Automated stellar variability classification using TESS light curves

Emma Chickles[1], Tansu Daylan[2], and Lindsey Gordon[3]

[1]Department of Physics, MIT, Cambridge, MA, USA
[2]Department of Physics and Kavli Institute for Astrophysics and Space Research, MIT, Cambridge, MA, USA
[3]Department of Astrophysics, University of Minnesota, Minneapolis, MN, USA

We use unsupervised feature learning and clustering to study stellar variability observed by the Transiting Exoplanet Survey Satellite (TESS) in 2-minute cadence. We use our census of stellar variability to calculate statistical descriptions of stellar parameters, such as mass and temperature, of eclipsing binary systems. After further validation, we will release a TESS Catalog of Stellar Variability for many variability types. Our unsupervised learning approach has the potential to find new members of known classes of variability, discover new classes, and investigate correlations between physical properties and variability properties.

Feature learning

For each of the 26 sectors, our *Mergen* pipeline extracts features from short-cadence PDC-SAP light curves using a Convolutional Autoencoder (CAE). The CAE is a convolutional neural network characterized by two components: an encoder that transforms input light curves into low-dimensional representative features, and a decoder that reproduces the light curves from the representative features. The CAE hyperparameters (including the number of neurons in the feature space representation, the activation function, and the batch size) were optimized with a simple grid search.

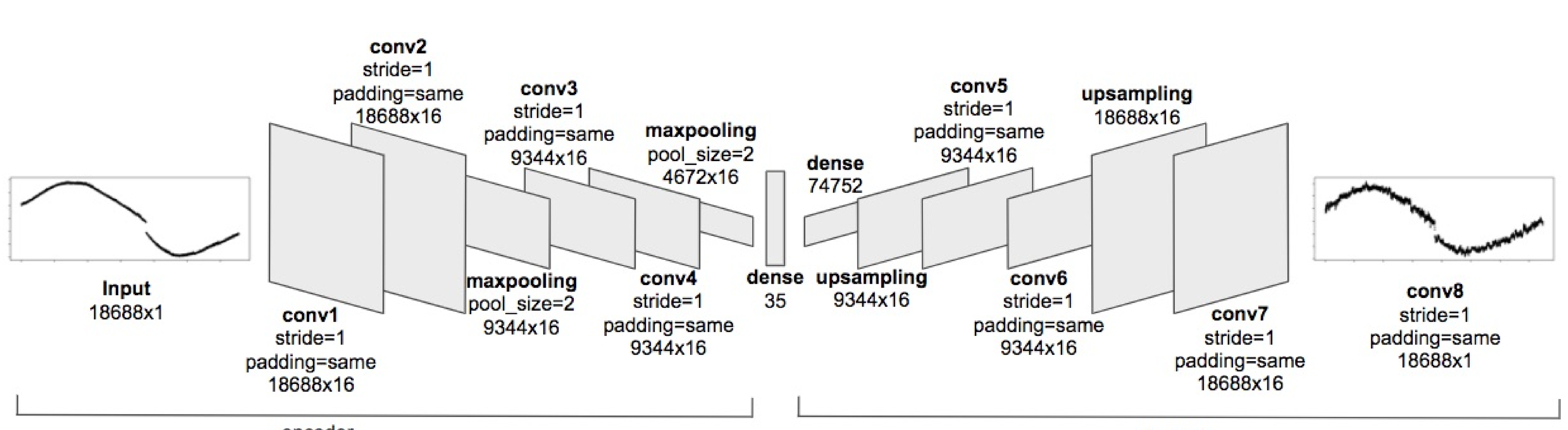

Figure 1: *CAE architecture for 2-minute cadence light curves in the TESS Observational Sector 1 (TESS observations from 25 July to 22 August, 2018)*

Clustering

We perform clustering analysis on the feature space extracted by the CAE. We use a Gaussian Mixture Model (GMM), which assumes that the features are generated from a finite number of Gaussian distributions. We assign each cluster a variability type based on existing classifications from the General Catalogue of Variable Stars (GCVS). We assume that light curves grouped into the same cluster share the same variability mechanism.

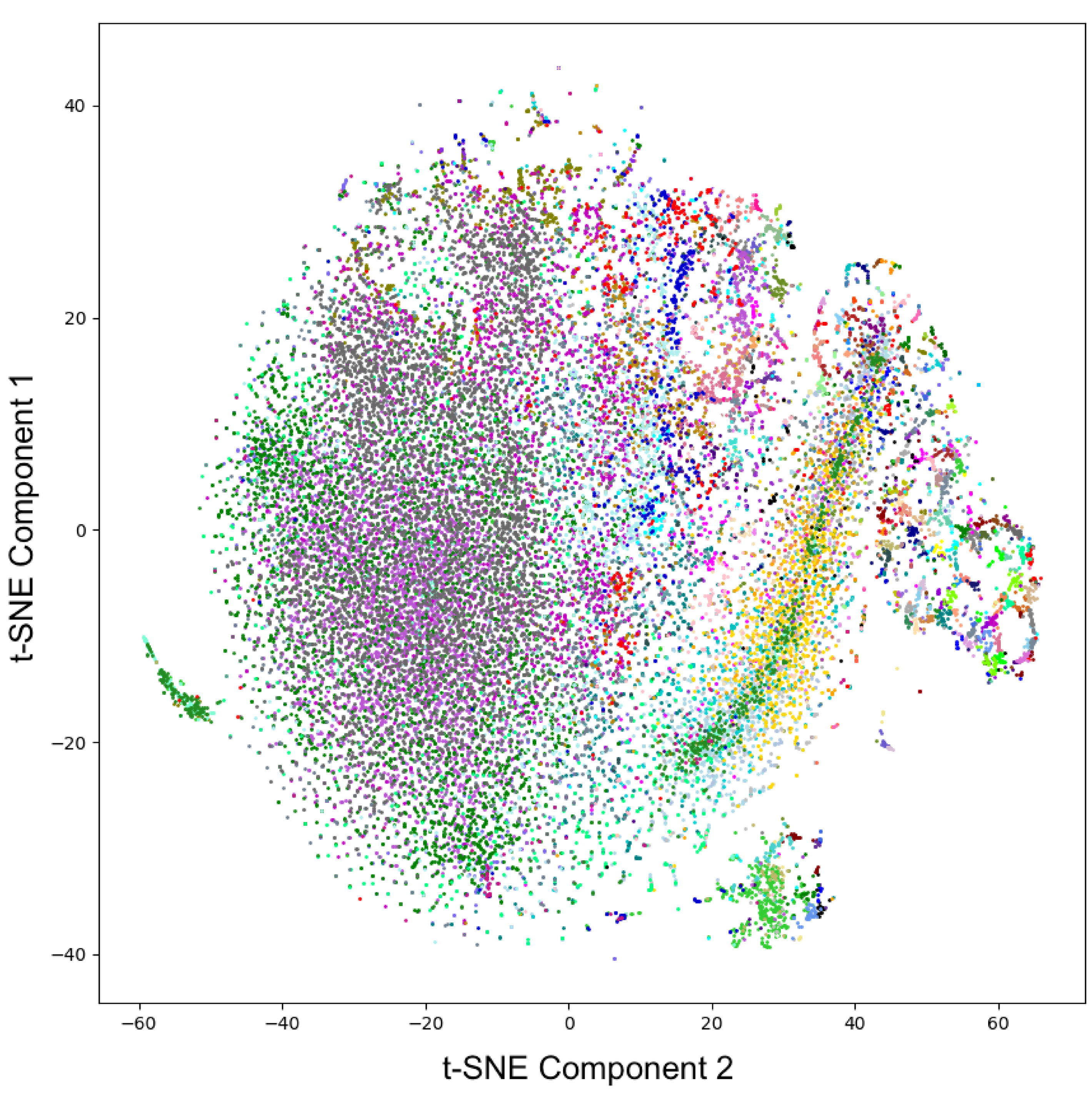

Figure 2: *2-dimensional visualization of the 35-dimensional feature space of Sector 1, colored with 200 distinct colors based on the 200 GMM clusters.*

Variability statistics for eclipsing binary systems

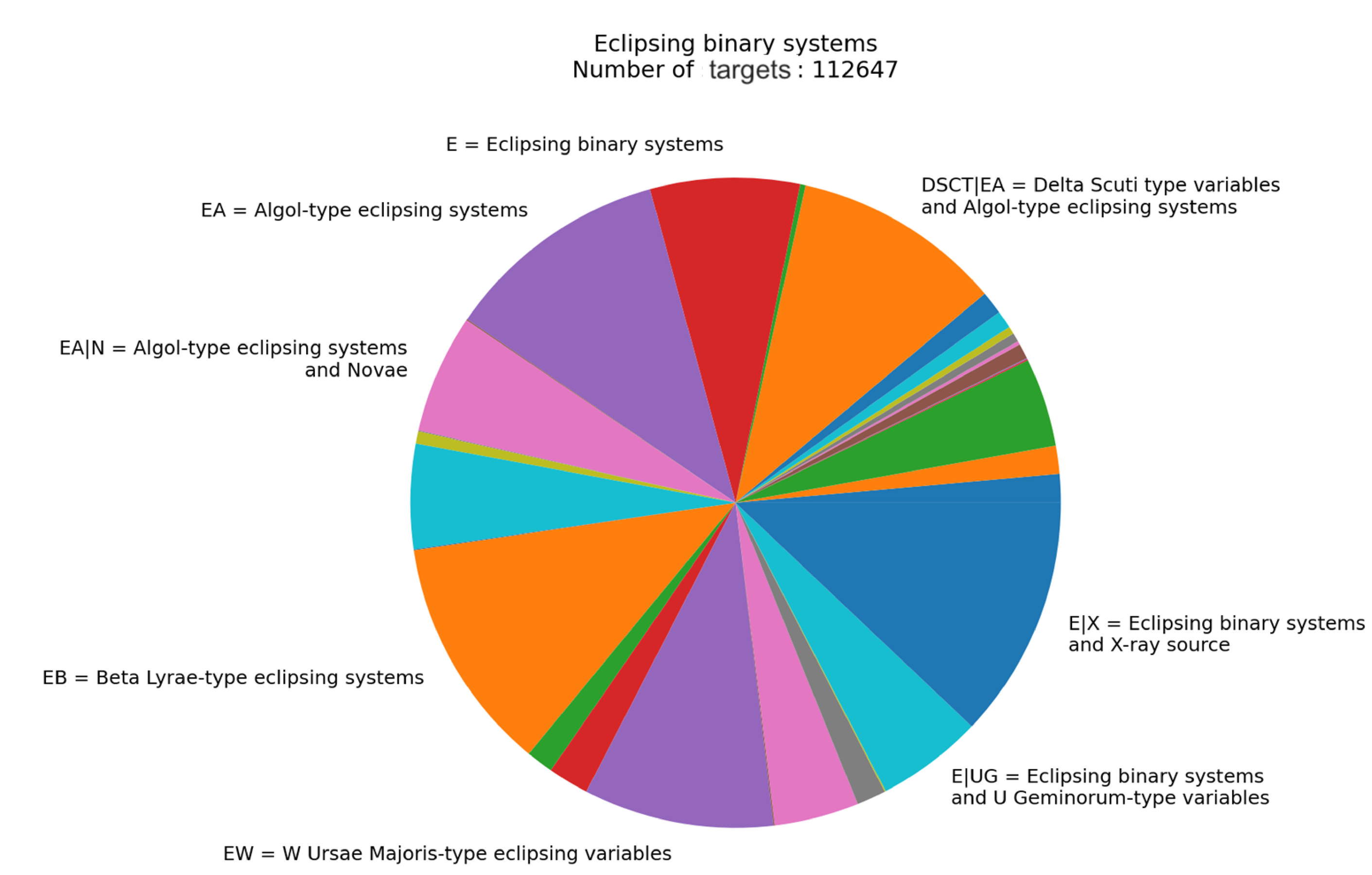

Figure 3: *8 of the most populated variability types within the 112,647 objects identified as eclipsing binary systems.*

Validation for low mass contact binaries

We validate our results by comparing the distributions of fundamental physical properties of our predicted and catalogued low mass contact binaries (or W Ursae Majoris variables). Since the distributions of the mass and temperatures of the predicted and the catalogued low mass contact binaries are similar, this is evidence that our clustering method can detect and group astronomical objects sharing this source of photometric variability.

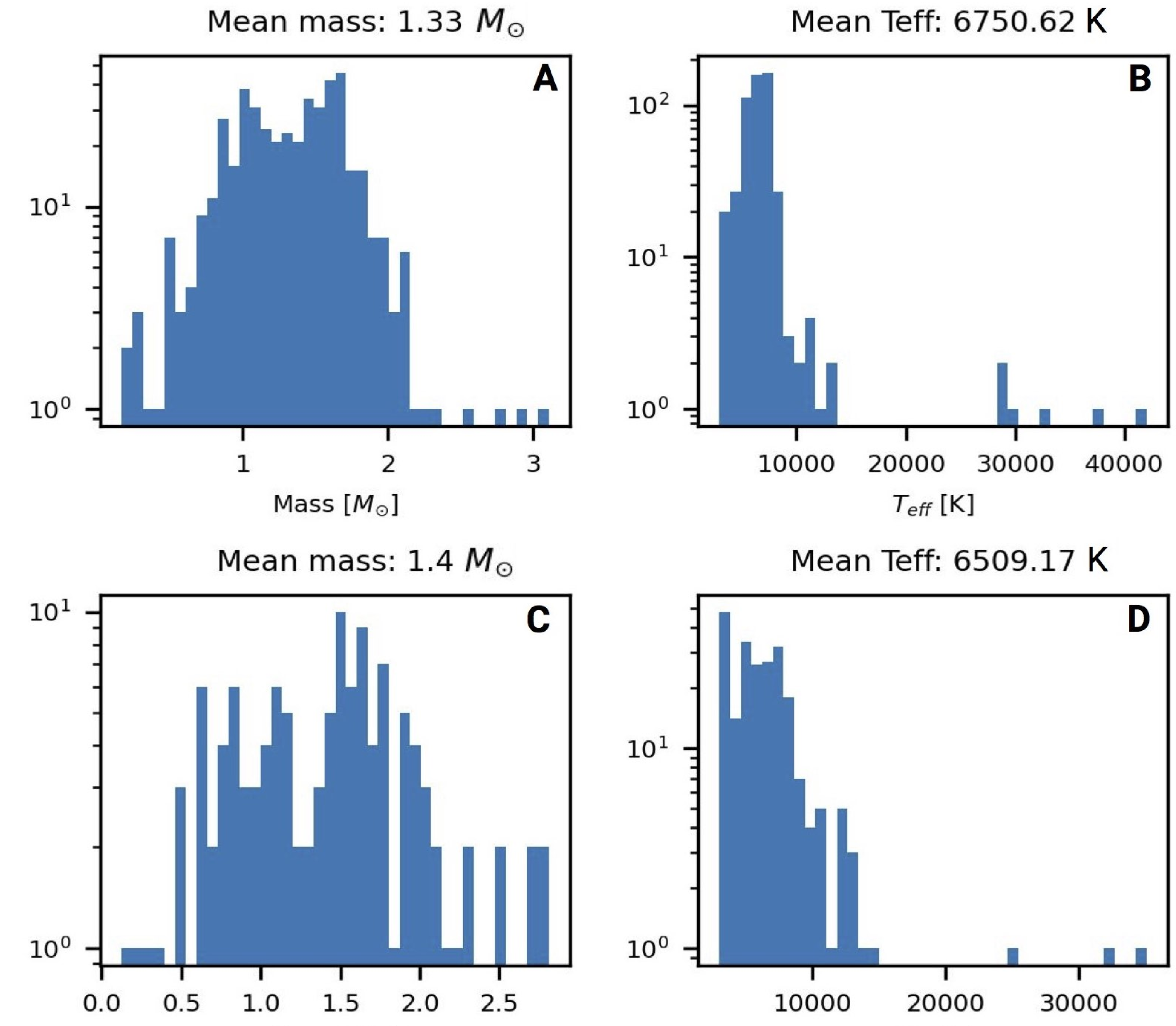

Figure 4: *(A) Distribution of mass of catalogued W Ursae Majoris variables (W UMas) in Sectors 1-26, (B) Distribution of temperature of catalogued W UMas in Sectors 1-26, (C) Distribution of mass of predicted W UMas in Sectors 1-26, (D) Distribution of temperature of catalogued W UMas in Sectors 1-26.*