Objective:

Galaxy morphological classification is a system used by [astronomers](https://en.wikipedia.org/wiki/Astronomer) to divide [galaxies](https://en.wikipedia.org/wiki/Galaxy) into groups based on their visual appearance. This has a been a manual process till date and extensive research is being to automate this process. Our aim is to predict the morphology of a galaxy by answering 11 questions in terms of probability.

Dataset Details:

The Galaxy Zoo Dataset used is taken from the Kaggle challenge. The dataset consists of 61,578 RGB galaxy images of 424 x 424 pixels. Few sample images from the dataset are shown in Fig. 1. Each of these images has been annotated by human volunteers by answering to the 11 questions and accordingly, probabilities are provided for a total of 37 answers. Due to computational limitations, we used 10,000 images and 70:30 splitting was done for training and testing respectively.

Proposed Methods:

* Regression: Linear, Lasso and Ridge Regression has been implemented for the purpose of prediction.

----------------Architecture---------------------

* Convolution Neural Network: Alexnet architecture, with 5 convolution layers and 3 fully connected layers has been used. It provides probabilistic estimation per answer for each of the 11 questions.

----------------Architecture---------------------

Feature Extraction:

We have used 3 features to describe the image: Histogram Of Gradient(HOG), Local Binary Pattern(LBP) and Image Moments.

---------------Images----------------------------

Feature Analysis:

For each of the proposed regression techniques, we analyzed the effect of each

feature on different questions separately.

Results:

The evaluation metric used for experiments was Root Mean Square Error(RMSE). We used the optimal feature for each question and applied regression technique and compared them with each other and also with results given by CNN.

--------------Image------------------

Conclusion:

We find that Ridge and Lasso regression perform better than both Linear regression and CNN. Image moment as a feature can be seen as the best for most of the questions. Primarily because morphology depends on shapial features and image moment describes that.