Abstract Submitted
for the APR12 Meeting of
The American Physical Society

Galaxy Image Processing and Morphological Classification Using Machine Learning JULIAN KATES-HARBECK, Stanford University — This work uses data from the Sloan Digital Sky Survey (SDSS) and the Galaxy Zoo Project for classification of galaxy morphologies via machine learning. SDSS imaging data together with reliable human classifications from Galaxy Zoo provide the training set and test set for the machine learning architectures. Classification is performed with hand-picked, pre-computed features from SDSS as well as with the raw imaging data from SDSS that was available to humans in the Galaxy Zoo project. With the hand-picked features and a logistic regression classifier, 95.21% classification accuracy and an area under the ROC curve of 0.986 are attained. In the case of the raw imaging data, the images are first processed to remove background noise, image artifacts, and celestial objects other than the galaxy of interest. They are then rotated onto their principle axis of variance to guarantee rotational invariance. The processed images are used to compute color information, up to 4^{th} order central normalized moments, and radial intensity profiles. These features are used to train a support vector machine with a 3^{rd} degree polynomial kernel, which achieves a classification accuracy of 95.89% with an ROC area of 0.943.

Julian Kates-Harbeck
Stanford University

Date submitted: 09 Jan 2012 Electronic form version 1.4