Please provide feedback

Please support the ScholarWorks@UMBC repository by emailing scholarworks-group@umbc.edu and telling us what having access to this work means to you and why it's important to you. Thank you.

Identifying Variable Stars from Kepler Data Using Machine Learning

J. Adassuriya, J. A. N. S. S. Jayasinghe and K. P. S. C. Jayaratne

ABSTRACT

Machine learning algorithms play an impressive role in modern technology and address automation problems in many fields as these techniques can be used to identify features with high sensitivity, which humans or other programming techniques aren't capable of detecting. In addition, the growth of the availability of the data demands the need of faster, accurate, and more reliable automating methods of extracting information, reforming, and preprocessing, and analyzing them in the world of science. The development of machine learning techniques to automate complex manual programs is a time relevant research in astrophysics as it's a field where, experts are dealing with large sets of data every day. In this study, an automated classification was built for 6 types of star classes Beta Cephei, Delta Scuti, Gamma Doradus, Red Giants, RR Lyrae and RV Tarui with widely varying properties, features extracted from training dataset of stellar light curves obtained from Kepler mission. The Random Forest classification model was used as the Machine Learning model and both periodic and non-periodic features extracted from light curves were used as the inputs to the model. Our implementation achieved an accuracy of 86.5%, an average precision level of 0.86, an average recall value of 0.87, and average F1-Score of 0.86 for the testing dataset obtained from the Kepler mission.

Keywords: Machine Learning, Random Forest, Short Period Variable Stars.

J. Adassuriya*
Astronomy Division, Arthur C Clarke Institute, Katubedda, Moratuwa, Sri Lanka.
(e-mail: adassuriya@gmail.com)
J. A. N. S. S. Jayasinghe
Department of Physics, University of Maryland Baltimore County, USA.
(e-mail: nirandj1@umbc.edu)
K. P. S. C. Jayaratne
Department of Physics, University of Colombo, Sri Lanka.
(e-mail: chandanajayaratne@gmail.com)

*Corresponding Author

I. INTRODUCTION

Stars with varying brightness or luminosity with the time are identified as variable stars and they are considered to be very important stellar objects in astrophysics. They are used in determining both external and internal properties such as distances, age, internal structure as well as their behavior over the time. Classification and analysis of stellar objects have been a major challenge in astronomy over the past few decades since the volume of the available data is increasing, obligations to many successful missions conducted by pioneers in space observations. Machine Learning and statistical analysis techniques such as Kohonen self-organizing maps, Bayesian mixture-model classifier, SVMs and Gaussian mixture models etc. have been using over the past decade in addressing this problem. This implementation was focused on developing a technique which is autonomous of the instrumentation specifications where it could be easily generalized for data sources which is extracted from different missions. Random Forest algorithm was more appropriate to use in time series analysis with minimum impact by instrumentation specificity towards the training.

Random Forest classification is a powerful and resilient framework which is based on decision trees. It is an efficient and scalable classifier that captures complicated patterns in the feature set without over-fitting into the training dataset and outliers. It fits several numbers of decision trees on partitions of the training dataset and takes the average to optimize the accuracy. The final prediction is determined by the majority votes which pick the most favorable output generated by the constructed trees. Definition of the random forest classifier is given as follows in [1].

Trees in the random forest are built upon bootstrap samples of equal size to the original dataset. As the sampling is done with replacements, same instances can be repeated in the sample while some are left out. These mislaid data instances called Out-of-Bag (OOB) [2] are used to evaluate the prediction error. Classification proceeds by partitioning the drawn sample recursively into more trees or nodes based on a randomly selected attributes from the previously chosen attribute list. This will repeat until a terminal node is reached, which contains a single type of object. Each of these constructed trees provides a prediction for the type, and the final prediction is determined to be the type with the highest aggregation.

Over the past decade, many studies have been undertaken to implement machine learning models using existing data repositories and investigate the idea and feasibility of introducing a generalized classifier that is independent of the data repository [3]-[11]. UPSILoN (Automated Classification

for periodic Variable Stars Using Machine Learning) [12] is one of the several software packages developed within the last few years. The UPSILoN is capable of classifying stars into different star classes, including variable and non-variable, using light curve data as the input irrespective of its survey specific characteristics. OGLE and EROS-2 variable star data repositories were used to train this classifier while tested on MACHO, ASAS and Hipparcos datasets. Machine Learning Classification of Variable Stars with Sparse and Noisy Time-Series Data [13] describes another automated machine learning framework that uses features of the light curves extracted using statistical and mathematical methods to identify the type of variability of a star. This identification can be made into sub-classes of the types as well. Random forest classifier algorithm has been used in classifications and to understand the correlations and importance of the used parameters.

II. TRAINING THE CLASSIFIER

Dataset was extracted from Kepler mission [14] in .dat and .pow text file format. Full dataset consists of 583 light curve data files of 6 different types of stars as mentioned in Introduction. These files were recorded in many different metric systems and different parameters were recorded. 540 out of 583 of data files were selected to carry out the project, which had all the necessary parameters (Time, Corrected Flux, and Corrected Flux error) recorded in acceptable metric systems.

TABLE I: DATA COMPOSITION

Star Classes	Number of Light curves
Beta Cephei (BC)	120
Delt Scuti (DS)	64
Gamma Doradus (GD)	127
Red Giants (RG)	113
RR Lyrea (RR)	90
RV Tarui (RV)	26
Total	540

Data pre-processing session was carried out to clean and refine the data set. Data fields were checked for unbalanced data columns, number of data points and the data points with a magnitude that were beyond 3σ level reference to its statistical mean, were dropped to clean out the outliers. Finally, the time scale was converted to units of normal calendar days except for different astronomical unit systems to record time; such as Baycentric Julian dates.

Both periodic and non-periodic features were used to do the classifications. Periodic features were determined based on the generalized Lomb-Scargle [15] which was built upon Fourier transformation and least square method. This analytical solution is applicable to time series data with uneven sample intervals [16],[17]. Main frequency or the period of the light curve is estimated as follows. For a time series with N number of observations of y_i at time t_i with error σ_i, the sinusoidal function fitted to the light curve is given by,

$$y(t) = a\cos(\omega t) + b\sin(\omega t) + c \qquad (1)$$

where, ω is the period of the time series which acquired by minimizing the residuals of the data points using least square

fitting. The χ^2 minimization is given by,

$$\chi^2 = \sum_{i=1}^{N} \frac{[y_i - y(t_i)]^2}{\sigma_i^2} = W \sum w_i[y_i - y(t_i)]^2 \qquad (2)$$

where

$$w_i = \frac{1}{W\sigma_i^2} \left(W = \sum \frac{1}{\sigma_i^2} \sum w_i = 1 \right)$$

Power spectrum for frequencies will be obtained as follows for sampled frequencies obtained by DFT (Discrete Fourier transform).

$$p(\omega) = \frac{\chi_0^2 - \chi^2(\omega)}{\chi_0^2} \qquad (3)$$

Periodic features are based on the period estimation of the time series dataset and its procedure. In order to do the period estimation with high accuracy, period was estimated in two ways; both periods were used in training and testing processes. Astropy Lomb-Scargle in-built module was used to estimate Period\textunderscore1 and Lomb-Scargle function coded from the scratch based on [15] was used for Period_2.

Light curve data are sequence of brightness observations, therefore non-periodic features can be derived from statistical time series analysis. Non-periodic features which have been used in this project are selected from both [18] and [19]. Following are the non-periodic features which were extracted from the light curves.

1. Mean: Average magnitude of the light curve.
2. STD: Standard Deviation
3. Median: Middle value of the sorted magnitudes.
4. Skew.
5. Kurtosis: measurement of the peak of the magnitude distribution.
6. Mean Variance
7. Stetson K: Robust kurtosis measure taken using magnitudes and errors [21],[22].

$$Stetson\ K = \frac{\left(\frac{\Sigma|\delta_i|}{N} \right)}{\sqrt{\frac{\Sigma \delta_i^2}{N}}} \qquad (4)$$

$$\delta_i = \sqrt{\frac{n}{n-1}} \frac{m_i - \bar{m}}{m_{err}} \qquad (5)$$

8. Slope: Maximum absolute slope between two consecutive observations.
9. Amplitude: Half the difference between the median of the maximum 5% and minimum 5% magnitudes.
10. Median absolute deviation.

$$median\ absolute\ deviation = median\ (|Magnitude - median(Magnitude)|) \qquad (6)$$

11. $M_{p\ 10}$: 10% percentile of slope of the light curve.
12. $M_{p\ 90}$: 90% percentile of slope of the light curve.
13. Variability Index: This parameter is used to identify

trends in data points or their independence. Following is the variability index for uneven sampled data [19].

$$\eta_e = \bar{w}(t_{N-1} - t_1)^2 \frac{\sum_1^{N-1} w_i (m_{i+1} - m_i)^2}{\sigma^2 \sum_1^N w_i} \qquad (7)$$

$$w_i = \frac{1}{(t_{i+1} - t_i)^2}$$

A. Classifier Model and Performance

Random Forest classifier in sklearn python module was used to define the classifier and were trained by 70% of the dataset and tested upon rest of the 30%. Attributes of the Random Forest algorithm were selected as 700 and 10 respectively for number of trees (t), no of features randomly selected at each node of trees (m).

There are metrics to measure the performance of a classifier based on the testing output of the trained classifier. They can be defined as follows:

1. Recall/ Sensitivity

$$Recall = \frac{TP}{TP+FN} \qquad (8)$$

2. Precision

$$Precision = \frac{TP}{TP+FP} \qquad (9)$$

3. Accuracy

$$Accuracy = \frac{TP+TN}{Total inputs} \qquad (10)$$

TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative.

Recall and precision are calculated for each class in this case for all 6 classes which are considered. Accuracy is calculated for all classes depending on the outcomes for testing data.

4. F_1 Score:

$$F_1 = 2 \times \frac{precision \times recall}{precision + recall} \qquad (11)$$

The value of F_1 is in the range of 0–1 and calculated for each class separately. This is used to measure the performance of the classifier.

III. EXTRACTED FEATURE ANALYSIS

Following observation were illustrated from the full dataset before splitting into test and train datasets. These results are based on light curve extracted feature values. Period is considered to be the most important of all features and it is determined using two ways based on Lomb-Scargle period estimation. Astropy [20] Lomb-Scargle in-built python module was more successful in estimating the period compared to the other method mentioned. This difference in estimation is caused from the selection of period range and the

step size for the Lomb-Scargle. But still there were instances in all types, where estimated value for period was not acceptable.

Fig. 1 visualizes the distribution of periods estimated from the astropy Lomb-Scargle function for the full dataset in a general scale. It shows a clear variation of range of the period depending on the type of the star as well as over estimations. According to the Fig. 1 and Fig. 2 DS, GD and RR tend to have period values distributed in a very narrow range while RV, RG and BC has a larger distribution in contrast.
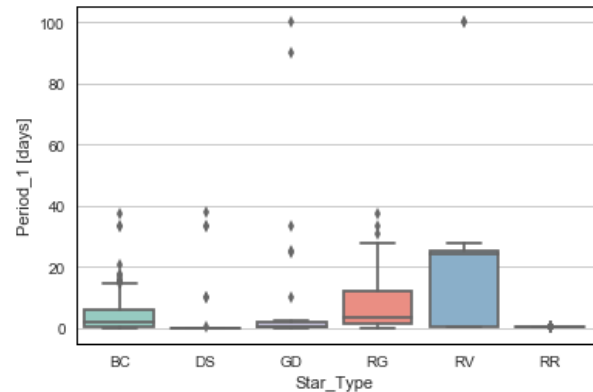

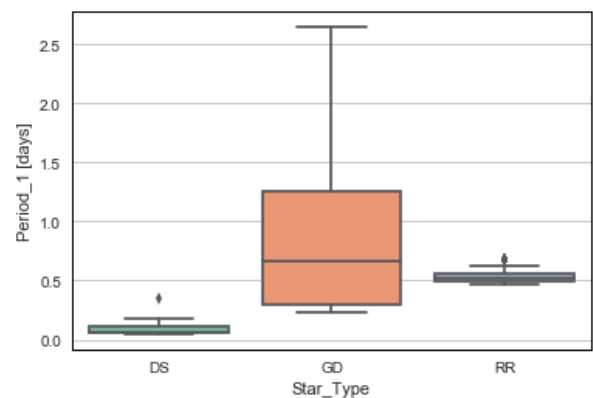Fig. 1. Period 1 distribution of all the types of the stars.

Fig. 2. Period 1 distribution for DS, GD, RR Stars.

TABLE II: PERIOD 1 (DAYS) STATISTICS

Statistic	BC	DC	GD	RG	RR	RV
Mean	4.96	1.99	3.13	7.08	0.54	20.61
STD	7.05	7.50	12.39	8.36	0.07	26.18
Minimum	0.12	0.05	0.22	0.19	0.47	0.51
Maximum	37.30	37.84	100.00	37.16	0.68	100.00
IQR	5.74	0.09	1.42	10.84	0.07	24.57

Table II gives a brief statistical description of the estimated period values versus star type. Inter-quartile range (IQR) shows the dispersion of estimated periods for each type of the star. RR and DS tend to have the smallest IQR, which indicates that the 50% of period data are concentric around its mean value, while RV has the highest IQR which means that data are distributed towards in a large range.

Second most important feature of the list is verified as the mean variance. Fig. 3, Fig. 4, and Table III describes the parameter values for each type of the star both visually and tabular form. Clear discrepancy in the distribution of the value can be observed in the boxplot of mean variance, between the types of RV, RR from the rest of types. RR and RV star types tend to have a higher mean variance value compared to the rest of the types in the Fig. 4.
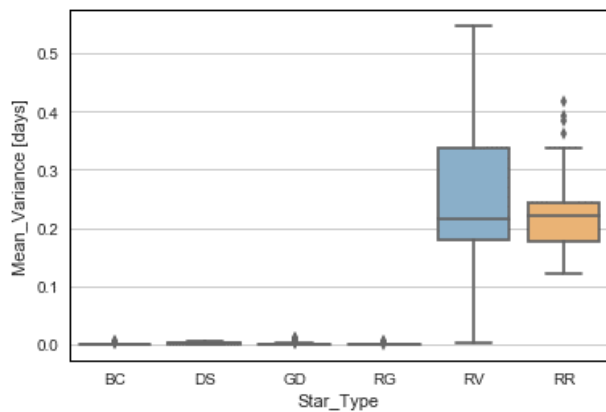
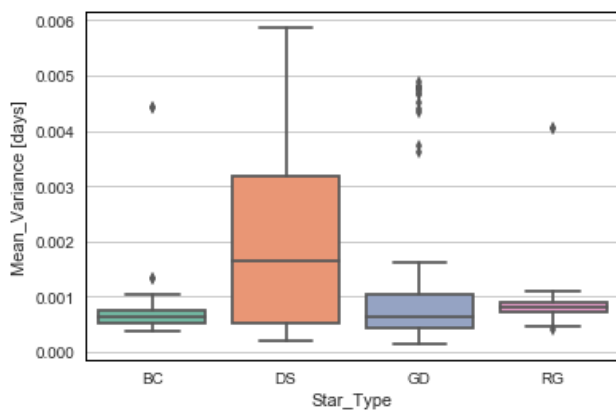Fig. 3. Mean Variance distribution of all the types of the stars.

Fig. 4. Mean Variance distribution for BC, DS, GD, and RG Stars.

For further clarifications among the types, Table III can be used. It clearly shows the ranges of the mean variance distribution based on the class. Ranges are [0.000383, 0.004431], [0.000214, 0.005860], [0.000149, 0.010612], [0.000405, 0.004043], [0.121603, 0.416183], [0.001751, 0.546282] corresponding to BC, DS, GD, RG, RR, RV types. RG type has the lowest distribution range while RV has the widest spread. According to mean value of this parameter, star types can be ordered in increasing mean variances as BC, RG, GD, DS, RR and RV.The distributions of skew, kurtosis and Stetson K parameters have the same distribution pattern as in Mean Variance; there is a clear difference in ranges of the distribution of RV and RR type among other types. They have to have larger ranges or wider spread compared to other 4 types of the star classes.

TABLE III: MEAN VARIANCE STATISTICS

Statistic	BC	DC	GD	RG	RR	RV
Mean	0.00071	0.00220	0.0012	0.00085	0.22	0.256
STD	0.00052	0.00170	0.0017	0.00045	0.06	0.127
Mini	0.00038	0.00020	0.0002	0.00041	0.12	0.002
Max	0.00443	0.00590	0.0106	0.00404	0.42	0.546
IQR	0.00230	0.00270	0.0005	0.00019	0.06	0.159

IV. CLASSIFIER RESULTS AND ANALYSIS

Feature importance was evaluated from the training process based on the training dataset. This does necessarily indicate the contribution from each feature to the classification process. Fig. 5 displays the feature importance bar chart with a numerical value for importance of the corresponding feature. According to this evaluation, which is

based on the Random Forest classifier, the most significant feature is Period with a significance value of 0.18 followed by other parameters in decreasing significance Mean Variance 0.16, Skew 0.10, Eeta 0.06, Amplitude 0.06 etc. This implies that period is the main factor when determining the type of the star. Also, the equal importance value represents that those features have equal contribution in determining the star type. This do not necessarily indicate that they are correlated features. The main advantage of evaluating this feature importance is to reduce the size of the dataset fed to the classifier by eliminating the least significant features. There are methods of doing this elimination process.
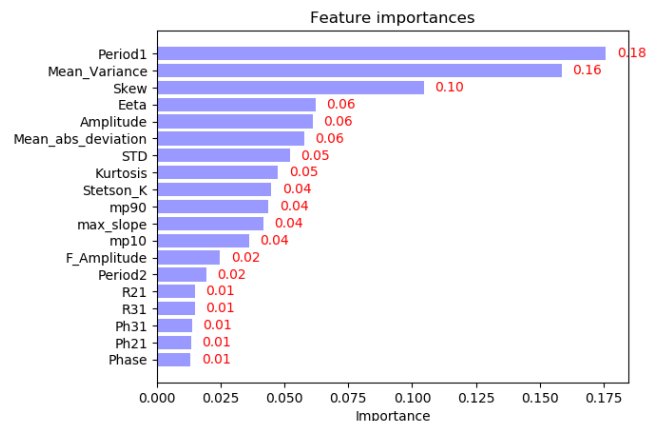

Fig. 5. Ranked list of features by decreasing importance measured using Random Forest.

540 light curves belonging to 6 types of stars are used in this project and out of that 377 light curves were used in the training process while the remaining 163 were used in the testing process. Testing phase is carried out to evaluate the performances of the trained classifier. The confusion matrix is represented in Fig. 6 and Table IV. This illustrates the number of correctly predicted classes and misclassifications. Diagonal elements of the grid give the number of correctly classified instances while off diagonals gives the misclassifications. 8 instances of BC stars were misclassified as 3 GD and 5 RG, 4 instances of GD were misclassified as 1 BC, 2 RG, 1 RR, 5 instances of RG were misclassified as BC, 2 instances of RR were misclassified as RV and 3 instances of RV were misclassified into RR. (see Table IV) These observations convey that there's a tendency to misclassify RR and RV into each other's classes or there are many similarities in RR and RV star classes.
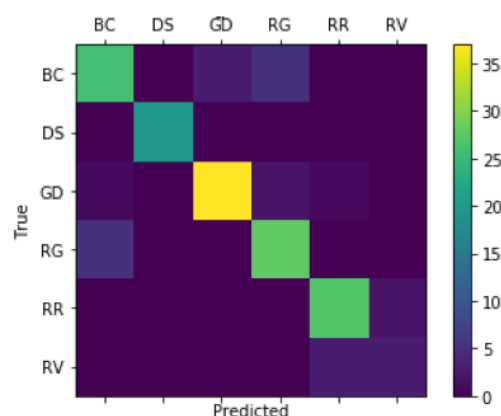

Fig. 6. Confusion matrix obtained for test dataset.

TABLE IV: CONFUSION MATRIX IN TABULAR FORM

Predicted / True	BC	DC	GD	RG	RR	RV
BC	26	0	3	5	0	0
DC	0	20	0	0	0	0
GD	1	0	37	2	1	0
RG	5	0	0	28	0	0
RR	0	0	0	0	27	2
RV	0	0	0	0	3	3

Precision, Recall, F1-Score; for each class as well as overall average values and accuracy achieved by the classifier is represented in Table IV. DS star class has achieved the 100% of highest precision, recall and F1 score among all the classes for this classifier which means that all the instances belong to this class were correctly classified. Lowest performance metrics were obtained by the RV class, and they are 0.60, 0.50, and 0.55, respectively. All the other classes have achieved precision ≥80%, recall ≥76%, F1-Score ≥79%. In general, all average performance metrics are above 85%. Overall Accuracy of the classifier is 86.5%.

Estimating the period of the star corresponding to the light curve data was the major issue which was faced in undertaking this project. Since the expected range of the periods of the time series data of many classes were different most in-built python modules and inappropriate for this project because there was no way to define a general range for all the types of classes. Still there were unexpected results for the period estimations and the classifier was trained including those results which directly impacts the accuracy of it. Period estimation should be enhanced further to extend the limits of the classifier to identify more star classes. Also, some instances had long gaps or data were missing for a considerable amount of time gap which may have caused shifts in the feature estimation.

There were some issues with gathering data and pre-processing stage since the available data were in different metric systems, various file types, missing data which reduced the dataset. Generally, more data means high accuracy and therefore using larger dataset improves the performance of the classifier.

In order to improve this machine learning method and expand this to next level, period estimation method has to be improved and a high time resolution dataset is needed. With these modifications and enhancements, the tool can be consistently use for the variable star classification of big datasets.

TABLE V: THE ARRANGEMENT OF CHANNELS

	Precision	Recall	F1	Support
BC	0.81	0.76	0.79	34
DS	1.00	1.00	1.00	20
GD	0.93	0.90	0.91	41
RG	0.80	0.85	0.82	33
RR	0.87	0.93	0.90	29
RV	0.60	0.50	0.55	6
AVg/Total	0.86	0.87	0.86	163
Accuracy =86.5%				

V. CONCLUSIONS

The machine learning technique developed can be used to identify star types such as BC, DS, GD, RG, RR and RV observed by the Keplar mission with an accuracy of 86.5%. Machine learning models which achieve accuracy beyond 80% were considered to be successful models. Program developed here achieved an average precision level of 0.86, average recall value of 0.87 and a F1-Score of 0.86 for the testing dataset of 163 stellar light curves obtained from the Kepler mission. This study could be extended into identifying more star classes by training it on extended dataset and further it can be improved as a web tool or an App.

REFERENCES

[1] L. Breiman, 2001. Random forests. *Machine learning*, *45*(1), Kluwer Academic Publishers, Netherlands, 2001, pp.5-32.
[2] L. Breiman, "*Bagging predictors. Machine learning*, 24(2), Kluwer Academic Publishers. Boston, USA, 1996, pp. 123-140.
[3] V. Belokurov, N. W. Evans and Y. Le Du, Light-curve classification in massive variability surveys–II. Transients towards the Large Magellanic Cloud. *Monthly Notices of the Royal Astronomical Society*, *352*(1), pp.233-242, 2004.
[4] V. Belokurov, N. W. Evans and Y. L. Du, Light-curve classification in massive variability surveys—I. Microlensing. *Monthly Notices of the Royal Astronomical Society*, *341*(4), pp.1373-1384, 2003.
[5] P. G. Willemsen andL. Eyer, A study of supervised classification of Hipparcos variable stars using PCA and Support Vector Machines. *arXiv preprint arXiv:0712.2898,* 2007.
[6] J. Debosscher, L. M. Sarro, C. Aerts, J. Cuypers, B. Vandenbussche, R. Garrido, and E. Solano, Automated supervised classification of variable stars-I. methodology. *Astronomy & astrophysics*, *475*(3), pp.1159-1183, 2007.
[7] A. Mahabal, S. G. Djorgovski, M. Turmon, J. Jewell, R. R. Williams, A. J. Drake, M. G. Graham, C. Donalek and E. Glikman, Automated probabilistic classification of transients and variables. *Astronomische Nachrichten: Astronomical Notes*, *329*(3), pp.288-291, 2008.
[8] L. M. Sarro, J. Debosscher, M. López and C. Aerts, Automated supervised classification of variable stars-II. Application to the OGLE database. *Astronomy & Astrophysics*, *494*(2), pp.739-768, 2009.
[9] J. Blomme, J. Debosscher, J. De Ridder, C. Aerts, R. L. Gilliland, J. Christensen-Dalsgaard, H. Kjeldsen, T. M. Brown, W. J. Borucki, D. Koch, and J. M. Jenkins, Automated classification of variable stars in the asteroseismology program of the Kepler space mission. *The Astrophysical Journal Letters*, *713*(2), p.L204, 2010.
[10] M. Paegert, K. G. Stassun and D. M. Burger, The eb factory project. i. a fast, neural-net-based, general purpose light curve classifier optimized for eclipsing binaries. *The Astronomical Journal*, *148*(2), p.31, 2014.
[11] P. Dubath, L. Rimoldini, M. Süveges, J. Blomme, M. López, L. M. Sarro, J. De Ridder, J. Cuypers, L. Guy, I. Lecoeur, and K. Nienartowicz, Random forest automated supervised classification of Hipparcos periodic variable stars. *Monthly Notices of the Royal Astronomical Society*, *414*(3), pp.2602-2617, 2011.
[12] D. W. Kim, C. A. Bailer-Jones, A package for the automated classification of periodic variable stars. *Astronomy & Astrophysics*, *587*, p.A18, 2016.
[13] J. W. Richards, D. L. Starr, N. R. Butler, J. S. Bloom, J. M. Brewer, A. Crellin-Quick, J. Higgins, R. Kennedy and M. Rischard, On machine-learned classification of variable stars with sparse and noisy time-series data. *The Astrophysical Journal*, *733*(1), p.10, 2011.
[14] NASA-Kepler/K2 Guest Observation Program. https://keplerscience.arc.nasa.gov/data-products.html#light-curve-files, April 29, 2018.
[15] M. Zechmeisterand M. Kürster, The generalised Lomb-Scargle periodogram-a new formalism for the floating-mean and Keplerian periodograms. *Astronomy & Astrophysics*, *496*(2), pp.577-584, 2009.
[16] N. R. Lomb, Least-squares frequency analysis of unequally spaced data. *Astrophysics and space science*, *39*(2), pp.447-462, 1976.
[17] J. D. Scargle, Studies in astronomical time series analysis. II-Statistical aspects of spectral analysis of unevenly spaced data. *The Astrophysical Journal*, *263*, pp.835-853, 1982.
[18] C. E. Ferreira Lopes and N. J. G. Cross, New insights into time series analysis. II-Non-correlated observations. *Astronomy and Astrophysics*, *604*, p.A121, 2017.

[19] I. Nun, P. Protopapas, B. Sim, M. Zhu, R. Dave, N. Castro and K. Pichara, Fats: Feature analysis for time series. *arXiv preprint arXiv:1506.00010,* 2015.

[20] J. VanderPlas, A. J. Connolly, Z. Ivezić, and A. Gray, Introduction to astroML: Machine learning for astrophysics. In *2012 conference on intelligent data understanding* (pp. 47-54). IEEE, 2012.

[21] P. B. Stetson, On the automatic determination of light-curve parameters for cepheid variables. *Publications of the Astronomical Society of the Pacific*, *108*(728), p.851, 1996.

[22] D. L. Welch and P. B. Stetson, Robust variable star detection techniques suitable for automated searches-New results for NGC 1866. *The Astronomical Journal*, *105*, pp.1813-1821, 1993.

J. Adassuriya was born in Colombo, Sri Lanka, on 29th May 1975. He obtained his B.Sc. with Physics major from the University of Colombo in 2002. He completed his M.Tech. Degree in Space and Atmospheric Science from Andhra University, Visakhapatnm, India, in 2015. He is reading for Ph.D. in short period variable stars.

He is currently working s a Senior Research Scientist at the Astronomy Division, Arthur C Clarke Institute, Sri Lanka. His research interests are short period variables, asteroseismology, and machine learning in astronomy.

Adassuriya, J., Ganesh, S., Gutiérrez, J.L., Handler, G., Joshi, S., Jayaratne, K.P.S.C. and Baliyan, K.S., 2021. Asteroseismology of SZ Lyn using multiband high time resolution photometry from ground and space. Monthly Notices of the Royal Astronomical Society, 502(1), pp.541-555.

Mr. Adassuriya is a life member of the Sri Lanka Association for the Advancement of Science and Institute of Physics, Sri Lanka. He won the gold medal for the first rank in P.G. Dip. (Space and Atmospheric Science).

Prof. K. P. S. C. Jayaratne was born in Colombo, Sri Lanka. He obtained his B.Sc. with Physics major from the University of Colombo in 1983. In 1986 he started his research on atmospheric physics and lightning with the Institute of High Voltage Research and Department of Meteorology in Uppsala University, Sweden. His Ph.D. work has led to the Construction of an Atmospheric Physics and Thunderstorm Electricity research laboratory in Sri Lanka. From 1992-1998 he was the Group Leader of the Atmospheric Physics and Lightning Research Group of Department of Physics, University of Colombo, and supervised several Ph.D. students in this field. He is the director of the astronomy and space science center of the department of physics, University of Colombo.

To his credit, he received in 1991 the "Third World Academy of Sciences Prizes for Young Scientists – The 1991 Prize for Physics awarded by NSF and in 1992, the GRC (General Research Committee of SLAAS) merit award for his contribution to Sri Lankan Science.

Herath, M., Gunesekera, S. and Jayaratne, C., 2021. Characterizing the possible interior structures of the nearby Exoplanets Proxima Centauri b and Ross-128 b. *Monthly Notices of the Royal Astronomical Society*, *500*(1), pp.333-354.

Prof. Jayaratneis is a life member of Sri Lanka Association for the Advancement of Science and Institute of Physics, Sri Lanka.
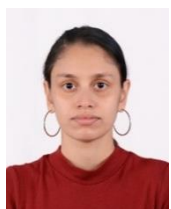
J. A. N. S. S. Jayasinghe was born in Kandy, Sri Lanka on 29th of March 1994. She obtained her B.Sc. with Computational Physics major from University of Colombo in 2019.

Currently she is reading for her PhD in Atmospheric Physics at University of Maryland Baltimore County, USA. She is interested in machine learning and data science application in satellite and ground based remote sensing instrumentation and data, climate modeling.