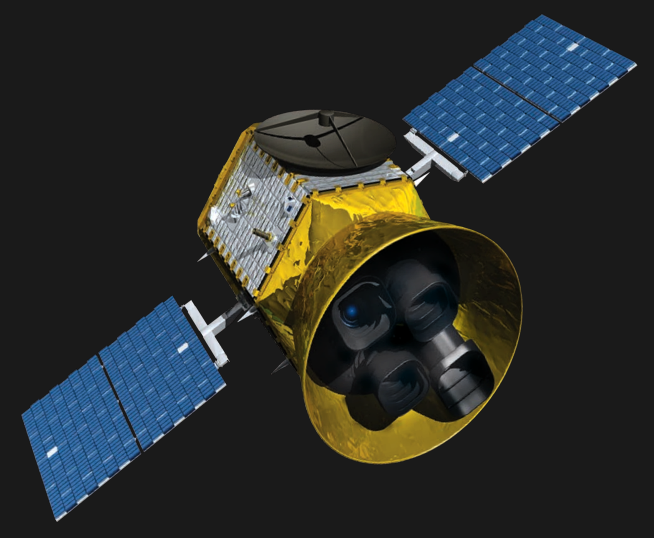
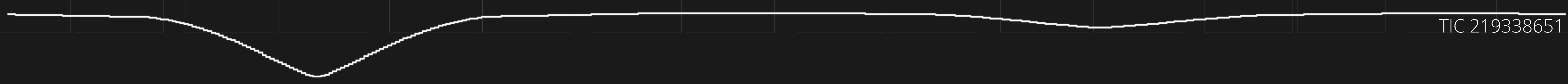


SYSTEMATIC CLASSIFICATION OF TESS ECLIPSING BINARIES

Jessica Birky (UW Seattle), James Davenport (UW Seattle), and Timothy Brandt (UC Santa Barbara)

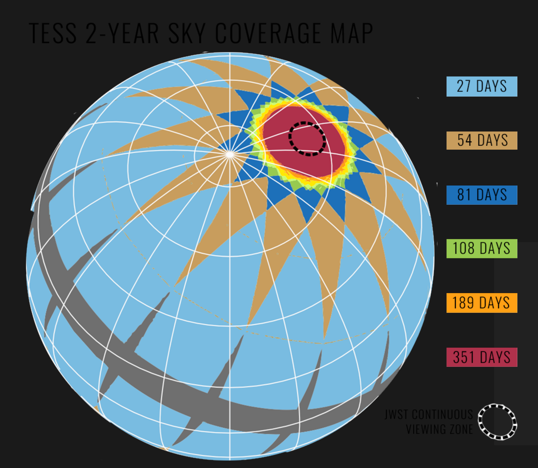


MOTIVATION

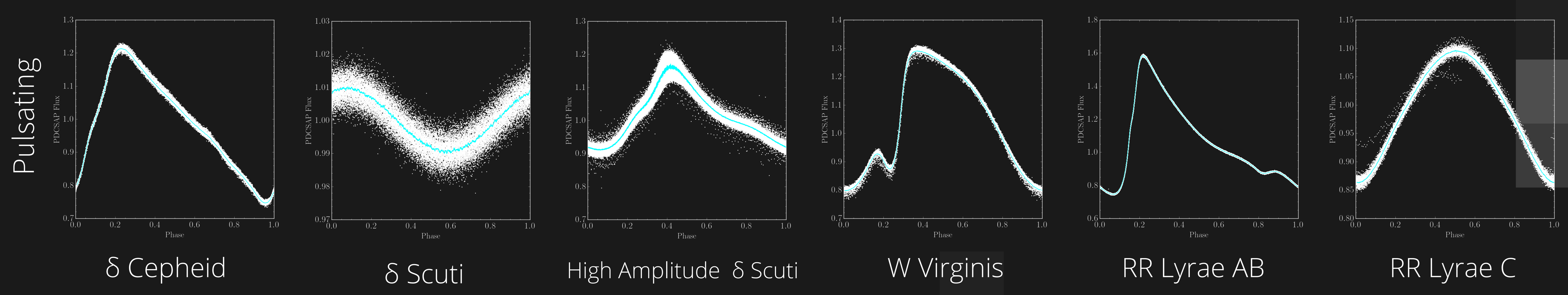
The Transiting Exoplanet Survey Satellite (TESS): ~85% sky coverage, high (1-3%) photometric precision, w/ timing baselines 27-351 days of ~400k sources (2 min cadence) and 20-150 million sources (30 min cadence).

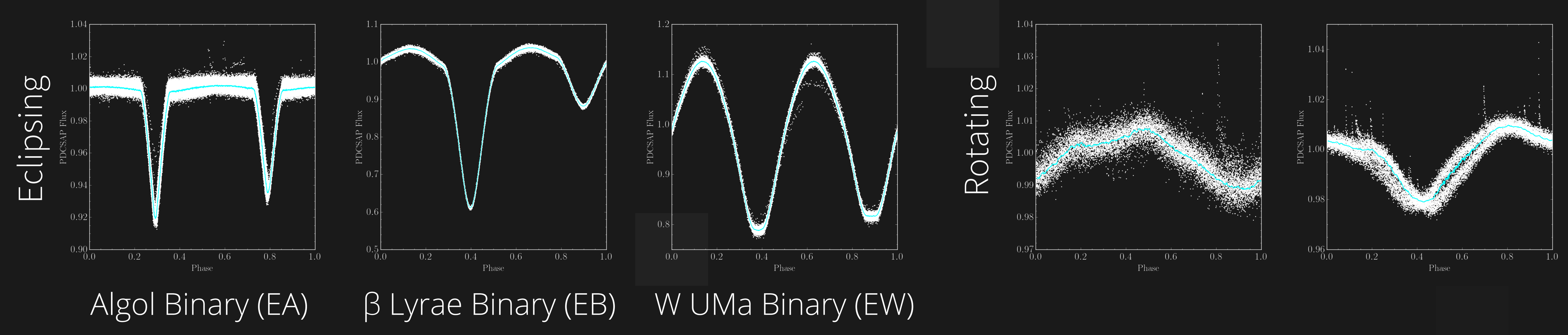
Population statistics (of stellar properties & orbital kinematics) from a comprehensive catalog of TESS EBs across Galactic environments are essential benchmarks for improving stellar evolution models, determining distances to callibrate the cosmic distance ladder, and tracing Galactic structure, among other applications.

Goal: Use supervised machine learning to accurately identify/ classify eclipsing binaries and determine precise periods.



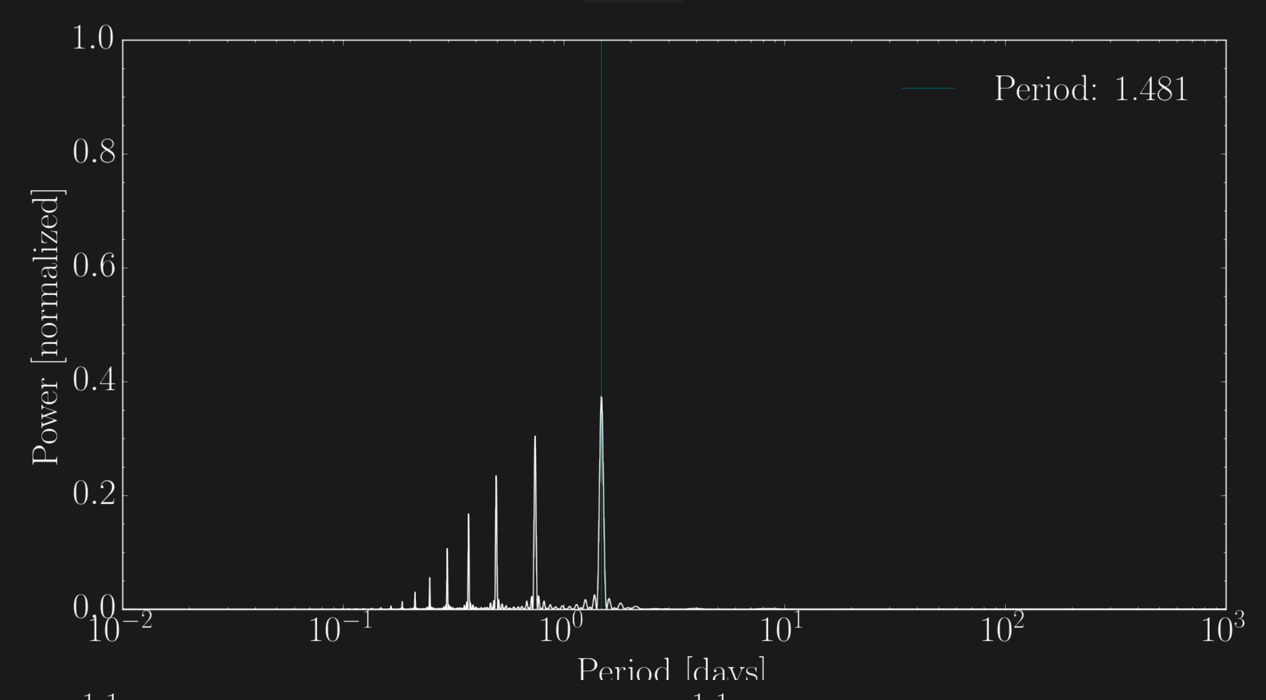
Examples of morphological classifications of periodically varying stars:

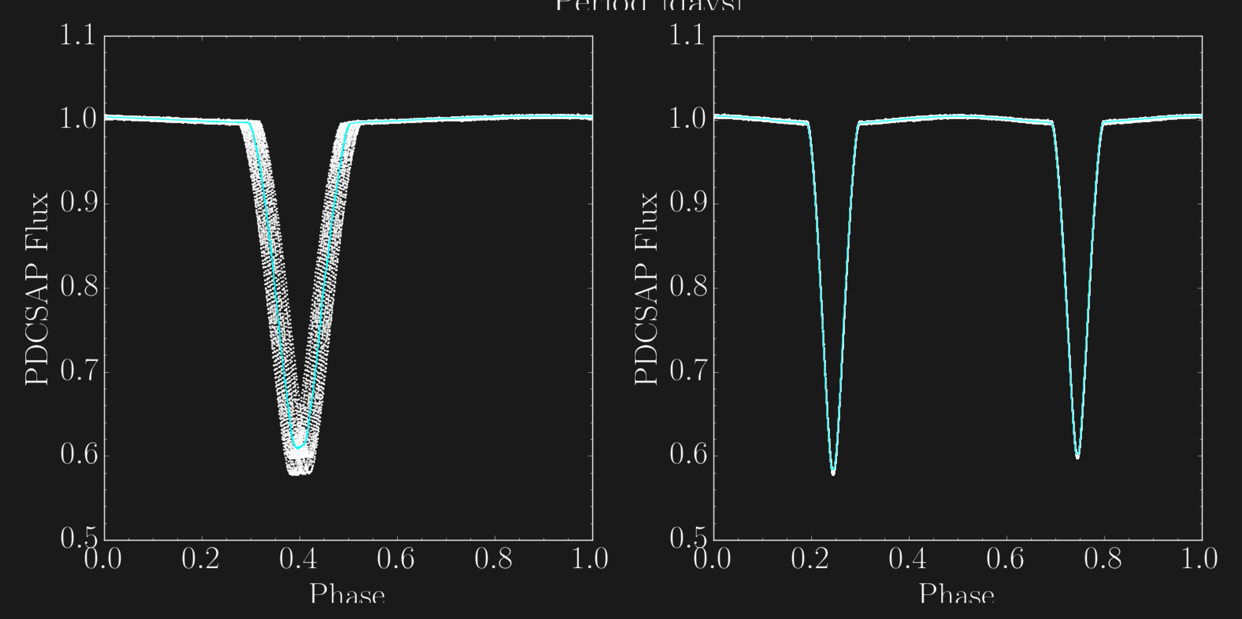




METHODS

(1) Period Finding / Phase Folding





LS: P = 1.4808 days

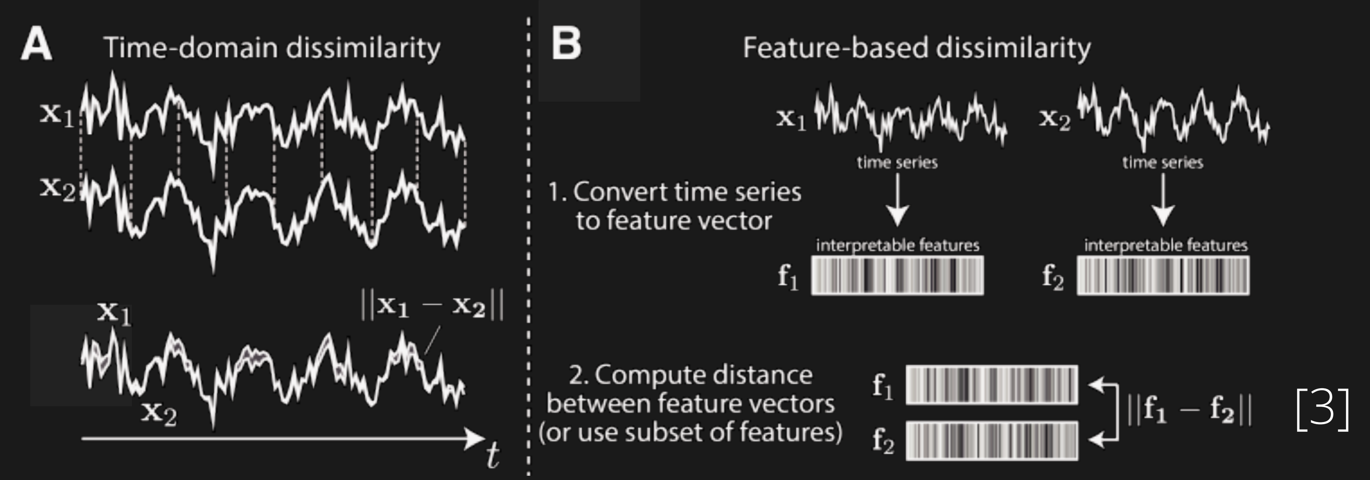
PDM: P = 2.9685 days

Initial period search using Lomb-Scargle periodogram, a variant of Fourier Transform applicable to unevenly sampled time-series data.

Period precision improved using Phase Dispersion Minimization: χ^2 between light curve points and a 100pt rolling-median smooth is minimized for the best period using `scipy.optimize` around factors of LS period.

(2) Dynamic Time Warping (DTW)

Approaches to time series classification w/ supervised learning:



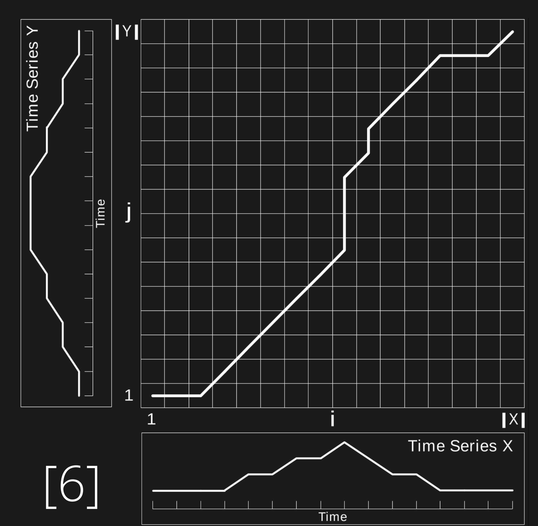
DTW [2]: distance metric for computing similarity between the shape of two time series. For time series $\{X_1,...,X_N\}$ & $\{Y_1,...,Y_M\}$, algorithm has three steps O(NM):

(1) Compute cost matrix (shown right) where

$D_{ij} = |X_i - Y_j| + \min(D_{i,j-1}, D_{i-1,j}, D_{i-1,j-1})$

(2) Use dynamic programming to solve for minimimum path across cost matrix which optimally align the two series in time.

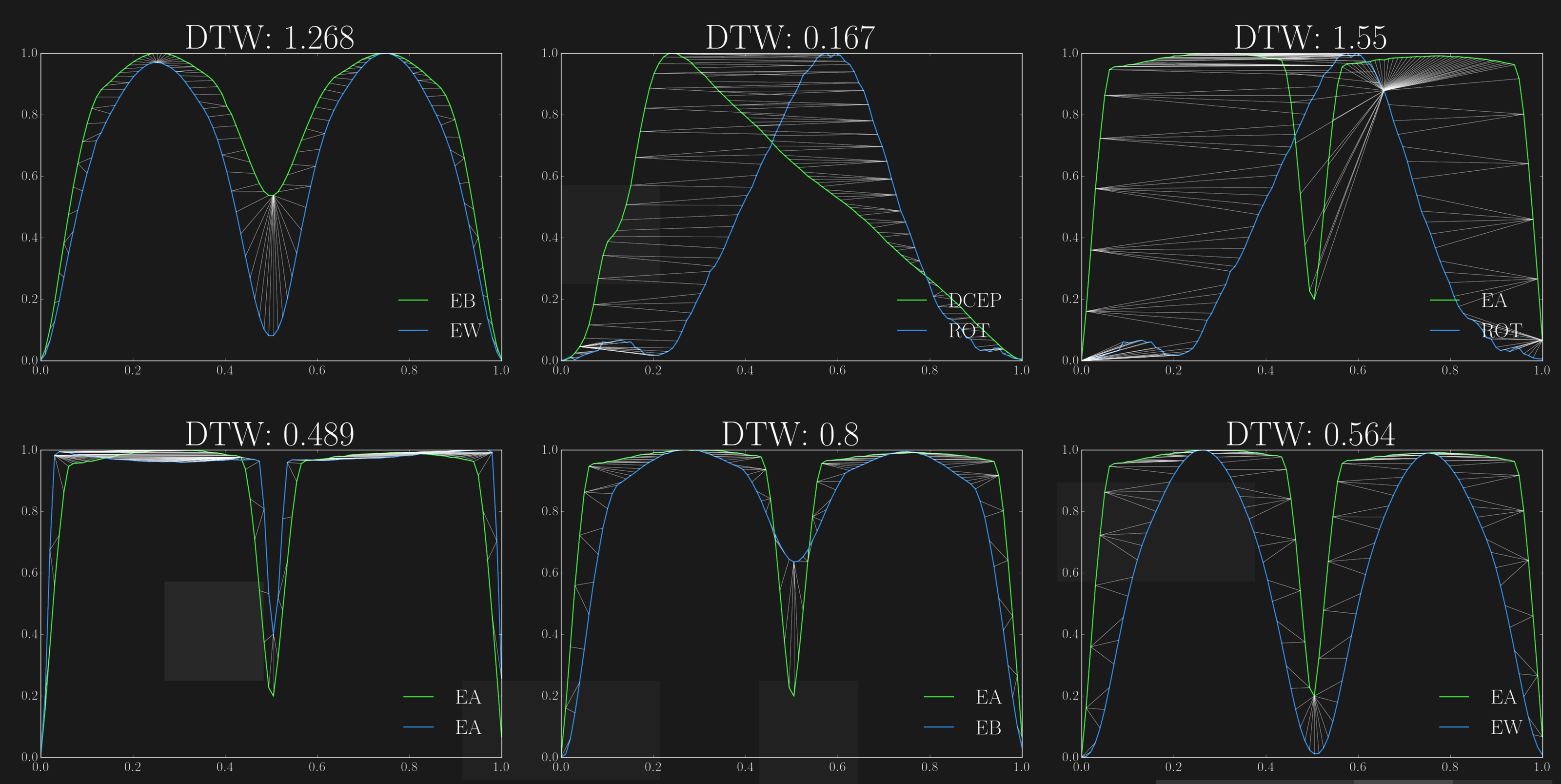
(3) DTW distance is the sum of path values.



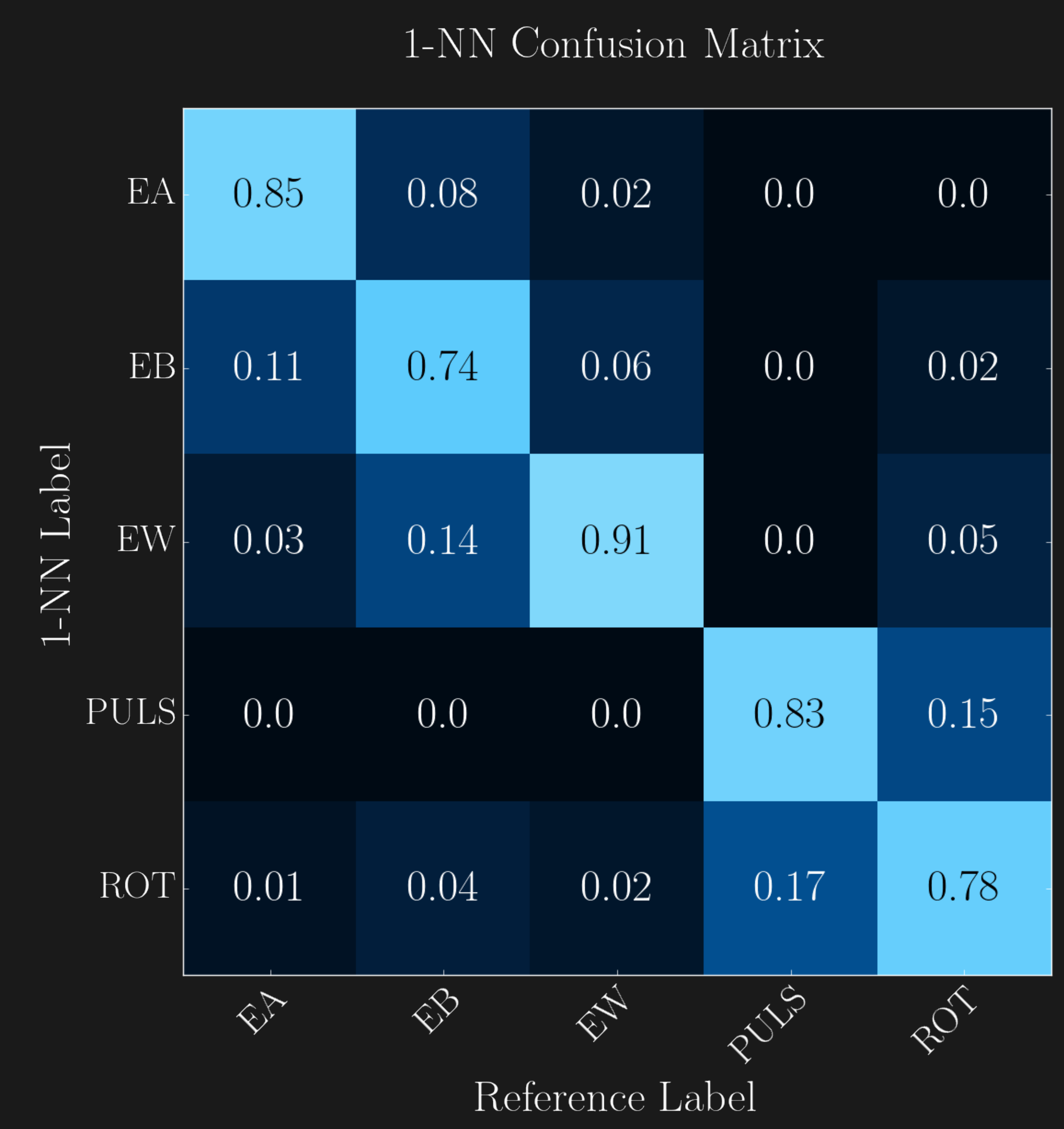
[6]

DTW has been used extensively for time series classification in machine learning literature [1] (on speech recognition, computer vision, biostats, etc.), but has seen few applications on astronomical time series.

Example optimal DTW path alignments (using `dtaidistance` [5]); lower distance is better:



(3) Classification



1-NN Label \ Reference Label	EA	EB	EW	PULS	ROT
EA	0.85	0.08	0.02	0.0	0.0
EB	0.11	0.74	0.06	0.0	0.02
EW	0.03	0.14	0.91	0.0	0.05
PULS	0.0	0.0	0.0	0.83	0.15
ROT	0.01	0.04	0.02	0.17	0.78

Confusion matrix comparing test labels and true labels using a 1-Nearest Neighbor (1NN) classifier using a training sample of 996 sources with labels from the ASAS-SN all-sky survey of variable stars [4] and light curves from TESS Cycle 1. Between two class labels, eclipsing binary types are distinguished from other periodic variables with a false-positive rate of 5% and false-negative rate of 2%.

FUTURE WORK

Address limitations:

- Eclipsing binaries with significant non-periodic variation
- Increase training sample size & representation to achieve better sub-class classification

Determine best uses for DTW:

- Compare to performance of feature-based methods
- Ensemble of classifiers

Apply classification routine to find new EBs

REFERENCES:
[1] Bagnall, et al. 2017, Data Mining and Knowledge Discovery, 31, 3
[2] Bellman, R. et al. 1959, IRE Transactions, 4, 2
[3] Fulcher, B. D. 2017, arxiv:1709.08055
[4] Jayasinghe, T. et al. 2019, MNRAS, 486, 2
[5] Meert, W. et al. 2019, "dtaidistance", v1.2.2, doi:10.5281/zenodo.1202378
[6] Salvador, S. et al. 2004, Intell. Data Anal., 11, 5

Acknowledgments: A Scialog grant supported by the Heissing-Simmons Foundation and NSF Graduate Research Fellowship Program.