

Comparison of Various Techniques to Classify Galaxies based on Morphology and to Detect Potential Exoplanets

Karishma Chavhan

Dept. of Computer Science and Engineering
Dayanada Sagar University, School of Engineering
Bengaluru, India

Sandesh Bhat

Dept. of Computer Science and Engineering
Dayanada Sagar University, School of Engineering
Bengaluru, India

Rahul Noronha

Dept. of Computer Science and Engineering
Dayanada Sagar University, School of Engineering
Bengaluru, India

Sahana M

Dept. of Computer Science and Engineering
Dayanada Sagar University, School of Engineering
Bengaluru, India

Abstract— In this paper, we are trying to examine which method is most suitable for classifying the galaxies based on their morphology into their various shapes Spiral, Elliptical and irregular. We are also trying to determine which method would work best to detect potential exoplanets.

Keywords—Galaxy morphology, Exoplanet Detection, ImageNet, Artificial Neural Network, Hubble type, Decision trees.

I. INTRODUCTION

We adopt a transfer learning approach and use the ResNet50 model on the crowdsourced Galaxy Zoo dataset. The different methods we compare for the potential exoplanet detection task are as follows: Tree-based, Naive Bayes, Logistic regression. Along with the machine learning models, we also make use of Deep learning models like a perceptron (Artificial Neuron) and compare their results.

A. Abbreviations

Abbreviations used:

ANN - Artificial Neural Network.

ResNet50 - Residual Neural Network (50 layers).

ResNet152 - Residual Neural Network (152 layers).

Xception - Extreme inception .

KNN - K-Nearest Neighbors.

RBF - Radial Basis Function.

II. PROBLEM STATEMENT

To determine which method is the best for classifying the galaxies based on their morphology and to examine for potential exoplanets detection which method would perform better. To use ANN for potential exoplanet detection and to try different ImageNet models like ResNet152, Xception, etc., and see how they compare with ResNet50.

III. LITERATURE SURVEY

We conducted a survey about the different methods available to perform classification of galaxies based on their

Morphology. Looking through the relevant research papers in potential exoplanet detection we identified some key techniques used, based on their time and resource usage. We will cover a few methods for each of these two tasks.

A. Galaxy Morphology

1) Rules-Based Approach

In the first method, we use a rules-based approach where we derive the Hubble type by following the Galaxy Zoo Decision Tree. What is the Galaxy Zoo 2 Project: In this crowdsourced project, the online participants are given an image starting with a question asking if the galaxy is simply smooth and rounded with no sign of a disk, depending on the responses the users give to the questions, another question is asked with the same image, until finally the Galaxy gets classified into spiral, elliptical or irregular shape. A small drawback is that non-expert labelling of the data may lead to human error.

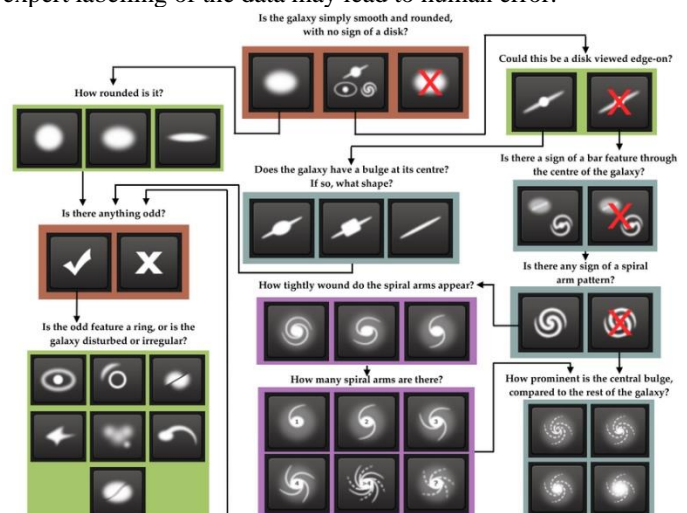


Fig. 1. The Hubble type decision tree.

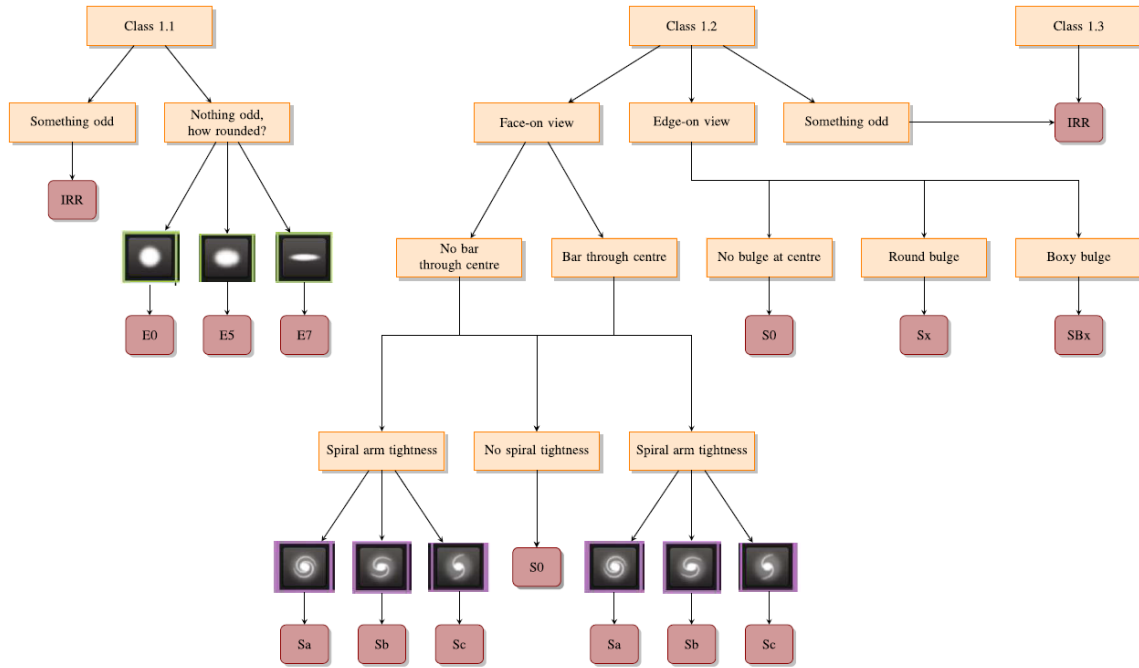


Fig. 2. Flowchart showing the rules-based mapping onto Hubble type

2) *Transfer Learning using ImageNet Models*

In the second method we use one of the pre-trained ImageNet models (ImageNet Models — Neural Network Libraries 1.25.0 documentation (nnabla.readthedocs.io)) we have available to us using Transfer learning approach. Since Deep learning training takes considerable time and resources for training an emerging technique, especially in this field has been the use of transfer learning. In transfer learning, a neural network trained for another set of images can be repurposed and

used for a different use case. This is done by removing the output layer of the image net and replacing it with just an output layer in the cases where we have less amount of unlabelled data, or replacing the last few, or even many of the Image net layers in case we have a large amount of labelled data with us. The ResNet50 is an Image net model that came around the time of Galaxy Zoo and that can be used effectively in the case of Galaxy morphology classification. There are also other Image Net models like ResNet152, Xception, etc that could potentially be used for this purpose.

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
		3×3 max pool, stride 2				
conv2_x	56×56	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Fig. 3. ResNet50 and ResNet152 architectures with other ResNet model architectures too.

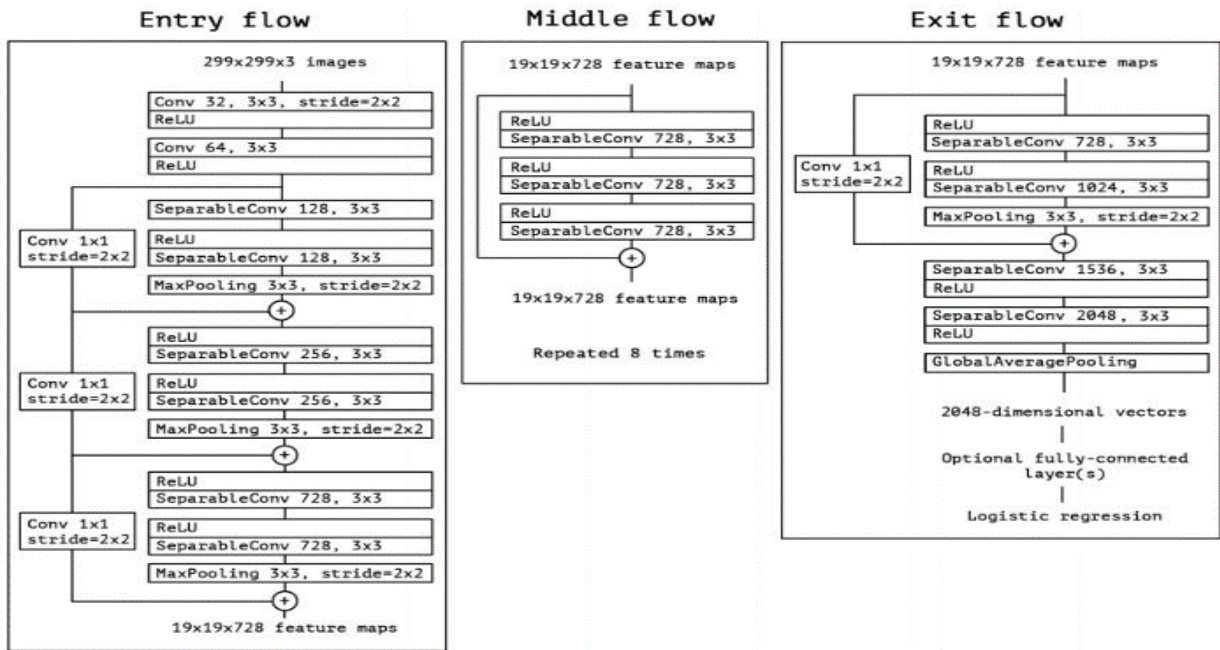


Fig. 4. Xception ImageNet architecture

3) *Hubble Tuning Fork method*

In the third method we use the Hubble tuning fork classification schemes for galaxies. Here techniques such as PCA (Principal component analysis) are used after which artificial neuron networks are trained using locally weighted regression.

But to map the 37 vectors used to the Hubble classification scheme we use the first two methods which are rules based and the Transfer learning approach.

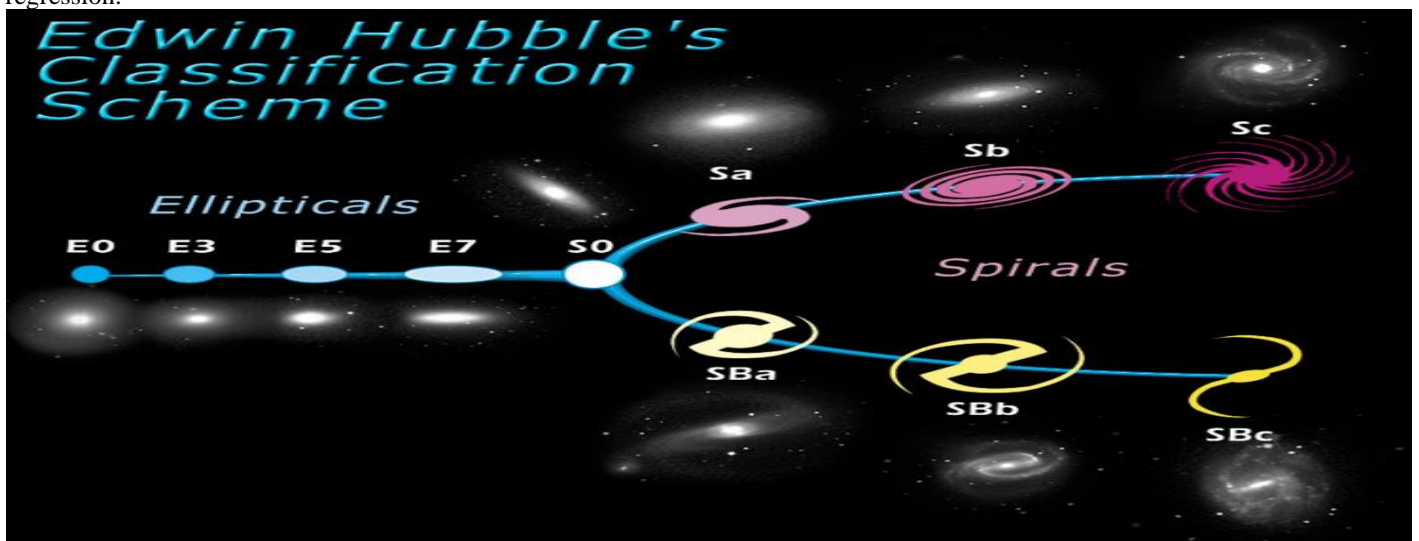


Fig. 5. Hubble Classification Scheme

B. Potential Exoplanet Detection

Planets orbiting stars outside our solar systems are called extrasolar planets or exoplanets. Several approaches have been proposed by astronomers for detecting them, being the fine-grained analysis of periodicities in star light-curves the most successful so far.

The methods present for Potential exoplanet detection are:

1) Transient Light curve analysis

The Hubble telescope captures light coming from a star, on seeing the time variance of the light reaching the telescope we can find out if an object such as an exoplanet has passed in front of the star by analysing the light curve. This method is resource intensive and requires more training time and resources. Usually for this light curve analysis we use CNNs.

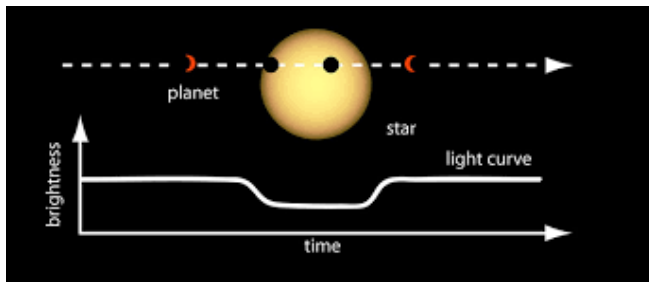


Fig. 6. Transient light curve.

2) c-SVM

For the c-SVM we use a Gaussian kernel as the Radial Basis Function (RBF) because it produces a more flexible decision boundary.

3) KNN

Load the data. Initialize K to your chosen number of neighbors. For each example in the data. Calculate the distance between the query example and the current example from the data. Add the distance and the index of the example to an ordered collection. Sort the ordered collection of distances and indices from smallest to largest (in ascending order) by the distances. Pick the first K entries from the sorted collection. Get the labels of the selected K entries. If regression, return the mean of the K labels. If classification, return the mode of the K labels.

4) Random Forest

One of the most important features of the Random Forest Algorithm is that it can handle the data set containing categorical variables as in the case of classification. It performs better results for classification problems.

5) Logistic Regression

Logistic Regression is useful analysis method for classification problems, where you are trying to determine if a new sample fits best into a category.

IV. CONCLUSIONS

The paper aims on performing comparative analysis and introduce different techniques. In the field of astronomy and combine it with machine learning to get more accurate results. Transfer Learning improves the training dataset and hence accuracy for galaxy images. Single or multi-layer perceptron is expected to give better results for exoplanets metadata. Among KNN, Random Forest, SVM and Logistic we adopt the model that gives better accuracy and save the model. The two models saved are used in a web app and deployed.

REFERENCES

- [1] M. Z. Variawa, T. L. van Zyl and M. Woolway, "A rules-based and Transfer Learning approach for deriving the Hubble type of a galaxy from the Galaxy Zoo data," 2020 IEEE 23rd International Conference on Information Fusion (FUSION), 2020, pp. 1-7, doi: 10.23919/FUSION45008.2020.9190462.
- [2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
- [2] H Domínguez Sánchez, M Huertas-Company, M Bernardi, S Kaviraj, J L Fischer, T M C Abbott, F B Abdalla, J Annis, S Avila, D Brooks, E Buckley-Geer, A Carnero Rosell, M Carrasco Kind, J Carretero, C E Cunha, C B D'Andrea, L N da Costa, C Davis, J De Vicente, P Doel, A E Evrard, P Fosalba, J Frieman, J García-Bellido, E Gaztanaga, D W Gerdes, D Gruen, R A Gruendl, J Gschwend, G Gutierrez, W G Hartley, D L Hollowood, K Honscheid, B Hoyle, D J James, K Kuehn, N Kuropatkin, O Lahav, M A G Maia, M March, P Melchior, F Menanteau, R Miquel, B Nord, A A Plazas, E Sanchez, V Scarpine, R Schindler, M Schubnell, M Smith, R C Smith, M Soares-Santos, F Sobreira, E Suchyta, M E C Swanson, G Tarle, D Thomas, A R Walker, J Zuntz, Transfer learning for galaxy morphology from one survey to another, Monthly Notices of the Royal Astronomical Society, Volume 484, Issue 1, March 2019, Pages 93–100, <https://doi.org/10.1093/mnras/sty3497>.
- [3] Ismael Araujo (2020). Using Machine Learning to Find Exoplanets with NASA's Data; <https://towardsdatascience.com/using-machine-learning-to-find-exoplanets-with-nasas-dataset-bb818515e3b3>.
- [4] Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetism Japan, p. 301, 1982].
- [4] L. Ofman, A. Averbuch, Adi Shlisselberg, Idan Benaun, David Segev, Aron Rissman (2021). Automated identification of transiting exoplanet candidates in NASA Transiting Exoplanets Survey Satellite (TESS) data with machine learning methods, Physics, Computer Science, 2021, doi: 10.1016/j.newast.2021.101693.