Data Augmentation in a Hierarchical-Based Classification Scheme for Variable Stars

Zafiirah Hosenie1, Robert Lyon3, Benjamin Stappers1, Arrykrishna Mootoovalloo2 & Vanessa McBride4

1. Introduction

- A major issue that impedes the successful automated classification of astronomical data is the imbalanced learning problem.
- This problem impacts classification of variable stars in particular, as some types of variable stars are rare, making it difficult to build automated machine learning (ML) systems.

2. Hierarchical Taxonomy

3. Methodology & Results

Stage 1: hierarchical tree classifiers

- We use the astrophysical properties of the various sources to construct a hierarchical-based structure to represent the different classes.
- For the HC, we use XGBoost and Random Forest.

Stage 2: level-wise data augmentation in HC

- Since the training set is still imbalanced after aggregating subclasses into superclasses, we use the three data augmentation techniques: SMOOTE, RASLE, GPFit.

Stage 3: feature extraction

- Our features are based on 6 intrinsic statistical properties relating to location (mean magnitude), scale (standard deviation), variability (mean variance), morphology (skew, kurtosis, amplitude), and time (period).

Stage 4: training with Bayesian optimization

- The training set moves through the first level in the HC scheme. The training examples are then augmented. The model (see dotted square at level 1) is trained using either RF or XGBoost classifier.
- Lastly, we evaluate our trained model on real LCs in the test set. The same concepts apply at different levels in the HC where real LCs move down the node, get augmented and classified in their respective classes.

4. Conclusion

- When using GPFit method, our RF implementation performs best at all HC levels when compared to H19.
- We found that both XGBoost and RF provide good performance for variable stars classification.
- We assess the consistency of the results using GPFit and RF by plotting the Receiver Operator Characteristic (ROC) curve for each class.