processes. The models will be designed to provide interpretable results and highlight the relevant features that differentiate anomalies from normal objects.The successful implementation of efficient deep learning models for anomaly detection in astronomical objects can significantly advance our understanding of the universe. It can enable the discovery of rare events, uncover new astrophysical phenomena, and aid in the exploration of uncharted territories in the realm of galaxies. Moreover, these models can serve as valuable tools for automated data analysis, facilitating the progress of large-scale astronomical surveys and accelerating the pace of scientific discoveries in the field of astronomy.

The work that follows is concerned with the classification of celestial objects and deep learning. Astronomical object classification: The following section will go through several recurrent classification issues in astronomy. In star/galaxy classification, as the name implies, a particular item from a set of images is categorized as either a star or a galaxy. This is relatively simple for bright objects but becomes more difficult as astronomical surveys penetrate further into the sky and fainter galaxies begin to appear as point like objects. To get more accurate estimations of the objects' true sizes and scales, stars must be separated from galaxies. The identification of two or more colliding galaxies occurs when merging galaxies are discovered from a collection of galaxy images. [5]

2 Literature Survey

David hand(2017)[6] discussed about A star-galaxy separator for the UKIRT Infrared Deep Sky Survey (UKIDSS) and a cutting-edge anomaly detection tool for cross-matched astronomical datasets are the two statistical methods for astronomical problems we discuss. Using prior knowledge from the source population, a statistical classification technique known as the star-galaxy separator produces class membership probabilities rather than class names. The performance of our classifier is evaluated using Deep Sloan Digital Sky Survey (SDSS) data from the multiplex imaged Stripe 82 region, and it outperforms the UKIDSS pipeline classification approach. The technique of anomaly detection addresses the issue of objects in cross-matched datasets having different sets of recorded variables. By avoiding approaches that can't handle missing values, direct object comparison is made possible.

Daniel Mortlock (2018) [7] discussed about When compared favourably to the UKIDSS pipeline classification method, the results of our classifier outperform it. They are validated using Data from the Stripe 82 region of the multiplexed Deep Sloan Digital Sky Survey (SDSS). The problem of objects in cross-matched datasets having different sets of recorded variables is addressed by the anomaly detection approach.. Direct comparisons between objects are made difficult, and it is not possible to employ techniques that do not support missing values. In order to provide an overall anomaly score, our method calculates anomaly scores for each source in sub spaces of the observed feature space[8]. The suggested approach is quite adaptable and ideal for non-astronomical applications. We demonstrate the features and efficiency of our method on both real-world and simulated datasets.

Sheikh Muhammad Saiful Islam (2019)[9] discussed about Group anomalies in astronomical data detection The algorithms in the Sloan Digital Sky Survey data collection are also used to look for group irregularities. Again, since this enormous data collection lacks labelling information, we must create artificial injections to gauge algorithm performance. To identify spatial clusters, we first create a graph that includes the edges between nearby galaxies. Then, we treat the network's connected elements as spatial clusters. Following this preprocessing, 518 spatial groups (7712 galaxies) were found with sizes between [10, 50]. The 500-dimensional feature Cs1 (normalised continuum) was then condensed to a 2-dimensional space using PCA while 95% of the overall variance was preserved. For testing, we gave the algorithms manufactured group anomalies.

M. Alazab discussed about[10] Five random points, each of which represents a different subject distribution, are then selected at random from the simplex's low-density area. The following are the unusual topic distributions: The next step is to create 10 injection groups by selecting galaxies at random from the same data set and aligning them with the chosen anomalous topic distributions (equivalent to around 2.5% of the normal data). Using a histogram-based methodology, we compared the DG and GMM models in this experiment. The histogram-based methods (H) imitate the injection method by quantizing the galaxies into several topics, computing the topic distributions for each group, and then identifying points in low-density regions on the simplex. An illustration of a transformation-based technique is this, Astronomical data has rapidly expanded in bulk and complexity. In order to process this massive amount of data and deliver results, along with any additional relevant data that the researchers require,Methods using machine learning and deep learning might be useful. In this study, the Inception V3 and Faster RCNN models are used to recognize galaxies in photographs and categorize galaxies based on their form. Astronet uses the Kepler time series data to evaluate astronomical objects and determine if they are exoplanets or not in order to find exoplanets. The SGD classifier, a machine learning technique, predicts the planetary habitability after analyzing PHL's exoplanet library[11].

3 Methodology

In this work we employ deep learning algorithms like CNN, Inception-v3 and VGG19 because the quantity and kind of data collected have changed over time. We now receive so much astronomical data, and so much brand-new data is generated every day, that it is physically impracticable to interpret it all alone by human labour. To increase the generalisation ability of our deep learning model, we train it using images. The model could recognize and categories astronomical targets on its own after training. We evaluate the performance of our framework using simulated data, and we find that it is nearly as good at identifying abnormalities as the traditional approach.

There is just a little amount of overlap between the three basic classes—elliptical, spiral, and edge-on—as seen by the confusion matrices for both pieces of training. As compared to the other classes, the DK and Merge classes perform badly since there are fewer annotations. Since it might be difficult to distinguish between a blended galaxy pair and two different neighboring galaxies, the merged class in the galaxy zoo is particularly ambiguous. Another factor contributing to the merged class's specific haziness is nearly galaxies.

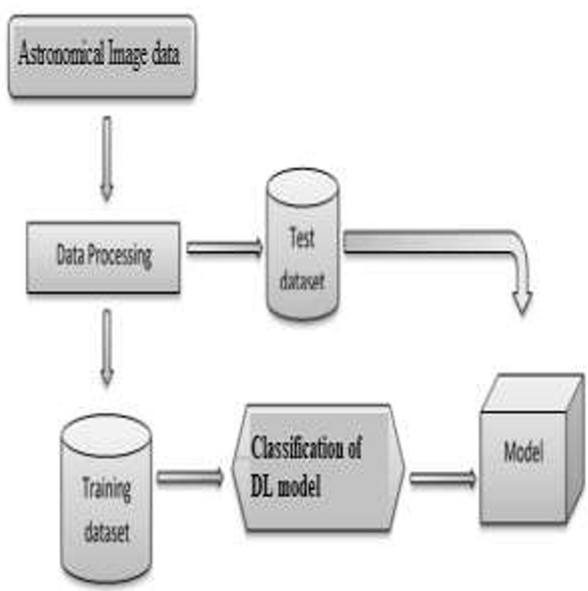

Fig. 1. Proposed System Architecture

Building a system architecture that can accurately identify and categorise unexpected or anomalous events in astronomical data is required for anomaly detection in astronomical objects of galaxies using deep learning. An overview of a typical system architecture for deep learning-based anomaly detection in astronomical objects is given below:

- Data Gathering: The first phase entails gathering pertinent astronomical data from observatories or space telescopes, such as galaxies' pictures or spectra. Observations made using different wavelengths, such as optical, infrared, or radio, may be included in this data.[12]
- Data preprocessing: To eliminate noise, account for instrument effects, and normalise the data, the obtained data frequently has to be preprocessed. Techniques like picture calibration, background subtraction, or noise reduction may be used in preprocessing.
- Feature Extraction: To identify key characteristics of the astronomical objects, significant features are extracted from the preprocessed data in this step. This could entail employing convolutional neural networks (CNNs) or other computer vision techniques to extract features from picture data. Relevant features for spectral data can be extracted using methods like Fourier transforms or wavelet transforms.
- Preparing Training Data: Training the deep learning model requires preparation of annotated training data. A portion of the data has to be labelled by specialists in the field or astronomers to show examples of typical or anomalous astronomical objects. The deep learning model will be trained using this labelled data.
- Design of the Deep Learning Model: Depending on the exact job of anomaly detection and the type of data being examined, the architecture of the deep learning model can change. Convolutional neural networks (CNNs), recurrent neural networks (RNNs), or more sophisticated architectures like autoencoders or generative adversarial networks (GANs) are a few examples of frequently used models.
- Training: Using the labelled training data created in the previous stage, the deep learning model is trained. The model picks up on the patterns and characteristics that separate typical items from oddities. During the training phase, the model's parameters are optimised by minimising a suitable loss function, often using backpropagation and gradient descent methods.
- Model evaluation: Using a different validation dataset, the deep learning model's performance is assessed after training. The effectiveness of the model in detecting anomalies can be evaluated using a variety of metrics, including accuracy, precision, recall, or F1 score.[13]
- Anomaly Detection: The deep learning model can be used to find anomalies in fresh, unexplored data after being trained and validated. Based on the recognised patterns and traits, the model determines if the observed astronomical object is normal or abnormal using the preprocessed data as input.[14]
- Post-processing and visualisation: Anomalies that are discovered can be examined and characterised in further detail. Similar anomalies can be grouped together or the feature space can be shrunk using post-processing techniques like dimensionality reduction or clustering.[15]

4 Model Efficiency

Accurate Detection: An anomaly detection model's primary goal is to find and categorize anomalies in astronomical objects. Metrics that can be used to evaluate a model's detection accuracy include precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC).. A more accurate model is more effective.

Computational efficiency: Deep learning methods can be computationally costly, particularly when working with large-scale astronomical information. Training time, inference time, and resource utilisation (CPU/GPU memory consumption) can all be used to assess efficiency. Models with great accuracy while being computationally efficient are preferred.

Scalability: To handle big and heterogeneous astronomical datasets, an efficient model should be scalable. It should be able to process a huge number of astronomical objects in an acceptable amount of time, accommodating the field's expanding supply of data.
Model Complexity: The efficiency of a deep learning model may be impacted by the model's complexity. For more complex models with a lot of parameters, longer training times and additional computational resources can be needed. Model complexity and performance must be balanced in order to achieve effective anomaly detection.

Generalisation: A good model should have high generalisation skills, which means it can find abnormalities in previously unseen or out-of-distribution data. It should be resistant to changes in data quality, observational conditions, and equipment properties. While not directly connected to efficiency, interpretability is a significant consideration in anomaly detection. Models that are effective should provide insights into the identified anomalies and the features that contribute to their classification. This aids astronomers in comprehending the underlying physical or astrophysical processes that are creating the anomalies.

Model Optimisation: To increase performance, efficient models may employ optimisation techniques. Model compression, parameter pruning, and quantization can all be used to minimise model size and inference time without sacrificing accuracy.

Table. 1. Object Detection Algorithm System

System	**Object Detection Algorithms**	**Detection Time (ms)**
Existing System	SVM, KNN, INCEPTION-V3 etc..,	25-30
Proposed System	VGG-19	05-10

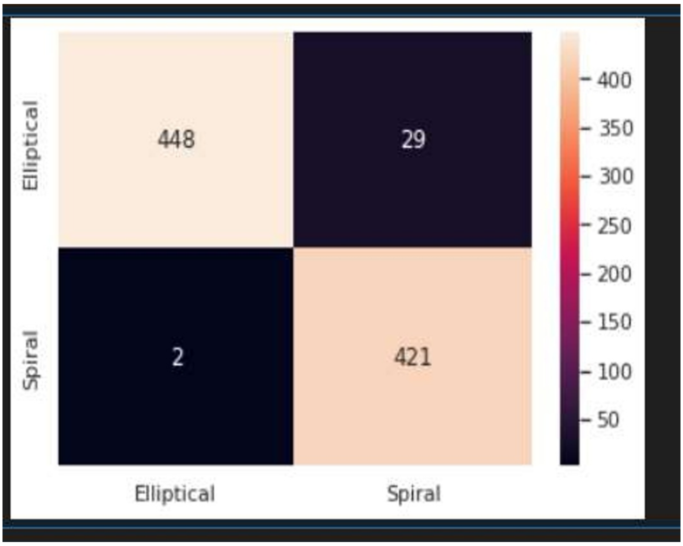

Fig. 2. Confusion Matrix

For both of the Object Detection methods utilised for the current system, the effectiveness is displayed in the table above in milliseconds (ms) and VGG-19 for Proposed System which we can observe that the efficiency is increasesA bar graph is used to graphically as shown below:

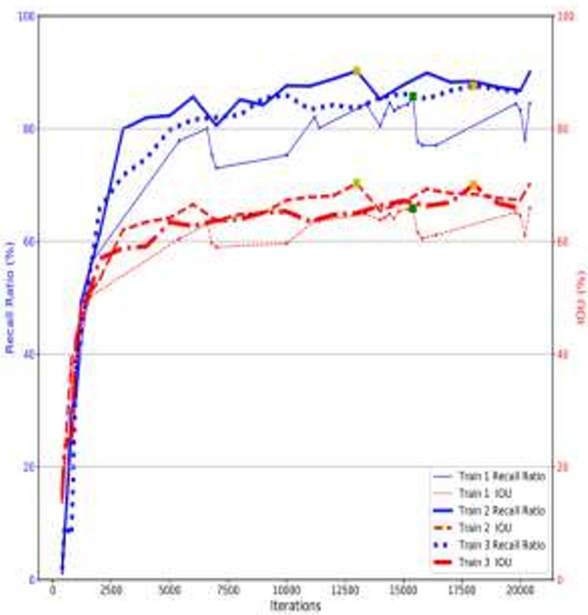

Fig. 3. Test Accuracy 0.935 (INCEPTION – V3)

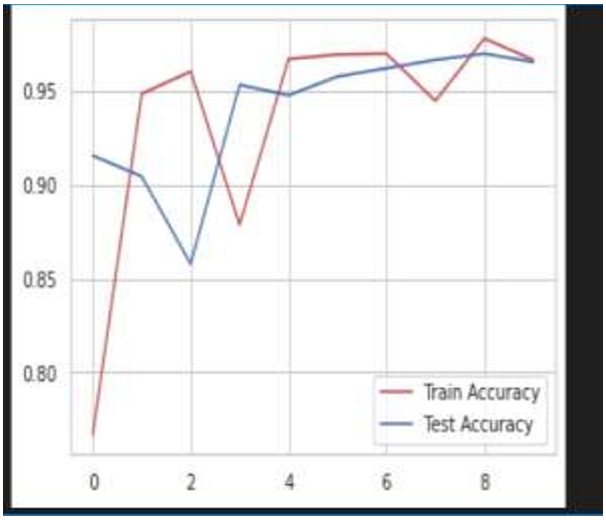

Fig. 4. Test Accuracy 0.965 (INCEPTION – VGG19)

5 Conclusion

Machine learning methods are used in this study to classify astronomical object data using space image data. Before classification, image processing and feature extraction are performed on picture data using a deep learning model such as Inception or VGG16. According to our results, the classification algorithm based on the VGG16 approach has the highest level of accuracy when compared to the other examined classification algorithms. On the other hand, the classification method's validity has not been examined when using the optimization default parameter value. There is now the potential for improving the categorization model. To improve the accuracy of the deep learning algorithm, parameter adjustments must be performed.

References

1. Shivani Dere, Maziya Fatima, Rutuja Jagtap and Nikhilkumar B Shardoor“Anomaly Detection in Astronomical Objects of Galaxies Using Deep Learning,”7th International Conference on Advanced Computing & Communication Systems (ICACCS),(2021)
2. A.Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenetclassification with deep convolutional neuralnetworks,” in Advances in Neural Information Processing Systems, (2012), pp. 1097–1105.
3. . S. Gidaris, P. Singh, and N. Komodakis, “Unsupervised Representation Learning by Predicting Image Rotations,” in International Conference on Learning Representations, (2018).
4. M. Caron, P. Bojanowski, A. Joulin, and M. Douze, “Deep Clustering for Unsupervised Learning of Visual Features,” in European Conference on Computer Vision, (2018).
5. C. Doersch, A. Gupta, and A. A. Efros, “Unsupervised Visual Representation Learning by Context Prediction,” (2015).
6. D. Hendrycks, M. Mazeika, S. Kadavath, and D. Song, “Using SelfSupervised Learning Can Improve Model Robustness and Uncertainty,” in Advances in Neural Information Processing Systems, (2019), pp. 15 663– 15 674

7. M. Raghu, C. Zhang, J. Kleinberg, and S. Bengio, “Transfusion: Understanding Transfer Learning for Medical Imaging,” in Advances in Neural Information Processing Systems, (2019).
8. A. Martinazzo, M. Espadoto, and N. Hirata, “Deep Learning for Astronomical Object Classification: A Case Study,” in VISAPP, (2020)
9. Venkateswara Reddy, L., Prema, K. (2020). Efficient Spatial Database Keyword Query Search. In: Kumar, A., Paprzycki, M., Gunjan, V. (eds) ICDSMLA (2019). Lecture Notes in Electrical Engineering, vol **601**. Springer, Singapore. https://doi.org/10.1007/978-981-15-1420-3_186
10. P. Venkateswarlu Reddy, D. Ganesh, L. S. Azeem, M. Padma, K. D. Lakshmi and L. Chakradhar, "Implementation of Latest Deep Learning Techniques for Brain Tumor Identification from MRI Images," (2023) 8th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, (2023), pp. 1166-1171, doi: 10.1109/ICCES57224.2023.10192620.
11. D. Ganesh et al., "Implementation of Novel Machine Learning Methods for Analysis and Detection of Fake Reviews in Social Media," 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, (2023), pp. 243-250, doi: 10.1109/ICSCDS56580.2023.10104856.
12. Venkateswara Reddy, L., Davanam, G., Pavan Kumar, T., Sunil Kumar, M., Narendar, M. (2023). Bio-Inspired Firefly Algorithm for Polygonal Approximation on Various Shapes. In: Rao, B.N.K., Balasubramanian, R., Wang, SJ., Nayak, R. (eds) Intelligent Computing and Applications. Smart Innovation, Systems and Technologies, vol **315**. Springer, Singapore. https://doi.org/10.1007/978-981-19-4162-7_10
13. Davanam, G.;; Singh, N.; Gunjan, V.K.; Roy, S.; Rahebi, J.; Farzamnia, A.; Saad, I. Multi-Controller Model for Improving the Performance of IoT Networks. Energies (2022), 15, 8738. https://doi.org/10.3390/en15228738
14. Ganesh, D., et al. "Improving Security in Edge Computing by using Cognitive Trust Management Model." 2022 International Conference on Edge Computing and Applications (ICECAA). IEEE, (2022).
15. Davanam, Ganesh, T. Pavan Kumar, and M. Sunil Kumar. "Novel Defense Framework for Cross-layer Attacks in Cognitive Radio Networks." International Conference on Intelligent and Smart Computing in Data Analytics: ISCDA (2020). Springer Singapore, (2021).
16. Ganesh, D., Kumar, T. P., & Kumar, M. S. (2021). Optimised Levenshtein centroid cross-layer defence for multi-hop cognitive radio networks. IET Communications, 15(2), 245-256.