


Classifying Morphological Galaxy Types from SDSS Images Using Convolutional Neural Networks

Syd Rothman, Data Scientist

Advancing NASA Taxonomy Fields



TX 11.4.1 and TX 11.4.2

- Volume of galaxies and by extension, galaxy data, requires automated morphological analysis
- Project qualifies as progression in state of the art (SOA) under TX 11.4.1 and TX 11.4.2
- Telescope projects gathering galaxy imaging data are active; classification of objects is currently evolving and a pain point in subject areas of navigational data, dark matter research, and galaxy formation research

Data Sourcing

Models are built with the goal of future application to related data from separate sources.



EFIGI Project

The EFIGI reference dataset contains SDSS images and morphological data from the RC3 catalogue for 4,458 galaxies.



RC3 Catalogue

Contains morphological data of 23,022 galaxies, extracted from database of 73,197 galaxies maintained by G. Paturel at Lyons Observatory.



SDSS

The Sloan Digital Sky Survey catalogues imaging data of over $\frac{1}{3}$ of the sky with three telescopes and multiple spectrograph instruments.

Data Attributes

Data Utilized

- Png image files of 4,458 galactic objects
- Galaxy type targets from numerical data files

Data Available for Further Project Iterations

- Numerical morphological data
- FITS image files separate by band (*g, i, r, u, z*)

Challenges

SOA in automated morphological analysis is evolving at a fast pace.

- *Imaging data is collected from three different instruments (SDSS telescopes)*
- *Models will perform differently on images filtered through various bands*
- *Data sources provide large amounts of numerical morphology data that will affect the accuracy of an image-only classification model*

Modeling Iterations

Base Model

A Sequential CNN with convolutional and dense layers

76%
accuracy

Pooling Model

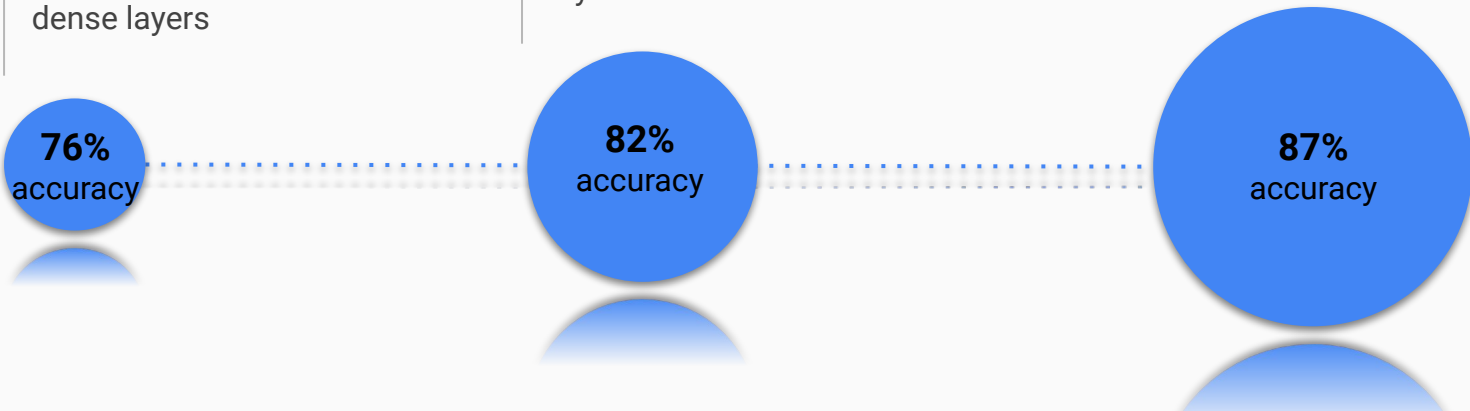
Pooling layers and additional convolutional layers added to base

82%
accuracy

Current Model

Convolutional, pooling, flatten, and dense layers with Dropout added to avoid overfitting

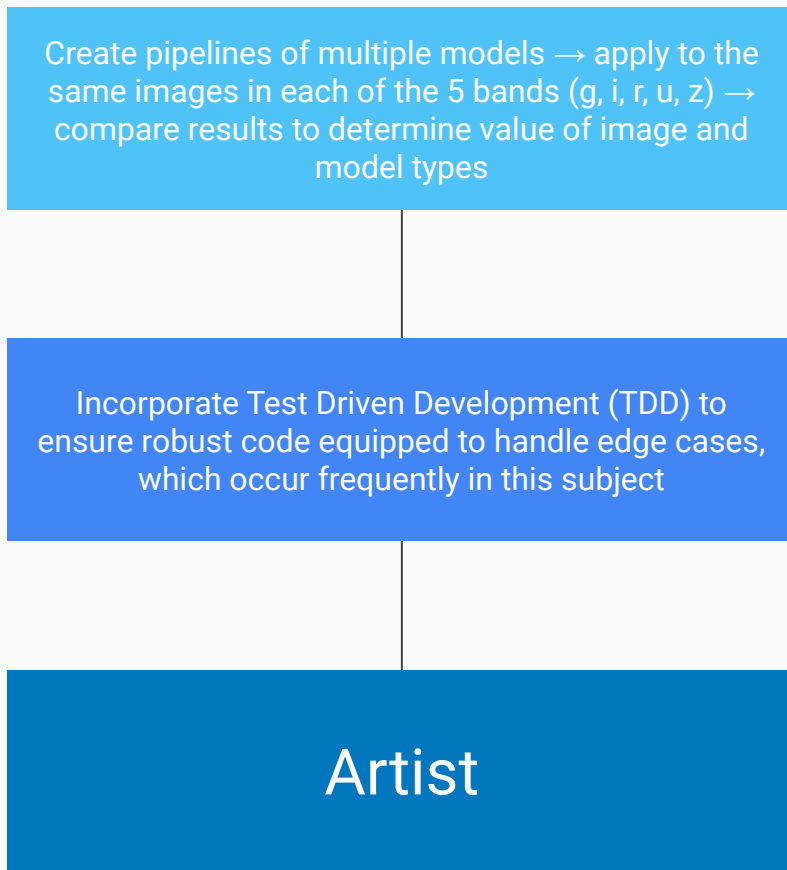
87%
accuracy



Future Development

The abundance of data available for future project iterations offers opportunities to explore varying model builds and results by image band, and to incorporate numerical morphological data.

Listed in the illustration on the right are three proposed next steps.



Why now?

Lorem ipsum dolor sit
amet, consectetur
adipiscing elit, sed do
eiusmod tempor
incidunt

