

# MOVIE GENRE CLASSIFICATION



Deep Learning Final Project

Nicolás Lozada  
Camilo Yate

201727313  
201630371

# Movies Datasets

Two different datasets were provided in order to predict the movie genre: The movie poster and its plot.

The dataset is formed by **11,278** movie posters and plots. The data set has been divided in training and testing set: **7,895** (70%) and **3,383** (30%) movies respectively.

In posters dataset each image is stored in two analogous version: full color and grayscale. For classification, in order to get better predictions full color version of each poster is going to be used.

For movie plots each movie has a complete description of the plot in natural language format. In this dataset text mining techniques will be applied for getting high performance in classification.

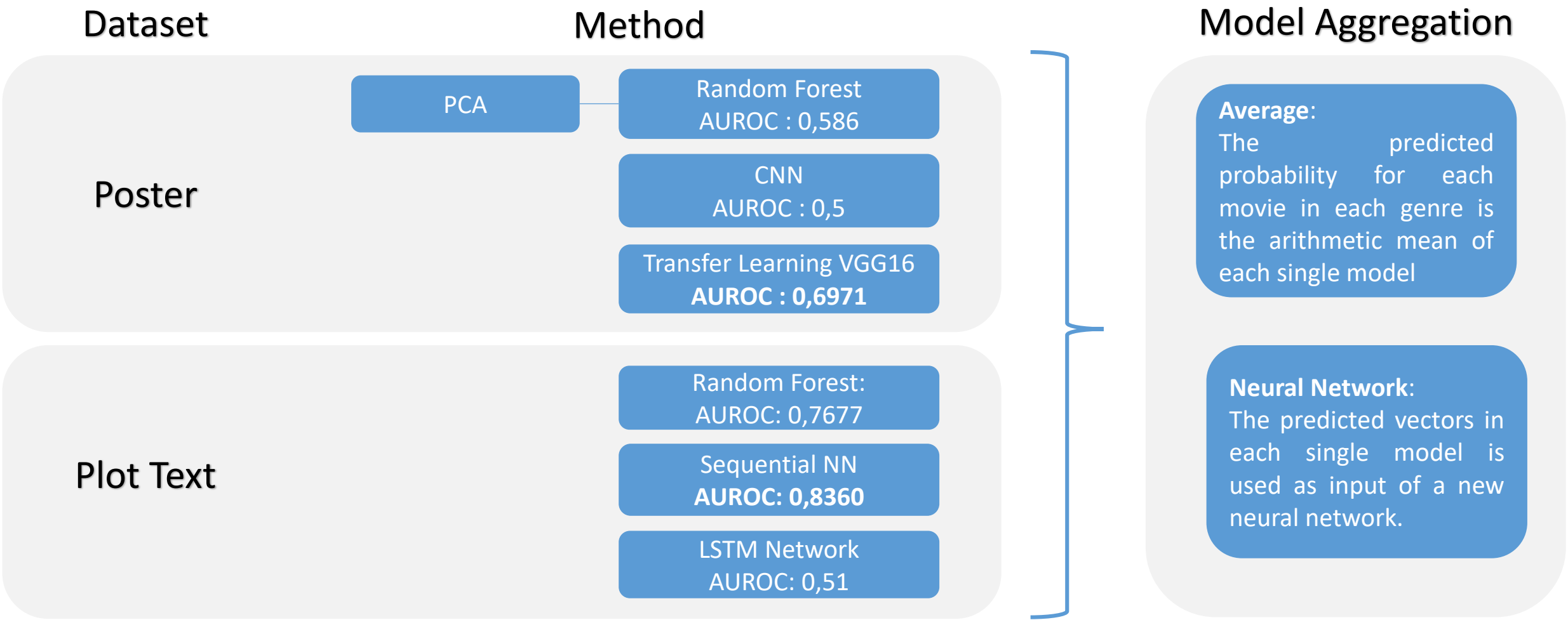
In training set, each poster is classified among **24** different genres (target variable). One movie can belong to several genres classification.

The proposed classification methodology will be able to predict the genres classification for a movie given its poster and plot.



# Genre Classification

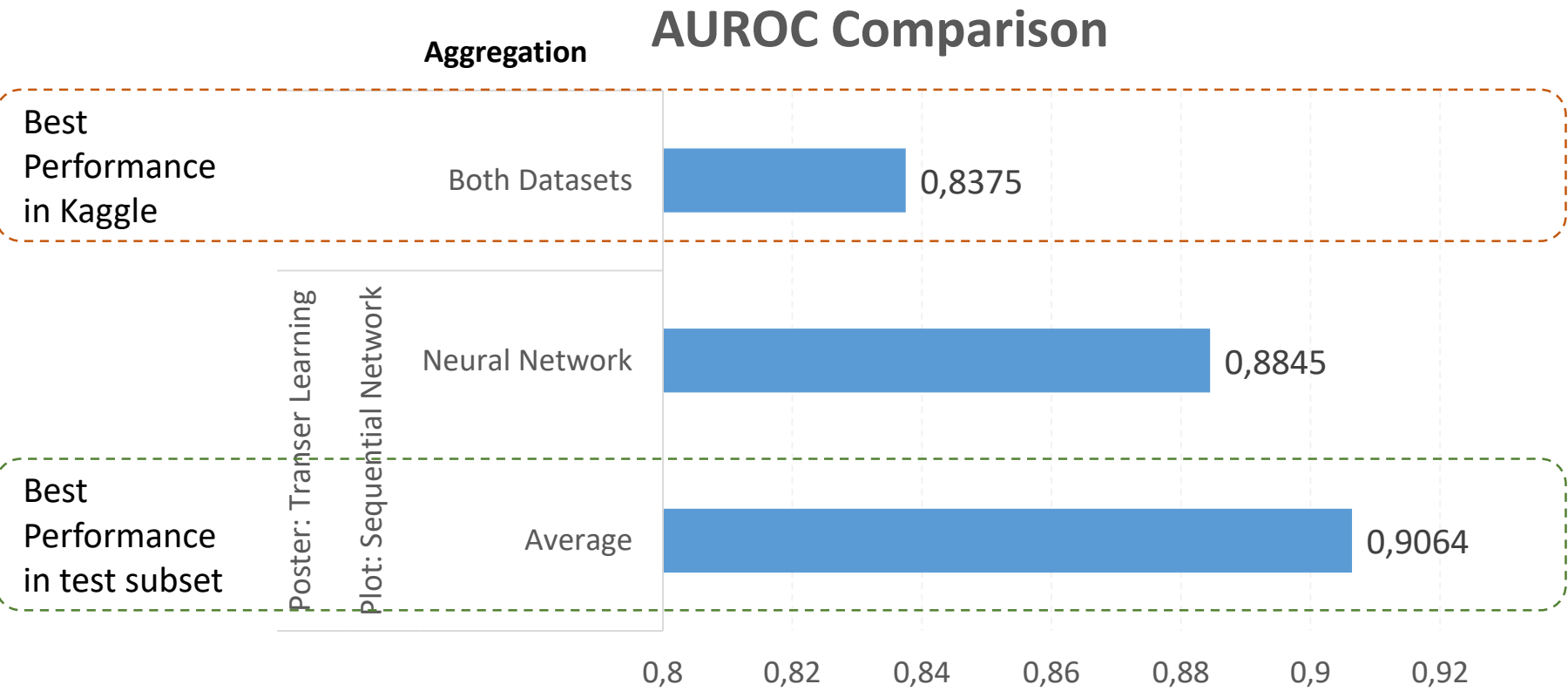
Genre classification could be divided in three stages: First each dataset will be used for a single classification, then both models will be joined in order to asses genres to each movie.



Models with the best performance are shown in **bold**

# Genre Classification

One final approach is build a neural network taking as input vectors the two data sets at once. The result of the different approaches are the following



In conclusion, the best model will be the average between transfer learning network for images and a sequential network for movie plot. However, the best model in Kaggle was the neural network taking as input vectors the two data sets at once



# Thanks