VaSt-ID: Novel Application for Variable Star Identification and Classification Using Machine Learning

Pan, Yida (School: William A. Shine Great Neck South High School)
Huang, Luke (School: William A. Shine Great Neck South High School)

Variable stars are critical distance indicators in the universe due to their characteristic period-luminosity relationships. They are a common method which astronomers use to find distances to celestial objects, and provide important information such as stellar age and structure. With a large influx of variable star data in the past decades, the demand for more efficient classification techniques has increased. This project aims to improve the identification and classification of variable stars using a machine-learning based classification model: random forest (RF). The model was trained with ~80,000 variable star light curves to classify the following variable classes: Delta Scuti, RR Lyrae, Cepheid, Type II Cepheid, Eclipsing Binary, Mira, Heartbeat, and non-variable, achieving an overall F1 score of 99%. This is also the first model capable of identifying Heartbeat stars, a recently discovered type of ellipsoidal variable, achieving an F1 score of 95%. We applied our model to the ASAS-SN database to find previously undiscovered Heartbeat stars and found 866 Heartbeat star candidates among ~600,000 variable stars, increasing existing data on Heartbeats by 87% if confirmed. We developed VaSt-ID, an application that implements our optimized model. It takes measurements of magnitude over time as input, and outputs the variability type. VaSt-ID also computes the distance to the inputted variable star by using period-luminosity relationships and the distance modulus. VaSt-ID is intended to be a user-friendly and highly accurate general-purpose variable star classifier that could be used by amateur and professional astronomers alike in discovering new variable stars.