Draft version April 1, 2021
Typeset using LaTeX **twocolumn** style in AASTeX63

TESS Data for Asteroseismology (T'DA) Stellar Variability Classification Pipeline: Set-Up and Application to the *Kepler* Q9 Data

J. Audenaert [1], J. S. Kuszlewicz [2,3], R. Handberg [3], A. Tkachenko [1], D. Armstrong [4,5], M. Hon [6,7], R. Kgoadi [8], M. N. Lund [3], K. J. Bell [9,10], L. Bugnet [11,12,13], D. M. Bowman [1], C. Johnston [1], R. A. García [14], D. Stello [7,15,3], L. Molnár [16,17,18], E. Plachy [16,17,18], D. Buzasi [19], C. Aerts [1,20,21] and the T'DA collaboration,

[1] *Institute of Astronomy, KU Leuven, Celestijnenlaan 200D, 3001, Leuven, Belgium*
[2] *Max-Planck-Institut für Sonnensystemforschung, Justus-von-Liebig-Weg 3, D-37077 Göttingen, Germany*
[3] *Stellar Astrophysics Centre, Department of Physics and Astronomy, Aarhus University, Ny Munkegade 120, DK-8000 Aarhus C, Denmark*
[4] *Department of Physics, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK*
[5] *Centre for Exoplanets and Habitability, University of Warwick, Gibbet Hill Road, Coventry, CV4 7AL, UK*
[6] *Institute for Astronomy, University of Hawai'i, 2680 Woodlawn Drive, Honolulu, HI 96822, USA*
[7] *School of Physics, The University of New South Wales, Sydney, NSW 2052, Australia*
[8] *College of Science and Engineering, James Cook University, Townsville, Australia, 4811*
[9] *DIRAC Institute, Department of Astronomy, University of Washington, Seattle, WA-98195, USA*
[10] *NSF Astronomy and Astrophysics Postdoctoral Fellow*
[11] *Flatiron Institute, Simons Foundation, 162 Fifth Ave, New York, NY 10010, USA*
[12] *IRFU, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France*
[13] *Université Paris Diderot, AIM, Sorbonne Paris Cité, CEA, CNRS, F-91191 Gif-sur-Yvette, France*
[14] *AIM, CEA, CNRS, Universit Paris-Saclay, Universit Paris Diderot, Sorbonne Paris Cit, F-91191 Gif-sur-Yvette, France*
[15] *Sydney Institute for Astronomy (SIfA), School of Physics, University of Sydney, Sydney, NSW 2006, Australia*
[16] *Konkoly Observatory, Research Centre for Astronomy and Earth Sciences, Eötvös Loránd Research Network (ELKH), Konkoly Thege Miklós út 15-17, H-1121 Budapest, Hungary*
[17] *MTA CSFK Lendület Near-Field Cosmology Research Group, 1121, Budapest, Konkoly Thege Miklós út 15-17, Hungary*
[18] *ELTE Eötvös Loránd University, Institute of Physics, 1117, Pázmány Péter sétány 1/A, Budapest, Hungary*
[19] *Department of Chemistry and Physics, Florida Gulf Coast University, 10501 FGCU Blvd. S., Fort Myers, FL 33965 USA*
[20] *Department of Astrophysics, IMAPP, Radboud University Nijmegen, NL-6500 GL, Nijmegen, the Netherlands*
[21] *Max Planck Institute for Astronomy, Koenigstuhl 17, 69117 Heidelberg, Germany*

(Received April 1, 2021; Revised April 1, 2021; Accepted April 1, 2021)

Submitted to ApJS

ABSTRACT

The NASA Transiting Exoplanet Survey Satellite (TESS) is observing more than ten million stars with time spans ranging from $\sim$ 27 days to about 1 year. This vast amount of data contains a wealth of information for variability, exoplanet, and stellar astrophysics studies but requires a number of processing steps before it can be fully utilized. In order to efficiently process all the TESS data and make it available to the wide scientific community, the TESS Data for Asteroseismology (T'DA) working group, as part of the TESS Asteroseismic Science Consortium (TASC), has created an automated open-source processing pipeline to produce light curves corrected for systematics from the short- and long-cadence raw photometry data and to classify these according to stellar variability type. This paper is the third in a series detailing the working process of the pipeline. We present our methodology for the automatic variability classification of space-based photometric data using an ensemble of supervised learners that are combined by means of the metaclassifier we developed. We successfully validate our method by means of a carefully constructed labelled sample of *Kepler* Q9 light curves with a 27.4

Corresponding author: J. Audenaert
jeroen.audenaert@kuleuven.be

days time span mimicking single-sector TESS observations, on which we obtain an overall accuracy of 94.9%. We demonstrate that our methodology can successfully classify stars outside of our labelled sample by applying it to all $\sim$ 167 000 stars observed in the *Kepler* Q9 observational Quarter.

Keywords: Asteroseismology, Machine learning, Supervised classification

1. INTRODUCTION

Stars deliver by far the largest fraction of visible light to the Universe, making them the main source of information for astrophysicists. Therefore, any astrophysical study that relies on the stellar component, either directly (e.g. detection and characterization of exoplanets, star formation, evolution, magnetism) or implicitly (e.g. cluster dynamics and evolution, galactic and extragalactic science), depends on how well we understand stars. In this context, the theory of stellar structure and evolution (SSE) is a critical ingredient of contemporary astrophysics, as it summarizes and represents our best understanding of how stars form, evolve, and ultimately die, with the evolution being largely driven by physical conditions in their interiors and atmospheres. Yet, SSE models are subject to large degeneracies due to the complicated physics being usually parameterized in 1D without observational test of the implemented theoretical prescriptions.

Stellar variability provides key insights for the determination of architectures of stellar and planetary systems alike. Both interior and surface properties of stars can be readily inferred from the temporal variability of their brightness. The field of asteroseismology relies on the detection, characterization, and interpretation of stellar oscillations in the form of these surface brightness variations to learn about the inner workings of stars (Aerts et al. 2010). The field is powerfully complemented with stellar spectroscopy that delivers precise measurements of atmospheric properties of stars as well as with binary stars that remain to be the prime source of precise and accurate measurements (i.e. model independent) of stellar masses and radii, as such serving as among the best calibrators of SSE models (Serenelli et al. 2020).

Both asteroseismology and stellar astrophysics in general have been revolutionized with the launch of space missions that delivered (and continue delivering) months- to years-long high-precision, high-cadence, and high-duty cycle brightness measurements for large amounts of stars. Following the MOST (Walker et al. 2003), WIRE (Buzasi 2004; Bruntt & Buzasi 2006) and CoRoT (Auvergne et al. 2009) space missions that were among the pioneers in the field of "space asteroseismology" (e.g. Aerts et al. 2010, for historical notes), *Kepler* (Borucki et al. 2010) observed around 160,000 stars in 30-minute (long-) and 1-minute (short-) cadence intervals for up to four years of time. After the failure of its two reaction wheels, the *Kepler* mission was turned into the *Kepler* Second Light (K2 Howell et al. 2014) mission that observed a large number of stars along the ecliptic plane during some 20 further campaigns, albeit with a shorter time span of some 93 days each. The TESS mission (Ricker et al. 2015) was launched in 2018 and is covering almost the entire sky. With millions of stars observed, it offers a multitude of *Kepler* in terms of targets, but only a smaller fraction in terms of light curve duration. The TESS targets in the Full Frame Images (FFIs) are observed at 30-minute cadence intervals while a pre-selected list of targets is observed at 2-minute cadence. In the extended mission the FFI cadence is reduced to 10 minutes, with an additional 20 sec cadence being introduced as well. The observing periods range from 27.4 d up to 352 d, depending on the position in the viewing zone. Given these large amounts of observations, it is intractable to manually process all light curves, and in particular to classify them according to their underlying variability type.

Automated variability classification based on light curves (and spectra) resulted from large-scale surveys such as by the Hipparcos mission, Optical Gravitational Lensing Experiment (OGLE[1]), All Sky Automated Survey (ASAS[2]), Sloan Digital Sky Survey (SDSS[3]), etc. The classifications varied in scale from general ones, e.g. Wyrzykowski & Belokurov (2008, OGLE), Pojmanski (2002, ASAS), and Ball et al. (2006, SDSS), to those focused on specific types of stars, e.g. Aerts et al. (1998) and Waelkens et al. (1998) from Hipparcos. Debosscher et al. (2007), Sarro et al. (2009), and Debosscher et al. (2011) presented an automated classification of light curves of variable stars in a supervised manner, employing Gaussian Mixtures and Bayesian Networks to classify OGLE, CoRoT, and *Kepler* Q1 data. Richards et al. (2011) also used a feature-based approach in combination with a Random Forest (RF) to classify variable stars in the OGLE and Hipparcos datasets. More recently, Kim & Bailer-Jones (2016) and Armstrong

[1] http://ogle.astrouw.edu.pl/
[2] http://www.astrouw.edu.pl/asas/
[3] https://www.sdss.org

et al. (2016) respectively used a RF, and Self-Organizing Maps (SOM) in combination with a RF, to perform classification of variable stars in the ASAS, MACHO (MAssive Compact Halo Objects), LINEAR (LincoIn Near-Earth Asteroid Research), and K2 (Fields 0-4) surveys. Pashchenko et al. (2018) apply Logistic Regression, Support Vector Machines, k Nearest Neighbours, Neural Networks, Random Forests, and Stochastic Gradient Boosting classifiers to a sub-set of the OGLE data and conclude that the Logistic Regression and k Nearest Neighbours methods provide a lower efficiency than the other tested classifiers. Naul et al. (2018) took a hybrid approach and reverted to automated feature learning by means of an unsupervised autoencoder in order to capture the stellar variability, and then subsequently used the latent layer as input into a RF. Jamal & Bloom (2020) extended this approach by making a comprehensive analysis of neural architectures suited for light curve classification. Unsupervised light curve classification is much less prevalent in the literature with a few application examples being by Valenzuela & Pichara (2018) and Modak et al. (2018).

1.1. *The TESS Data for Asteroseismology (T'DA) pipeline*

Coping with the large volume of data obtained by various space-missions, and in particular by the currently operational TESS mission, requires a coordinated effort. To that end, the TESS Data for Asteroseismology (T'DA[4]) coordinated activity has been created within the TESS Asteroseismic Consortium (TASC[5]). The major task of the T'DA unit is to serve the community with optimal processing of TESS data (both short cadence and full frame images), which includes raw light curve extraction, corrections of the extracted light curves for systematics, and their automated classification into variability classes. The overall scheme of the T'DA operations in depicted in Fig. 1, and includes the data processing and interpretation pipeline itself as well as the ways our data products are made available to the astrophysical community. The steps of the light curves extraction and their optimal corrections for systematic effects are described in detail in Handberg et al. (2021) and Lund et al. (2021), respectively.

This paper is the third in a series of the T'DA papers and concerns the automated stellar variability classification. The corresponding component within the T'DA pipeline structure is highlighted by the red dashed box in Fig. 1, while the classification scheme itself is depicted in Fig. 2. It comprises two major steps: (i) "top-level classification" that is based solely on the information encoded in the light curves themselves; and (ii) "second-level classification" that involves using extra information, such as Gaia parallaxes, photometric colours, etc. This latter classification step also involves using unsupervised methods for variability classification that will help us identify potential contamination and to search for new (sub-)groups of variable stars within our predefined general variability classes, and will be the subject of a separate study. Here we present a method for the supervised classification of light curves into broad variability classes as depicted by the blue boxes in Fig. 2 ("top-level classification"). We discuss the feature engineering and the collection of the training set, including a detailed description of each variability class, in Sections 2 and 3, respectively. Our individual base classifiers are described in Section 4 while their testing and validation is presented in Section 5. The individual learners form the basis for the metaclassifier that is tested and validated in Section 6 and is ultimately applied to the truncated 27.4 d segment *Kepler* Q9 data to mimick the single sector TESS case (Section 7). We close the paper with the discussion, conclusions, and an outline of future prospects in Section 8.

2. CLASSIFICATION FEATURES

The optimally extracted and corrected light curves are subject to a parameterization, a step that is often referred to as *feature engineering*. The T'DA classification pipeline provides the means for an automated feature extraction that is tuned to the needs of the individual classification base algorithms (cf. Sect. 4). Three main types of features are extracted and used in the process: (i) Fourier-based features; (ii) features computed in time domain; and (iii) image-type features.

2.1. *Fourier-based features*

An efficient way of extracting periodic signal from a time series of data is to take its Fourier transform. We employ the Lomb-Scargle periodogram method (Lomb 1976; Scargle 1982) to represent input light curves in the Fourier domain and perform classical iterative prewhitening (Degroote et al. 2009; Antoci et al. 2019) to extract individual frequencies with their corresponding amplitudes and phases. In this process, stellar flux is represented as

$$F(t) = C + \sum_{i=1}^{n}\sum_{j=1}^{m}\left(a_{ij}\sin\left(2\pi f_i j t\right) + b_{ij}\cos\left(2\pi f_i j t\right)\right), \tag{1}$$

with C representing the mean value of flux, n and m are the number of extracted frequencies f_i and their

[4] https://tasoc.dk/tda/

[5] https://tasoc.dk/

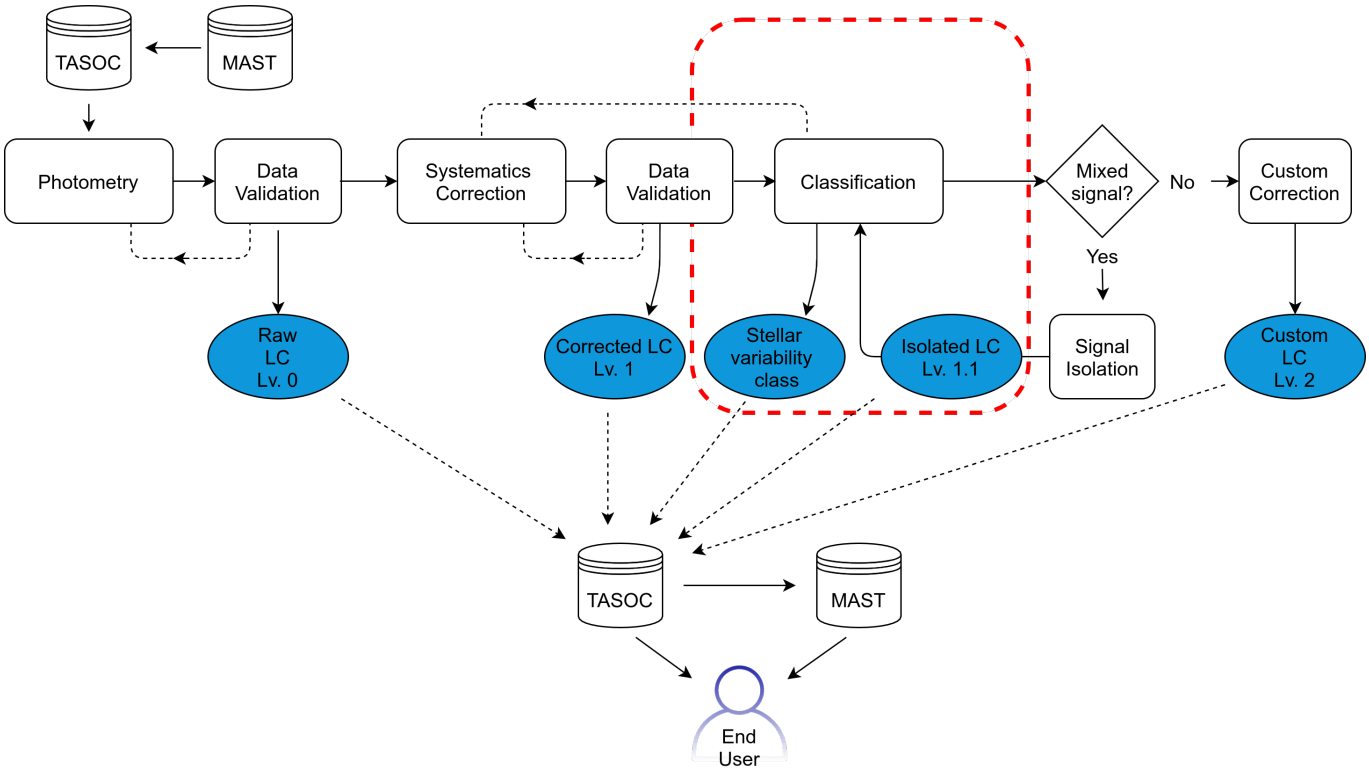

Figure 1: The overall structure of the full T'DA pipeline, with modules given as rectangular boxes, data products as ellipses, and "TASOC" and "MAST" indicate the databases hosting the data products. Dashed lines between modules indicate that an iteration might take place. The part enclosed by the red dashed line indicates the pipeline component described in this paper. The "photometry" part of the pipeline is described in Handberg et al. (2021), while the "correction" is detailed in Lund et al. (2021).

harmonic terms, and a_{ij} and b_{ij} are the Fourier coefficients. These coefficients are not unique, hence they are converted into time-translation invariant frequency amplitude (A_{ij}) and phases (ϕ_{ij}). We refer the reader to Degroote et al. (2009) for a detailed mathematical description of the conversion.

Iterative prewhitening assumes obtaining and subtracting the optimal fit for the frequency f_i and its m harmonic terms from the flux $F(t)$, and repeating the procedure until n frequencies are extracted from the data. The total number of extracted frequencies varies between individual time series and is determined by a significance criterion. This criterion can be based on a comparison between the amplitude optimized during the prewhitening process and the amplitude of the signal in the original Fourier transform. Additional criteria based on the signal-to-noise level of a frequency rather than its amplitude can also be used, as in e.g. Pápics et al. (2012). Such an approach prevents the extraction of spurious frequencies which are the residual signal from the preceding prewhitening steps. We refer the reader to Van Reeth et al. (2015a) and Antoci et al. (2019) for a detailed discussion on the method. The obtained set of frequencies, amplitudes, and phases form a basis for calculation of the Fourier-based classification features whose overview is provided in Table 1.

2.2. *Image-type features*

The Fourier-based feature selection assumes periodic signals as a good representation of the light curve. This is however not suitable for stochastic or non-periodic variability. For that reason, some of our base algorithms work with image-like features where either the power density spectrum (PDS) or the time series itself are represented as images. The PDS is the dominant frequency frequency analysis method for stochastically excited oscillations (Hekker & Christensen-Dalsgaard 2017; García & Ballot 2019a). The multiclass solar-like oscillation shape hunter algorithm (multiSLOSH, see Sect. 4.1 for details) therefore performs image recognition on the PDS of variable stars. The random forest general classification algorithm (RFGC, see Sect. 4.2 for details) employs the location of a star on the self-organising map (SOM) as one of the features in its classification scheme. The SOM location is obtained by comparing light curve shapes after folding them on the dominant extracted period, essentially treating the

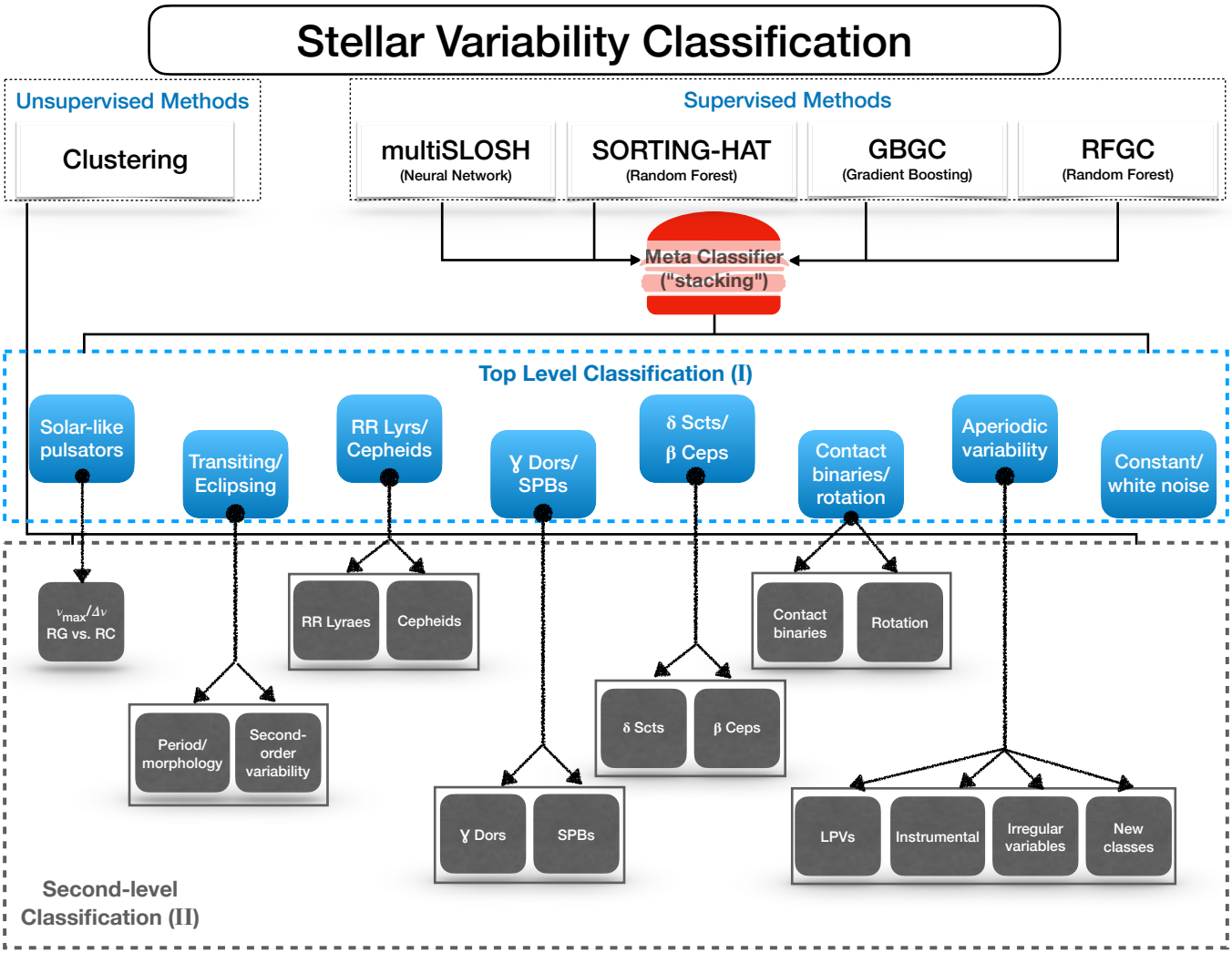

Figure 2: Graphical representation of the TASC classification scheme that encompasses two major stages: "Level 1" is the most general, largely light curve based classification, while "Level 2" stands for a detailed classification based on external features, such as parallaxes, colours, effective temperatures, etc. Rather than only relying on supervised learning, we also make use of unsupervised learning algorithms in Level 2.

phase-folded light curves as images. The respective features are listed in Table 1 with an indication of the base classifier(s) that employ them for the classification.

2.3. *Time-domain features*

Other classification features are extracted directly from the time series and are a certain type of statistical measure of the distribution of data points in the said time series. Some of those are well-known, general statistics features (e.g., skewness and variance). Below we provide a short description of features that are less intuitive and hence require a certain level of insight. All time-domain features are listed in Table 1 with the reference to classifiers that use them.

The *zero-crossings* parameter is computed from the "clipped" time series $Z_{k,i}$ defined as

$$Z_{k,i} = \begin{cases} 1 & \text{if } \mathrm{X}_{k,i} \geq 0 \\ 0 & \text{if } \mathrm{X}_{k,i} < 0, \end{cases} \tag{2}$$

where $X_{k,i}$ stands for the input time series comprising N data points. The number of zero-crossings D_k is then computed directly from the "clipped" time-series and is given by

$$D_k = \sum_{i=2}^{N} \left(Z_{k,i} - Z_{k,i-1}\right)^2 . \tag{3}$$

We normalize the number of zero-crossing to the total number of points N in the light curve to account for a possibly different length of the time series for the indi-

Table 1: Overview of classification features employed by the individual algorithms.

Algorithm/Feature	SORTING-HAT	SLOSH	RFGC	GBGC	Notes
PDS		x			power density spectrum
f_i, jf_i[a]	x		x	x	frequencies and their harmonics
A_{ij}				x	amplitudes
$\frac{A_{21}}{A_{11}}$, $\frac{A_{31}}{A_{11}}$			x		amplitude ratios
ϕ_{ij}				x	phases
$\phi_{i1} - \phi_{11}, i = 2, 3$			x		phase differences
FliPer (F_p)[b]					mean power in a given frequency range
F_p07,7,20,50			x		0.7, 7, 20, 50μHz onwards
SOM_loc			x		location on the trained self-organizing maps
ϕ_p2p_98			x		point-to-point difference, 98_{th} percentile,
p2p_98			x		ϕ refers to the phase-folded light curve
ϕ_p2p_mean			x		mean of the point-to-point difference,
p2p_mean			x		ϕ refers to the phase-folded light curve
D_k			x		number of zero-crossings in a light curve
ψ^2			x		coherency parameter
η_e[d]				x	variability index
skewness[c]	x			x	light curve skewness
MAD[e]			x	x	median absolute deviation
Rcs[f]				x	range of the cumulative sum of the fluxes
σ^2				x	variance
varrat[g]	x				variance ratio
NSH	x				number of significant harmonics of f_1
FR	x				flux ratio
$h(x)$	x				differential entropy
MSE					multiscale entropy
MSE avg,std,max,pow	x				mse mean, stdev, max and power

[a] $i \in [2, 6]$ and $j \in [1, 10]$; the number of frequencies and harmonics used is algorithm-dependent
[b] $F_{p,f_i} = \overline{\text{PDS}[f \to f_{\max}]} - P_n$, where P_n is the photon noise computed by considering the averaged power at high frequencies (Bugnet et al. 2018).
[c] skewness is defined by skew $= \frac{m_3}{m_2^{3/2}}$, where $m_r = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^r$ is the r_{th} moment about the mean $\overline{x}$
[d] variability index η_e is computed as ratio of the mean square of successive differences to the variance of the data points
[e] MAD = median($|X_{0,i} - \text{median}(X_0)|$), where X_0 stands for the whole time series while the subscript i refers to a single data point in the time series X_0
[f] cumulative sum of the fluxes is defined by $S_i = S_{i-1} + (x_i - \overline{x}), i \in [1, N]$, where $\overline{x}$ is the mean flux
[g] varrat $= (\sigma^2_{init} - \sigma^2_{sines})/\sigma^2_{init}$, where $\sigma^2_{sines} = \sum_{i=1}^{j} A_i^2$, A the amplitude and j the number of harmonics

vidual targets. Setting $k = 0$ gives the number of zero-crossing in the original light curve while $k > 0$ refers to the number of zero-crossing in the time series of higher-order differences. The kth order differences is defined by

$$X_{k,i} = X_{k-1,i} - X_{k-1,i-1}. \tag{4}$$

For example, the 1st order differences $X_{1,i}$ is given by the point-to-point differences in the original time series X_0, the 2nd order differences $X_{2,i}$ is given by the point-to-point differences in the time series X_1, and so on (Kuszlewicz et al. 2020).

The *coherency* parameter ψ^2 is a measure of the coherence (or stochasticity) of the signal in a time series and is computed from the zero-crossings of the higher-order differences in the time series. It is given by

$$\psi^2 = \sum_{k=0}^{5} \frac{(\Delta_k - \phi_k)^2}{\phi_k}, \tag{5}$$

where Δ_k gives the rate of change (i.e. increments) of the number of zero-crossings in the time series of higher-order differences and $\phi_k = (0.167, 0.066, 0.038, 0.025, 0.018)$ are the increments

computed from simulated time series of white noise (Kuszlewicz et al. 2020).

The *flux ratio* is the ratio of the sum of squared residuals of the fluxes either brighter or fainter than the mean flux (Kim & Bailer-Jones 2016) and is meant to capture eclipse-like variability. It is defined as

$$\mathrm{FR} = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2}{\frac{1}{M}\sum_{j=1}^{M}(x_j - \overline{x})^2}, \tag{6}$$

where $\overline{x}$ is the mean flux of the light curve and x_i and x_j the fluxes respectively brighter or fainter than the mean flux. For sinusoidal light curves the ratio will be close to 1, while for light curves with eclipses the steep flux gradients will cause it to be larger than 1.

The *differential entropy* is an extension of the Shannon Entropy (Shannon 1948) to the continuous domain. It is a measure of the average uncertainty of a variable, and thus a quantification of its unpredictability. The Shannon entropy $H(x)$ of a discrete random variable x is defined as

$$H(x) = -\sum_{i=1}^{n} p(x_i)\log p(x_i) = -E[\log p(x_i)] \tag{7}$$

where E is the expected value.

We use the differential entropy because, although the light curves are not continuous, they can typically take on a large range of values, causing the number of discrete states to equal the number of samples. This could distort the calculation in the discrete case, so we therefore opted to use the differential entropy. As an alternative one could also opt to use a binned version of the Shannon entropy. The differential entropy $h(x)$ of a continuous random variable x is defined as

$$h(x) = -\int \mu(x)\log\Big(\mu(x)\Big)dx \tag{8}$$

where $\mu(x)$ is the density function.

The entropy $h(x)$ can be calculated for a light curve or power spectrum, where in the latter case it essentially becomes the spectral entropy. Although both are strongly correlated, they complement each other in specific areas. The calculations of $h(x)$ are done with the Python-based Non-parametric Entropy Estimation Toolbox (NPEET)[6], which uses the Kozachenko-Leonenko estimate (Kozachenko & Leonenko 1987) to calculate the differential entropy as defined in Kraskov et al. (2004).

The sample entropy (Richman & Moorman 2000) is a different type of entropy metric that allows us to assess the complexity of a time series. The Sample entropy S_E of a signal is defined as

$$S_E(m, N, r) = -\ln\frac{A}{B} = \ln\frac{\sum_{i=1}^{N-m} n_i^m}{\sum_{i=1}^{N-m} n_i^{m+1}} \tag{9}$$

where m is the number of consecutive data points or the embedding dimension, r the tolerance, N the number of data points and n_i the number of vectors close to a basis vector, i.e. $d[u_i^m, u_i^m] \leq r$.

In practice we calculate the sample entropy by first identifying all unique sequences consisting of m consecutive data points, where each data point is written as $x_i + r$, with r a tolerance margin usually set to a factor of 0.15 of the standard deviation. We then count how many times a sequence or template vector of length m occurs and subsequently extend the template vector to length $m+1$ and count how many times that occurs. The calculations are repeated for each of the next m and $m+1$ template vector to determine the ratio between the total number of m and $m+1$ component templates, A and B respectively in Eq. (9). The sample entropy is the natural logarithm of this ratio and represents the probability that a sequence matching each other for the first m data points also match for the next $m+1$ data points.

The *Multiscale entropy* (MSE, Costa et al. 2005) attempts to take advantage of the fact that stellar variability is active on multiple time scales. Rather than calculating one entropy metric for the full series, we calculate the entropy at each time scale, allowing us to capture the full complexity. More specifically, we first coarse-grain the signal and then calculate the sample entropy for each of these new signals. This allows the MSE to assign minimum values to both deterministic/predictable signals and random/unpredictable signals. Given a time series $x_1, ...x_i, ...x_N$, the coarse-graining is achieved by dividing the time-series into non-overlapping windows of length τ. Each element in this time series is then calculated as

$$y_j^\tau = \frac{1}{\tau}\sum_{i=(j-1)\tau+1}^{j\tau} x_i, \quad 1 \leq j \leq \frac{N}{\tau}, \tag{10}$$

where τ is the window length, N the time series length and j the index after coarse-graining.

For $\tau = 1$, the time series $\{y_j^\tau\}$ is simply the original series. For each coarse-grained time series we then calculate the sample entropy given by Eq. (9) and plot it as a factor of the scale. The different types of complexity will then be represented by different types of MSE curves. In general we can say that 1) if for most values of τ the entropy is higher for one signal than for

[6] https://github.com/gregversteeg/NPEET

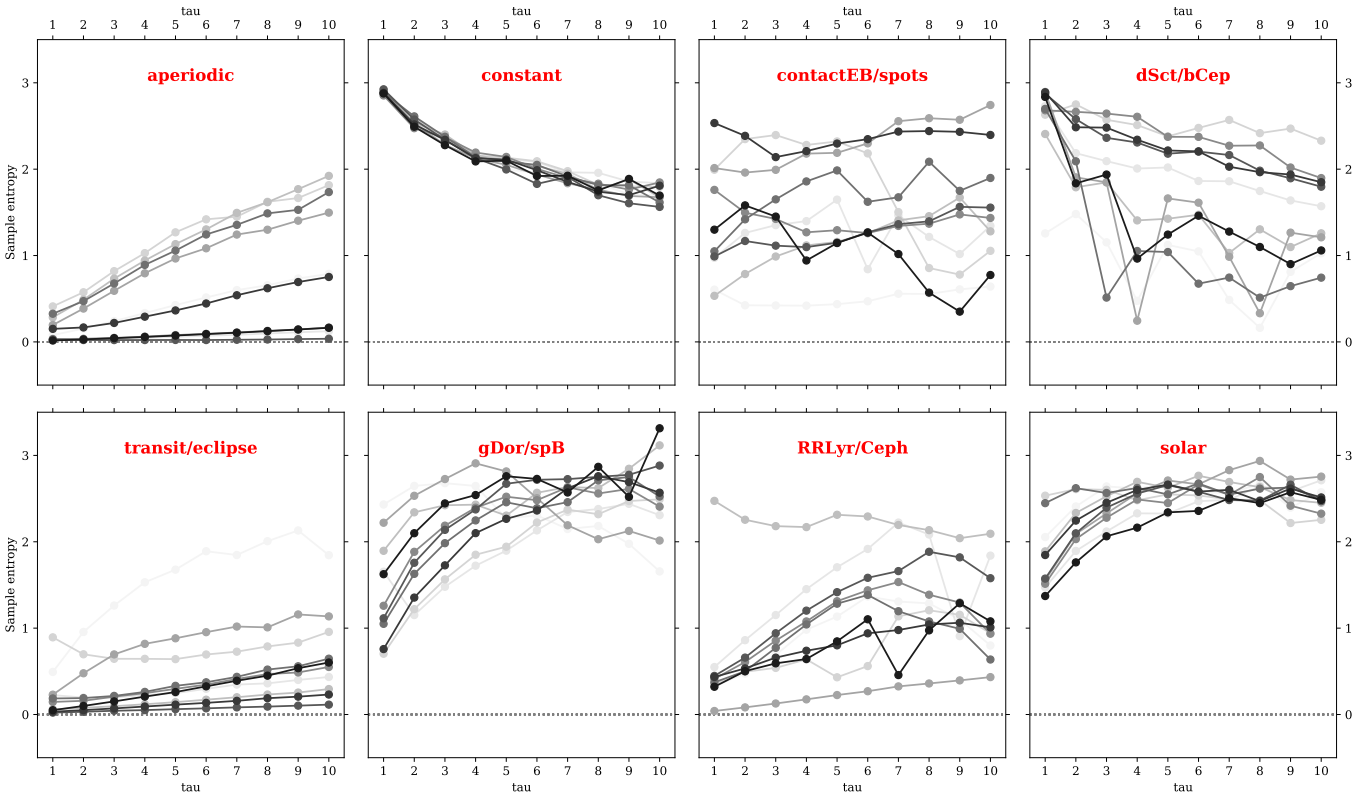

Figure 3: Examples of the Multiscale Entropy (MSE) curves for 10 random samples per variability class in our classification scheme as defined in Table 2.

another, that signal is considered more complex, and 2) that a monotonic decrease of the entropy indicates that the signal only contains information on the shortest time scale.

In order to obtain consistent Sample Entropy values it is suggested to have 200 data points per window at the minimum (Busa & van Emmerik 2016). Given that the shortest light curves observed by the TESS nominal mission will have a time span of ∼27.4 days, consisting of slightly over 1300 data points we set $\tau = 10$. This means that for the majority of the coarse-grained time series we have more than 200 data points, where at the smallest window length, i.e. when the scaling factor reaches 10, we have around 130 data points, which is still acceptable in terms of stability. We also did experiments with $\tau = 20$, and those provided good results as well. Fig. 3 shows the MSE curves for ten random samples per variability class. The figure illustrates the MSE's separating capacity, in particular for constant stars, solar-like oscillators and pulsators. Due to complexity associated with implementation of the full curves, we parametrize MSE through its maximum, mean, standard deviation and power, and use these as classification features.

3. VARIABILITY CLASSES & TRAINING SET

Table 2: Description of the training set.

Class label	Type	Size
aperiodic (Sect. 3.1)	Aperiodic stars	830
contactEB/spots (Sect. 3.2)	Contact binaries and rotational variables	2 260
dSct/bCep (Sect. 3.3)	δ-Sct and β-Cep stars	772
transit/eclipse (Sect. 3.4)	Eclipsing binaries	974
gDor/SPB (Sect. 3.5)	γ-Doradus and SPB stars	630
RRLyr/Ceph (Sect. 3.6)	RR Lyraes and Cepheids	62
solar (Sect. 3.7)	Solar-like pulsators	1 800
constant (Sect. 3.8)	Constant stars	1 000

The scientific needs of the TESS Asteroseismic Consortium (TASC) drive the selection of the main variability classes (schematically represented in Fig. 2) and hence our selection of the training set. Below we provide a short description of each of the variability classes listed in Table 2 alongside the selection criteria that were used to select stars into the respective classes. For all but one ("constant", see below for details) variability classes, we

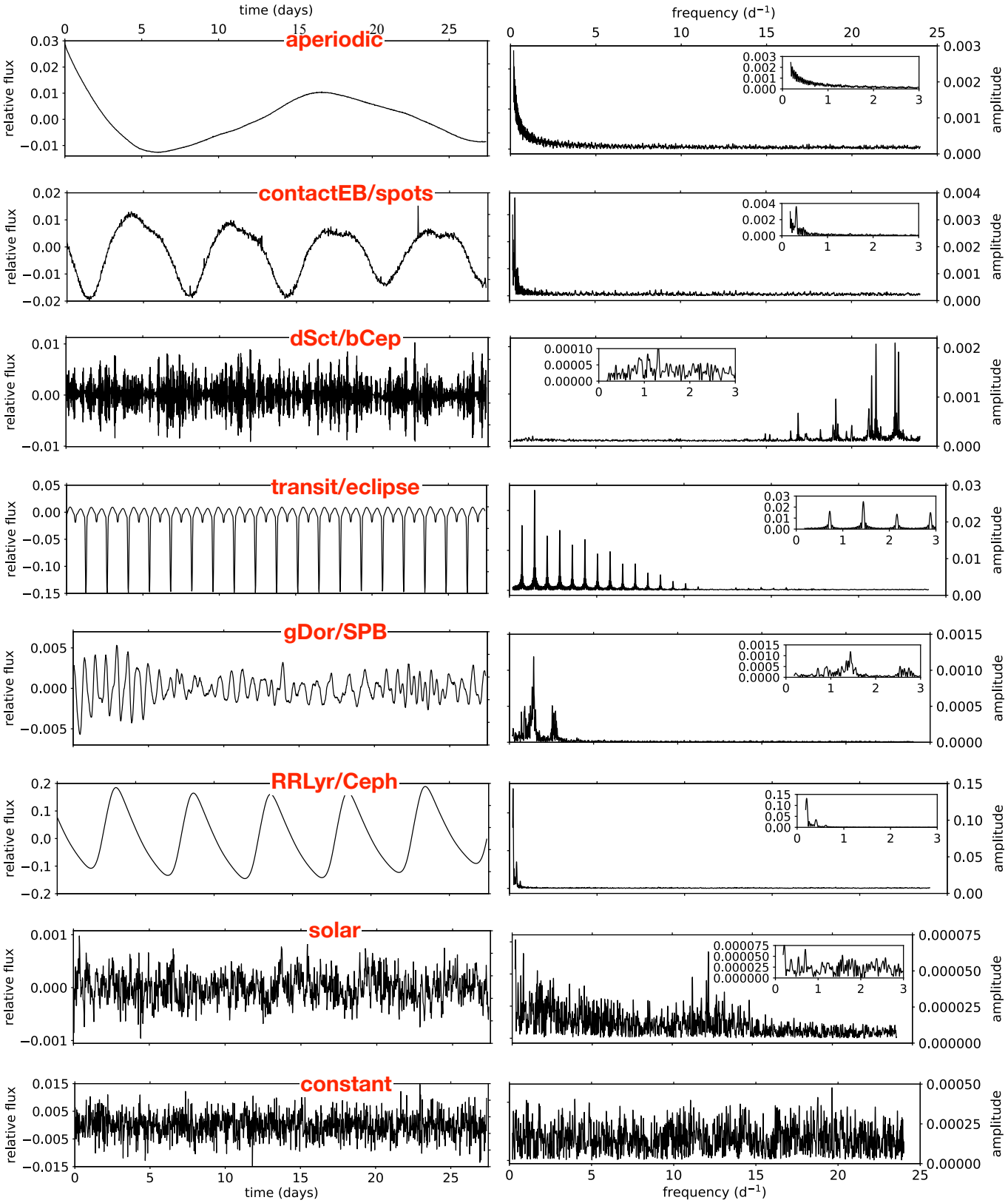

Figure 4: Examples of the light curves (left column) and the respective amplitude spectra (right column) from the training set as defined in Table 2. The inset in the amplitude spectrum panel (where provided) shows a zoom-in into the low-frequency domain of 0 to 3 d^{-1}. Note the different scale on the Y-axis.

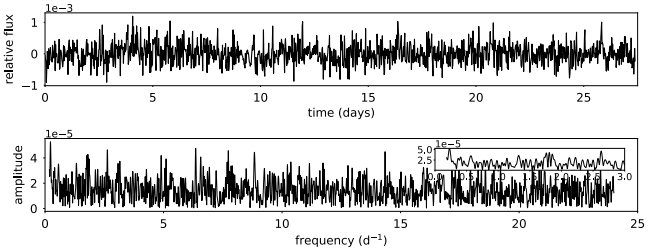

Figure 5: Light curve and amplitude spectrum of KIC 3448787 that was removed from the training set. We refer the reader to Fig. 4 for a light curve example of a typical RR Lyr/Cepheid pulsator.

make use of the latest Kepler data release 25[7] (Thompson et al. 2016), specifically the first 27.4 days of the Q9 PDCSAP data. Our choice of the 27.4 days total time base is dictated by the majority of TESS nominal mission targets being observed during its one targeting sector whose duration amounts to 27.4 days – two full orbits of the satellite around the Earth. The choice of the total length of the light curve and building the training set from white-light space-based *Kepler* photometric data enables a smooth knowledge and methodology transition to the TESS data. The choice for Q9 was made because this particular quarter of data is known to be the most complete in the sense of minimum number of gaps in the data.

3.1. *Aperiodic variables*

Aperiodic variability ("aperiodic") is a class introduced to account for targets whose variability (for one reason or another) appears to be lacking periodicity on the time scale of 27.4 days. For example, these can be Mira long-period variables whose variability remains unresolved on the time scale of 27.4 days as only a small fraction of the variability cycle is being captured. Similarly, a fraction of rotational variables may also appear as aperiodic stars due to their rotation periods being much longer than the length of the data set. Another example of aperiodic variables is Be stars, whose light curves are often irregular due to interplay of complex physical phenomena (such as mass-loss and rapid rotation) observed in these stars.

Our selection of aperiodic variables is based on the catalog of long-period variables compiled by Yu et al. (2020). The selection consists of 830 objects with the *Kepler* Q9 data available and having periods longer than 13.7 days so that less than two variability cycles are covered on the time scale of 27.4 days. An example of the light curve and amplitude spectrum of a *Kepler* aperiodic variable is shown in Fig. 4 (first row).

3.2. *Contact binaries & rotational variables*

Contact binaries & rotational variables ("contactEB/spots") is a combined class of i) contact binary systems, and 2) objects whose light curves show signatures characteristic of surface inhomogeneities modulated by stellar rotation over time. Contact binaries are short period gravitationally bound systems of two stars that both fill their Roche-lobes, and are therefore in contact at the Lagrangian point L_1. An example of a rotational variable is that of chemically peculiar (CP) BAF-stars that show anomalies in their surface chemical composition often associated with a non-uniform distribution of chemical elements. These surface inhomogeneities of either enhanced or depleted abundances of certain chemical elements are often termed "spots" as they appear to a distant observer as darker/brighter regions with respect to the bulk of the star due to significantly modified local opacities (Preston 1974). Atomic diffusion is believed to be at the origin of chemical peculiarities in CP stars, a mechanism that is driven by a competition between radiative levitation and gravitational settling (Michaud 1970; Richer et al. 2000).

The term "surface spots" in application to BAF stars with radiative envelopes should not be confused with surface spots observed in cooler stars that have extended convective envelopes, e.g. in the Sun. In the latter case, these are regions of reduced surface temperature associated with the contribution of a magnetic field to the total pressure, reducing the gas pressure. Solar-type spots are typically short-lived and vary in their appearance on the time-scales ranging from a day to a few months. Depending on the level of stellar magnetic activity, such "temperature spots" can be covering up to a few percent of the stellar surface, hence notably modulating the light curves of the respective stars (e.g. Namekata et al. 2019). Because many spots with varying temperature gradients and surface areas can be formed at the same time, light curves of cool active stars are typically much more complex than those of BAF CP stars whose "chemical abundance spots" are long-lived (time-scales ranging from years to decades).

Our selection of rotational variables of later spectral types is based on the catalog by McQuillan et al. (2014). The catalog contains rotation periods measured for over 30 000 *Kepler* main-sequence stars and selected to have KIC($T_{\rm eff}$)$<$ 6 500 K. In order to make sure at least two rotation cycles are covered with the 27.4 days data, we restricted our selection to systems whose rotation periods are shorter than 13.7 days. A total of

[7] https://archive.stsci.edu/missions-and-data/kepler/documents/data-release-notes

907 objects all having *Kepler* Q9 light curves were selected this way. The training set was enriched with rotational variables of earlier spectral types, i.e. with stars having KIC($T_{\rm eff}$)$\geq$ 6 500 K. To that end, we used the catalogs by Nielsen et al. (2013) and Hümmerich et al. (2018) which were cross-matched with the lists of "dSct/bCep"and "gDor/SPB" variables (see below) to check for and remove possible duplicates. Furthermore, we excluded stars that do not have Kepler Q9 data and/or whose rotational modulation signal is nowhere near the dominant signal in the light curve/amplitude spectrum. A total of 656 objects passed the above selection criteria and were added to the list of 907 rotational variables with KIC($T_{\rm eff}$)$<$ 6 500 K. An example of the light curve and amplitude spectrum of a rotational variable is shown in Fig. 4 (second row).

By analogy with the "transit/eclipse" class (see below), we queried the *Kepler* Eclipsing Binary Catalog for stars that have *Kepler* Q9 data and whose light curve morphology parameter is larger than 0.6 (high probability contact systems according to Matijevič et al. 2012). All 1054 systems selected that way were subject to visual inspection to remove contaminants of semi-detached type, resulting in the final selection of 697 contact binaries. Altogether, the "contactEB/spots" class comprises 2 260 objects, of which some 70% are rotational variables.

3.3. *δ Scuti & β Cephei stars*

δ Sct & β Ceph stars ("dSct/bCep") stars are two classes of variable stars pulsating in radial and low-order non-radial or gravity (g-) modes which are excited by means of the κ mechanism acting on the zone of partial ionization of helium (δ Sct stars) and of iron-group elements (β Cep stars, Aerts et al. 2010). The instability regions of the δ Sct and β Cep stars (partially) overlap in the HR-diagram with those of γ Dor and SPB stars (see below), respectively, giving rise to hybrid pulsators that exhibit both low-order modes and high-order g-modes simultaneously. β Cep stars have masses between some 8 and 18 $M_\odot$ and the periods of their pulsations range from about 2 to 8 hours. Less massive δ Sct stars cover the mass range from 1.5 to 2.5 $M_\odot$ and have periods from some 18 minutes to about 8 hours (Aerts et al. 2010).

Thanks to their short-period variability, δ Sct stars were among the first non-radial pulsators studied in detail from ground-based observations (Breger 2000). Significant observational progress in the study of δ Sct stars was achieved with the launch of CoRoT and *Kepler* space-missions, with a few highlights being: 1) the detection of pulsation amplitude modulations in the majority of the class members (Bowman et al. 2016; Bowman 2017); 2) the detection of semi-empirical relations between characteristic frequency spacing (an equivalent of $\Delta\nu$ in solar-like pulsators) and mean density of the star (Suárez et al. 2014; García Hernández et al. 2015, 2017); and 3) the detection of regular sequences of high-frequency p-modes and their exploitation in the modelling of δ Sct stars (Bedding et al. 2020). We refer the reader to Bowman & Kurtz (2018) for a comprehensive overview of observational properties of the δ Sct.

An overview of β Cep stars as a class was presented by Stankov & Handler (2005). Because massive stars were purposely avoided by the *Kepler* mission due to its core science of searching for (exo)planets around solar-type stars, β Cep stars are generally under-represented in space-based photometry. Hence, the vast majority of asteroseismic studies of the class members are associated with ground-based (multi-colour) photometric and (high-resolution) spectroscopic data (e.g., Aerts et al. 2003b,a; Aerts & De Cat 2003; Handler et al. 2006; Briquet et al. 2007, 2012; Tkachenko et al. 2014), with just a few class members being observed with the space missions WIRE (e.g., Cuypers et al. 2002), MOST (e.g., Handler et al. 2009; Tkachenko et al. 2016) and BRITE (e.g., Walczak et al. 2017; Handler 2017; Zocłońska 2018; Pigulski et al. 2018).

It proves to be difficult to distinguish between δ Sct and β Cep stars solely based on their light curve information. Hence, we introduce a joint class of coherent low-order mode (δ Sct/β Cep) pulsators in our classification scheme. Our selection of the training set for this class is based on the catalog compiled by Bowman et al. (2016). All 983 objects from that catalog were cross-matched with the catalogs we used to select g-mode pulsators (see below) to search for and remove possible duplicates. Light curves of the remainder of stars were subject to a visual inspection in order to exclude objects with pronounced signatures of rotational modulation as well as stars whose dominant pulsation signal was found to be in the g-mode regime (those hybrid pulsators were included in the class of g-mode pulsators; see below). Ultimately, we selected 772 objects into the class of p-mode pulsators, among those are stars exhibiting p-modes only and hybrid pulsators whose dominant signal is in the p-mode frequency domain. An example of the light curve and amplitude spectrum of a *Kepler* δ Sct p-mode pulsator is shown in Fig. 4 (third row).

3.4. *Eclipsing binaries and transit events*

Eclipsing/Transiting (transit/eclipse) systems are a class of objects that exhibit extrinsic variability in the form of (in the case of two or more events) periodic tran-

sits/eclipses. The latter occur due to a partial or total obscuration of the stellar disk by the companion that can be of either stellar (eclipses) or a planetary (transit) mass. We do not make a distinction between transits and eclipses, neither do we intend to distinguish between binary/multiple stellar systems with different Roche geometries (e.g., detached vs. semi-detached configurations). That said, we introduce a general class of eclipsing/transiting objects in our classification scheme which is also likely to contain members whose stellar components are intrinsically variable stars. Many of eclipsing and transiting systems have been discovered in the *Kepler* space-photometry in recent years, with a variety of orbital and stellar/planetary configurations. The most up-to-date overview of the detections in the *Kepler* field can be obtained from the *Kepler* Eclipsing Binary Catalog[8] and from the NASA Exoplanet Archive[9].

Our selection of the training set for the "transit/eclipse"class is based on the latest release of the *Kepler* Eclipsing Binary Catalog (Prša et al. 2011; Slawson et al. 2011; Kirk et al. 2016; Abdul-Masih et al. 2016). We started by selecting all systems with the morphology parameter smaller than 0.6 which allows us to filter out contact binaries while keeping the majority of detached and semi-detached systems (Matijevič et al. 2012). The light curves of all those 1679 objects were subject to a visual inspection in order to remove i) systems whose eclipses are hidden in the noise or any other astrophysical signal and are not traceable in the time domain without aggressive cleaning of the light curve; and 2) (long-period) systems that do not show a single eclipse event in the first 27.4 days segment of their *Kepler* Q9 light curve. Our final training set for the class comprises 974 objects; an example of the light curve and amplitude spectrum of a *Kepler* eclipsing binary is shown in Fig. 4 (fourth row).

3.5. *γ Doradus & Slowly Pulsating B stars*

γ Doradus & Slowly Pulsating B (SPB) stars (gDor/SPB) are members of a class of high-order gravity (g-)mode pulsators whose oscillations are excited by means of the flux blocking mechanism at the base of their convective envelope (γ Dor stars, Guzik et al. 2000) and by means of the κ mechanism operating on the zone of partial ionization of iron-group elements (SPB stars, Aerts et al. 2010). First detection of the γ Dor-type variability in the class prototype was reported by Cousins et al. (1989); Cousins (1992), followed by the detailed characterization of the variability class by Kaye et al. (1999). The term "SPB star" was first introduced by Waelkens (1991) who presented a study of seven variable stars of spectral types B3–B9. Although γ Dor and SPB stars occupy different locations in the HR-diagram representing F- (mass range between some 1.2 and 2.0 $M_\odot$) and B- (with masses from some 2.5 to 8 $M_\odot$) type stars, respectively, their light curves are remarkably similar. The light curves of γ Dor and SPB stars are shaped by an ensemble of g-mode pulsations whose periods range from ∼0.2 to ∼3 days. Gravity modes in these stars were also shown to exhibit characteristic quasi-equidistant period spacing patterns whose exact properties depend on the physical conditions in the near-core regions of the star, notably the rotation (Dupret et al. 2005; Bouabid et al. 2013). Following the first detection of a period spacing pattern in a B-type star observed with CoRoT (Degroote et al. 2010), g-mode period spacings are being actively exploited nowadays to deduce the interior physics of intermediate-mass stars (e.g., Van Reeth et al. 2015a,b; Pápics et al. 2015; Keen et al. 2015; Bedding et al. 2015; Moravveji et al. 2015; Van Reeth et al. 2016; Murphy et al. 2016; Moravveji et al. 2016; Pápics et al. 2017; Aerts et al. 2017, 2018, 2019; Li et al. 2020).

Our selection of g-mode pulsators is based on several intermediate- to large-scale studies of F- and B-type stars in the *Kepler* field. The sample of lower mass γ Dor stars was adopted from Tkachenko et al. (2013); Van Reeth et al. (2015a,b, 2016); Li et al. (2020), making sure to cross-match between the catalogs to exclude possible duplicates. In addition, the catalog of δ Sct stars compiled by Bowman et al. (2016) was employed to complement pure g-mode pulsators with stars that exhibit both g- and p-modes simultaneously, the so-called hybrid pulsators. We selected only those hybrid pulsators from Bowman et al. (2016) whose dominant variability was found in the g-mode frequency domain. Finally, the training set of g-mode pulsators was enlarged with SPB stars from Pápics et al. (2017) and Pedersen et al. (2020), providing us with a total of 694 stars, of which 630 objects have *Kepler* Q9 data. Because of the similar observational properties of their light curves, we do not distinguish γ Dor stars from their higher-mass SPB counterparts and combine them into a joint class of g-mode pulsators in our classification scheme. A typical light curve and amplitude spectrum of a g-mode pulsator is shown in Fig. 4 (fifth row).

3.6. *RR Lyrae and Cepheid stars*

Classical pulsators (RRLyr/Ceph class) are low- to high-mass evolved stars whose intrinsic pulsation variability is driven by the opacity (κ) mechanism acting on the partial ionisation zone of helium. The majority of

[8] http://keplerebs.villanova.edu

[9] https://exoplanetarchive.ipac.caltech.edu

these stars pulsate in a single dominant radial mode and have characteristic non-sinusoidal light curves. However, a small fraction of these objects show two or even three radial modes with comparable amplitudes. Variability of RR Lyrae stars occurs at periods shorter than one day, while Cepheids cover a much larger period range, from half a day to several months. Thanks to their famous period-luminosity relation established by Leavitt & Pickering (1912), Cepheids and RR Lyrae stars are used as galactic and extragalactic distance indicators.

About 50 RR Lyrae stars were identified in the *Kepler* field during the mission (Szabó 2018). In Q9, 42 of those were observed: 34 fundamental-mode and 8 first-overtone pulsators. No double-mode RR Lyrae stars have been targeted in the field, and only two Cepheids have been confirmed: a classical Cepheid, V1154 Cyg, and a medium-period, type II Cepheid, DF Cyg (Szabó et al. 2011; Derekas et al. 2017; Kiss & Bódi 2017; Vega et al. 2017; Plachy et al. 2018). From the list of 44 RR Lyrae stars and Cepheids, we excluded one object whose 27.4 days segment of the Kepler light curve and the Fourier transform computed from it do not display any significant signal (see Fig. 5). To increase the training sample, we collected 19 further Cepheids from the K2 observations, and created artificial light curves for them. We extrapolated the Fourier decomposition of the light curves to the Q9 time stamps and added appropriately scaled white noise to the data. Together, the 19 simulated Cepheid-type light curves and 43 Kepler Q9 RR Lyrae/Cepheid light curves provide us with a total of 62 objects in the final training set for the class. We do not differentiate between RR Lyrae stars and Cepheids in our classification scheme, but consider them as being members of the joint class of classical radial pulsators. Fig. 4 (sixth row) shows an example of a *Kepler* light curve of a RR Lyrae star along with its amplitude spectrum.

3.7. *Solar-like pulsators*

Solar-like pulsators (solar class) are intrinsically variable stars exhibiting oscillations driven by turbulent convective motions near their surfaces. Any star with an outer convective zone is expected to show such stochastically excited oscillations. Indeed, following the detection of solar-like oscillations in a number of main-sequence and evolved stars from ground-based data (e.g., Kjeldsen et al. 1995; Merline 1999; Bedding et al. 2001; Frandsen et al. 2002; Bedding & Kjeldsen 2003; Carrier et al. 2005; De Ridder et al. 2006; Bedding & Kjeldsen 2007), space-based photometry with the Hubble Space Telescope (HST), WIRE, MOST, SMEI, and in particular CoRoT and *Kepler*, revealed a treasure of pulsational variability in stars with outer convective regions and enabled extraordinary probes of their interiors and improvement of the respective models (e.g., Buzasi et al. 2000; Barban et al. 2007; Tarrant et al. 2007; Gilliland 2008; Stello et al. 2008; Kallinger et al. 2008; Stello & Gilliland 2009; De Ridder et al. 2009; Hekker et al. 2009; Miglio et al. 2009; Carrier et al. 2010; Kallinger et al. 2010; Bedding et al. 2010; Mosser et al. 2011; Beck et al. 2012; Davies et al. 2016; Brun et al. 2017; Buldgen et al. 2018; Eggenberger et al. 2019; Weinberg & Arras 2019; García & Ballot 2019b; Zhou et al. 2020; Yu et al. 2020). Stochastically driven solar-like oscillations are well characterized with two global asteroseismic quantities, namely the frequency of maximum power $\nu_{\rm max}$ and the large frequency separation $\Delta\nu$, which were shown by Kjeldsen & Bedding (1995) to scale with mass, radius, and effective temperature of the star. We do not provide an estimate of the global asteroseismic parameters of solar-like pulsators in our classification scheme, hence no differentiation is made between different evolutionary stages of stars.

Our selection of a sample of solar-like pulsators for the training set is based on the latest release of the APOKASC Catalog (Pinsonneault et al. 2018). A total of 1 800 objects were selected in a random way but making sure each of the targets had a Q9 Kepler light curve and oscillations detected with the CAN pipeline (Kallinger 2019). An example of the light curve and amplitude spectrum of a solar-like pulsator is shown in Fig. 4 (seventh row).

3.8. *Constant stars*

Constant stars (constant) are a class of objects that do not show any statistically significant variability on the time scale of 27.4 days. We made a random selection of 1 000 objects from the TESS Input Catalog[10] (Stassun et al. 2019) with the TESS noise estimates and simulated their light curves on the 27.4 days *Kepler* timestamps. An example of the light curve and amplitude spectrum is shown in Fig. 4 (last row).

4. METHODS – BASE CLASSIFIERS

We first train four individual base classifiers each using different feature sets and learning algorithms. In the next step we then combine these different models using stacked generalization by means of a meta classifier. The benefit of using this stacked ensemble of models is that by using a diverse set of base classifiers, we can leverage

[10] https://tess.mit.edu/science/tess-input-catalogue/

the individual strengths and weaknesses of each classifier to come to the optimal combination of models and obtain a better predictive performance compared to using just one classifier. In the following subsections we discuss each base classifier.

4.1. *Multiclass **S**olar-**L**ike **O**scillation **S**hape **H**unter (multiSLOSH)*

multiSLOSH uses image recognition via deep learning to visually determine the presence of the desired signal on a 2D plot of the power density of a star. This is the multiclass generalisation of the method described by Hon et al. (2018). The only difference from Hon et al. (2018) is that rather than inferring whether or not there are solar-like oscillations, we have allowed the classifier to classify all the classes of star in our sample. To summarize, a 128x128 grayscale image of the star's power density in log-log space is used as the input into a 2D deep learning network. The log-log representation of the power density is used because it easily shows the main distinguishing features of most stars. For example, in the case of a solar-like oscillator, which are the convective granulation background and the 'Gaussian-like' power excess containing the oscillation modes.

While the original method has shown to be effective in classifying red giants at any observation length obtained by TESS (Hon et al. 2018), SLOSH can be very easily generalized towards stars only observed in short-cadence, for example, dwarf (MS) or subgiant (SG) stars. This can be done by modifying the training set that the networks use to learn. To allow for the detection of signals in MS or SG stars, the plotting range in the 2D image has to be modified. The range in frequency, $f_{\rm range}$, and power density, $P_{\rm range}$, for the different evolutionary states are defined by the following:

$$f_{\rm range}(\mu{\rm Hz}) = \begin{cases} [3, 283] & \text{for long-cadence} \\ [40, 4160] & \text{for short-cadence} \end{cases}$$
$$P_{\rm range}({\rm ppm}^2\mu{\rm Hz}^{-1}) = \begin{cases} 3 \times [10^1, 10^7] & \text{for long-cadence} \\ 1 \times [10^{-1}, 10^5] & \text{for short-cadence} \end{cases} \quad (11)$$

Much of the effectiveness of deep learning comes from its ability to learn complex relations from large quantities of data. The red giant classifier by Hon et al. (2018) used a training set comprising ∼16000 oscillating giants from Yu et al. (2018) and ∼15000 randomly selected dwarf targets that did not show oscillations in the *Kepler* long cadence data. Due to the relatively small size of the training set, it is not possible for the deep learning network to have 'seen' and learned all types of stellar variability shown from different stars. This means that while it is highly efficient at identifying solar-like oscillations, it can at times get confused by other signals such as those from a δ Scuti star or a detached eclipsing binary. It is for this reason that we choose not to include the original SLOSH of Hon et al. (2018) but to instead create a multi-class version that can turn this observed weakness and turn it into a strength. As we shall see later in section 5 the new multi-class version of SLOSH does very well at classifying all of the stars in our sample.

4.2. ***R**andom **F**orest **G**eneral **C**lassification (RFGC)*

The RFGC uses a hybrid self-organising-map (SOM, Kohonen 1990; Brett et al. 2004) and Random Forest (RF, Breiman 2001) classifier, as previously demonstrated on data from the *K2* satellite (Armstrong et al. 2015, 2016). A full methodological description is provided in Armstrong et al. (2016). While the underlying methodology is the same, the features used here have been updated to better account for the new datasets and variability classes considered. Below we give an overview of the method.

Light curves are initially phase folded, using 64 equal width bins, on the dominant frequency as extracted in Section 2.1. We also test each light curve using half the dominant frequency, and if the resulting phase-folded light curve shows significantly reduced dispersion, the half-frequency is used. This test corrects typical eclipsing binary light curves, where the presence of primary and secondary eclipses often results in the dominant frequency being double the true binary orbital frequency.

The training set of phase-folded light curves is then used to train a SOM with shape (1,400) using 300 training iterations and a learning rate of 0.1. Training a SOM involves creating a set of template 'pixels' which steadily approach similarity to underlying shapes in the input data. In the end the pixels contain representations of various common and uncommon shapes seen in the training set. The index of the closest matching pixel to a test input is then a powerful feature for parameterizing the phase-folded light curve shape.

The actual classification is performed by a RF, implemented through scikit-learn (Pedregosa et al. 2011). The 22 features used are listed in Table 1, including the SOM location described above. We set the parameters of the RF by optimising the out-of-bag score. This led to a RF with 1000 component decision trees, 3 features maximum considered at each node split, 2 samples minimum required to split an internal node and a maximum tree depth of 15. We use the Gini impurity to select the best splits at decision tree nodes.

4.3. *Supervised randOm foRest variabiliTy classIfier using high-resolution pHotometry Attributes in TESS data (SORTING-HAT)*

The SORTING-HAT is a RF classifier with an architecture similar to RFGC. It does not use a SOM, but relies on a set of 13 carefully constructed features in the entropy, Fourier and time-domain, as described in Table 1. The usage of entropy metrics allows it to differentiate light curves based on their unpredictability and complexity.

The set of hyperparameters is the same as in RFGC, but was independently confirmed by optimising the weighted F_1 score in an initial version of the classifier, through a general randomized grid search followed by a narrow but complete grid search. This led to a RF with 1000 decision trees, 15 maximum tree depth, 2 samples per split at the minimum and the usage of the Gini impurity measure. The hyperparameters seem to be very robust as changes only marginally affect classifier performance.

4.4. *Gradient Boosting General Classification (GBGC)*

Similar to the RFGC discussed in section 4.2, GBGC is a tree-based ensemble method whose trees were constructed with Gradient Boosted Machines (GBMs, Friedman 2001). In contrast to RFGC, the *GBGC* is an adaptive method of constructing a model where the classifier aims to correct previous trees in the ensemble by assigning higher weights to the incorrectly predicted samples. The efficiency and generalisation abilities of the GBGC classifier were established using a sample of labelled light curves from the OGLE's catalog of variable stars (Udalski et al. 2008, 2015) in the LMC and *Kepler/K2* mission by Kgoadi et al. (2019). Eight hyperparameters were adjusted to improve the performance of the classifier. In addition to the number of trees in the ensemble (`n_estimators`) and the optimal depth of the trees (`max_depth`), the fraction of samples in the training set (`subsample`) and features (`colsample_bytree`) during were training were also tuned. To ensure convergence was reached in a timely manner, the learning rate of the gradient descent (`learning_rate`) was tuned once the `n_estimators` were determined. Adjustment of the hyperparameters was done to prevent over-fitting and reduce running time complexities. Optimal hyperparameters were established using a grid search with 10−fold cross validation. The resulted in 500 estimators, 6 maximum tree depth, 0.8 training sample ratio, 0.7 feature sample ratio, and a 0.1 learning rate.

The finalized GBGC classifier was trained on a subset of features from Table 1 that were selected using the feature selection strategy RFECV-CBFS introduced in Kgoadi et al. (2019). This is a two step feature selection process where recursive feature elimination (Granitto et al. 2006, RFE) is applied to select features that best describe light curves and can be mapped to the star classes. To reduce redundancy, the Pearson's correlation coefficients were used to remove correlated features from the selected subset (Hall 1999). Given that CBFS was RFE, a correlation threshold of 0.65 was applied to remove features. Furthermore, to accommodate the uneven distribution of the classes in our training set, feature selection was done with cross-validation. The GBGC classifier was constructed using `XGBoost` (Chen & Guestrin 2016) as the base estimator of the model.

5. TESTING & VALIDATION (INDIVIDUAL CLASSIFIERS)

The individual classifiers are tested and validated in two different ways to ensure that they are not overfitting the training data.

For a given training set, we take 20% from the start as a holdout set to test the trained ensemble of base classifiers (this is also used to test the metaclassifier which we will come back to in Section 6). We split the remaining data set into k folds and produce class probabilities for each by cross validation. When partitioning the data—including the creation of the holdout set—we stratify so that each class is represented by the same proportion in each subset.

The performance of the classifiers trained on each cross-validation fold is the first way we validate them. The accuracy of each classifier should be approximately the same over each fold if they are not overfitting the training set.

- multiSLOSH: $92.39 \pm 0.89\%$
- RFGC: $93.41 \pm 0.27\%$
- SORTING-HAT: $93.79 \pm 0.26\%$
- GBGC: $91.70 \pm 0.83\%$

Here we report the mean of the accuracy over the 5 cross-validation folds and the uncertainty as the standard deviation. The mean scatter of $\sim 0.5\%$ over the cross-validation is due to the small size of the some of the classes and the initial training set (0.5% corresponds to ~ 8 stars). This will become clearer when we look at the confusion matrices of the classifiers later in this section.

All of the classifiers perform well on the training set, with SORTING-HAT performing the best. As we shall see in Section 6, we are not concerned with a single

classifier performing better than all the others but more so about the classifiers being uncorrelated with one another. It is important that the individual classifiers have different strengths and weaknesses and each perform best on different parts of the training set if we are to leverage this information in a meta-classification stage.

The individual classifiers are also assessed on the 20% holdout test set as the second way to validate them. Whilst validating the classifiers over cross-validation folds gives us a good idea of how well the classifiers generalize to unseen data, testing the classifiers on a holdout set will give an idea of pure performance and accuracy.

- multiSLOSH: 91.48%
- RFGC: 92.56%
- SORTING-HAT: 93.70%
- GBGC: 91.36%

Overall the holdout set accuracy of the classifiers is comparable to their mean accuracy over the cross-validation folds, as for most classifiers the holdout set accuracy lies almost within one standard deviation of the mean cross-validation accuracy. For RFGC alone, we notice that the holdout set accuracy is about 0.6% lower than the left uncertainty bound. In absolute numbers however this is still very small and only represents ±10 stars. Given that the accuracies on both sets are of the same order, we can safely assume that the individual classifiers are fitting the data well and not doing any significant overfitting.

We use SHAP (SHapley Additive exPlanations; Lundberg & Lee 2017; Lundberg et al. 2020), a unified approach that connects game theory with local explanations to explain the output of a machine learning model, to compute the feature importance scores. The feature importance plot for SORTING-HAT is displayed in Fig. 6, where the hatched regions indicate the most important feature per class. It is clear from this plot that the Multiscale entropy (MSE) together with the first fundamental frequency and skewness are the most important attributes in the classification process of SORTING-HAT. The feature importance plots for RFGC and GBGC can be found in Appendix A. Due to the fact that updated features have been used in the training of RFGC, the feature importances will be different to those reported in Armstrong et al. (2016). We do not plot the feature importance scores for multiSLOSH given that it is a neural network classifier that does not rely on a set of predefined features, but rather learns a set of weights that define the importance of each region in the power spectrum image.

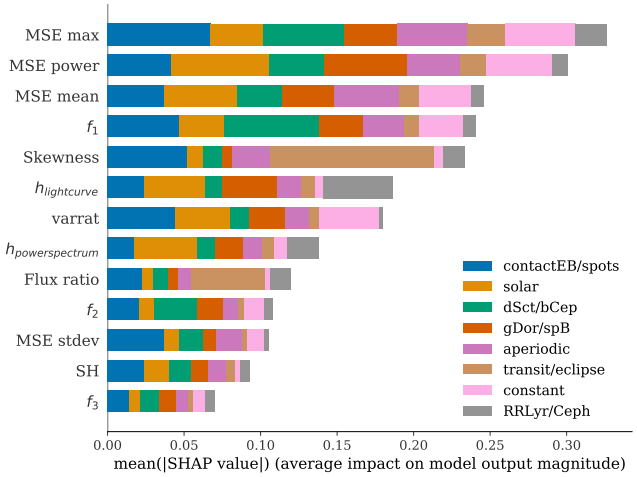

Figure 6: SORTING-HAT feature importances from SHAP. The hatched regions indicate the most important features per class.

6. THE METACLASSIFIER

Each of the individual classifiers described in Section 4 predicts the class probabilities for each light curve. We combine the predictions from this ensemble of models using stacked generalization (Wolpert 1992), in which we turn to another classifier that takes the probabilities outputted by the individual classifiers as its features to produce overall class probabilities for each light curve. This "metaclassifier" accounts for the relative strengths of the individual strong learners in the model ensemble.

6.1. *Training the Metaclassifier*

The stacked nature of our overall classification scheme could lead to overfitting and poor generalization to the unseen TESS data if the model is not trained carefully. Our approach to training the metaclassifier and individual base classifiers together closely follows Algorithm 19.7 in Aggarwal (2014). We represent our application of this training algorithm graphically in Figure 7.[11]

As explained in Section 5, for a given training set, we take 20% from the start as a holdout set to test the trained ensemble of base classifiers. We split the remaining data set into k folds and produce class probabilities for each by cross validation. We predict the class probabilities for each fold using the classifiers that are trained on the other $k-1$ folds. We assume that the performance of the classifiers trained on each cross validation fold approximates the performance of the models trained on all

[11] Inspired by the illustration at http://rasbt.github.io/mlxtend/user_guide/classifier/StackingCVClassifier/

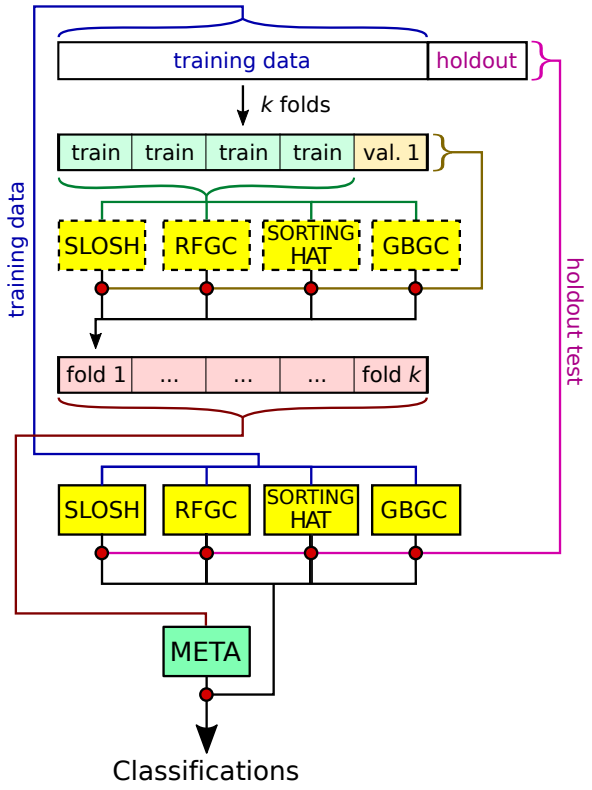

Figure 7: Depiction of the classifier training and testing procedure. 80% of the data set is split into k stratified folds for cross-validation. Class probabilities for data in each fold are predicted by the supervised base classifiers trained on the other $k-1$ folds. These predictions are used to train the metaclassifier. The base classifiers use to characterize the unseen data are trained on all of the training data. The success of the overall classification is tested by classifying the holdout data with the base classifiers, and then classifying the output of the supervised classifiers with the metaclassifier.

of the training data. The cross-validated class probabilities from each of the base classifiers on the training data are the inputs used to train the metaclassifier. The performance of the metaclassifier is finally tested on the holdout data by using the holdout set class probabilities predicted by the base classifiers trained on the training data (indicated in blue on Fig. 2) as input.

The algorithm we use for the metaclassifier is a RF with a similar architecture to RFGC (see Section 4.2), but with the number of estimators and maximum tree depth constrained to respectively 100 and 7, in order to avoid overfitting. This is chosen over a simpler scheme such as majority/soft voting because we want to leverage the potential correlations between classes. The meta classifier, like the base classifiers, predicts the class probabilities per star.[12] It is important to note that the class probabilities predicted by the metaclassifier Random Forest are well-calibrated, but not perfect. It might thus be better to interpret them as a confidence score rather than in a purely probabilistic fashion. If the metaclassifier assigns a confidence of 0.8 to 100 predictions, we should not expect that exactly 80 of those are correct. However, if we have a star with a confidence of 0.3 and a star with a confidence of 0.7, we can safely assume that the second one has a much higher probability of belonging to the class than the first one.

6.2. *Metaclassifier Testing & Validation*

The metaclassifier obtains an accuracy of 94.90%, meaning that overall it makes a high number of correct predictions. However, given that there are class imbalances in our training set the accuracy only provides a limited amount of information. We therefore also look at the confusion matrix as this gives a more detailed view on which classes the metaclassifier is performing better and worse. The confusion matrix for the metaclassifier is shown in Fig. 9. The classification rates (or recall) per class range from 89% for the γ Dor/SPB class to near a perfect score for the constant class. A detailed look reveals that the lower score for γ Dor/SPB is mostly caused by confusion with the δ Sct/β Cep and contactEB/spots class. Our visual analysis shows that the former can be explained by the presence of hybrid pulsators in the training sample, while the latter is caused by γ Dor/SPBs containing either some rotational signal or low frequencies that resemble those of contactEB/spots. We also notice some confusion between the aperiodic and contactEB/spots classes, this is mostly caused by the fact that both classes can mimic each other on the short time scale of 27.4 days. Lastly, there is a fraction of solar-like oscillators being predicted as aperiodic variables, where we find that the contaminants all have low $\nu_{\rm max}$ values, hence their light curve and power spectrum properties are similar to those of aperiodic stars. The high percentage for RRLyr/Ceph miss-classifications is due to the small class size and in absolute numbers only concerns one star.

In Fig 8 we show the feature importance plot for the metaclassifier, which allows us to analyze the contribution of each individual classifier towards the final prediction. This reveals that the 'multiSLOSH_solar' probability is the most important feature in the classification of

[12] We use the default *soft voting* scheme for the metaclassifier in which the predicted class probabilities are equal to the average predicted probabilities over all decision trees, where the class probability of a single tree is equal to the fraction of samples of the same class in a leaf.

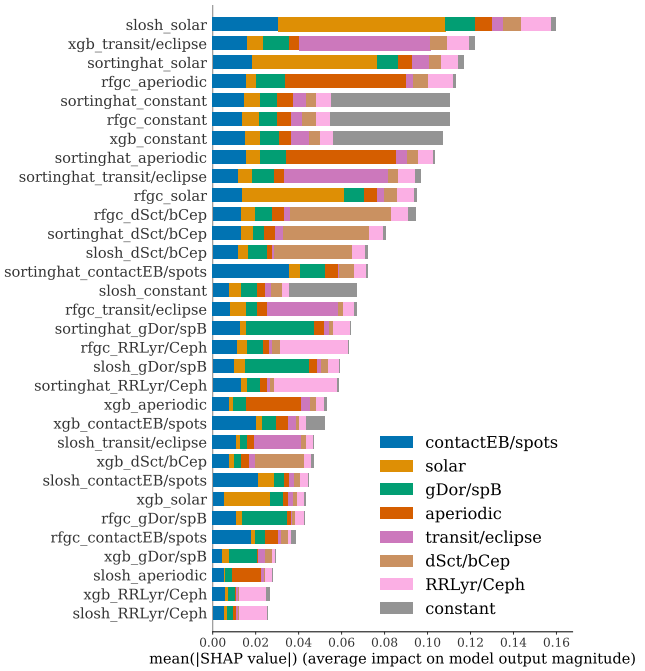

Figure 8: Metaclassifier feature importance from SHAP. The hatched regions indicate the most important feature per class.

solar-like oscillators, followed by SORTING-HAT. This could be expected given that SLOSH was initially designed to classify this type of star and in the case of SORTING-HAT the entropy features allows it to capture the stochastic nature of the signal. GBGC is most important in classifying transit/eclipse signals, followed by SORTING-HAT. In the aperiodic case it is RFGC also followed by SORTING-HAT. As for the constant class, we notice that the three feature-based classifiers, RFGC, SORTING-HAT and GBGC, are making nearly equal contributions, with RFGC being the most important one. The same true for the dSct/bCep class. For the contactEB/spots class SORTING-HAT is most important. What is interesting in this case however, is that it is followed by the multiSLOSH probability of being a solar-like star. By plotting the shap values of every feature for every star specifically for the solar class we can analyze the impact of each feature on the model output (i.e. the probability of being classified as solar). This reveals that a high 'multiSLOSH_solar' probability lowers the predicted probability of being a contactEB/spots star and vice-versa. Lastly, we see that SORTING-HAT followed by multiSLOSH are the primary classifiers for gDor/spB stars and RFGC followed by SORTING-HAT for RRLyr/Ceph. The feature importance plot clearly shows that the metaclassifier gets its strength by combining the different base classifier results.

We also assess performance by looking at the receiver-operator characteristic (ROC) curves. The ROC illustrates the diagnostic ability of a classifier by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) for different classification thresholds. This allows us to assess the performance of the classifier at each threshold. We calculate the ROC curves for each class using a one-vs-rest methodology (Fawcett 2006). The ROC curves per class are shown Fig 10. Ideally, the ROC curve is as close to the top left corner (0,1) as possible, as for this threshold on the curve the classifier is making a high number of correct classifications with a small amount of false positives.

Final class labels are commonly assigned to the class with the highest probability, which is equivalent to using a probability threshold of $1/C$, where C is the number of classes. In case more than one class exceeds this threshold, the class with the highest probability is chosen. When dealing with class imbalance however, this $1/C$ approach will often not lead to the optimal results (Provost 2000). We therefore opt to fine-tune the classification threshold by choosing the threshold that maximizes the True Positives Rate (TPR) and minimizes the False Positive Rate (FPR), which is essentially the point on the ROC curve that is closest to the top left corner. This point can be determined by finding the threshold that maximizes Youden's J statistic (Youden 1950), which is the difference between the TPR and FPR. Given that we have one ROC curve per class, this implies that we also have a different threshold for each, reflecting the classifier's differing ability in identifying the class members of each variability class. The obtained thresholds per class are given in Table 3.

As an aggregate performance measure across all probability thresholds used in the ROC curve, we can measure the Area Under the ROC curve (AUROC). The AUROC represents the probability that the classifier assigns a random positive example a higher probability than a random negative example. It is thus a measure of how well the classifier predicts the correct class. Given that we are working with a one-vs-rest methodology here, it means that the respective ROC class is the positive class and are other classes represent belong to the negative class. The confusion between the gDor/spB and contactEB/spots class causes their AUROC values to be slightly lower compared to the other classes.

7. *KEPLER* Q9 DATA

We validate our classification scheme by applying it to all 167 243 stars observed in *Kepler* Q9, but with light curves cut to the first 27.4 days. We start with the "base" scenario as described in the previous sec-

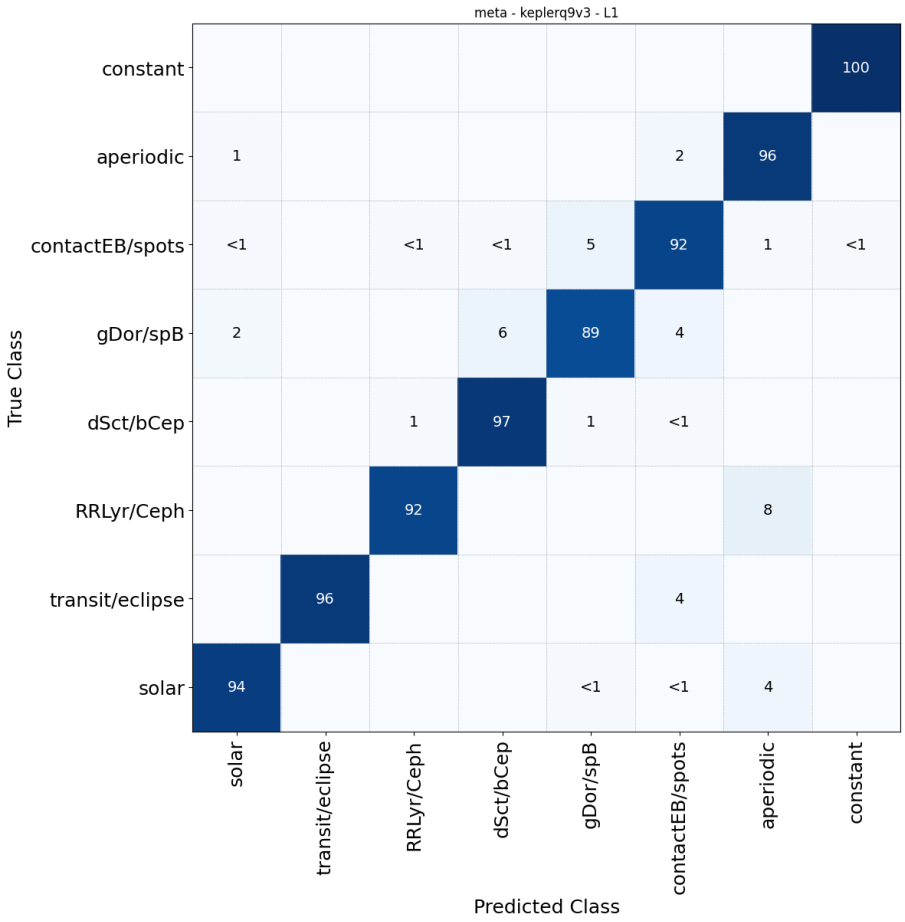

Figure 9: Normalized confusion matrix of the metaclassifier for the holdout set in percentages. Each element shows the fraction of stars that were predicted as positive for a particular class (column) over the total number of stars that truly belong to that class (row). The diagonal shows the fraction of stars that the classifier correctly predicted as positive for that class (i.e. the recall rate = $\frac{TP}{TP+FN}$, where TP is the number of True Positives and FN the number of False Positives).

Table 3: Classification thresholds per class

Class label	Probability threshold
aperiodic	0.116
constant	0.573
contactEB/spots	0.274
dSct/bCep	0.048
transit/eclipse	0.065
gDor/spB	0.138
RRLyr/Ceph	0.111
solar	0.338

Table 4: *Kepler* Q9 classification summary: number of stars per class for each thresholding method

Class label	# stars per threshold type	
	1/C	Youden's J
aperiodic	3 711	3 711
constant	5 061	0
contactEB/spots	140 566	137 377
dSct/bCep	1 758	1 758
transit/eclipse	1 563	1 563
gDor/spB	2 263	2 263
RRLyr/Ceph	96	96
solar	12 225	12 027
unknown		8 448

tions, test the effect of linear detrending in the "linfit" scenario, and ultimately assess the advantage of introducing an "instrument" class in the "base+instrument" scenario. For each of those scenarios, we assess both the results on the holdout set and on the *Kepler* Q9 data set. The assessment is achieved by the summary statis-

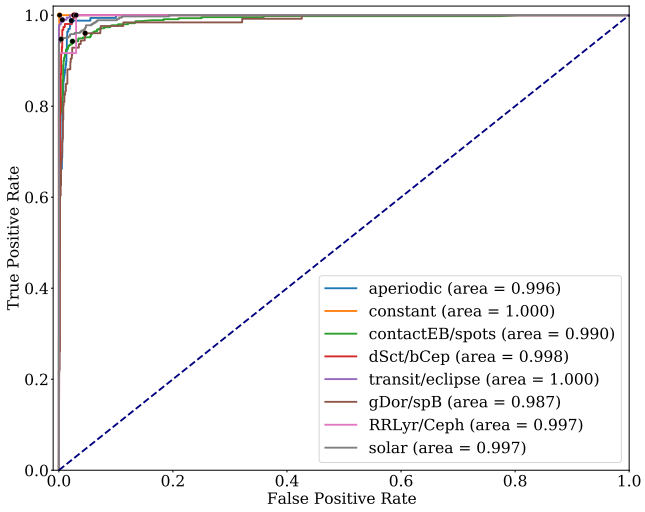

Figure 10: Receiver-Operator Characteristic (ROC) curves of the metaclassifier for each class. The Area Under the ROC (AUROC) curve is indicated next to the classes in the legend and the dots represent the TPR and FPR for the chosen probability thresholds. The dashed line indicates the 'random chance' curve.

tics of both data sets and by visually inspecting random sub-samples of 1000 light curves in each class for the *Kepler* Q9 classification results.[13]

7.1. *The "base" scenario*

The results for the "base" scenario are summarized in Table 4. The left column lists the numbers for the label assignments being made according to the highest probability, while the right column gives those according to the optimized probability thresholds. Overall, we see that all predicted classes, apart from "contactEB/spots", are rather pure and are in the same order as in the confusion matrix in Fig. 9. The high number of stars in the "contactEB/spots class" can be explained by the fact that light curves that are not assigned to any of the other classes, for example with the dominant instrumental signal in the low-frequency domain, end up in this bin. A careful visual inspection of the light curves and amplitude spectra of random subsamples of 1000 light curves in each class strengthens the above conclusion; below we present a concise summary of our visual analysis.

[13] Before taking the random sample we first removed the stars that were included in the training set.

The aperiodic variables are identified robustly by our methodology with an overall low number of misclassifications. After inspecting 1000 light curves randomly selected from the respective class, we confirm that some 97% of the light curves indeed exhibit aperiodic type of variability as demonstrated in the first row in Fig. 4. The most common misclassifications (about 3% in total) belong to the "contactEB/spots"class and are the light curves resembling rotational modulation and/or binary ellipsoidal signals, in many cases with the coverage of a single rotation/orbital period. The median probability value for misclassifications is found to be $p(x) \approx 0.50$. The worst-case scenario misclassification light curve is shown in Fig. 11 (first row) where likely a close eclipsing binary got (mis)classified as an aperiodic variable.

The "contactEB/spots" class suffers the most from misclassifications and partially resembles properties of "miscellaneous" classes often employed by some other light curve classification methods (e.g., Debosscher et al. 2011). Fig. 12 (orange line) shows the probability density function for the "contactEB/spots' class in the "base" classification scenario. One immediately notices an excess of objects in the low probability regime ($p(x) \lesssim 0.55$) as well as a double-peak feature at high probabilities ($p(x) \gtrsim 0.65$). Owing to this distribution, we divide the "contactEB/spots" class into three probability bins and visually inspect 500 randomly selected light curves in each of them: 1) $p(x) < 0.65$, 2) $0.65 < p(x) < 0.75$, and 3) $p(x) > 0.75$. We find that the lowest probability bin ($p(x) < 0.65$) contains some 97% misclassifications, among those the dominant fraction (about 90%) are light curves that exhibit some sort of an "instrumental" signal (see the second row in Fig. 11). The latter can be either truly instrumental in origin or due to inferior processing of data. The intermediate-probability bin that is associated with the first peak in the kernel-density plot ($0.65 < p(x) < 0.75$, Fig. 12) is also found to be rich in misclassifications (overall about 92%). Yet, the major difference with the low-probability regime is that the fraction of light curves that exhibit pure "instrumental" signal is significantly lower, some 55%. In the rest of the light curves, the "instrumental" and the true astrophysical signals are found to co-exist, e.g. as shown in the third row in Fig. 11. In this particular example, a weak astrophysical signal that is (on the time scale of 27.4 days) aperiodic in nature co-exists with the low-frequency signal due to inferior processing of the data. Lastly, the highest probability bin associated with the tallest peak in the probability density function ($p(x) > 0.75$, Fig. 12) contains some 16% misclassifications that are pure "instrumental" in nature and about 34% are the light curves that show both "in-

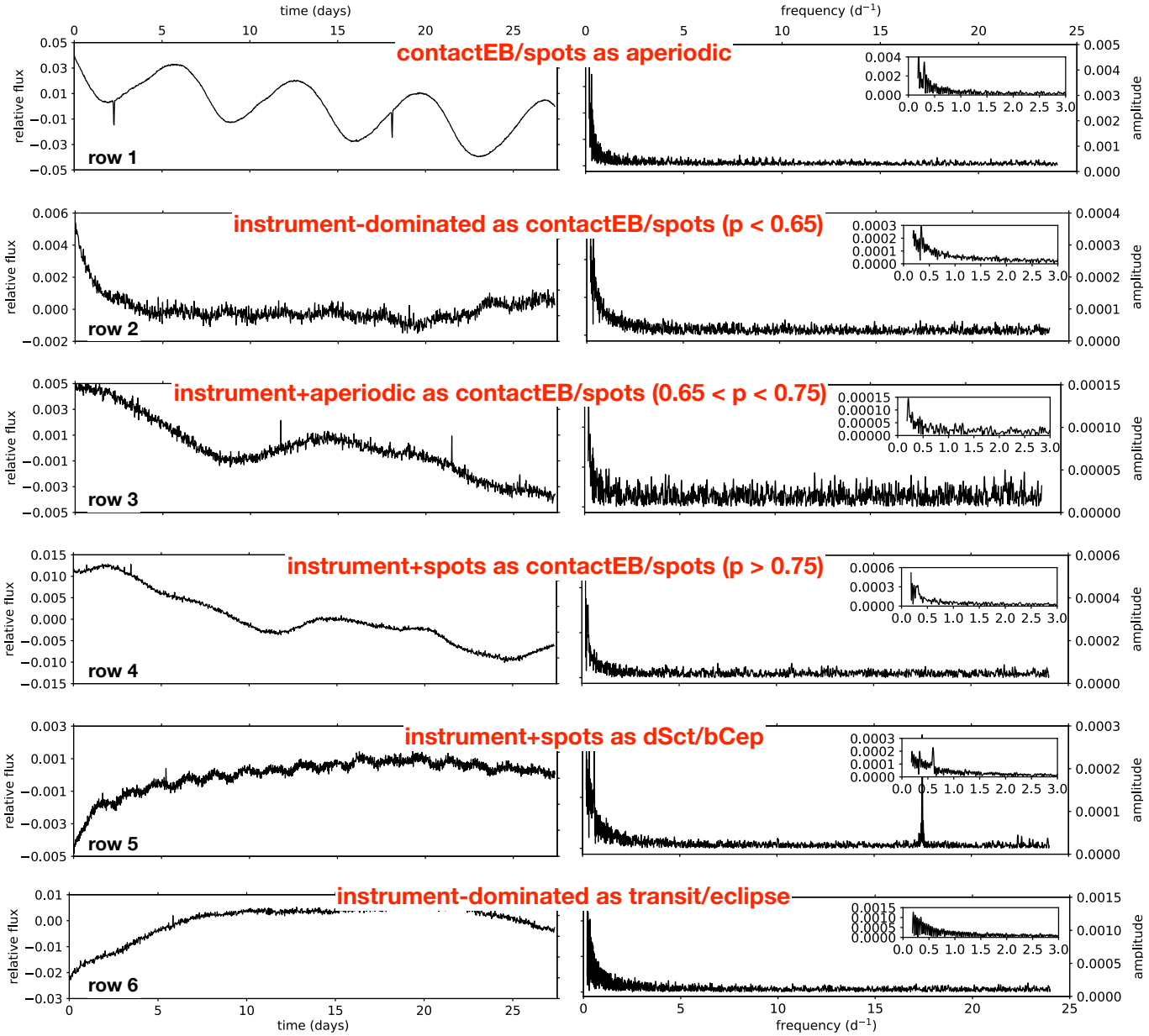

Figure 11: Examples of the misclassified *Kepler* Q9 light curves in the "base" scenario. Left and right columns show the light curves and the amplitude spectra, respectively. Note the different scale on the Y-axis of the plots.

strumental" and astrophysical signals at the same time. We note that the latter are not necessarily misclassifications, it is just that we visually identify the "instrumental" signal as being the dominant one in the respective light curves (the fourth row in Fig. 11). Finally, we note that pure astrophysical contamination is dominated by aperiodic variables and is at the level of some 17%. Many of those seemingly aperiodic signals might in fact be rotational variables with periods longer than 13.7 days and therefore spanning less than half of the rotation cycle, hence are visually classified by us as "aperiodic" stars. That said, we recommend a probability threshold of $p(x) \gtrsim 0.75$ for the high-confidence selection of "contactEB/spots' variables from the respective class (an example is shown in the second row in Fig. 4) while the results with probabilities $p(x) \lesssim 0.75$ should be taken with caution and keeping in mind that the number of genuine astrophysical signals in this particular class drops substantially towards low probability values.

The "dSct/bCep" variables are identified with high confidence by our methodology, where the overall fraction of misclassifications amounts to some 3.5%. The vast majority of misclassifications are due to spurious

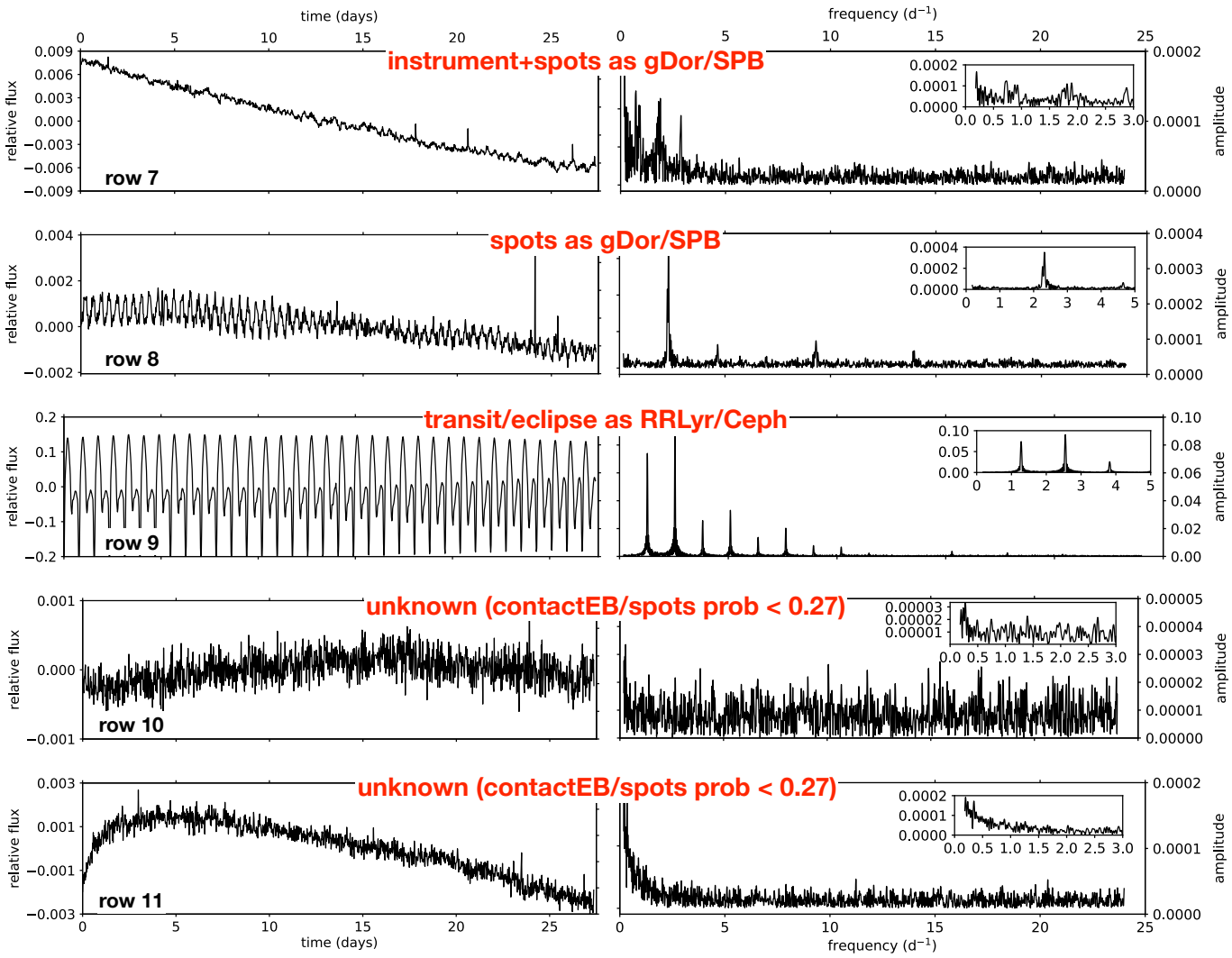

Figure 11: Continued

peaks found in the high-frequency domain (the fifth row in Fig. 11). The median probability value for the misclassified light curves is $p(x) \approx 0.45$. A considerable fraction of the identified δ Sct stars also exhibit low-frequency variability (either due to gravity-mode oscillations or rotational modulation), yet the high-frequency component is significant in all the detections and is the dominant one in the majority of them.

The "transit/eclipse" class is among the cleanest identified with our method, containing some 10% misclassifications overall. All misclassifications look alike and are due to imperfections in the data processing mimicking a flux drop in the light curve, most often at the beginning/end of the dataset. A typical example of the "transit/eclipse" missclassification is shown in the sixth row in Fig. 11. We also note that the type of light curve shown in the fifth row in Fig. 11 has high chances of being misclassified as a "transit/eclipse" variable, subject to the absence of the high-frequency peak. We find the median probability value for the misclassifications to be $p(x) \approx 0.45$.

The "gDor/SPB" class suffers from some 30% contamination, either from astrophysical signal of different origin or from low-frequency signal due to imperfections of data processing. The median probability value for misclassifications appears to be $p(x) \approx 0.52$ and the most common astrophysical contaminants are stars that belong to the "contactEB/spots" class otherwise. We show a typical example of the misclassified light curve in the seventh and eights rows in Fig. 11. The Fourier transform of the light curve reveals a rich variability spectrum at low-frequencies, possibly with the harmonic structure. Owing to the frequency range of gravity-mode oscillations observed in γ Dor/SPB stars and to the short time span of light curves that are being classified (see the fifth row in Fig. 4), "contactEB/spots" class members are indeed the primary candidates for an astrophysical contamination in the gDor/SPB class.

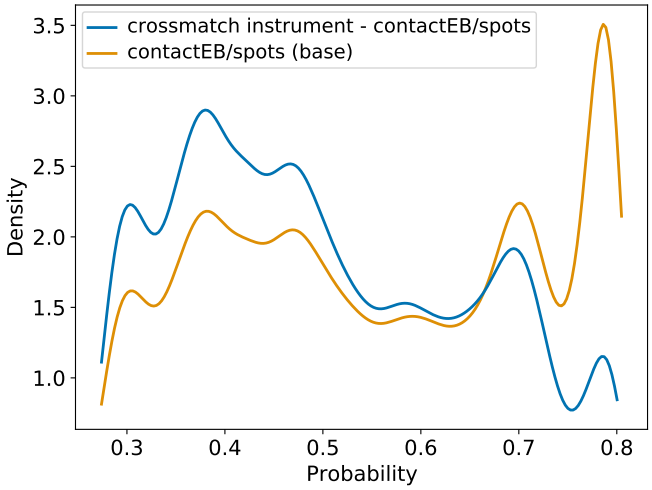

Figure 12: Kernel density estimate (KDE) plot comparing $p^{base}_{contactEB/spots}(x)$ of instrument stars from the "base-instr" contained in the contactEB/spots "base" scenario class against the complete contactEB/spots "base" class, with class assignments based on Youden's J statistic. Stars from the training set have been subtracted from both sets.

The "RRLyr/Ceph" class of classical pulsators is found to be small (see Table 4) which is expected given the location of the *Kepler* field. Contamination amounts to some 60%, is mostly astrophysical in origin, and is dominated by the "contactEB/spots" or "transit/eclipse" class members. All three types of objects (including "RRLyr/Ceph") have their dominant signals in the low-frequency domain and will often show a harmonic structure. However, owing to the characteristic shapes of their light curves (as shown in the sixth row of Fig. 4), "RRLyr/Ceph stars" are usually readily distinguished from other classes of variable stars. Indeed, only a small fraction of binaries and/or rotational variables with light curves that closely resemble those of the "RRLyr/Ceph" class are expected to exist and hence contaminate the class of classical pulsators. The high relative contamination from binaries and rotational variables that we find in this study is very likely the result of the "contactEB/spots" and "transit/eclipse" classes being at least two orders of magnitude larger than the "RRLyr/Ceph" class itself, hence increasing the chance of contamination. An example of the light curve and amplitude spectrum of the eclipsing binary contaminant is shown in the ninth row in Fig. 11.

The "solar" class stands for stars exhibiting solar-like oscillations and is the second largest class that we identified. Contamination is found to be small in this class, amounting to some 4%, and is mostly instrumental in origin. The median probability value for misclassifications is found to be $p(x) \approx 0.55$.

The "unknown" class contains objects that do not satisfy the Youden's J statistics-based thresholds per class as listed in Table 3. By comparing the class sizes before and after applying the thresholds (see Table 4) one notices that the "unknown" class inhibits the entire class of "constant" stars as well as a small fraction of the lowest probability objects from the "contactEB/spots" class. 99% of objects found in the "unknown" class do not exhibit any significant astrophysical signal, with the two typical examples being shown in the two bottom rows (10 and 11) in Fig. 11.

7.2. *The "linfit" and "base+instrument" scenarios*

In our next step, we test the "linfit" scenario where we automatically remove the linear trend from all *Kepler* 27.4 days light curves prior to computing the Fourier- and time-domain features from them. As can be seen in Fig 11 (e.g., rows 2 and 11) the light curves often display a complex overarching shape due to imperfect processing of the data. Yet, some of the light curves display a seemingly linear trend (e.g., rows 3 and 4 in Fig. 11) whose removal may have a positive effect on the ultimate classification of the respective light curves. We explicitly note that detrending the light curves with a higher degree polynomial or a higher order spline will have undesirable effects on the classification as i) the original light curves may be significantly distorted, and ii) the signal of long-period variables may be largely filtered out during the process. Hence, we preform the linear detrending of the light curves and follow exactly the same steps of the "base" scenario described in Sect. 7.1 to validate the results. The holdout set and the *Kepler* Q9 results show that both scenarios achieve nearly equal performance leading us to stick to the "base" scenario, as we do not see obvious advantages of performing linear detrending of the light curves.

One of the key features in our "base" scenario is that the "contactEB/spots" class is largely overpopulated, with a clear tendency for a large pollution towards the low probability values by the light curves containing some sort of an "instrumental" signal. We note that we use the term "instrumental signal" to mark a signal that is either truly instrumental in origin or is the result of sub-optimal detrending/correction of the data. To overcome the above-mentioned drawback of the "base" classification scenario, we opt to introduce an "instrumental" class with the properties resembling those of light curves affected by the instrumental trends and/or sub-optimal data processing. We use a sub-sample of the "contactEB/spots" class light curves in the "base"

scenario whose probability values were found to be of $p(x) \lesssim 0.65$ to manually select a training set for the "instrumental" class based on the visual inspection of the light curves. To preserve the balance with other variability classes in the training set, in total some 1100 light curves are selected as members of the "instrumental" class. We subsequently follow the same steps of the "base" scenario to train the classifiers and to validate the obtained results. The most notable differences between the results of the "base" and "base+instrumental" scenarios are i) a considerable reduction of the size of the "contactEB/spots" variability class by about a factor 3.5, and ii) a much smaller size of the "unknown" class that is also reduced by about two orders of magnitude. The latter result owes to the fact that the originally low-probability (lower than the respective thresholds reported in Table 3) objects in the various classes in the "base" classification scenario are classified with high confidence as members of the newly introduced "instrumental" class in the "base+instrumental" classification scenario, hence there are considerably less candidates to feed the "unknown" class in the latter. Furthermore, we cross-match the newly obtained "instrumental" class with the "contactEB/spots" class in the "base" scenario and find some 70% of overlap between the two classes. The probability density plot for the cross-matched sample is shown in Fig. 12 (blue line) where the distribution is evidently skewed towards low probabilities. Therefore, we conclude that introducing an "instrumental" class does not necessarily improve the overall performance of the method, instead a considerable fraction of light curves that receive low confidence values in their respective classes in the "base" scenario are moved to the class of miscellaneous variables in the "base+instrumental" scenario. While not having clear advantages, the disadvantages of introducing an "instrumental" class is that it is not only instrument-dependent but is also extremely sensitive to the way data from a given instrument are being processed. Therefore, an "instrumental" class proves impractical as it has to be re-designed each time the data from a given instrument are being reprocessed and/or the methodology is being applied to data from a different instrument. A much more practical solution is the one outlined and employed by us in Sect. 7.1, i.e. a recommended probability-based threshold to separate high-confidence detections of genuine "contactEB/spots" variables from their low-confidence counterparts that are most likely not astrophysical in origin.

8. DISCUSSION AND CONCLUSIONS

The TESS Data for Asteroseismology (T'DA) pipeline is being designed by us for a largely automated processing and high-level interpretation of the TESS space-based photometric data. As depicted in Fig. 1, the first two modules of the pipeline are designed for the extraction of light curves from the TESS Full Frame Images (Handberg et al. 2021) and for their subsequent optimal correction for systematic effects (Lund et al. 2021). In this work, we have designed a third module of the T'DA pipeline that performs an automated classification of the corrected light curves according to the type of variability encoded in them.

We employ the method of a supervised classification of photometric time-series that is based exclusively on the information contained in the light curves themselves. We utilize several supervised learning algorithms, each one with its own strengths and weaknesses, and ultimately combine them into a metaclassifier that employs the class probabilities returned by the base classifiers as its own classification features. We show that by stacking the predictions of a set of different base classifiers we can accurately classify light curves according to their general variability type, without relying on any extra information other than the light curves themselves. The key ingredient of our method lies in the fact that, when combining the predictions from our set of well-established classifiers with the predictions of a new classifier making use of a novel feature set, we make sure that the predictions from each classifier are as uncorrelated as possible. Such an approach allows the metaclassifier to obtain optimal performance by combining the relative strengths of the individual learners. In total, four individual base classifiers are employed by us in this work, of which the RFGC (Armstrong et al. 2016, Sect. 4.2) and SLOSH (Hon et al. 2018, Sect. 4.1) algorithms have been previously published and modified by us according to the needs of the metaclassifier, while the SORTING-HAT (Sect. 4.3) and GBGC (Sect. 4.4) learners were additionally developed by us for the needs of the T'DA pipeline classification module.

Although inspired by the amount of TESS data being currently collected, our ultimate goal is to design an automated pipeline for the end-to-end processing of high-cadence and duty-cycle space-based photometric data, irrespective of whether these come from the retired CoRoT and *Kepler*/K2 missions, currently operational TESS mission, or future space-missions such as PLATO (Rauer et al. 2014). Hence, in this work, we make use of the *Kepler* mission legacy, both in terms of the available high-precision, cadence, and duty-cycle data and the published catalogs of variable stars, to build a training set for our classifiers as well as to perform their test

and validation. The training set is carefully built from the existing catalogs with a subsequent vetting of light curves in all eight variability classes employed in our classification scheme. All individual learners as well as the metaclassifier are trained on 80% of the compiled training set, while the remaining 20% are kept as a holdout set to test and validate the method. We obtain an overall accuracy of 94.9% on the holdout set with some small differences between the different classes.

We further apply the designed classification scheme to the *Kepler* Q9 data set that has been truncated into the 27.4 days light curves upfront. This reduction in the length of the data set by close to a factor of three is motivated by the fact that we want to i) have a broad sense of the performance of the method on short time base light curves and understand its (astro)physical and technical limitations, and ii) ease the transition from *Kepler* to TESS in the future applications of the method. We test three different classification scenarios, "base" (classification of the original light curves without applying any extra corrections to them), "linfit" (all light curves are linearly detrended prior to applying the classification method), and "base+instrument" (classification of the original light curves with an extra "instrumental" class to isolate light curves dominated by the instrumental signal). We show that though the latter scenario allows for a significantly lower contamination by the sub-optimally processed light curves in some of the classes, it has the disadvantage that the "instrumental" class has to be re-designed each time the method is applied to the re-processed data from the same space-mission and/or data obtained by another mission. We also note that our training set is built from the catalogs that are based on all available *Kepler* data and so are the class assignments in the training set. Therefore, one of the expected and detected (astro)physical limitations of our method are apparent "misclassifications" of objects whose variability on the time scale of 27.4 days does not necessarily resemble their true origin. A common example is not fully resolved rotational variability in cool stars that gives rise to an overdensity of low frequencies in the Fourier transform of the light curve causing a confusion with the class of g-mode pulsators and/or aperiodic variables. Another example is the flux drop in a light curve due to sub-optimal data processing which mimics a single transit/eclipse event in the time-domain and gives rise to a missclassification as a transiting/eclipsing object. Other than that, we find that all the predicted classes are rather pure and are in the same order as in the confusion matrix based on the hold-out validation set (see Fig. 9).

We make both the methodology and the results of its application to the *Kepler* Q9 27.4 days data in the finally adopted "base" scenario publicly available to the community. Our training set, individual base learners, and the metaclassifier can be accessed through the dedicated GitHub repository[14] as well as through the TESS Asteroseismic Consortium (TASOC) Wiki pages[15]. We also release the predicted class probabilities and class labels for the *Kepler* Q9 27.4 days "base" classification scenario in electronic format; a snippet of the class probabilities table is shown in the Appendix (Table. 5).

8.1. *Future prospects*

With the machinery built, our immediate future prospects include:

- Classification of all *Kepler* stars based on i) 1-year data to mimic TESS Continuous Viewing Zone (CVZ) operations and enable direct comparison with the results presented in this work; ii) 2-year data to mimic PLATO Long Pointing Field (LPF) operations enabling an important set of tests for the PLATO Consortium; and iii) 4-year data to provide a full *Kepler* classification catalogue and quantitatively assess performance of our method on ultra-high precision data by cross-matching with the existing *Kepler* catalogues. At this step, we will consider employing extra information, such as photometric colours, Gaia parallaxes, etc., in order to break the existing degeneracies within and between the individual variability classes. This particular step also covers our intended "second-level classification" (as depicted in Fig. 2) where we aim to distinguish between different evolutionary states of solar-like pulsators (RGB stars vs. red-clump stars), between subgroups of g- (γ Dor vs. SPB variables) and p-mode (δ Sct vs. β Cep stars) pulsators, etc.

- Inclusion of a learning algorithm capable of identifying transient phenomena, such as stellar flares. For this, we will consider existing algorithms such as STELLA[16] (Feinstein et al. 2020) which will be adapted to the needs of our metaclassifier similarly to the RFGC and multiSLOSH methods.

- Inclusion of an unsupervised learning algorithm to help identify contaminants and search for overdensities in the feature space within the identi-

[14] https://github.com/tasoc/starclass
[15] https://tasoc.dk/tda/
[16] https://archive.stsci.edu/hlsp/stella

fied with the supervised classification module variability classes. This particular step is depicted in Fig. 2 as the "unsupervised methods" box and will strengthen our classification scheme by allowing the detection of additional variability (sub)classes.

- Transition to TESS data that are processed with the corresponding T'DA pipeline light curve extraction and systematics correction modules. At this step, we also envision an iteration between all three modules of the T'DA pipeline, in particular to inform the light curve correction algorithms on the variability time-scales that should be preserved rather than removed for specific classes of objects. In terms of the corresponding data releases, we plan them jointly with the light curves themselves on the per sector basis and will make our results publicly available through the MAST and TASOC databases. The accompanying TESS classification papers are also foreseen and will be based on the full year of TESS data, i.e., per TESS observational hemisphere.

ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement N°670519: MAMSIE), from the KU Leuven Research Council (grant C16/18/005: PARADISE), from the Research Foundation Flanders (FWO) under grant agreement G0H5416N (ERC Runner Up Project), as well as from the BELgian federal Science Policy Office (BELSPO) through PRODEX grant PLATO. D.J.A acknowledges support from the STFC via an Ernest Rutherford Fellowship (ST/R00384X/1). R.A.G. acknowledges the support from the GOLF and PLATO CNES grants. K.J.B. is supported by the National Science Foundation under Award AST-1903828. J.S.K and K.J.B. were supported by funding from the European Research Council under the European Communitys Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement no. 338251 (StellarAges). L.M. was supported by the Premium Postdoctoral Research Program of the Hungarian Academy of Sciences. The research leading to these results has been supported by the Hungarian National Research, Development and Innovation Office (NKFIH) grant KH_18 130405 and the Lendület LP2014-17 and LP2018-7/2020 grants of the Hungarian Academy of Sciences. D.B. acknowledges support from the NASA TESS Guest Investigator Program under award 80NSSC19K0385.

This paper includes data collected by the TESS mission, which are publicly available from the Mikulski Archive for Space Telescopes (MAST) and described in Jenkins et al. (2016). Funding for the TESS mission is provided by NASAs Science Mission directorate. This research has made use of NASAs Astrophysics Data System, as well as the NASA/IPAC Extragalactic Database (NED) which is operated by the Jet Propul- sion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration. Funding for the TESS Asteroseismic Science Operations Centre is provided by the Danish National Research Foundation (Grant agreement no.: DNRF106), ESA PRODEX (PEA 4000119301) and Stellar Astrophysics Centre (SAC) at Aarhus University. We thank the TESS team and staff and TASC/TASOC for their support of the present work.

This paper includes data collected by the Kepler mission. Funding for the Kepler and K2 mission was provided by NASA's Science Mission Directorate. The authors acknowledge the efforts of the Kepler Mission team in obtaining the light curve data and data validation products used in this publication. These data were generated by the Kepler Mission science pipeline through the efforts of the Kepler Science Operations Center and Science Office. The Kepler light curves are archived at the Mikulski Archive for Space Telescopes.

REFERENCES

Abdul-Masih, M., Prša, A., Conroy, K., et al. 2016, AJ, 151, 101

Aerts, C., Christensen-Dalsgaard, J., & Kurtz, D. W. 2010, Asteroseismology

Aerts, C. & De Cat, P. 2003, SSRv, 105, 453

Aerts, C., Eyer, L., & Kestens, E. 1998, A&A, 337, 790

Aerts, C., Lehmann, H., Briquet, M., et al. 2003a, A&A, 399, 639

Aerts, C., Mathis, S., & Rogers, T. M. 2019, ARA&A, 57, 35

Aerts, C., Molenberghs, G., Michielsen, M., et al. 2018, ApJS, 237, 15

Aerts, C., Thoul, A., Daszyńska, J., et al. 2003b, Science, 300, 1926

Aerts, C., Van Reeth, T., & Tkachenko, A. 2017, ApJL, 847, L7

Aggarwal, C. C. 2014, Data Classification: Algorithms and Applications, 1st edn. (Chapman & Hall/CRC)

Antoci, V., Cunha, M. S., Bowman, D. M., et al. 2019, MNRAS, 490, 4040

Armstrong, D. J., Kirk, J., Lam, K. W. F., et al. 2016, Monthly Notices of the Royal Astronomical Society, 456, 2260

Armstrong, D. J., Kirk, J., Lam, K. W. F., et al. 2015, Astronomy and Astrophysics, 579, A19

Auvergne, M., Bodin, P., Boisnard, L., et al. 2009, A&A, 506, 411

Ball, N. M., Brunner, R. J., Myers, A. D., & Tcheng, D. 2006, ApJ, 650, 497

Barban, C., Matthews, J. M., De Ridder, J., et al. 2007, A&A, 468, 1033

Beck, P. G., Montalban, J., Kallinger, T., et al. 2012, Nature, 481, 55

Bedding, T. R., Butler, R. P., Kjeldsen, H., et al. 2001, ApJL, 549, L105

Bedding, T. R., Huber, D., Stello, D., et al. 2010, ApJL, 713, L176

Bedding, T. R. & Kjeldsen, H. 2003, PASA, 20, 203

Bedding, T. R. & Kjeldsen, H. 2007, Communications in Asteroseismology, 150, 106

Bedding, T. R., Murphy, S. J., Colman, I. L., & Kurtz, D. W. 2015, in European Physical Journal Web of Conferences, Vol. 101, European Physical Journal Web of Conferences, 01005

Bedding, T. R., Murphy, S. J., Hey, D. R., et al. 2020, Nature, 581, 147

Borucki, W. J., Koch, D., Basri, G., et al. 2010, Science, 327, 977

Bouabid, M.-P., Dupret, M.-A., Salmon, S., et al. 2013, MNRAS, 429, 2500

Bowman, D. M. 2017, Amplitude Modulation of Pulsation Modes in Delta Scuti Stars

Bowman, D. M. & Kurtz, D. W. 2018, MNRAS, 476, 3169

Bowman, D. M., Kurtz, D. W., Breger, M., Murphy, S. J., & Holdsworth, D. L. 2016, MNRAS, 460, 1970

Breger, M. 2000, in Astronomical Society of the Pacific Conference Series, Vol. 210, Delta Scuti and Related Stars, ed. M. Breger & M. Montgomery, 3

Breiman, L. 2001, Machine Learning, 45, 5

Brett, D. R., West, R. G., & Wheatley, P. J. 2004, Monthly Notices of the Royal Astronomical Society, 353, 369

Briquet, M., Morel, T., Thoul, A., et al. 2007, MNRAS, 381, 1482

Briquet, M., Neiner, C., Aerts, C., et al. 2012, MNRAS, 427, 483

Brun, A. S., Strugarek, A., Varela, J., et al. 2017, ApJ, 836, 192

Bruntt, H. & Buzasi, D. L. 2006, Mem. Soc. Astron. Italiana, 77, 278

Bugnet, L., García, R. A., Davies, G. R., et al. 2018, A&A, 620, A38

Buldgen, G., Reese, D. R., & Dupret, M. A. 2018, A&A, 609, A95

Busa, M. A. & van Emmerik, R. 2016, Journal of Sport and Health Science, 5, 44

Buzasi, D., Catanzarite, J., Laher, R., et al. 2000, ApJL, 532, L133

Buzasi, D. L. 2004, in ESA Special Publication, Vol. 538, Stellar Structure and Habitable Planet Finding, ed. F. Favata, S. Aigrain, & A. Wilson, 205–213

Carrier, F., De Ridder, J., Baudin, F., et al. 2010, A&A, 509, A73

Carrier, F., Eggenberger, P., & Bouchy, F. 2005, A&A, 434, 1085

Chen, T. & Guestrin, C. 2016, in 22nd ACM sigkdd Int Conf Knowl Discov data Min (ACM), 785 – 794

Costa, M., Goldberger, A. L., & Peng, C.-K. 2005, Phys. Rev. E, 71, 021906

Cousins, A. W. J. 1992, The Observatory, 112, 53

Cousins, A. W. J., Caldwell, J. A. R., & Menzies, J. W. 1989, Information Bulletin on Variable Stars, 3412

Cuypers, J., Aerts, C., Buzasi, D., et al. 2002, A&A, 392, 599

Davies, G. R., Silva Aguirre, V., Bedding, T. R., et al. 2016, MNRAS, 456, 2183

De Ridder, J., Barban, C., Baudin, F., et al. 2009, Nature, 459, 398

De Ridder, J., Barban, C., Carrier, F., et al. 2006, A&A, 448, 689
Debosscher, J., Blomme, J., Aerts, C., & De Ridder, J. 2011, A&A, 529, A89
Debosscher, J., Sarro, L. M., Aerts, C., et al. 2007, A&A, 475, 1159
Degroote, P., Aerts, C., Baglin, A., et al. 2010, Nature, 464, 259
Degroote, P., Aerts, C., Ollivier, M., et al. 2009, A&A, 506, 471
Derekas, A., Plachy, E., Molnár, L., et al. 2017, MNRAS, 464, 1553
Dupret, M.-A., Grigahcène, A., Garrido, R., et al. 2005, MNRAS, 360, 1143
Eggenberger, P., Deheuvels, S., Miglio, A., et al. 2019, A&A, 621, A66
Fawcett, T. 2006, Pattern Recognition Letters, 27, 861, rOC Analysis in Pattern Recognition
Feinstein, A. D., Montet, B. T., Ansdell, M., et al. 2020, AJ, 160, 219
Frandsen, S., Carrier, F., Aerts, C., et al. 2002, A&A, 394, L5
Friedman, J. H. 2001, The Annals of Statistics, 29, 1189
García, R. A. & Ballot, J. 2019a, Living Reviews in Solar Physics, 16, 4
García, R. A. & Ballot, J. 2019b, Living Reviews in Solar Physics, 16, 4
García Hernández, A., Martín-Ruiz, S., Monteiro, M. J. P. F. G., et al. 2015, ApJL, 811, L29
García Hernández, A., Suárez, J. C., Moya, A., et al. 2017, MNRAS, 471, L140
Gilliland, R. L. 2008, AJ, 136, 566
Granitto, P. M., Furlanello, C., Biasioli, F., & Gasperi, F. 2006, Chemom Intell Lab Syst, 83, 83
Guzik, J. A., Kaye, A. B., Bradley, P. A., Cox, A. N., & Neuforge, C. 2000, ApJL, 542, L57
Hall, M. A. 1999, PhD thesis, University of Waikato Hamilton
Handler, G. 2017, in Second BRITE-Constellation Science Conference: Small satellites big science, Proceedings of the Polish Astronomical Society volume 5, held 22-26 August, 2016 in Innsbruck, Austria. Edited by Konnstanze Zwintz and Ennio Poretti. Polish Astronomical Society, Bartycka 18, 00-716 Warsaw, Poland, pp.151-157, ed. K. Zwintz & E. Poretti, 151–157
Handler, G., Jerzykiewicz, M., Rodríguez, E., et al. 2006, MNRAS, 365, 327
Handler, G., Matthews, J. M., Eaton, J. A., et al. 2009, ApJL, 698, L56
Hekker, S. & Christensen-Dalsgaard, J. 2017, A&A Rv, 25, 1
Hekker, S., Kallinger, T., Baudin, F., et al. 2009, A&A, 506, 465
Hon, M., Stello, D., & Zinn, J. C. 2018, The Astrophysical Journal, 859, 64
Howell, S. B., Sobeck, C., Haas, M., et al. 2014, PASP, 126, 398
Hümmerich, S., Mikulášek, Z., Paunzen, E., et al. 2018, A&A, 619, A98
Jamal, S. & Bloom, J. S. 2020, arXiv e-prints, arXiv:2003.08618
Kallinger, T. 2019, arXiv e-prints, arXiv:1906.09428
Kallinger, T., Guenther, D. B., Matthews, J. M., et al. 2008, A&A, 478, 497
Kallinger, T., Mosser, B., Hekker, S., et al. 2010, A&A, 522, A1
Kaye, A. B., Handler, G., Krisciunas, K., Poretti, E., & Zerbi, F. M. 1999, PASP, 111, 840
Keen, M. A., Bedding, T. R., Murphy, S. J., et al. 2015, MNRAS, 454, 1792
Kgoadi, R., Engelbrecht, C., Whittingham, I., & Tkachenko, A. 2019, arXiv preprint arXiv:1906.06628
Kim, D.-W. & Bailer-Jones, C. A. L. 2016, Astron Astrophys, 587, A18
Kirk, B., Conroy, K., Prša, A., et al. 2016, AJ, 151, 68
Kiss, L. L. & Bódi, A. 2017, A&A, 608, A99
Kjeldsen, H. & Bedding, T. R. 1995, A&A, 293, 87
Kjeldsen, H., Bedding, T. R., Viskum, M., & Frandsen, S. 1995, AJ, 109, 1313
Kohonen, T. 1990, Proceedings of the IEEE, 78, 1464
Kozachenko, L. F. & Leonenko, N. N. 1987, Probl. Peredachi Inf., 23, 95101
Kraskov, A., Stögbauer, H., & Grassberger, P. 2004, Phys. Rev. E, 69, 066138
Kuszlewicz, J. S., Hekker, S., & Bell, K. J. 2020, MNRAS, 497, 4843
Leavitt, H. S. & Pickering, E. C. 1912, Harvard College Observatory Circular, 173, 1
Li, G., Van Reeth, T., Bedding, T. R., et al. 2020, MNRAS, 491, 3586
Lomb, N. R. 1976, Ap&SS, 39, 447
Lundberg, S. M., Erion, G., Chen, H., et al. 2020, Nature Machine Intelligence, 2, 2522
Lundberg, S. M. & Lee, S.-I. 2017, in Advances in Neural Information Processing Systems 30, ed. I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Curran Associates, Inc.), 4765–4774

Matijevič, G., Prša, A., Orosz, J. A., et al. 2012, AJ, 143, 123
McQuillan, A., Mazeh, T., & Aigrain, S. 2014, ApJS, 211, 24
Merline, W. J. 1999, in Astronomical Society of the Pacific Conference Series, Vol. 185, IAU Colloq. 170: Precise Stellar Radial Velocities, ed. J. B. Hearnshaw & C. D. Scarfe, 187
Michaud, G. 1970, ApJ, 160, 641
Miglio, A., Montalbán, J., Baudin, F., et al. 2009, A&A, 503, L21
Modak, S., Chattopadhyay, T., & Chattopadhyay, A. K. 2018, arXiv e-prints, arXiv:1801.09406
Moravveji, E., Aerts, C., Pápics, P. I., Triana, S. A., & Vandoren, B. 2015, A&A, 580, A27
Moravveji, E., Townsend, R. H. D., Aerts, C., & Mathis, S. 2016, ApJ, 823, 130
Mosser, B., Barban, C., Montalbán, J., et al. 2011, A&A, 532, A86
Murphy, S. J., Fossati, L., Bedding, T. R., et al. 2016, MNRAS, 459, 1201
Namekata, K., Maehara, H., Notsu, Y., et al. 2019, ApJ, 871, 187
Naul, B., Bloom, J. S., Pérez, F., & van der Walt, S. 2018, Nature Astronomy, 2, 151
Nielsen, M. B., Gizon, L., Schunker, H., & Karoff, C. 2013, A&A, 557, L10
Pápics, P. I., Briquet, M., Baglin, A., et al. 2012, A&A, 542, A55
Pápics, P. I., Tkachenko, A., Aerts, C., et al. 2015, ApJL, 803, L25
Pápics, P. I., Tkachenko, A., Van Reeth, T., et al. 2017, A&A, 598, A74
Pashchenko, I. N., Sokolovsky, K. V., & Gavras, P. 2018, MNRAS, 475, 2326
Pedersen, M. G., Escorza, A., Pápics, P. I., & Aerts, C. 2020, MNRAS, 495, 2738
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, Journal of Machine Learning Research, 12, 2825
Pigulski, A., Cugier, H., Handler, G., & Kgoadi, R. 2018, in 3rd BRITE Science Conference, ed. G. A. Wade, D. Baade, J. A. Guzik, & R. Smolec, Vol. 8, 77–79
Pinsonneault, M. H., Elsworth, Y. P., Tayar, J., et al. 2018, ApJS, 239, 32
Plachy, E., Bódi, A., & Kolláth, Z. 2018, MNRAS, 481, 2986
Pojmanski, G. 2002, AcA, 52, 397
Preston, G. W. 1974, ARA&A, 12, 257
Provost, F. 2000, in Proceedings of the AAAI2000 workshop on imbalanced data sets (AAAI Press)
Prša, A., Batalha, N., Slawson, R. W., et al. 2011, AJ, 141, 83
Rauer, H., Catala, C., Aerts, C., et al. 2014, Experimental Astronomy, 38, 249
Richards, J. W., Starr, D. L., Butler, N. R., et al. 2011, Astrophys J, 733 [1101.1959]
Richer, J., Michaud, G., & Turcotte, S. 2000, ApJ, 529, 338
Richman, J. S. & Moorman, J. R. 2000, American Journal of Physiology-Heart and Circulatory Physiology, 278, H2039, pMID: 10843903
Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2015, Journal of Astronomical Telescopes, Instruments, and Systems, 1, 014003
Sarro, L. M., Debosscher, J., López, M., & Aerts, C. 2009, A&A, 494, 739
Scargle, J. D. 1982, ApJ, 263, 835
Serenelli, A., Weiss, A., Aerts, C., et al. 2020, arXiv e-prints, arXiv:2006.10868
Shannon, C. E. 1948, Bell System Technical Journal, 27, 623
Slawson, R. W., Prša, A., Welsh, W. F., et al. 2011, AJ, 142, 160
Stankov, A. & Handler, G. 2005, ApJS, 158, 193
Stassun, K. G., Oelkers, R. J., Paegert, M., et al. 2019, AJ, 158, 138
Stello, D., Bruntt, H., Preston, H., & Buzasi, D. 2008, ApJL, 674, L53
Stello, D. & Gilliland, R. L. 2009, ApJ, 700, 949
Suárez, J. C., García Hernández, A., Moya, A., et al. 2014, A&A, 563, A7
Szabó, R. 2018, in The RR Lyrae 2017 Conference. Revival of the Classical Pulsators: from Galactic Structure to Stellar Interior Diagnostics, Vol. 6, 119–123
Szabó, R., Szabados, L., Ngeow, C. C., et al. 2011, MNRAS, 413, 2709
Tarrant, N. J., Chaplin, W. J., Elsworth, Y., Spreckley, S. A., & Stevens, I. R. 2007, MNRAS, 382, L48
Thompson, S. E., Caldwell, D. A., Jenkins, J. M., et al. 2016, Kepler Data Release 25 Notes, Kepler Science Document KSCI-19065-002
Tkachenko, A., Aerts, C., Pavlovski, K., et al. 2014, MNRAS, 442, 616
Tkachenko, A., Aerts, C., Yakushechkin, A., et al. 2013, A&A, 556, A52
Tkachenko, A., Matthews, J. M., Aerts, C., et al. 2016, MNRAS, 458, 1964
Udalski, A., Szymański, M. K., Soszyński, I., & Poleski, R. 2008, AcA, 58, 69
Udalski, A., Szymanski, M. K., & Szymanski, G. 2015, AcA, 65, 1

Valenzuela, L. & Pichara, K. 2018, MNRAS, 474, 3259

Van Reeth, T., Tkachenko, A., & Aerts, C. 2016, A&A, 593, A120

Van Reeth, T., Tkachenko, A., Aerts, C., et al. 2015a, A&A, 574, A17

Van Reeth, T., Tkachenko, A., Aerts, C., et al. 2015b, ApJS, 218, 27

Vega, L. D., Stassun, K. G., Montez, Rodolfo, J., Boyd, P. T., & Somers, G. 2017, ApJ, 839, 48

Waelkens, C. 1991, A&A, 246, 453

Waelkens, C., Aerts, C., Kestens, E., Grenon, M., & Eyer, L. 1998, A&A, 330, 215

Walczak, P., Daszyńska-Daszkiewicz, J., Pamyatnykh, A., Handler, G., & Pigulski, A. 2017, in Second BRITE-Constellation Science Conference: Small satellites big science, Proceedings of the Polish Astronomical Society volume 5, held 22-26 August, 2016 in Innsbruck, Austria. Edited by Konnstanze Zwintz and Ennio Poretti. Polish Astronomical Society, Bartycka 18, 00-716 Warsaw, Poland, pp.173-179, ed. K. Zwintz & E. Poretti, 173–179

Walker, G., Matthews, J., Kuschnig, R., et al. 2003, PASP, 115, 1023

Weinberg, N. N. & Arras, P. 2019, ApJ, 873, 67

Wolpert, D. H. 1992, Neural Networks, 5, 241

Wyrzykowski, Ł. & Belokurov, V. 2008, in American Institute of Physics Conference Series, Vol. 1082, Classification and Discovery in Large Astronomical Surveys, ed. C. A. L. Bailer-Jones, 201–206

Youden, W. J. 1950, Cancer, 3, 32

Yu, J., Bedding, T. R., Stello, D., et al. 2020, MNRAS, 493, 1388

Yu, J., Huber, D., Bedding, T. R., et al. 2018, The Astrophysical Journal Supplement Series, 236, 42

Zhou, Y., Asplund, M., Collet, R., & Joyce, M. 2020, MNRAS, 495, 4904

Zocłońska, E. 2018, in 3rd BRITE Science Conference, ed. G. A. Wade, D. Baade, J. A. Guzik, & R. Smolec, Vol. 8, 94–98

APPENDIX

A. FEATURE IMPORTANCE PLOTS

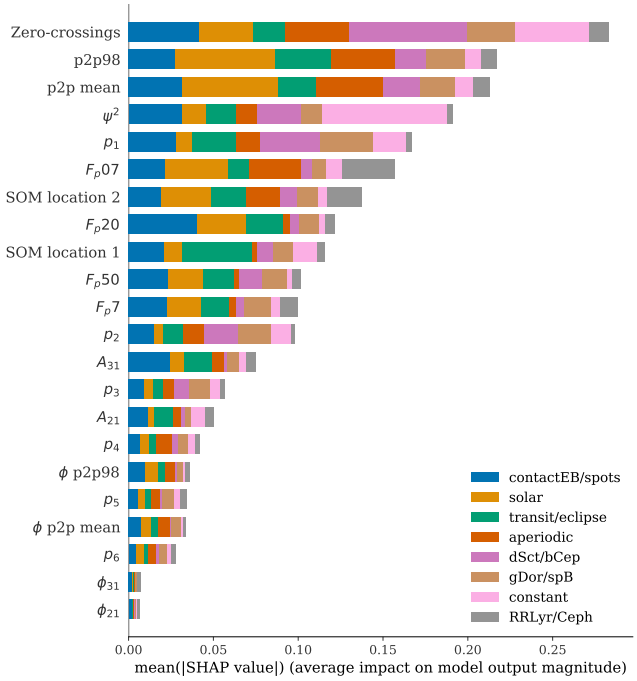

Figure 13: RFGC feature importances from SHAP. The hatched regions indicate the most important feature per class.

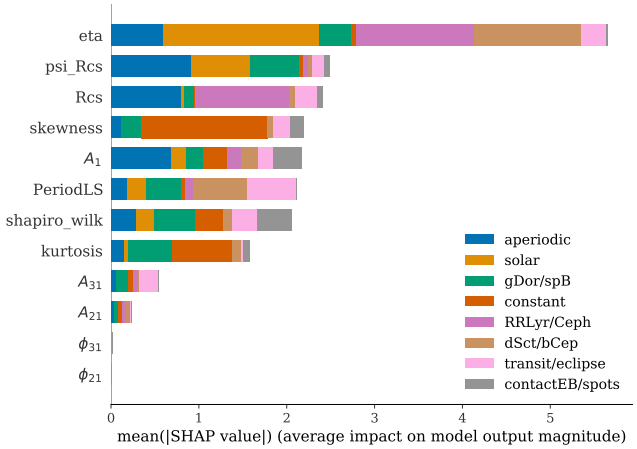

Figure 14: GBGC feature importances from SHAP. The hatched regions indicate the most important feature per class.

B. CLASS PROBABILITIES *KEPLER* Q9

Table 5: First five rows of the electronic table with class probabilities and assigned labels.

kic	p(aperiodic)	p(constant)	p(contactEB/spots)	p(dSct/bCep)	p(eclipse/transit)	p(gDor/spB)	p(RRLyr/Ceph)	p(solar)	Label	Label (Youden)
11513597	0.0641	0.0221	0.5526	0.0354	0.0327	0.1211	0.1078	0.0642	contactEB/spots	contactEB/spots
9266835	0.0299	0.0121	0.5285	0.0639	0.0819	0.1530	0.1001	0.0305	contactEB/spots	contactEB/spots
11241837	0.02449	0.0012	0.7854	0.01607	0.0259	0.0477	0.0858	0.0134	contactEB/spots	contactEB/spots
9591826	0.0362	0.0849	0.3630	0.1165	0.0542	0.1967	0.0971	0.0514	contactEB/spots	contactEB/spots