An Effective Galaxy Classification Using Fractal Analysis and Neural Network

Priyanka S. Radhamani
Computer Science and DT, ACE
UEL, University way, E16 2RD,
London, UK.
u2123552@uel.ac.uk

Mhd Saeed Sharif
Computer Science and DT, ACE
UEL, University way, E16 2RD,
London, UK.
s.sharif@uel.ac.uk

Wael Elmedany
College of Information Technology,
University of Bahrain, P.O. Box 32038
Kingdom of Bahrain.
welmedany@uob.edu.bh

***Abstract*—Astronomy is always in a quest of revealing the mysteries of our Universe. There is a vast amount of astronomical data collected and this information comes from stars, galaxies and other celestial objects. While exploring this type of astronomical data, we can identify some complex self-similar patterns. Such self-similar patterns are shown in our own galaxy and are called fractals. This research work has been developed for finding such self-similarity that can be measured from galaxy clusters and this feature can be learned through a suitable neural network. This research work gives an insight about calculating the fractal dimension of galaxy images using a box counting algorithm and training the images using LeNet - 5. The box counting fractal dimension is a specified range of values for each particular class of galaxy. By using the fractal dimension as a primary feature of different classes of galaxy and with the help of LeNet-5 network model classifying the galaxy images into ten specified classes according to its morphological properties. The model produced an accuracy of 74% when implemented with the baseline algorithm. When implemented with LeNet- 5 it produced an accuracy of 96%. The precision recall and f1-Score value of the LeNet-5 model was also calculated. The precision recall and f1-Score value for class 1, class 2, class 4 and class 6 were higher than those of the other classes.**

Keywords—Fractal Analysis, LeNet_5, Box counting, Neural Network

I. INTRODUCTION

When looked into the complex structure of the universe, many fractal patterns or some self-similar patterns can be seen. 'Father of Fractals' described in his famous book [1] "Clouds are not spheres, mountains are not cones, coastlines are not circles, and bark is not smooth, nor does lightning travel in a straight line ." In these lines it is clear that some unknown shapes and patterns also exist in the universe. When we consider a galaxy cluster which contains lakhs of galaxies, that means that repeatedly similar patterns can be identified from one place to another in the cosmos also. Especially a group of galaxies [2], halos and its clusters will be apt for this fractal geometry. Using this property, many fractals geometry can be extracted from a galaxy cluster aiming for the classification.

This paper attempts a morphological classification [3] of galaxies using fractal dimension and neural network. By using the fractal dimension, the topological and special properties of an object, which includes the knowledge about the object's space, can be analyzed. After finding the fractal dimension of different types of galaxies, it can be used as an attribute for the classification of the different classes of galaxies. Like many other fields, Astronomy is enriched with large quantities of data. It is mainly because of the advanced technology development in telescopes in recent years. Many citizen projects like Galaxy Zoo, Sloan Digital Sky survey [4] are used for the morphological classification of galaxies into different classes.

In 1926, Edwin Hubble used a galaxy classification method which is well known called Hubble tuning-fork [5]. The scheme of tuning-fork is shown in figure 1.This method classified the galaxies into three basic classes, which is called elliptical galaxies, spiral galaxies and irregular galaxies. The classification of galaxies is the first stage of understanding their formation. By doing the galaxy classification manually needed experience and it is a tedious job. But automated classification can classify a lot of images in seconds. After feature extraction, the classification method can be done and a good neural network method, the model can predict the specified class.

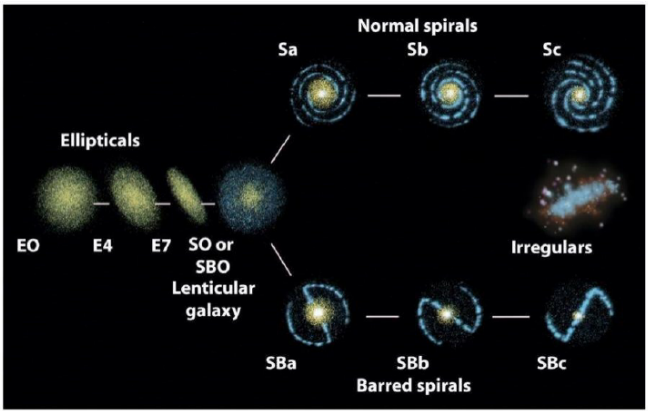

Fig. 1. Edwin Hubble tuning fork scheme[5]

This paper is organized into five sections. The first section describes the Introduction. The second section is the Background and Inspiration which helps for the background of study. The third section deals with the proposed methodology. The result and analysis is described in the fourth section. Finally, the conclusion is included in the fifth section.

II. BACKGROUND AND INSPIRATION

Fractal cosmology is an important part of our universe. Fractal cosmology [6] is the study of fractals that appear in the universe or cosmos. Fractality can be seen in both observational and theoretical basis in the study of cosmology. Benoit Mandelbrot described the fractality distribution in the galaxy. More detailed observation of fractal nature in the universe was found by Luciano Pietronero and his team in 1987.Rather than this first observation of fractal geometry in cosmology was found by Andrei Linde'. The book Discovery of Cosmic Fractals(Yurij Baryshev and Pekka Teerikorpi) reviews the full story of cosmology from ancient times to present. All these are evidence of fractals in the cosmos and galaxy.

In the research work [7], they are automatically classifying the galaxy based on morphology by analysis of fractal dimension. They calculate the experimental Haussdourt - Besi - covich fractal dimension for doing the galaxy morphology classification. They are calculating the fractal dimension of different types of galaxies including ellipticals, spirals and irregular. After calculation they are using that as a feature for the classification. They are also performing the Principal Component Analysis for finding other attributes in the image. They got 100% accuracy for spiral galaxies.

The research paper [8] gives a detailed study of the effect of compression (jpeg) in the classification of remote sensing images. It was found that little bit compromising the accuracy, so they used a method of fractal analysis and deep learning for the classification problem. It is the most important work which gives a good groundwork for my proposal. This paper describes the effect of the compression which will affect remote sensed images while on classification. So they proposed a method for classification using fractal analysis. They are proposing another method which is multiple kernel learning. First, they convert the images into fractal dimension matrices and extract the features. After feature extraction done in three image scales apply the machine learning technique to train a model for prediction. Differential box counting method is used for finding fractal dimensions.

Reviewing the paper on classification of galaxies using the fractal signature [9], this paper is done with the support of the Indian Space Research Organization. They are finding the fractal signature of the nearest galaxy for classification. They found that fractal signature with a specified range is a good parameter for the classification programme. They used the neural network for the classification. By using this the efficiency increased from 92% to 95%.

Self - supervised Learning for Astronomical Image Classification, In the research work of [10], unlabeled astronomical image to pertain the deep CNN to extract the domain specific areas and which may improve the classification with little data which is labeled. The data they used is a portion of the Southern Photometric Local Universe Survey. They are using a robust supervised learning method from astronomical images which are not labeled. They compared the result with Image Net properties. In majority cases their model worked better than ImageNet.

Multiple Kernel Learning algorithms [11] use different kernels instead of using a single one. Multiple kernels use different means of similarity or information from different sources. Some multiple kernel learning algorithms use fewer kernels and support vectors for optimizing. It can improve accuracy, decrease complexity and decrease learning. When reviewed an entirely different paper based on traffic image classification[12], a method of image classification using fractal dimension. They took the fractal dimension for all small and complex images of traffic. It proves that fractal dimension can be a good parameter for the image classification like a person's perspective.

Multiple kernel learning [13] algorithms are used with combined multiple kernels instead of using a single one. Different kernels contain data from different sources. They conducted ten experiments on real data sets using simple linear kernels and eight experiments on three regal data sets using complex kernels. When using simple linear kernels, the accuracy is a bit compromising but when we use a combination of kernels, the accuracy is found to be high.

The paper, [14], studied the galaxy distribution, they found that in a certain range they are fractals and have a fractal dimension of 2.0. Some others have values as 1.2,1.4. It is because of the different luminosities of galaxies. That is the galaxies with different luminosity produce different fractal dimensions.

III. THE PROPOSED METHOD

In this research work, fractal structures that are residing in galaxies, are used for the correct classification of ten classes of galaxy images. Each type of galaxy has a different fractal dimension. Fractal dimension will change according to the different scale, size, shape and intensity levels in images. So by taking the fractal feature as a suitable component for the classification can predict the classes it belongs to. By extracting fractal dimension as a primary feature and using a suitable neutral network it can be learned. By inputting the different features selected, the final output will be produced and they will classify the images into different classes.

A. Dataset Explanation

Data is considered as the most vital part of machine learning. The data set was collected from the galaxy 10 DEcals data set. Galaxy 10 is rich with 17736 images which are 256x256 pixels. In the galaxy 10 decals data set, the galaxy images of 10 different classes are available. Images of galaxy 10 come from DESI Legacy imaging survey and labels are from Galaxy Zoo [15]. Galaxy 10 decals [16] consist of DEcals images and DEcals campaign a, b, c. From the original Galaxy 10 they avoided 17 images of Edge on with Boxy Bulge and formed Galaxy 10 DEcals. The figure 2. shows the classes of the Galaxy 10 DEcals dataset. The accuracy of the labels cannot be guaranteed. Furthermore, Galaxy10 is not a balanced dataset, but it should only be used for research purposes and utilized for educational studies.

Galaxy10 DECals images come from DESI Legacy Imaging Surveys and labels come from Galaxy Zoo.

```
Galaxy10 dataset (17736 images)
├─ Class 0 (1081 images): Disturbed Galaxies
├─ Class 1 (1853 images): Merging Galaxies
├─ Class 2 (2645 images): Round Smooth Galaxies
├─ Class 3 (2027 images): In-between Round Smooth Galaxies
├─ Class 4 ( 334 images): Cigar Shaped Smooth Galaxies
├─ Class 5 (2043 images): Barred Spiral Galaxies
├─ Class 6 (1829 images): Unbarred Tight Spiral Galaxies
├─ Class 7 (2628 images): Unbarred Loose Spiral Galaxies
├─ Class 8 (1423 images): Edge-on Galaxies without Bulge
└─ Class 9 (1873 images): Edge-on Galaxies with Bulge
```

Fig. 2. Classes of Galaxy 10 DEcals [16]

B. Box Counting Method

The calculations of fractals can be done by box counting. The image can be put in an equally spaced box or grid. Then after counting the boxes, which are needed to cover the entire pattern. The fractal dimension can be calculated using the box counting algorithm [16] given below. For analyzing the complexity of certain geometry, fractal dimension can be used as a parameter. The basic idea of a Box counting method is to divide the image with small boxes or grids. Then count the number of boxes covered in the image and then repeat the same pattern with finer or smaller grids or boxes. By decreasing the size of grids repeatedly, the structure of the pattern can be calculated in the image in a more accurate way.

Using the box counting method [17], the fractal dimension can be calculated using the formula (1). When we plot the value of log (M) on the y- axis against the value of log (Ɛ) on x-axis, the fractal dimension will be a slope. In box counting, N is the number of boxes that cover the image and Ɛ is the inverse of box size.

$$D = \lim_{\varepsilon \to 0} \frac{\log N(\varepsilon)}{\log (1/\varepsilon)} \quad (1)$$

C. Feature Extraction and Classification

As the image is in the r, g , b channel, it is essential for doing image analysis methods for finding fractal dimensions. Here a threshold value is being used for putting the boxes or grids for finding the box counting, fractal dimension. The correct threshold value for getting a good dimensional value cannot be predicted. When the value is less than 1, it does not provide a good dimension. The image turned into an ink blob.

It is very important that some particular range of fractal dimensions value will predict the same class of galaxy images. While for some other classes, the box counting algorithm will produce some other particular values. When a threshold of '50' and the image type of 'completely smooth round' galaxy was used, it produced a box-counting fractal dimension of 0.940 as shown in figure 3. Here, the fractal dimension only for 2D images is being calculated.

Fig. 3. Fractal dimension of completely smooth round galaxy

When the category 'Disk, Face on, Medium Spiral' category was taken into consideration, the box counting fractal dimension with a threshold value of 50 was found, it produced a range of value like 1.009. It is shown in figure 4.

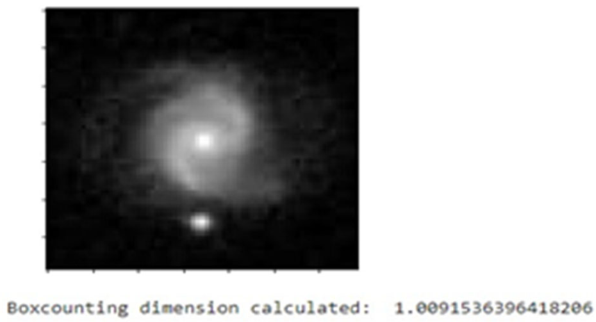

Fig. 4. Fractal dimension of Disk, Face on, Medium Spiral

By analysing the above result, it is clear that the box counting fractal dimension is different for different classes of galaxy images. For some class of galaxy images the fractal dimensions produced will be inside a particular range of values. From these it is clear that finding fractal dimension from a galaxy is a good way of extracting a signature feature. It can predict the class of galaxies with less error rate.

D. Baseline Model

For building a model, First, a baseline model[18] was used. Baseline model is the basic or a reference for our real training model. The baseline model is used because it will predict how good our actual model is at evaluation scale. Baseline model is the critical model for the real training model.Before using the real model, first the Baseline model was compared to the LeNet-5. If the baseline model performs well, the struggle of creating a more complicated model could be avoided. If a baseline model produces a lot of errors and problems in the data, then the more complicated model will be the worst one. By a baseline model the behavior of data could be observed.

E. LeNet-5 Implementation

The LeNet -5 [19] is a convolution neural network, which consists of 7 layers in which 3 convolutional layers, 2 subsampling and 2 fully connected layers. LeNet - 5 is an old classic convolutional neural network architecture developed by Yann Andre Lecum, Yoshua Bengio Et.al . It is first used for handwritten MNIST number classification.

Later this LeNet - 5 became a foundation model like alexNet and VGG. The basic LeNet-5 architecture is illustrated in figure 5.

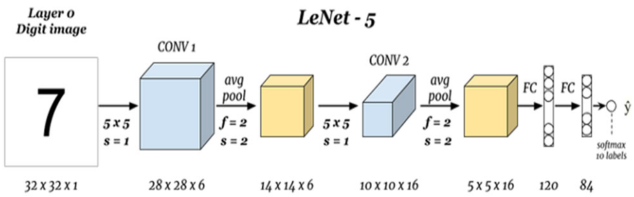

Fig. 5 LeNet - 5 Architecture [19]

In the first layer of the CNN, this is not considered to be a layer of learning because nothing is learnt in this layer. The input layer is designed as if to take 32x32 and these are the information of images that are given to the following layers. As we know the MNIST dataset image is considered as 28x28 in dimension, for receiving the same dimension of MNIST dataset to the input layer, images are added to 28x28.

The fractal dimension found by the box counting algorithm can be used for more crystal clear classification of images. Along with the named labels, the fractal dimension can also be taken as an interesting feature thus by double ensuring, the class using the LeNet -5 would be predicted. The figure 6, shows the architecture of the proposed model.

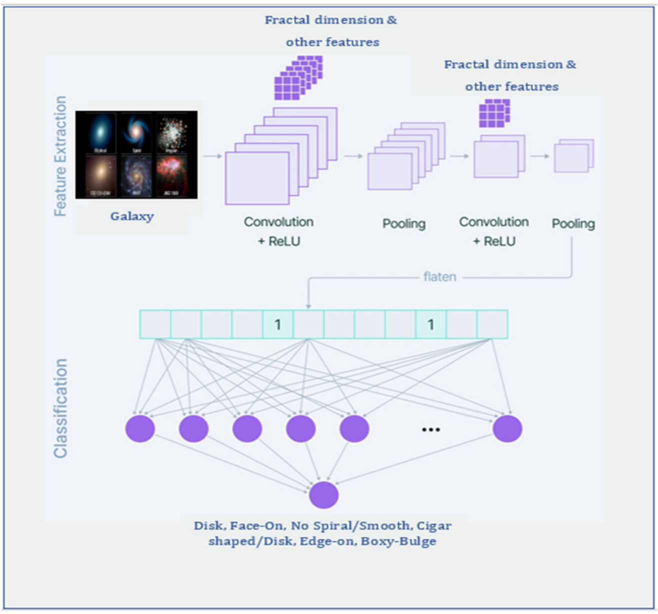

Fig. 6. Architecture diagram of proposed model

III. RESULT AND ANALYSIS

The galaxy 10 DEcals dataset consists of 17736, 256x256 pixel coloured images. The command below shows the loading of the dataset using the astroNN using Python 3.10.6

```
! pip install astroNN
```

The figure 7 shows the random selection of images present in the Galaxy 10DEcals dataset. It will randomly show about 20 images, which are of different classes out of 10 classes of galaxy images. It shows five subplots images in four rows and five columns. The total 20 images and its specified in each class.

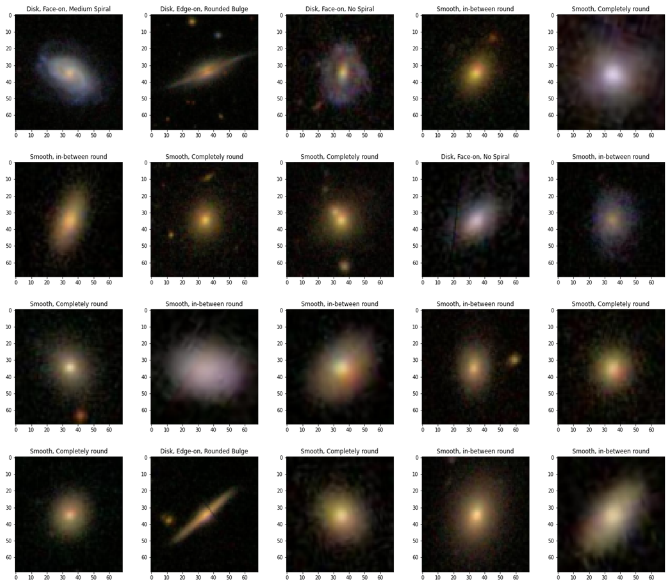

Fig. 7. Random selection of 20 images in galaxy 10

A. Accuracy prediction of Baseline Model

The baseline model with EPOCH 15 will produce a final accuracy of 74 percent as shown in figure 8. It is actually good for a classification problem. As the baseline model produced 74% accuracy, another model can be considered for the classification. So, LeNet -5 was implemented after the baseline model. The graphical representation for Baseline model accuracy is shown in figure 9 and model loss is shown in figure 10.

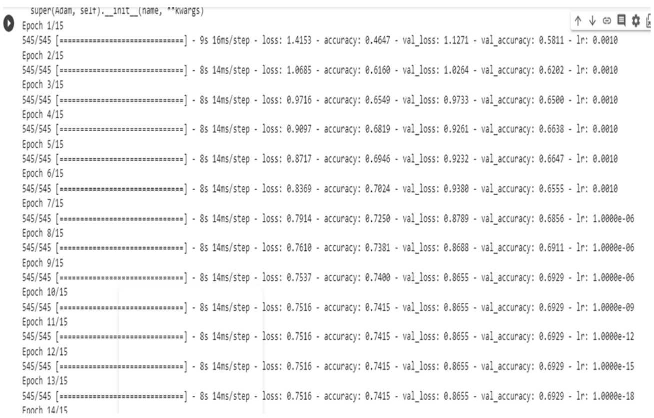

Fig. 8. Accuracy Parameter of Baseline model

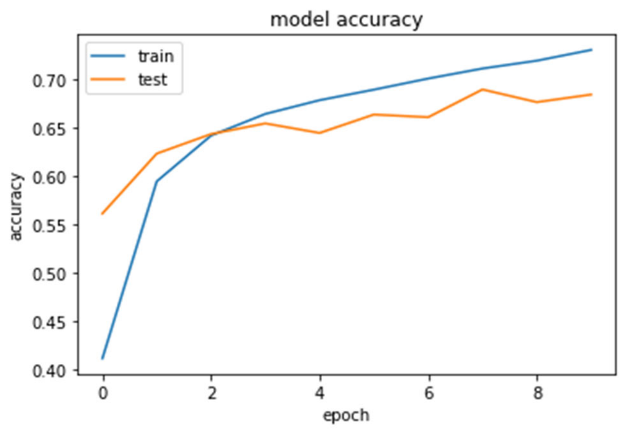

Fig. 9. Baseline Model Accuracy

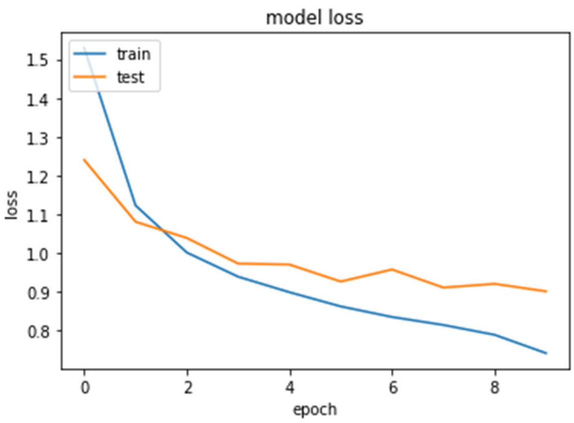

Fig. 10. Baseline model loss

B. Accuracy prediction of Lenet-5 Model

The accuracy produced for this model is 96%. It is a very good accuracy score for the LeNet - 5 classification of galaxy images. The figure 11, shows the accuracy and loss gained through the training process of LeNet - 5. The accuracy is 96% in this dataset. So, the prediction ability for the LeNet - 5 model for the classification of galaxy images was checked using the 'predict' function. The graphical representation LeNet - 5 model accuracy is shown in figure 12 and model loss is shown in figure 13.

```
Epoch 1/15
545/545 [==============================] - 76s 139ms/step - loss: 1.3495 - accuracy: 0.4741 - val_loss: 0.9995 - val_accuracy: 0.6491 - lr: 0.0010
Epoch 2/15
545/545 [==============================] - 74s 136ms/step - loss: 0.8797 - accuracy: 0.6860 - val_loss: 0.8334 - val_accuracy: 0.6947 - lr: 0.0010
Epoch 3/15
545/545 [==============================] - 73s 135ms/step - loss: 0.7361 - accuracy: 0.7304 - val_loss: 0.7568 - val_accuracy: 0.7308 - lr: 0.0010
Epoch 4/15
545/545 [==============================] - 76s 140ms/step - loss: 0.6445 - accuracy: 0.7631 - val_loss: 0.7209 - val_accuracy: 0.7358 - lr: 0.0010
Epoch 5/15
545/545 [==============================] - 74s 136ms/step - loss: 0.5665 - accuracy: 0.7896 - val_loss: 0.8237 - val_accuracy: 0.7165 - lr: 0.0010
Epoch 6/15
545/545 [==============================] - 74s 136ms/step - loss: 0.5021 - accuracy: 0.8173 - val_loss: 0.7353 - val_accuracy: 0.7457 - lr: 0.0010
Epoch 7/15
545/545 [==============================] - 74s 136ms/step - loss: 0.4358 - accuracy: 0.8416 - val_loss: 0.7207 - val_accuracy: 0.7501 - lr: 0.0010
Epoch 8/15
545/545 [==============================] - 74s 135ms/step - loss: 0.3696 - accuracy: 0.8688 - val_loss: 0.7581 - val_accuracy: 0.7324 - lr: 0.0010
Epoch 9/15
545/545 [==============================] - 73s 134ms/step - loss: 0.3171 - accuracy: 0.8886 - val_loss: 0.8182 - val_accuracy: 0.7363 - lr: 0.0010
Epoch 10/15
545/545 [==============================] - 74s 135ms/step - loss: 0.2582 - accuracy: 0.9127 - val_loss: 0.8233 - val_accuracy: 0.7455 - lr: 0.0010
Epoch 11/15
545/545 [==============================] - 73s 135ms/step - loss: 0.2120 - accuracy: 0.9307 - val_loss: 0.8604 - val_accuracy: 0.7443 - lr: 0.0010
Epoch 12/15
545/545 [==============================] - 74s 136ms/step - loss: 0.1725 - accuracy: 0.9426 - val_loss: 0.9668 - val_accuracy: 0.7260 - lr: 0.0010
Epoch 13/15
545/545 [==============================] - 75s 137ms/step - loss: 0.1442 - accuracy: 0.9544 - val_loss: 0.9969 - val_accuracy: 0.7345 - lr: 0.0010
Epoch 14/15
545/545 [==============================] - 75s 137ms/step - loss: 0.1072 - accuracy: 0.9676 - val_loss: 1.0210 - val_accuracy: 0.7345 - lr: 0.0010
Epoch 15/15
```

Fig. 11. Accuracy parameter of LeNet - 5

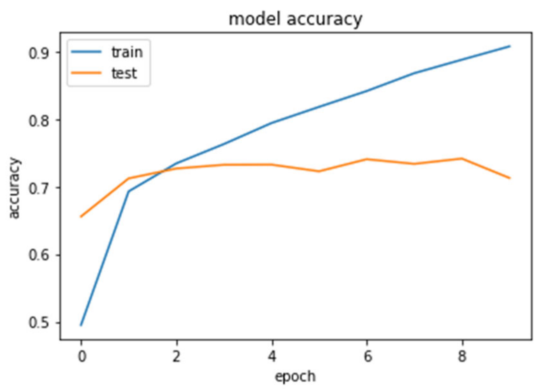

Fig. 12. Lenet 5 Model Accuracy.

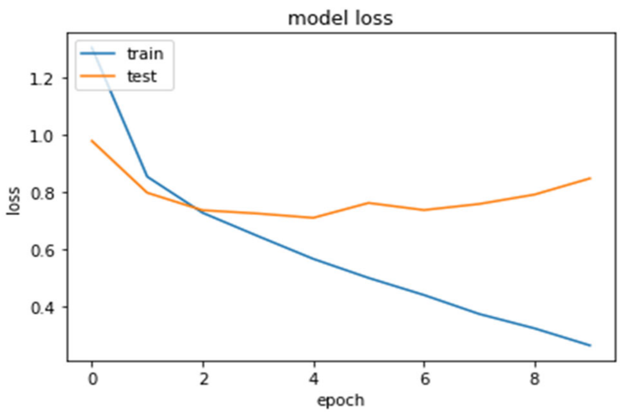

Fig. 13. Lenet -5 Model Loss

C. Precision, Recall of model

Table I shows the precision, recall and f1 score of the model. From this information, it is clear that for classifying the classes 0, 1, 2, 4 has a high precision. Not only the precision for these classes, the model got high recall value also. Specially for the class one and two have maximum precision and recall value.

While taking classes five, eight and nine, the precision and accuracy is very less. Among these, the precision and recall value of class five became zero. It is because ' Disk, Edge on, Boxy Bulge' has only seventeen testing samples in the dataset. That is the reason for the prediction of the model getting confused with classes having less number of data samples in it.

D. Confusion Matrix

If each class of data has a required amount of samples, the model will predict the classification in a more accurate way. This model is good for predicting an unbalanced dataset like galaxy 10 DEcals. For getting a clear idea, the confusion matrix could be used.

The confusion matrix will give a clear idea for the prediction of the six classes. i.e. class 1, class 2, class 0, class 4, class 7 and class 8. The figure 14, shows the confusion matrix for the predicted model.

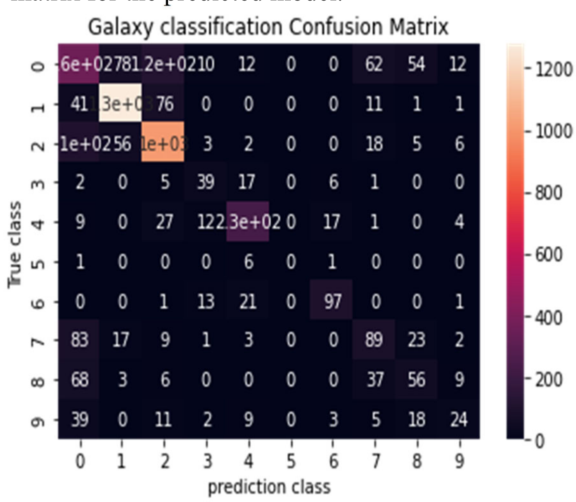

Fig. 14. Confusion Matrix for the predicted model.

Table I: Precision, recall and f1 score of LeNet - 5

Classes	Precision	Recall	F1-score	Support
0	0.56	0.48	0.51	711.00
1	0.85	0.92	0.88	1400.00
2	0.83	0.85	0.84	1266.00
3	0.38	0.35	0.37	66.00
4	0.74	0.81	0.77	290.00
5	0.00	0.00	0.00	2.00
6	0.71	0.76	0.73	119.00
7	0.45	0.50	0.47	226.00
8	0.36	0.23	0.28	176.00
9	0.48	0.31	0.37	101.00
Accuracy			0.74	4357.00
Macro Avg.	0.54	0.52	0.52	4357.00
Weighted Avg.	0.73	0.74	0.73	4357.00

From the above confusion matrix, the model is very good in the prediction of class one and class two. Class one which is ‘ smooth, completely round’ which has an image of 6997, class two which consists of 6292 images and which belongs to the class ‘ smooth, in between, round’. The classes which have more samples have predicted good results in this model. But when we look at the confusion matrix, the class zero and class 4 are also far apart in the prediction. It is because it also has more samples in galaxy10 DEcals. Like class zero has 3461 and class 4 has 1534 data samples.

IV. Conclusion

Machine learning models can perform an accurate classification with an effective time factor. If classification is done by humans it takes more time whereas the automated ML classification is not time consuming. Moreover, it is difficult for a human to distinguish the different galaxy classes in their correct category because of their similarity level. It is very important in this case that a good parameter is chosen for the accurate classification. The fractal dimension distinguishes each class correctly because the calculated fractal dimensions for each class have different values. For each class it produces a specified range of values which is always decimal point values. When the dataset was analysed, there was a chance for mixing up with class 2 and class 3. Class 2 consists of ‘round, smooth’ galaxies and class 3 consists of ‘in between, round, smooth’ galaxies. These two class images look somewhat similar. The proposed model finds the box counting dimension of each class as a parameter for the classification. The box counting dimension produces a specified range of values, so it can differentiate each class accurately when compared to other existing methods. The proposed model predicted a 96% accuracy, which was higher than expected and distinguished between these two classes.

As a future enhancement, this idea could be used with other complex networks by taking the fractal dimension as a feature. When a very good dataset is employed with a good number of samples in it, the accuracy will be higher. It is important that the threshold value is taken into consideration when calculating the fractal dimension by using the algorithm of box counting. The proposed method uses a threshold value of 50 for the box counting dimension calculation.

References

[1] Benoit B. Mandelbrot, “The fractal geometry of nature, ” W.H. Freeman and Company: United States,January 1983

[2] R.De La Fuente Marcos; C.De La Fuente Marcos Monthly Notices of the royal astronomical Society: OUP, 2006

[3] Jorge de la Calleja, Elsa m. de la alleja, Hugo Jair Escalante, “Fractal dimension analysis for automatic morphological galaxy classification”,June 2017

[4] Brue Margon, “The Sloan Digital Sky Survey”: University of Washington, 1998, https://www.sdss.org/

[5] Roberto G. Abraham, “Explorations in Hubble Space: A quantitative tuning fork”, Department of Astronomy, University of Toronto,Canada. fig 1:Hubble tuning fork,August 2000

[6] Jonathan J. Dickau., “Fractal cosmology”, Alternative Cosmology Group, Poughkeepsie, United States, 2009.

[7] De la Calleja, Jorge, Fuentes, Olac, “Automated Classification of Galaxy Images”, September 2004

[8] Zhenzhong Chen ,Ye Hu, Yingxue Zhan, “Effects of compression on remote sensing image classification based on fractal analysis”, IEEE, 2019

[9] S. Lekshmi., Et al “Galaxy classification using fractal signature”,University of Kerala, India, 2003

[10] Ana Martinazzo, Mateus Espadoto, Nina S. T. Hirata, “Self-supervised learning for astronomical image classification”, Institute of Mathematics and Statistics, April 2020

[11] Mehmet Gonen, Ethem Alpaydın, “Multiple kernel learning algorithms” ¨ GONEN@BOUN.EDU.TR, July 2011

[12] CAO Wen-lun', SHI ZHong -ke 1, FENG Jian -hu, “Traffic Image Classification Method Based on Fractal Dimension”, April 2006

[13] Mehmet Gonen , Ethem Alpaydın, “Localized algorithms for multiple kernel learning”, Department of Computer Engineering, Bogazici University, Turkey,March 2013

[14] Wen ZhengDeng Zu-ganLiu Yong-zhenXiaXiao-yang, “Fractal structure in the large-scale distribution of galaxies, February 2022, Billkeel,https://blog.galaxyzoo.org/category/paper/, February 2022

[15] Henry Leung,, “astroNN Documentation” https://astronn.readthedocs.io/en/latest/galaxy10.html,June 2017 Fig. 2. Classes of Galaxy 10 DEcals

[16] Song Xue, Xinsheng Jiangg, vJimiao Duan, “A new box-counting method for image fractal dimension estimation” IEEE, 2017

[17] Moonzarin Reza, “Galaxy morphology classification using automated machine learning”, Department of Physics and Astronomy, A&M University, USA, 2021

[18] Tsung-Han Chan, Kui Jia, Shenghua Gao, Jiwen Lu, Zinan Zeng, and Yi Ma , “PCANet: A Simple Deep Learning Baseline for Image Classification”, 2014.