# Intrinsic Variable Star Classification using Python

E.J Grainger
*Level 3 MSci. Laboratory, School of Physics, University of Bristol.*

(Dated: February 20, 2024)

This experiment focuses on categorising intrinsically variable stars using the LightKurve Library. The results were accurately calculated and verified through a Convolution neural network and a Gaussian process. The results gathered match with literature values and accurately categorise the stars in the Hertz-Russel Diagram.

## INTRODUCTION

The Transiting Exoplanet Survey Satellite core objectives involved collecting star flux data for $200,000$ to $400,000$ stars. The goal was to use this flux data to detect the transit of exoplanets [9] as they pass through the observation cone of the satellite, causing deviation in flux. The wide scope in individual star measurements guaranteed observations of multiple stellar phenomenon; exoplanets, solar flares and binary pairs. In this analysis, we are focusing on stars that exhibited little to no stellar phenomena. These stars were captured in relatively inactive periods of their lives, and as such are great candidates for observing intrinsic variability. These are typically oscillations with a steady-state sinusoidal flux.

This experiment had three aims. One, to determine the periodicity of identified intrinsically periodic stars through the use of the LightKurve python library. Two, implement machine learning methods to generate a data pipeline for mass star analysis. The justification for this is to validate the TESS library methods using separate models for accuracy. Often programmatic systemic errors can be missed in large libraries, and as such should be checked. Three, to use analysed data to predict star type. This can again be used to test the rigour of LightKurve by comparing against literature values.

The LightKurve Library contains an extensive list of star data and analysis functions. As such, the focus is on optimising data through pre-processing and considering function contribution for the desired output. The majority of period analysis comes from Fourier transforms and phase folding, all of which are built in functions.

The justification for the use of machine learning models is in part outlined in the paper [2]. We have interpreted this philosophy and adapted our experiment accordingly. The use of machine learning in this project aims at solving the problem of manual classification for astronomical data. As astronomical data sets grow exponentially, approaching the level of a petabyte [2], the need to automate the classification of stellar objects is essential for handling increasing data backlogs and improving accuracy. In this experiment we implement a mass classifier in the form of a convolution neural network like that of the [2] experiment. The model is fed images of time-series data, and categorised into either intrinsic variability (excluding solar flares) versus all other forms of stellar variation. There is also the implementation of a Gaussian process, in which time-series data is mapped to a continuous function and the period is extracted.

Predicting star type exclusively from period analysis is a non-trivial task. Stellar objects of all different kinds evolve under a vast amount of different mechanics. Cepheid variable and Mira variable stars, typically red giants, have a history of keen astronomical observation, as their periods give guidance in determining stellar distances. We focus our attention on stars that fall under the category of bright, cool dwarf stars [9] which reside in the lower right hand band of the Hertz-Russel Diagram.

## THEORY

Steady-state stars' primary mechanic contributor arises from the dynamic equilibrium shifts between fusion and gravity forces. Hydrostatic equilibrium is the reason stars maintain their structure. Given the violent nature of stars, variability is intrinsically linked to deviations about this point from gradual fluctuations of strength form each opposing force. However, it is not complete to say that this is the variability that has been observed in the TESS missions. In the following section, the possible other mechanics, star-spots and convection currents, will be outline. It is important to mention that a much greater amount of data is required to make sensible predictions about a stars exact (or combination) of mechanics. Most of the mechanics we are interested in evolve on the time scale of years.

### Variable Star Mechanics

Intrinsic Variability is the oscillation in flux of a star over some time period which arises from a property that is directly linked to the star. To specify, this variability is completely determined by internal star mechanics[1]. This doesn't include variability due to extra-terrestrial interference like exoplanets or gas clouds. It is observed through increases and decreases in intensity as the number of photons(flux) increases and decreases.

Stellar convection occurs under special conditions of pressure and temperature gradients[8]. These currents disrupt the normal propagation of energy towards the surface of the star, and instead matter follows currents at a non perpendicular angle to the core. In cooler stars, such as the ones we are observing, these currents can alter the normal transition of photons from the core to the centre. Mean photon transmission results in uniform distribution of photons (a star which appears spherically bright across all surfaces). This creates a differential in energy flow to the surface, and periodically decreasing and increasing surface luminosity. As the star rotates, it would appear to the observer as regular periods of brightness and dimness.

Star-spots are local cooler regions on star surfaces. A stars magnetic fields form bands within the surface following a spherical shape. Due to the great size of stars, the matter moving at the radius moves far quicker that that at the poles, as star matter rotates, these bands overlap and combine. These differential currents are attributed with stellar convection. Eventually, the fields snap and rise above the surface [3], ejecting matter and cooling the surface. These cooled areas are known as star-spots. Often multiple spots merge to from areas known as an active region. When these star spots appear/rotate to the side of observation, the reduced flux is associated with an increase is star surface opacity. These star-spots do decrease flux, but also have a lifetime directly proportional to its size[6]. The movement, density and size can indicate stellar variability, but the mean lifetime is so large that to compare this to normal flux, far more data is needed.

### Convolution Neural Networks and Gaussian Processes

The nature of convolution neural networks are to classify images through the identification of spacial hierarchies on any grid like object, in our case time-series images. Through a system of layers, input data (pixels) are condensed into features through convolution and pooling layers. These form decision trees acting to point input data towards some classified region based on the pixel properties. The last layers of the neural network act to associate the identified structures with our pre processed data, this is our identifier complete.

These identified features are part of the hidden layers of the neural network. These topological abstraction cannot be identified by eye, hence the need for this model. For a well trained model, it is expected to massively out scoring humans in terms of correct classifications in a large dataset. [5].

In this experiment, 433 time-series images were used to classify between intrinsic variability and extrinsic/non variability.

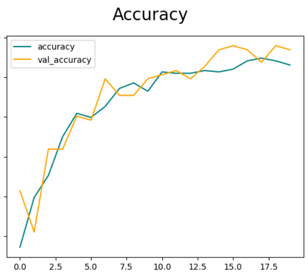

FIG. 1. Accuracy of the Neural Network

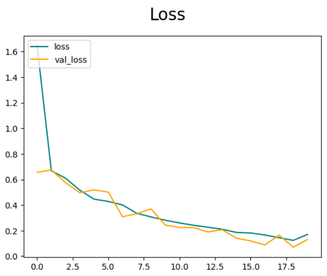

FIG. 2. Loss of the Neural Network

The Gaussian process is another form of machine learning technique. It is used to approximate an underlying function given a set of points in some n-dimensional coordinate system. For our time series data, we have normalised flux against time, with a discrete subset of the time axis( where the discontinuity comes from data gathering errors). Due to the complexity in gathering astronomical data, it is often expected that there will be gaps or deviation from the true measurements due to mechanised errors, the Gaussian process will override this and fill in gaps in its data. We postulate that there exists some function f(t) that passes through all points and can accurately describe how flux varied.The Gaussian process then involves forming a Gaussian distribution across each known value of the independent variable(s) in the x-y plane, where its value is determined by the mean of a Gaussian distribution based on the variance associated at that point. This outputs is a complete function of x and y that accommodates all space. We then can analyse properties of this function to make estimations about the underlying data, and even extrapolate into the future. The bands can be observed as the grey regions surrounding the black line in fig[3].

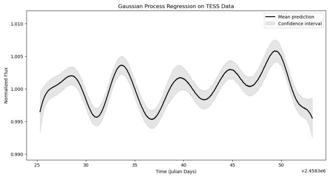

FIG. 3. GP fit

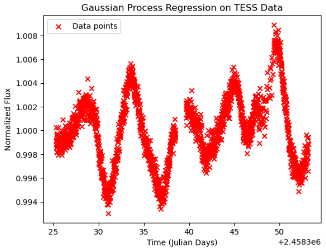

FIG. 4. LightKurve Data

### Time Series Analysis

Extracting periodicity from time-series data involves a multitude of steps, our main focus in on the LightKurve Periodogram function. We start with a timeseries graph that has been scattered.

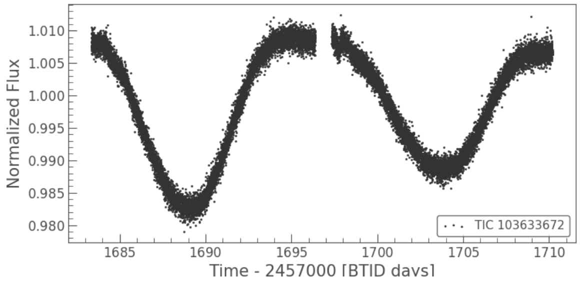

FIG. 5. LightKurve Scatter Graph

This function gives a power spectrum of different signal strengths at different frequencies. The most prominent frequency is most likely the underlying one, so this is what is extracted. The discrete Fourier transform follows the generation equation:

$$P(f) = \left| \sum_{n=0}^{N-1} x[n] e^{-i2\pi f n} \right|^2 \tag{1}$$

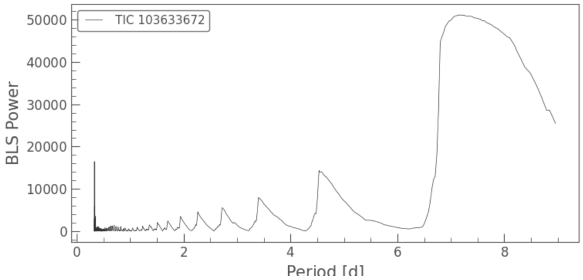

FIG. 6. Fourier Transformation of LightKurve

This function is applied using the Box-Least-Square method, following the general form

$$\log L(P, \tau, t_0) = \frac{1}{2} \sum_{in} \frac{(y_n - y_{in})^2}{\sigma_n^2} - \frac{1}{2} \sum_{out} \frac{(y_n - y_{out})^2}{\sigma_n^2} + c \tag{2}$$

In some cases, there is one clear underlying frequency(fig[6]), although often enough there are major contributions from several peaks. Using the ”period at max power” function, we can extract the greatest peak, and fold the data on top of the original time-series. We observe the quality of the fold, and apply the $Chi^2$ fitness test to gather matching. The phase folding follows the general equations:

$$\phi = \frac{t \bmod P}{P} \tag{3}$$

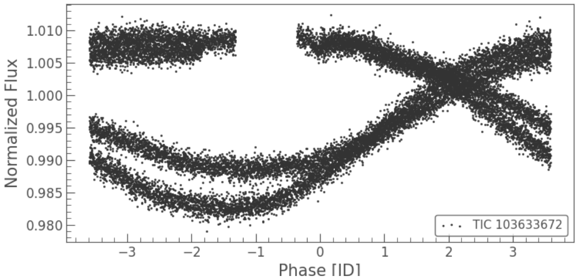

FIG. 7. Period fold over LightKurve data

We use the relation $m - M = 5\log_{10}(d) - 5$ to calculate the absolute magnitude to plot on the HV diagram.

## EXPERIMENTAL DETAILS

The Programmatic flowchart fig[8] shows the flow of data through the diagram

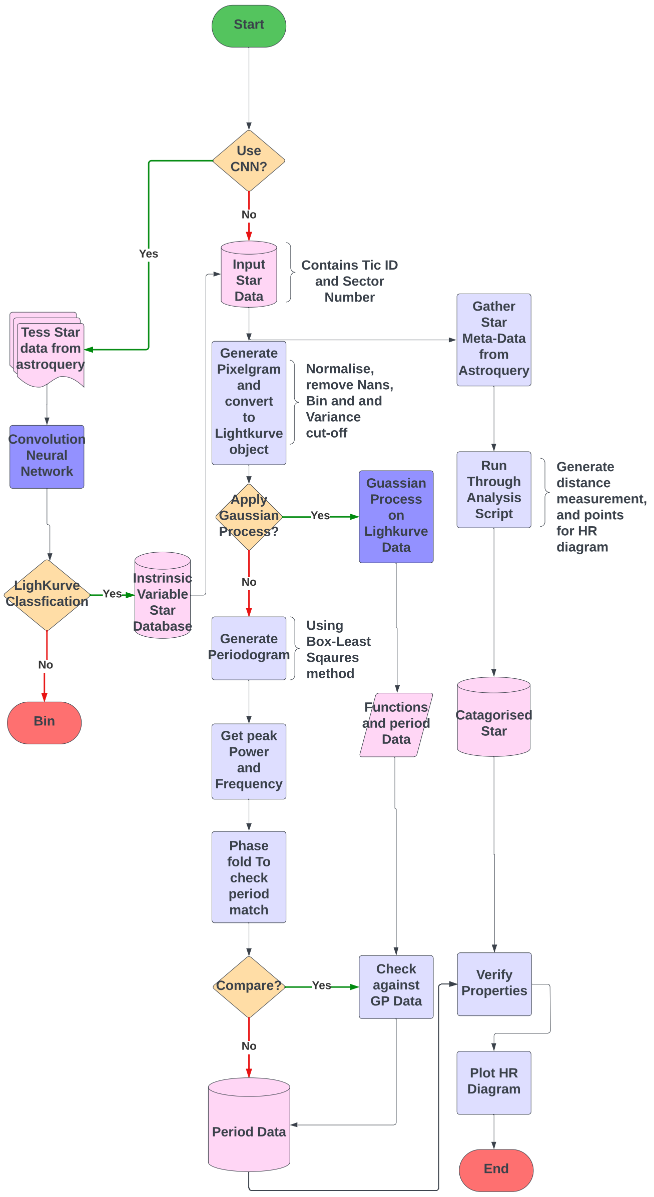
Start
Use CNN?
Yes
No
Input Star Data
Contains Tic ID and Sector Number
Tess Star data from astroquery
Gather Star Meta-Data from Astroquery
Generate Pixelgram and convert to Lightkurve object
Normalise, remove Nans, Bin and and Variance cut-off
Convolution Neural Network
Apply Gaussian Process?
Yes
Guassian Process on Lighkurve Data
Run Through Analysis Script
Generate distance measurement, and points for HR diagram
No
LighKurve Classfication
Yes
Instrinsic Variable Star Database
No
Generate Periodogram
Using Box-Least Sqaures method
Bin
Functions and period Data
Catagorised Star
Get peak Power and Frequency
Phase fold To check period match
Compare?
Yes
Check against GP Data
Verify Properties
No
Plot HR Diagram
Period Data
End

FIG. 8. Flowchart of Python Script

## RESULTS

From a combination of LightKurve analysis and the Astroquery database[4], we gather these results

TABLE I. Combined properties of TIC IDs with periods

| TIC ID | Period | cdpp | GP Period | Vmag | Ms | Kmag | Teff | Distance (pc) |
|---|---|---|---|---|---|---|---|---|
| 23609565 | 3.4207 | 143.46 | 4.4323 | 8.33 | 1.32 | 7.067 | 6485.87 | 166.304 |
| 103633672 | 7.1738 | 413.33 | 4.840 | 10.586 | 0.87 | 8.575 | 5134 | 84.5772 |
| 126947245 | 5.2137 | 153.54 | 0 | 10.021 | - | 6.116 | 3983 | 1260.09 |
| 279614617 | 7.1418 | 283.20 | 4.496 | 9.867 | - | 7.884 | 5522.77 | 279.505 |
| 341849173 | 2.7481 | 212.42 | 4.259 | 9.186 | 2.38 | 8.602 | 9384 | 652.782 |
| 441420236 | 4.8577 | 227.44 | 4.729 | 8.81 | 0.662074 | 4.529 | 9722.1 | - |

## DISCUSSION

The results gathered from this Hertz-Russel diagram clearly indicate that all stars selected for observation can be classified as main sequence dwarf stars. These stars have masses around that of the suns(0.87 - 2.38), relatively low intensities and surface temperature. This is also what is expected when considering which stars in the HR diagrams are under the most inactive periods of their life-cycles out of the total main sequence structure. The results for periodicity can be seen ranging between 3-7 days, which matches stated values of [7] around 3-8 days. There is a clear correlation between the period data set and the HR classification data set. It can also be seen that the Gaussian process roughly estimates a period that is consistently in the middle of the expected range. It is expected that the generalisation of the kernel function for a Gaussian will result in periodicity error, as kernel function accuracy greatly determines the quality of outcome.

We can further reinforce our estimates without further period analysis by looking at the HR diagram of a large dataset of CNN classified stars. We can generalise our matching between the two methods by considering the distribution of stars which have been classified as variable. It can be found in fig[9] that all stars are in the lower region which mathces to the cool, bright dwarf variable star. For completeness the Combined Differential Photometric Precision is also included, the greater the value, the greater the noise associated with the data. This is removed with the flatten function, but in this experiments case we loose all generality when flattening, so it is used to indicate uncertainties in the values. All stars excluding Tic 23609565 had one or two significant frequency contributions, so the strongest power peak was a suitable match against the lightkurve data. The star aforementioned has the greatest associated uncertainty.

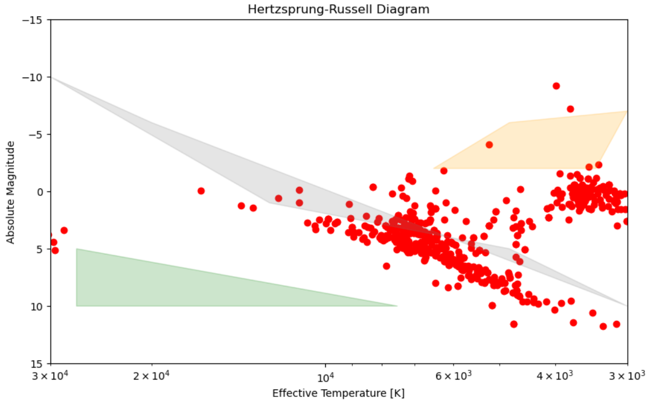

FIG. 9. Hertz-Russel Diagram of Identified Stars

## CONCLUSIONS

Although the periods that are calculated are not intrinsically indicative of star classification, the unification with data from the Astroquery database and the plotting on the Hertz-Russel diagram allows us to identify what kind of stars we have observed. All objectives were met successfully, accurate period analysis as referenced by [7] show consistency in the LightKurve Library. Both machine learning models helped reinforce and expand our certainty in the star classifications. To improve upon this experiment, a servery of singular stars would be more appropriate for intrinsic mechanics, as more data is required to make these calculations. It would also be useful to consider other calculation methods of Fourier analyses that had multiple contributions of significant power peaks, as a more accurate representation would involve greater peak number consideration.

## REFERENCES

[1] E. H. Anders, D. Lecoanet, M. Cantiello, K. Burns, Benjamin A. Hyatt, and Emma Kaufman. The photometric variability of massive stars due to gravity waves excited by core convection. *Nature Astronomy*, 2023.

[2] John A. Armstrong and Lyndsay Fletcher. Fast solar image classification using deep learning and its importance for automation in solar physics. *Solar Physics*, 2019.

[3] Harold D. Babcock and Horace W. Babcock. The topology of the sun's magnetic field and the 22-year cycle. *The Astrophysical Journal*, 1961.

[4] Adam Ginsburg, Brigitta M. Sipőcz, C. E. Brasseur, and Cowperthwaite. astroquery: An Astronomical Web-querying Package in Python. , 158(5):235, November 2019.

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *arXiv: Computer Vision and Pattern Recognition*, 2015.

[6] Kristof Petrovay. Solar cycle prediction. *Living Reviews in Solar Physics*, 2020.

[7] J. De Ridder, Conny Aerts, V. Ripepi, and C. Aerts. Gaia data release 3. pulsations in main-sequence obaf-type stars. *Astronomy and Astrophysics*, 2022.

[8] Da run Xiong. Convection theory and related problems in stellar structure, evolution, and pulsational stability ii. turbulent convection and pulsational stability of stars. *Frontiers in Astronomy and Space Sciences*, 2021.

[9] Keivan G. Stassun, Ryan J. Oelkers, Joshua Pepper, Martin Paegert, Nathan De Lee, Guillermo Torres, and Guillermo Torres. The tess input catalog and candidate target list. *The Astronomical Journal*, 2018.

## STRIKE ACTION IMPACT

Please state any impact the strike action on your experiment even if there was none, e.g. reduced supervision, quantity of data, quality of analysis, etc. This is a compulsory section and the only section which should appear on page 5.