Stanford University

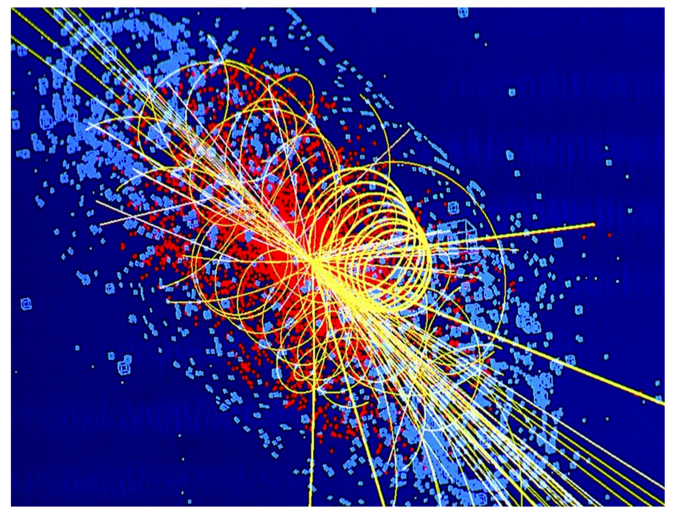
Identifying the Higgs Boson

Guy Amdur, Anton Apostolatos, Leonard Bronner

CERN

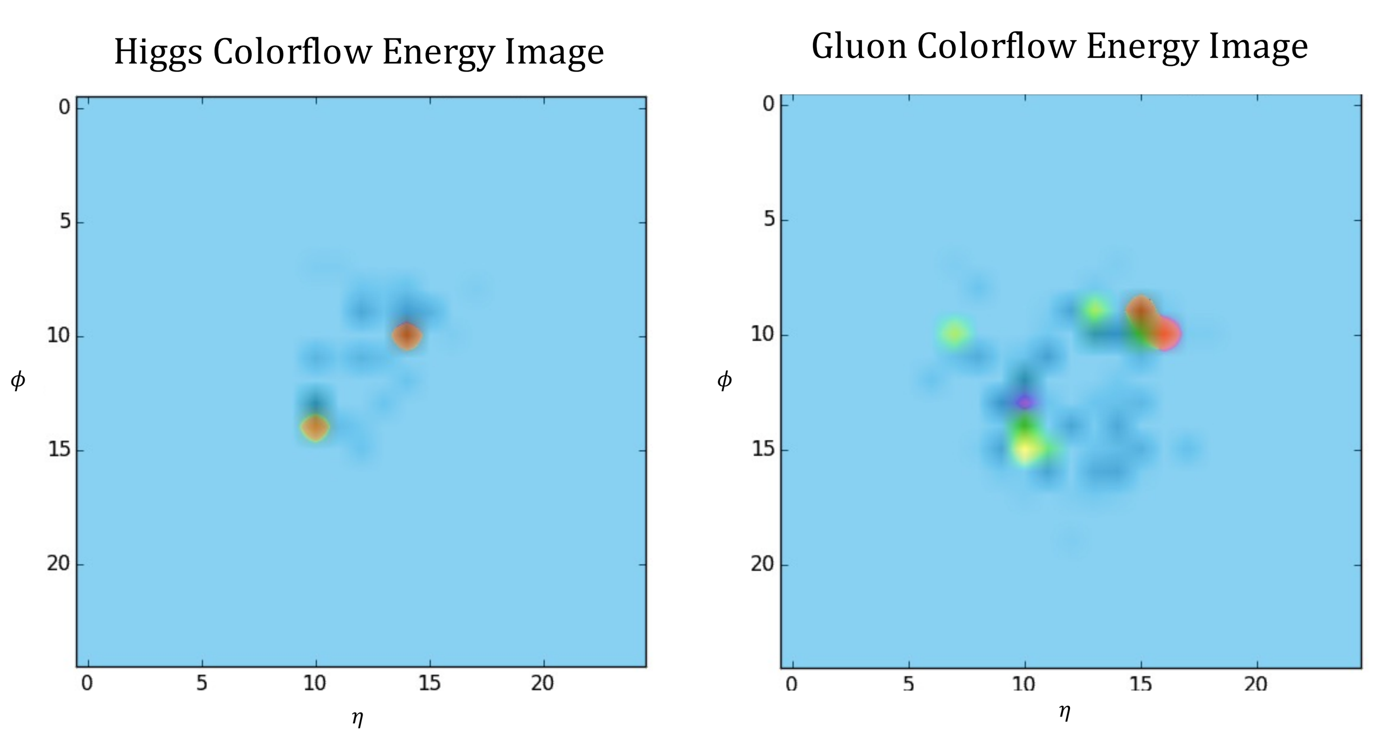
ATLAS EXPERIMENT

LHS and the Higgs Boson



The ATLAS detector at the Large Hadron Collider, recording energy snapshots of 40 million proton-proton collisions a second, is tasked to finding the elusive Higgs boson subatomic particle. In a super-charged state, the particle decays in observable ways. By measuring and analyzing the energy patterns of this decay we can identify the subatomic particle itself. Discovering the Higgs boson would validate a generation of scientific thought in particle physics and alter our understanding of mass as a physical property.

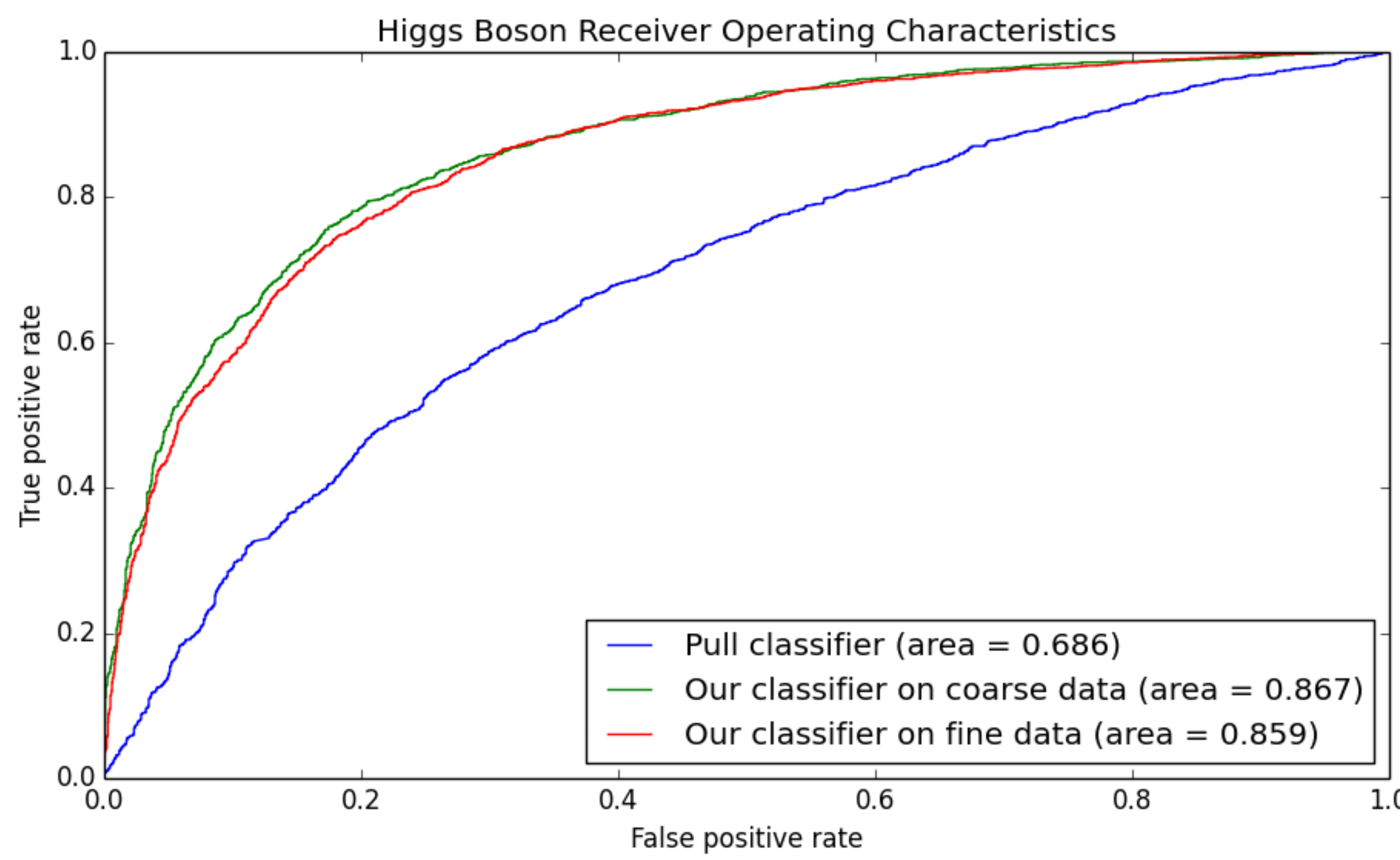
Data



In collaboration with ATLAS we were given access to energy images for both Higgs boson and gluon decay. We had 50,000 25x25 and 100x100 pixel images in η and ϕ cylindrical coordinates. These were pre-processed so that the jet center is at the center of an image, with resonance being kept constant in every sampled data point. The nature of the electromagnetic mechanism of jet pull energy observation leads the larger images to less accurately detect charged particles. Thus, there is an inherent balance between higher resolution and lower quality of information.

Results

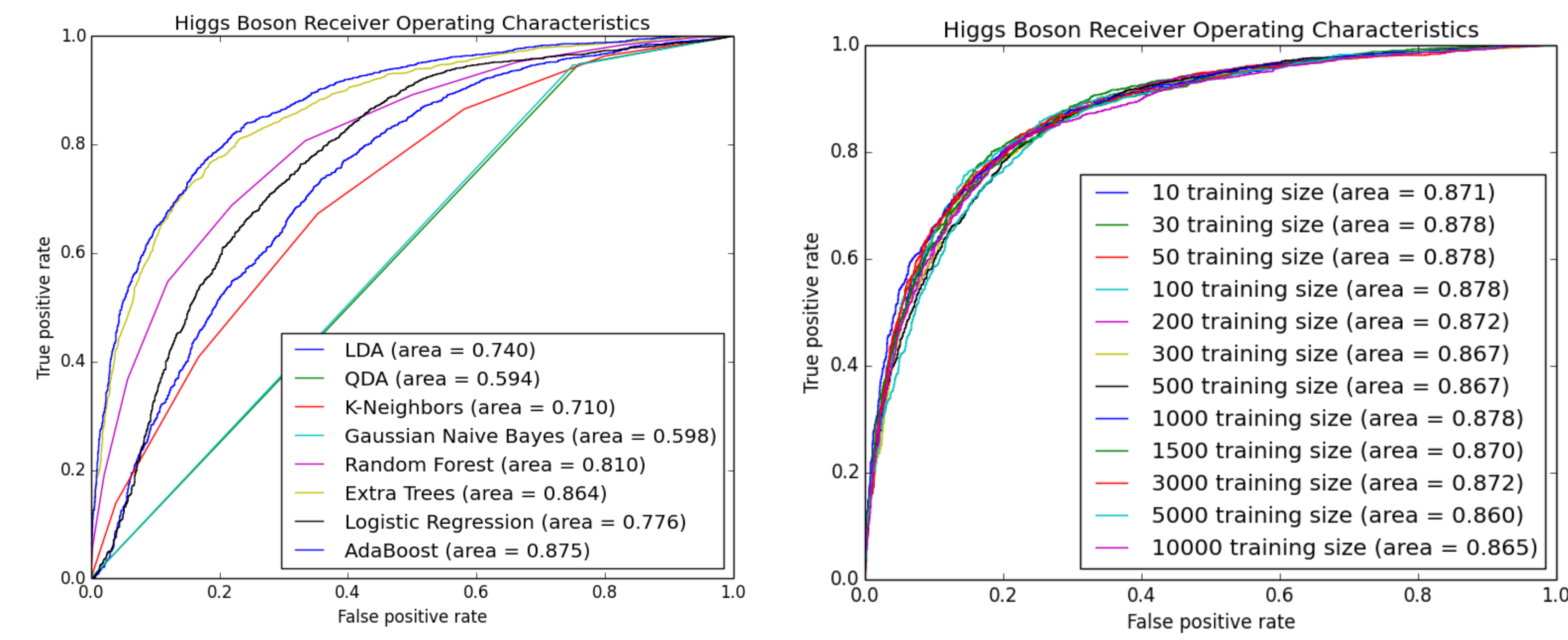
Current state-of-the-art runs Fisher Discriminant Analysis on pull data, a feature class which characterizes the superstructure of the particle collision events. This classification method has a AUC of 0.686.



We achieved an AUC of 0.867 by running an AdaBoost tree algorithm with its base estimator being a Random Forest classifier on the vectorized images. The AdaBoost classifier is an adaptive model that runs weaker learning algorithms, tweaking them and combining them into a weighted sum classifier that is sensitive to outliers and noise. This method of classification is significantly better than current state-of-the-art models, as is evident in the figure above.

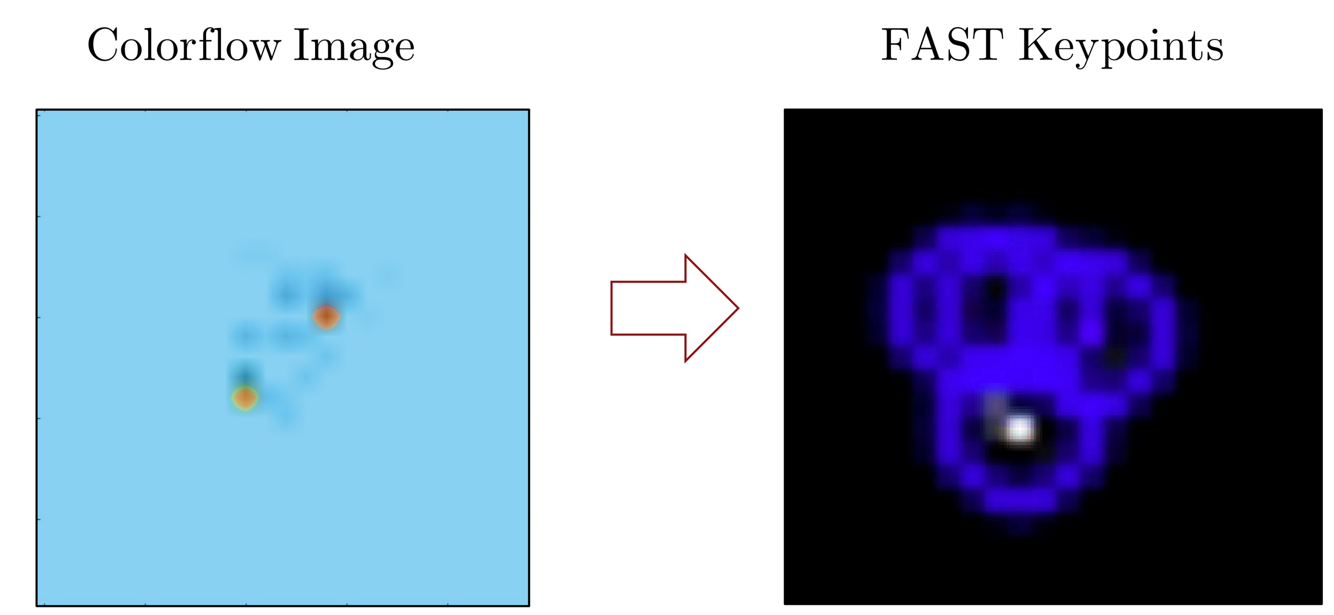
The finer granularity dataset did not lead to better classification, indicating that the higher resolution did not offset the loss of information that came as a result of larger images.

Optimizations and Experiments



We evaluated many different supervised learning classification methods for this project and ended up finding decision tree models to be the most successful. Interestingly, the final model excelled even with very little training data, indicating that even just a handful of samples have enough information to make these decision tree classifiers reach their full potential.

Vision



We experimented with various image descriptors, including ORB, BRIEF, and SIFT for the higher granularity images, with unsatisfactory results, achieving an AUC of only 0.550. These descriptors weren't able to capture many of the nuances of the sparse pixel matrices. Image histograms of pixel density also proved unsuccessful, resulting in an AUC score of 0.660.