Galaxy10 classification with DeepCNN

Anusha S Rao
PES1UG19CS077
PES University

Prajwal A
PES1UG19CS334
PES University

Sammed T
PES1UG19CS426
PES University

Abstract -A Convolutional Neural Network are a special kind of multi-layer neural networks, designed to recognize visual patterns directly from pixel images with minimal preprocessing. Several such networks have been implemented in this work on Galaxy10 dataset. Networks such as EfficientNet, VGG19, LeNet , AlexNet have been implemented in this paper . As a result of these models a remarkable increase in the accuracy have been achieved . Having applied these above techniques on the datatset we are able to classify the 10 sub-classes of the galaxies. Thus, we have obtained higher classification accuracies of 89.7 percent and 89 percent using VGG19 and EfficientNet respectively between 10 sub-classes of galaxies.

I. INTRODUCTION

Finding clues about the origin and the evolution of the universe remains a considerable challenge for astrophysicists. Morphological classification of Galaxy is a system used to divide galaxies into groups based on their visual appearance.A huge amount of effort and time is required for any manual classification and analysis of these datasets. There are various approches used by which galaxies can be classified according to their morphologies , the most famous being the Hubble sequence[Fig.1]. However, galaxy classification and morphology are now largely done using computational methods and physical morphology .

The different classes of galaxies are elliptical , spiral , irregular, lenticulars. Elliptical galaxies are smooth, featureless objects, appearing as ellipses in the images . Spiral galaxies have disc-like structures and generally have 2 spiral arms. The third category from the Hubble galaxy classification scheme is irregular galaxies, which have very distorted shapes without the spiral arms or galactic bulge of spiral galaxies .

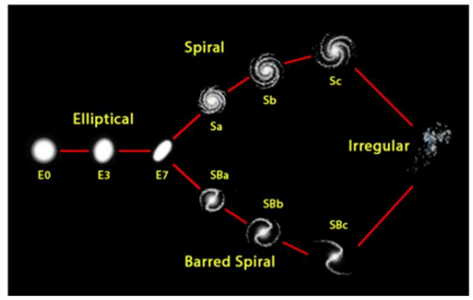


Fig 1. Hubble’s Original Classification

Various attempts have been made to examine galaxies and categorize galaxies into various shapes . We have many attempts of classification based on main classes of the galaxies . But we have only few works on sub-class classification of galaxies . This paper focuses on 10-class classification of galaxies based on their morphology as follows

(i) elliptical galaxies (E0, E1, E2, E3, E4, E5, E6, E7)

(ii) spirals (Sa, Sb, Sc), barred spiral (SBa, SBb, SBc) and

(iii) irregulars.

II. KEY WORDS

CNN , EfficientNet, VGG19, LeNet , AlexNet

III. DATSET

Galaxy10 dataset was created with Galaxy Zoo (GZ) . The dataset contains 17736 256x256 pixels colored galaxy images (g, r and z band) separated in 10 classes . Galaxy10 SDSS(Sloan Digital Sky Survey) images come from Sloan Digital Sky Survey and labels come from Galaxy Zoo. The class distribution for the 10 sub-classes[fig 2.]

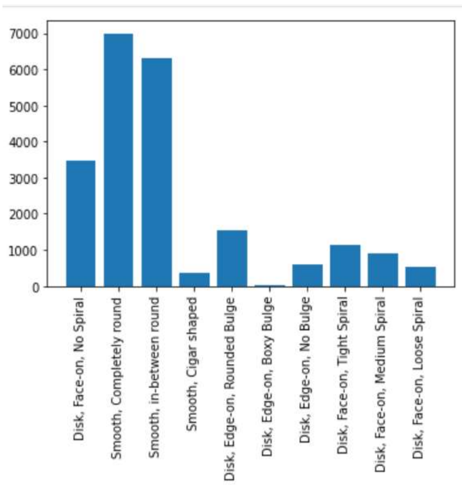


Fig 2.Class Distribution of Morphologies of galaxies

IV. RELATED WORK

Deep neural networks (DNNs) with a step-by-step introduction of inputs, which is constructed by imitating the somatosensory system in human body, known as SpinalNet. This algorithm for example is an architecture which mimics the natural way of reacting. With the idea of SpinalNet[fig 3.] which is reported to have achieved in most of the DNNs , appling the same to the Galaxy Zoo dataset to classify the different classes and/or sub-classes of the galaxies ,the paper obtains higher classification accuracies of 98.2 percent , between elliptical and spiral and 95 percents between the former and irregulars , and 82 between 10 sub-classes of galaxies[1] .

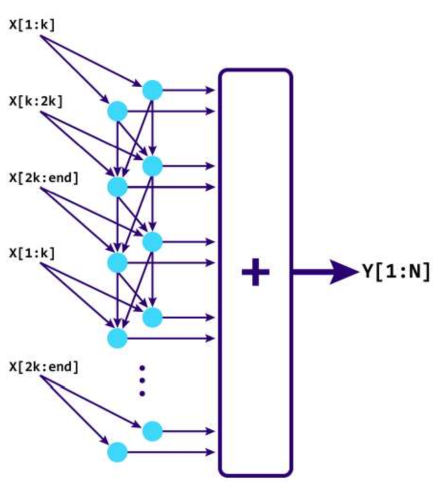


Fig 3. Architecture of SpinalNet DNN

Classification of galaxy based on its features into main three categories Elliptical, Spiral, and Irregular. The proposed deep galaxies architecture consists of 8 layers, one main convolutional layer for features extraction with 96 filters, followed by two principles fully connected layers for classification. It is trained over 1356 images and achieved 97.272% in testing accuracy[2].This work does not focus on 10 – class classification of galaxies .

EfficientNets and their applications can be used to classifiy galaxies based on their morphologies . The usage of EfficientNets is done 79,975 testing images from the Galaxy Zoo 2 challenge on Kaggle.The paper does Evaluation of the model using the standard competition metric i.e, rmse score and rank among the top 3 on the public leaderboard with a public score of 0.07765. Paper proposes a fine-tuned architecture using EfficientNetB5 to classify galaxies into seven classes - completely round smooth, in-between smooth, cigarshaped smooth, lenticular, barred spiral, unbarred spiral and irregular. The network along with other popular convolutional networks are used to classify 29,941 galaxy images. Different metrics such as accuracy, recall, precision, F1 score are used to evaluate the performance of the model along with a comparative study of other state of the art convolutional models to determine which one performs the best. an accuracy of 93.7% was obtained on classification model with an F1 score of 0.8857 [3]. Therefore EfficientNets can be applied to large scale galaxy classification.

Computer-based approach to galaxy morphology done using an unsupervised machine learning system that can deduce the visual similarities between sets of images and reconstruct morphological sequences of galaxies. The analysis is performed such that the algorithm determines the network of similarities between the different morphological classes automatically, and without human guidance [4].

V. PROPOSED METHODOLOGY

This section describes the approaches and implementation details of Different Models used.

A. EfficientNet

EfficientNet is a convolutional neural network architecture and scaling method that uniformly scales all dimensions of depth/width/resolution using a compound coefficient. Using EfficientNet accuracy of 89 percent and 88 percent on training and validation set is achieved respectively.

B. ResNet

Residual neural network (ResNet) is an artificial neural network (ANN). Residual neural networks utilize skip connections, or shortcuts to jump over some layers. Typical ResNet models are implemented with double- or triple- layer skips that contain nonlinearities (ReLU) and batch normalization in between. Using EfficientNet an accuracy of 85 percent and 83 percent on training and validation set respectively is achieved respectively.

C.VGG-19

VGG-19 is a convolutional neural network that is 19 layers deep. This network can classify images into different categories .Using VGG -19 along with data augmentation accuracy of 89.7 percent and 86.6 percent on training and validation set is achieved respectively .

D. LeNet

LeNet refers to LeNet-5 and is a simple convolutional neural network. Convolutional neural networks are a kind of feed-forward neural network whose artificial neurons can respond to a part of the surrounding cells in the coverage range and perform well in large-scale image processing . Using LeNet accuracy of 84 percent and 75 percent on training and validation set is achieved respectively.

E .AlexNet

The architecture is comprised of eight layers in total, out of which the first 5 are convolutional layers and the last 3 are fully-connected. The first two convolutional layers are connected to overlapping max-pooling layers to extract a maximum number of features. The third, fourth, and fifth convolutional layers are directly connected to the fully-connected layers. All the outputs of the convolutional and fully-connected layers are connected to ReLu non-linear activation function. The final output layer is connected to a softmax activation layer, which produces a distribution of 1000 class labels . Using EfficientNet accuracy of 75 percent on training set was achieved .

V. RESULTS

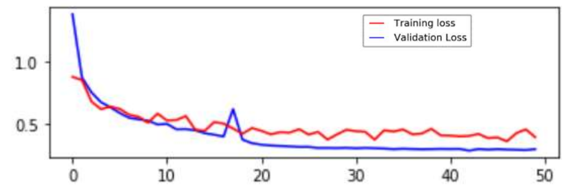
1.TRAINING AND VALIDATION LOSSES

1.1EffcientNet and ResNet

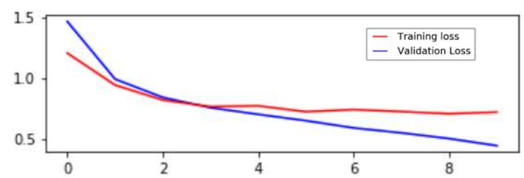
1.1.1Training Loss

1.1.2.Validation Loss

1.2 VGG-19

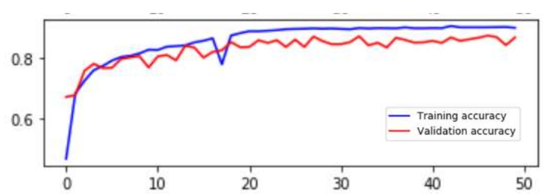


1.3 LeNet

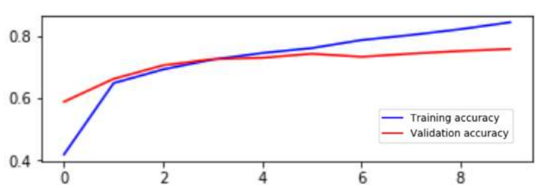


2. ACCURACIES OF DIFFERENT MODELS

2.1 VGG-19



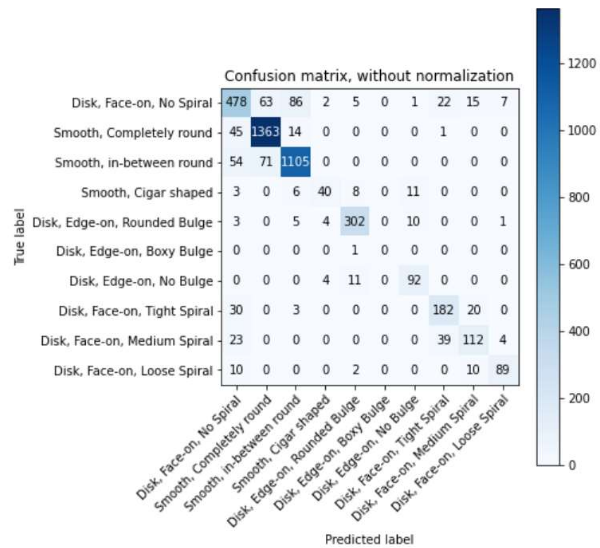
2.2 LeNet

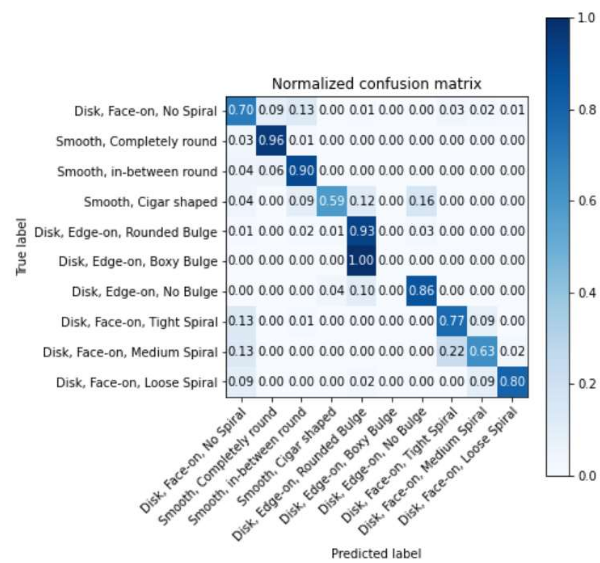


3. CONFUSION MATRIX

Therefore all the models used have achieved accuracy which is more than the base paper [1].The base paper have used the method of SpinalNet whereas different DeepCNN have been used along with augmentation technique to achieve higher accuracies .

The highest accuracy is achieved using VGG-19 model along with data augmentation technique . Accuracy of 89.7 percent and 86.6 percent on training and validation set is achieved The confusion matrix for the same is represented below





VI. CONCLUSION AND FUTURE SCOPE

The project has achieved 10-class classification of different morphologies of galaxy . Therefore Deep learning models such as, Convolutional Neural Network (CNNs)works well in image classifications. Libraries like TensorFlow Keras Tuner, embrace the possibility of quick optimization of the classifier with millions of combinations in hyperparameters.

Future scope would require focusing on implementing the different newer methods to perform the classification .

VII. REFERENCES

[1] Shaiakhmetov, D., Mekuria, R. R., Isaev, R., & Unsal, F. (2021, November). Morphological Classification of Galaxies Using SpinalNet. In *2021 16th International Conference on Electronics Computer and Computation (ICECCO)* (pp. 1-5). IEEE.

[2] Khalifa, N. E. M., Taha, M. H. N., Hassanien, A. E., & Selim, I. M. (2017). Deep galaxy: Classification of galaxies based on deep convolutional neural networks. *arXiv preprint arXiv:1709.02245*.

[3] Kalvankar, S., Pandit, H., & Parwate, P. (2020). Galaxy morphology classification using efficientnet architectures. *arXiv preprint arXiv:2008.13611*.

[4] Schutter, A., & Shamir, L. (2015). Galaxy morphology—An unsupervised machine learning approach. *Astronomy and Computing*, *12*, 60-66.

[5] Kabir, H. M., Abdar, M., Jalali, S. M. J., Khosravi, A., Atiya, A. F., Nahavandi, S., & Srinivasan, D. (2020). Spinalnet: Deep neural network with gradual input. *arXiv preprint arXiv:2007.03347*.