# AstronomicAL: an interactive dashboard for visualisation, integration and classification of data with Active Learning

Grant Stevens, Sotiria Fotopoulou, Malcolm N. Bremer, Oliver Ray

Motivation

By formalising and simplifying the process of presenting and labelling heterogeneous datasets, AstronomicAL provides a unique workflow that allows for the creation of reliable datasets and robust classifiers. This removes the need for multiple specialised but disjoint tools. AstronomicAL has been developed to address the growing issues of incorrect or missing labels and imbalanced class sizes by removing the barrier to the adoption and effectiveness of active learning in research. AstronomicAL enables experts to visualise domain-specific plots and key information relating both to broader context and details of a point of interest drawn from a variety of data sources, ensuring reliable labels. By exporting a simple configuration file, entire layouts, models, and assigned labels can be shared with the community. This allows for complete transparency and ensures that the process of reproducing results is effortless.

Active Learning with AstronomicAL
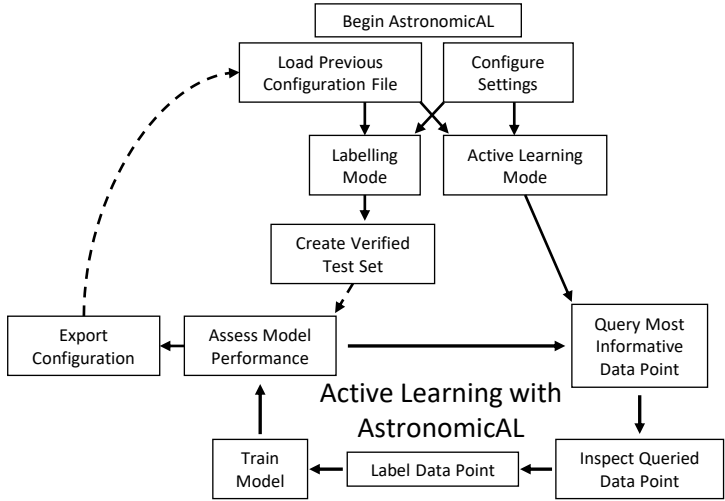
The Interactive Dashboard
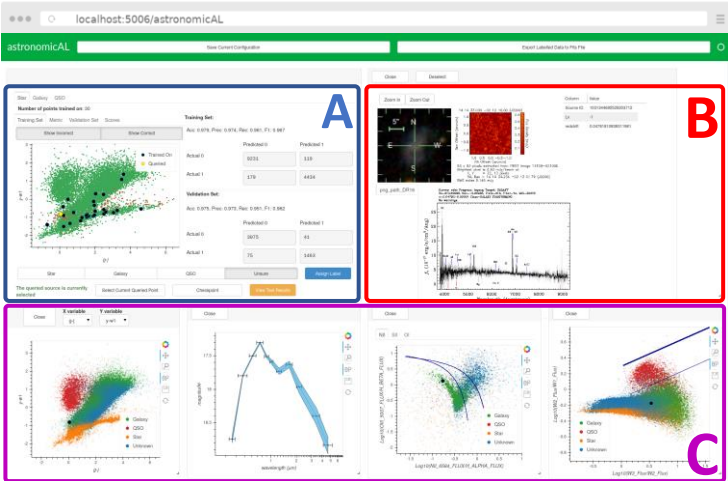
The dashboard is set up in 3 main areas:

A) Active learning and labelling panel which controls the machine learning functionality. Users can train separate models for each class and see how each performs as additional points are added to the model. AstronomicAL provides functionality to explore all aspects of the training process, including custom models and query strategies.

B) User-defined information shown for the currently active data point, such as specific columns, local images and online cutout services.

C) Customisable domain-specific plots that can render millions of data points and are fully interactive. Plots are also interconnected, with any updates projected to all panels simultaneously, allowing analysis of how the properties of a particular point fit in the context of the whole dataset.

Active Learning

Active learning utilises methods for predicting which part of the search space of unseen examples would be most likely to improve classification accuracy. By querying these key examples and presenting them directly to the user for inspection and labelling, active learning removes the reliance on non-verified data. In the optimal case, it provides users with simple classifications that have a high impact on model performance, leading to a dramatic reduction in the amount of data required for training a classifier.
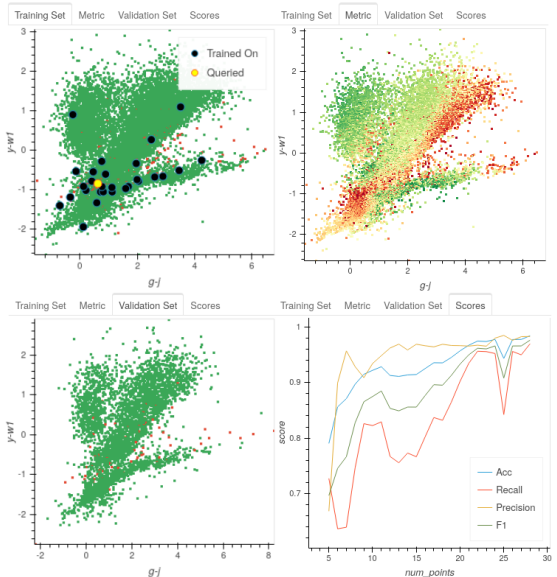
AstronomicAL allows users to view all key information during training, including visualising the model's confidence of each data point (top right). As shown above, we were able to classify stars with a validation accuracy of 98% when training on only 30 data points.