Classification of Galaxy Morphologies using Artificial Neural Network

Manish Biswas[1], Ritesh Adlak[2],Swapnil Sonawane[3], Venu Bajaj[4], Shubham Parate[5], Prof. A M Magar[6]
[1,2,3,4,5]Department of Information Technology, SPPU, Maharashtra, India
[6]Professor, Department of Information Technology, SPPU, Maharashtra, India

***Abstract*—From the beginning of the modern astronomy, there has been a huge collection of image data set from various projects and surveys like Galaxy Zoo and Sloan Digital Sky Survey, which is used for the study of the universe. But they do not provide classification information about the galaxy images it contained. During the past decade the manual classification of the objects in these images are becoming obsolete due to the use machine learning methods which makes this process automated. This tool makes use of artificial neural network to classify the galaxy morphologies into three classes. This algorithm classifies galaxies into spiral, elliptical and irregular.**

***Index Terms*—Artificial Neural Network, Data Mining, Computer Vision, Galaxy Morphologies, Classification, Astronomy**

I. INTRODUCTION

There has been a huge collection of data in the field of science after the advancement in the field of computing and hardware used to make observation. Due to the introduction of new telescope, sensors and charge couple device it has pushed the boundaries of astronomy making new scientific opportunities and challenges. Now large amount of data is being collected from surveys, which are more precise, clean and imaging data in high definition [1]. Survey such as Sloan Digital Sky Survey uses the Apache point observatory located in the New Mexico; it provides image data with photometric and spectral parameter as well. Each night this survey produces 200 Giga Bytes of data and contains a catalogue of approximately 500 million objects [1]. To this day million field images have been taken containing more than 200 million galaxies, which forms the major subset of the survey catalogue.The Galaxy Zoo is a citizen science project, which invites people to assist in the morphological classification of large number of galaxies. Its data set is made up of million galaxies imaged by Sloan Digital Sky Survey. Many discoveries have been made using galaxy zoo data providing hints for establishing correlation processes governing galaxy evolution [10].

Fig 1: NGC 4414 Spiral Galaxy

These data set provided by the project involve many important information valuable to an astronomer. It is difficult to quantize the morphologies of galaxy and therefore it does not provide any classification information [1]. Since modern astronomical sky, survey contains millions of objects of interest (Galaxy). Analysis of these data manually is therefore not feasible and time consuming. Such types of problems require an automated process in order to classify galaxy according to its morphology. Therefore, machine-learning technique can be used to automate such classification problem, where algorithm learns the characteristics of the categories and build a model based on this data this model can be used to predict other objects having no classification information. This will help to build large catalogues of galaxy based on their categories, which will have various application according to the astronomer.

The artificial neural networks are data processing system used for the analysis of data. it is based on the neural network present in the cortex of an animal brain. Artificial neural network is one of the most popular learning algorithms that has been used for solving many real-life problems this paper applies artificial neural network to learn galaxy images from the image data set provided by Galaxy Zoo based [3]. Section 2 provides introduction to morphological classification of galaxy. Section 3 provides description of support vector machine. Section 4 contains methodology and numerical analysis. Conclusion is further given in the section 5.

II. GALAXY MORPHOLOGICAL CLASSIFICATION

In the field of astronomy, classification is an important part in understanding observed phenomena. As astronomers have observed galaxies, they have noticed certain patterns in their morphologies, or physical structure. They have found that it is possible to categorize galaxies as belonging to a certain group. One of the most widely used classification schemes today divides galaxies into three broad categories known as spiral, elliptical, and irregular [1]. Their flattened disk characterizes spiral galaxies with the presence of spiral arms. Elliptical galaxies have a much smoother appearance and are primarily in the shape of an ellipse. The irregular category contains galaxies that cannot be placed easily in either of the two categories. These galaxies may contain distortions or have no definite structure. In many cases, such a galaxy is the result of a merging or collision of two or more galaxies. It is possible to break each of these categories down into further detailed categories. For example, the presence of a barred center or the ellipticity of a galaxy can be used to create further categories. For the purpose of this paper, only the three primary categories will be considered.

(a) (b)

Figure 2. An example of two morphology categories: (a) the spiral galaxy M101 (credit: NASA, European Space Agency (ESA), K. Kuntz (Johns Hopkins University), F. Bresolin (University of Hawaii), J. Trauger (Jet Propulsion Lab), J. Mould (National Optical Astronomy Observatory), Y.-H. Chu (University of Illinois, Urbana), and Space Telescope Science Institute (STScI)), and (b) the elliptical galaxy NGC 1132 (credit: NASA, ESA, and the Hubble Heritage Team (STScI/Association of Universities for Research in Astronomy (AURA))-ESA/Hubble Collaboration).

Classifying these galaxies into categories has great value to astronomers. By studying how structures of galaxies in the same category are similar, a better understanding of the processes that created these structures can be gained. In the past, large catalogues of classified galaxies have had many practical applications. Astronomers have used such catalogues to test theories about the universe. For instance, this has been useful in one such study that examined the relationship between spin direction and type of galaxies. Other studies have required large numbers of classified galaxies to analyse the properties of dust in elliptical galaxies. Several different machine-learning algorithms have been applied to photometric data extracted from galaxy images in order to automate classification [9]. One of the most prominent examples of this is the use of neural networks. This has been one of the most successful attempts at automatic classification to date. This method has produced some of the highest classification accuracies so far. Other algorithms that have been applied to galaxy morphologies include the use of regression, which involves building an eigenimage for each galaxy image. This eigneimage consists of a set of eigenvectors that completely represents the data [1]. Shapelets are another approach to using machine learning with this problem. The idea behind a shapelet is to decompose an image of a galaxy into a set of basis vectors [1].

III. ARTIFICIAL NEURAL NETWORK

First introduction to working of neurons in ANN was made in 1949 by neurophysiologist Warren McCulloch and mathematician Walter Pitts. They modeled a simple neural network using electrical circuits.Neurons are single computing system but are capable of solving complex problem. The essential features of neurons include Massive Parallelism, high interconnection with each other. The artificial neuron can receive one or more inputs and produce sums This project aims to use Feed Forward Network, which includes perceptron arranged in layers with first layer taking in inputs and the last layer producing outputs. The middle layers are called as Hidden Layers. to bring out an output. Usually each input has different weight, and the sum is then passed through a non-linear function known as an activation function or transfer function. [8]. The Learning Rate modifies the parameters of neural network to favour the favoured output with minimum number of errors. ANNs incorporate the two fundamental components of biological neural networks Nodes – Neurons, Weights – Synapses.

Single layer Neural Network consist of single hidden layer with Input layer taking as input and output layer producing output. This Network are known to handle linear decision surfaces but if we want to capture non-linear functions then we go for Multi layered Neural Network. [9]

Multi layered Neural Network consists of input layer, output layer and more than one hidden layers which are capable of producing more accurate results. Main advantage is that it can represent any Boolean function and continuous function as long as hidden units are sufficient and appropriate activation function is used.

The method for updating weights in Neural Network to achieve the accuracy is called as Back Propagation. It is aimed such a way that it results with minimum errors. The error is visible only at output layer and that error is propagated back to previous layers. The error is propagated back to number of layers present in Neural Network. At last, the new weights are updated and the iteration is again followed. As large is the magnitude of error at output layer, the same amount (proportion) of error is propagated back to previous layer.

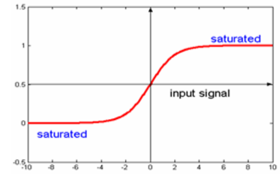
The perceptron consists of weighted sum, activation function and threshold processor (Bias). The Bias can be considered as the propensity (a tendency towards a way of behaving) of the perceptron's. The weighted sum is calculated with the help of following formula.

$$\sum_{i=1}^{m} bias + (w^i x^i)$$

Fig 3: Weighted Sum

The perceptron models a neuron by taking the input as weighted sum and sends the output as 1 if the sum is greater than some threshold value, called as Activation Function.

Activation Function also known as Transfer function in neural network is a mathematical function, which gives the corresponding values for the input values of neuron. It also predicts the behavior of the component. The Activation Function uses Sigmoid Function. This function is useful because it offers the ability to apply differentiation technique using Sigmoid Curve, which is used in Backpropagation technique.



$$f(x) = \frac{1}{1+e^{-\beta x}}$$

(a) (b)

Fig 4. (a) Sigmoid Curve (b) Sigmoid Function

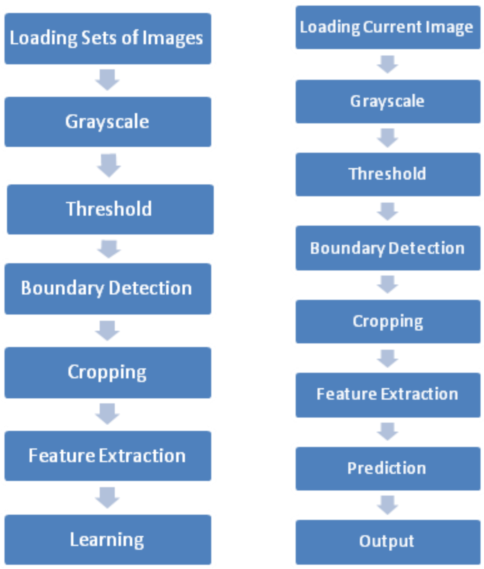
The errors, which are observed at output layer, are calculated as. [8]

Delta Output[i] = Output[i] * (1.0 – Output[i]) * (Target[i] – Output[i])

IV. METHODOLOGY

The working of system gets divided into two parts i.e. training of images and predicting morphologies of galaxies. The project implementation aims to provide three provisions in front end designed using Java swing. They are object localization; Training and Testing. Object localization is responsible for only Blob detection which is recognizing object of interest from image input. In training phase, it consists of loading sets of galaxy images from the surveys and projects like galaxy zoo. It also provides choice to set threshold, minimum height and minimum width. The path for dataset is specified in backend which consist of three folders namely Circular, Spherical and Others. The known images are used to train ANN model. After loading the images then they are grayscaled using averaging formula and binary images are created using thresholding. Global thresholding is used in project. The region of interest is converted into white color pixel and rest are converted to black pixels.

After this blob of white pixels are generated in pictures this can be detected and further used for classification. After this stage the images undergo feature extraction. Features included to detect the type of galaxy are Aspect Ratio, Black Pixel percentage and White Pixel percentage also giving white pixel count and black pixel count on console.



(a) (b)

Fig 5: (a) Training and (b) Prediction of Morphology of Galaxy

Once the features are extracted from input, ANN algorithm is applied on input images with certain number of iterations.

In testing phase the unknown image to be identified is loaded. The image goes through the same methods like Grayscale, Thresholding, Cropping, etc. as mentioned in training phase except the method named learning. Instead, it predicts the image using ANN algorithm. Once the image is identified, it is then known to user.

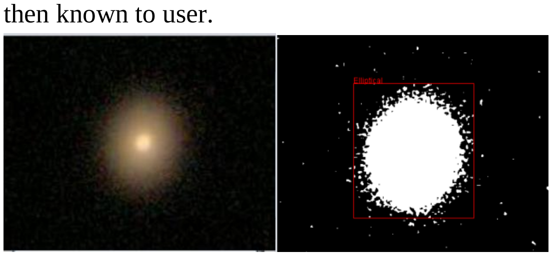


Fig 6: Prediction of Morphologies of Galaxy

V. NUMERICAL RESULTS

A total of six experiments were run. three data samples were used for each one these. Each training sample contained 3,000 galaxies, with 1,000 objects from each category. A separate testing sample for each of the training samples was used to determine the classification accuracy. These testing samples also contained 3,000 objects.

The first three experiments 1, 2 and 3 respectively measured the ability of the ANN to distinguish between each category within the training data which was again given for testing. The accuracy was calculated by testing the training model by feeding the same training data or sample to test by checking whether it was trained to classify the data within those three classes to its highest accuracy. In experiment 4,5 and 6 respectively were performed to calculate the testing of the ANN to test the testing sample which were not classified and not included during the training of the model.

The results of experiment 1-3 are presented in Table 1. The table summarizes the average accuracies and runtimes of the five samples in three categories. The ANN was able to distinguish between spiral and elliptical at the highest accuracy than that of irregular. The ANN was able to distinguish ellipticals than the other galaxies, with their featureless disks, are much easier to characterize in the feature set. The classification required the least average runtime of 4.0 minutes with an iteration of 1000. There was a correlation between the run times and the accuracy of the classification. Lower classification accuracy corresponded with a higher runtime. The least accurate classification between spiral and irregular had the highest runtime.

The summary results from the different implementations of the three class ANN in experiments 4- 6 are presented in Table 2. The overall accuracy is presented as well as the percentage correctly classified for each of the three categories. Separating the spiral category and elliptical category first both resulted incomparable performance in terms of accuracy. The total percent of objects correctly classified at the highest with 83.4 % accuracy respectively and least accuracy of 80.7 %, though it still provided a useful accuracy. The primary reason for the noticeable drop in percent of irregulars correctly classified is that the ANN has difficulty in performing a classification with irregulars. Due to the broad range of feature values of the irregular galaxy are difficult to distinguish between the two other categories.All three multi-classifications implementations resulted in very high accuracies for the elliptical category. This further shows the ability of the ANN to distinguish this category from the others. The spiral category also achieved percentages as high as 88.9%, though this is noticeably lower than the elliptical category. The irregular category contained the lowest percentages out of the three.

Table 1. Test result for training samples after 1000 Iterations:

Experiment	Accuracy (%)	Spiral (%)	Elliptical (%)	Irregular (%)	Time (m)
1	90.2	90.5	91.2	88.8	4.7
2	88.7	92.3	90.9	83.0	4.0
3	90.9	92.5	94.1	86.1	4.3

Table 2. Test result for testing samples after 1000 Iterations

Experiment	Accuracy (%)	Spiral (%)	Elliptical (%)	Irregular (%)	Time (m)
4	80.7	79.7	91.1	71.3	4.6
5	82.6	84.5	94.3	69.0	4.2
6	83.4	88.9	87.5	73.7	4.9

VI. CONCLUSION

Artificial Neural Network proves a trustworthy method to the problem of classifying galaxy morphologies. As modern sky surveys, COSMOS surveys and surveys like Dark Energy Surveys (DES) and upcoming Large Synoptic Survey Telescope (LSST) continue to produce more and more data, machine-learning algorithms such as the ANN will be required to analyze the data. The automation of the galaxy classification process through these tools will save several hours that are required to manually classify astronomical objects. Astronomers will be able to make use of the addition of this information to these catalogues of galaxies to test many theories about the universe and gain a better understanding about the evolution of galaxies. Many Research Students will be able to use these tool for their projects in exploring new astronomical objects. There can also be extension to these project by classifying the morphology of galaxies into more than three categories.

Additionally, the enthusiastic ones can add more astronomical objects like Stars, Nebulae etc. for classification. The results of two-class separation between the spiral and elliptical categories are particularly noteworthy.

The results will demonstrate the challenge of using classification algorithm on the irregular category of galaxies. As the category contains galaxies with no definite structure that do not meet the requirements of either of the other categories, it is difficult to characterize the category within the feature set, and it is less diagnostic. It is therefore frequently confused with other objects.

Future scope and research can be done to focus on improving the runtime and accuracy of the algorithm. When the system is dealing with either massive datasets form surveys or telescope, the accuracy of the algorithm is highly dependent on its efficiency. Efficiency might be improved by increasing the iteration of algorithm.

REFERENCES

[1] Matthew Freed and Jeonghwa Lee. Application of Support Vector Machines to the Classification of Galaxy Morphologies. International Conference on Computational and Information Sciences.

[2] John Jenkinson, Arytom Grigoryan, and Mehdi Hajinoroozi, Raquel Diaz Hernandez, Hayde Peregrina Barreto, Ariel Ortiz Esquivel, Leopoldo Altamirano, and Vahram Chavushyan. Machine Learning and Image Processing in Astronomy with Sparse Data Sets.IEEE International Conference on Systems, Man, and Cybernetics October 5-8 2014, San Diego, CA, USA.

[3] Jan Kremer, Kristoffer Stensbo-Smidt, Fabian Gieseke, Kim Steenstrup Pedersen, and Christian Igel,Big Universe, Big Data: Machine Learning and Image Analysis for AstronomyUniversity of Copenhagen.

[4] Eduardo Machado, Marcello Serqueira, Eduardo Ogasawara, Ricardo Ogando, Marcio A. G. Maia, Luiz Nicolaci da Costa, Riccardo Campisano, Gustavo Paiva Guedes, Eduardo Bezerra. Exploring Machine Learning Methods for the Star/Galaxy Separation Problem.2016 International Joint Conference on Neural Networks (IJCNN)

[5] NGC 4414 Spiral Galaxy: - http://astronomy.swin.edu.au/cosmos/S/Spiral+Galaxy

[6] The spiral galaxy M101:- https://en.wikipedia.org/wiki/Pinwheel_Galaxy

[7] The elliptical galaxy NGC 1132:- https://www.nasa.gov/image-feature/goddard/hubble-views-a-young-elliptical-galaxy

[8] Gao Ningning. Artificial Neural Network Management System. 978-1-4244- 7618-3 /10/$26.00 ©2010 IEEE.
[9] Guoqiang Peter Zhang. Neural Networks for Classification: A Survey. IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART C: APPLICATIONS AND REVIEWS, VOL. 30, NO. 4, NOVEMBER 2000.
[10] K. N. Abazajian. The seventh data release of the sloan digital sky survey. Astrophysical Journal Supplement Series, 182, 2009.