The verification of periodicity with the use of recurrent neural networks

N. Miller [1]★ P. W. Lucas [1] Y. Sun,[2] Z. Guo,[1,3,4,5] W. J. Cooper [1,6] and C. Morris [1]

[1]*Centre for Astrophysics Research, University of Hertfordshire, College Lane, Hatfield, Hertfordshire AL10 9AB, UK*
[2]*Centre for Computer Science and Robotics Research, University of Hertfordshire, College Lane, Hatfield, Hertfordshire AL10 9AB, UK*
[3]*Instituto de Física y Astronomía, Universidad de Valparaíso, ave. Gran Bretaña, 1111, Casilla 5030, Valparaíso, Chile*
[4]*Núcleo Milenio de Formación Planetaria (NPF), ave. Gran Bretaña, 1111, Casilla 5030, Valparaíso, Chile*
[5]*Departamento de Física, Universidad Tecnicá Federico Santa María, Avenida España 1680, Valparaíso, Chile*
[6]*Istituto Nazionale di Astrofisica, Osservatorio Astrofisico di Torino, Strada Osservatorio 20, I-10025 Pino Torinese, Italy*

ABSTRACT

The ability to automatically and robustly self-verify periodicity present in time-series astronomical data is becoming more important as data sets rapidly increase in size. The age of large astronomical surveys has rendered manual inspection of time-series data less practical. Previous efforts in generating a false alarm probability to verify the periodicity of stars have been aimed towards the analysis of a constructed periodogram. However, these methods feature correlations with features that do not pertain to periodicity, such as light-curve shape, slow trends, and stochastic variability. The common assumption that photometric errors are Gaussian and well determined is also a limitation of analytic methods. We present a novel machine learning based technique which directly analyses the phase-folded light curve for its false alarm probability. We show that the results of this method are largely insensitive to the shape of the light curve, and we establish minimum values for the number of data points and the amplitude to noise ratio.

Key words: Machine Learning – Numerical Methods – Data Methods – Algorithms – Light curves – Variable Stars.

1. INTRODUCTION

The identification of periodic variable stars is not a trivial task; well-understood statistical measures can be used to identify variability in time-series but not so easily periodic variability. The Stetson variability index 'I' (Stetson 1996) compares the variability of each observation with its neighbour and their errors. The Von Neumann eta index 'η' (Neumann 1941) represents the ratio of the mean of the successive differences squared, to the variance of the light curve. Both of these methods are reasonably robust in detecting variability in time-series. More simplistic methods, such as a comparison between some measure of scatter (inter quartile range, standard deviation σ, or median absolute deviation) and the uncertainty, have also been shown to be useful (Sokolovsky et al. 2017). Using tools such as the Lomb–Scargle method (Lomb 1976; Scargle 1982) and Phase Dispersion Minimization (PDM, Stellingwerf 1978), we can construct a periodogram to probe for periodic variability. Nevertheless, extrema in the periodogram are likely to be present regardless of whether or not the source is truly periodic. These extrema can scale with the amplitude of the periodic signal such that periodograms of periodic sources become distinct from truly random variability. However, in cases where a light curve features aperiodic or secular variability, ambiguities can arise (Park et al. 2021). This is of particular issue when dealing with stars which can feature multiple sources of variability, such as asymptotic giant

branch stars (Templeton, Mattei & Willson 2005), whose long-term periodicity could be undifferentiable to that of secular variability by periodogram analysis alone. Furthermore, their values do not scale universally (i.e. the peak value for an aperiodic source may be the same as that for a periodic source).

In cases where extrema are not present, this could be interpreted as an indication of insufficient periodogram coverage or the lack of periodic variability.

Thus, we do not automatically obtain a universal measure of periodicity from a periodogram. If a periodogram shows candidate periods, then for smaller selections of sources, it is feasible to manually verify the periodicity of each. This is typically performed by visual inspection of the phase-folded light curve. Looking forward, in the current and future age of survey astronomy with surveys such as LSST (Ivezić et al. 2019), ZTF (Bellm et al. 2019), *Kepler* (Borucki et al. 2003), and *TESS* (Ricker et al. 2015), we anticipate time-series catalogues of sizes that render sufficient manual inspection increasingly non-viable. Hence, a reliable and robust metric for identifying periodicity is required.

It is not a guarantee that a large survey will feature high cadence sampling. Surveys such as VISTA Variables in the Via Lactea (VVV, Minniti et al. 2010; Saito et al. 2012), the NEOWISE mission of the Wide Field Infrared Survey Explorer (Wright et al. 2010; Mainzer et al. 2014), and *Gaia* (Gaia Collaboration 2021) have catalogues which can also feature large sample sizes for which rigorous human inspection is impractical. These surveys contain relatively few observations for each source, an issue that is also very common in small, targeted observing projects. The sparse

★ E-mail: n.miller4@herts.ac.uk

sampling makes it harder to confirm periodicity with classical methods.

The metric for determining periodicity in a time-series is commonly referred to as a false alarm probability (FAP). Previous work on determining an accurate FAP has largely been directed towards the analysis of the constructed periodogram. These methods, such as the method proposed by Baluev (2008), employ extreme value statistics to determine an upper bound for the FAP of a Lomb–Scargle periodogram. This has the clear limitation that the method is designed to distinguish sinusoidal variations from Gaussian white noise, not accounting for stochastic variability, non-Gaussian errors, imprecise error estimates, and non-sinusoidal periodic variations. Baluev (2009) extended their earlier work to the case of multi-harmonic light curves but this is only a partial solution to the above issues. Bootstrapping is another commonly used technique where the periodogram of a light curve that has been randomly shuffled N times to create N aperiodic periodograms is compared with that of the unshuffled light curve. The FAP in this case is the percentage of times the peak of an aperiodic periodogram is larger than that of the peak from the suspect periodic periodogram.

In Stellingwerf (1978), a statistical analysis of the constructed PDM periodogram is used to obtain a metric of FAP (p-value). This method assumes that photometric errors are perfectly estimated Gaussians. The absence of any other aperiodic variability is also assumed. There is also no treatment of spurious artificial periodic signals, which can occur with unevenly and sparsely sampled light curves. Many surveys feature these periods at varying rates of incidence. It is of particular note for ground-based surveys with semi-regular observing patterns, such as the VVV survey. Methods such as PDM that bin the phase-folded light curve to construct their periodogram are also limited by imperfections in the model. This can become increasingly significant as sampling decreases. This issue exists even with the binless approach to PDM presented by Plavchan et al. (2008). Separately, heuristic methods based on reduced χ^2 statistics have been employed to distinguish true and false periodic variable star candidates (e.g. Irwin et al. 2009). This explicitly acknowledges the effects of an imperfect light-curve model and imprecise photometric uncertainties. In this work, we show how we can utilize neural networks to differentiate between true and false periodic variable star candidates without the need for a prior light-curve model.

2. METHOD

In our approach, the analysis of the light curve is achieved via a recurrent neural network (RNN, Hochreiter & Schmidhuber 1997). An RNN was chosen because they are designed and used for serially correlated data, such as astronomical light curves. Previous efforts in their use with light curves have shown their applicability and ability to parse astronomical time-series data (Zhang & Zou 2018; Burhanudin et al. 2021). This network is trained on pre-labelled periodic and aperiodic phase-folded light curves of variable stars. The network was trained for 96 epochs[1] with an Adam optimizer (Kingma & Ba 2014) and with 20 per cent of the training data used as a validation set. Early stopping was used to halt training as soon as the incremental change in the validation loss function, $\Delta L < 10^{-5}$.

The model is constructed with 13 gated recurrent unit (GRU, Cho et al. 2014) layers, 1024 nodes per layer, and a binary cross entropy loss model. The choice of GRUs over long-short term

---

[1]An 'epoch' here is an iteration over the whole training set.

memory (Hochreiter & Schmidhuber 1997) was motivated by the calculated loss, which was lower for GRUs. The RNN was written in KERAS (Chollet et al. 2015).

Ablative testing has shown that the specifics of the architecture of the network are not crucial and having 'enough' GRUs is sufficient for operability.

2.1 Data preparation

The input training data consist of the magnitude (m_i), phase (ϕ_i), and change in phase (i.e. $\Delta\phi_i = \phi_i - \phi_{i-1}$). Magnitude errors were not used for this method as they commonly do not fully represent the true photometric uncertainty. We tested various combinations of features and removing the magnitude error consistently improved performance with lower loss and higher accuracy. Instead, we reject any points with a large magnitude error ($m_{i,\mathrm{err}} \geq 0.1$ in this case). We also reject points with a high DoPHOT (Schechter, Mateo & Saha 1993) 'Chi' parameter, which indicates a poor fit to a stellar profile.

The input also includes a feature that is derived from an interpolated fit of the time-series with 200 evenly spaced points, performed by an inverse distance-weighted K-nearest neighbours (KNN) regressor (Fix & Hodges 1951) which was taken from SCIKIT-LEARN (Pedregosa et al. 2011). This was performed as a form of smoothing in an attempt to more clearly display variability with evenly spaced data.

A randomly variable light curve will have an interpolated fit that tends towards a straight line. Each of these features were added after ablative testing (i.e. features were added and removed iteratively and the combination of features that produced the highest accuracy and lowest loss was used). Each light curve was either cut to 200 data points in size or padded with zeroes to a length of 200.

The same light curve is phase-shifted randomly 10 times by an amount between 0 and 2π and each version is shown to the neural network. This is done in an attempt to remove a dependency on the starting position of the light curve. This is similar to the methodology for contrastive learning (Chen et al. 2020a). We do not want the network to care about the absolute phase value.

Alternatively, we could ensure the light curve is always ordered from a set point in the light curve, such as the turning points. However, we found this step to be unreliable with noisy data. A single unfiltered outlier or otherwise erroneously extreme point would cause such an approach to fail as the light curve's minima could be incorrectly identified.

3. DATA

The training data used for training the neural network FAP (NN FAP) is a combination of both real and synthetic light curves.

In the trained model used for this paper, there were 20 000 real and 60 000 synthetic light curves with half of each corresponding to periodic or aperiodic. This means that an FAP of 0 was given to the 10 000 real and 30 000 synthetic periodic light curves and an FAP of 1 was given to the other 10 000 real and 30 000 synthetic aperiodic light curves. The synthetic light curves were split evenly across each of the five listed equations (1–4).

Through the development of this method, it was found that a small number of mislabelled light curves can have a large impact on the abilities of this method (i.e. an aperiodic light curve being labelled as periodic or vice versa).
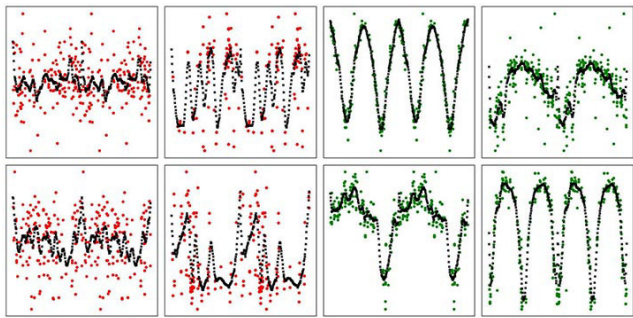
Figure 1. Some example real aperiodic (red, left) and periodic (green, right) light curves used for training. The black points represent the evenly spaced fit provided by the KNN regressor.
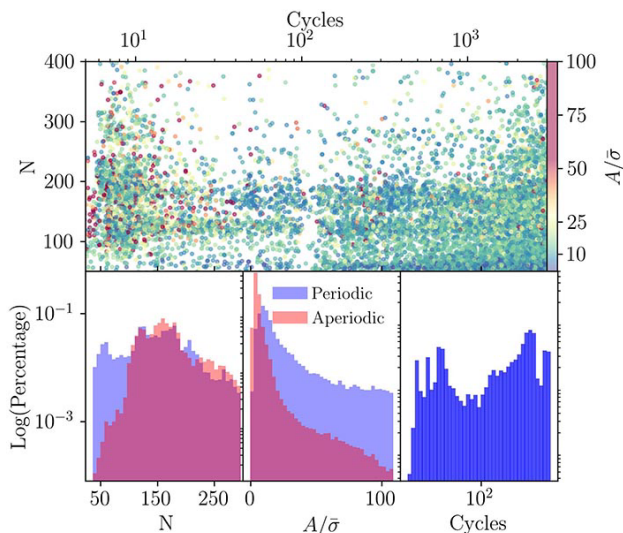
Figure 2. Showing the distribution of the number of data points (N), signal-to-noise ratio, and cycles (light curve length/period) for the training set used. Top panel: Cycles versus the number of data points with colour axis as signal-to-noise ratio for periodic variables in the training data. Bottom panel: Histograms for each of these values.

3.1 Real training data

The training data are VVV light curves whose periodic nature was supported by classification from two optical surveys. A set of 10 000 known real periodic light curves were identified by eye (by co-authors NM, CM, and WC) after cross-matching data from the VVV survey [and a pre-release version of its time-series catalogue (VIRAC 2 – Smith et al. 2018; VIRAC 2-β – Smith et al., in preparation)] with other known periodic variable star catalogues, namely the ZTF catalogue of periodic variable stars and the ASAS-SN catalogue of variable stars (Pawlak et al. 2019; Chen et al. 2020b). The cross-matching was performed to generate a list of suspect periodic and aperiodic variable stars.

All of the 10 000 aperiodic light curves were identified by eye as rejected periodic variables.

Fig. 1 shows a random selection of real training light curves and their interpolated fit. Both the interpolated fit and the raw magnitude measurements are given to the RNN.

Fig. 2 shows the distributions of number of data points, signal-to-noise ratio, and number of cycles in the time-series for the training data. The real data are drawn directly from this distribution. We show
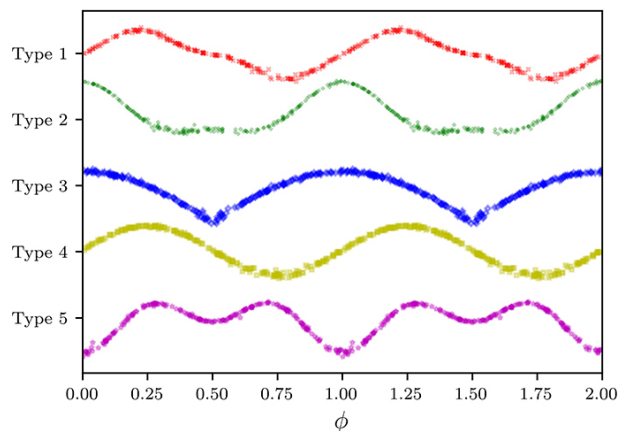
Figure 3. Examples of each of the forms of the light curves used for testing and training the NN FAP.

the number of cycles as opposed to period because the neural network is trained exclusively with phase-folded light curves. As VVV light curves can vary in length, it is the product of the light-curve length with the stars frequency that affects the structure of the phase-folded light curve.

3.2 Synthetic training data: periodic light curves

Synthetic light curves were created via the use of a real light curve with a periodic signal injected, similar to the work by Graham et al. (2013). An overview of the steps taken is as follows:

(i) Remove all photometric information from a real light curve, retaining only the time stamps.

(ii) Inject periodic signal into 'blank' light curve (see equations 1–4).

(iii) Generate the errors by sampling from those associated with the real photometry, using a look-up table.

(iv) Scatter light curve based on the injected error.

Training the neural network on exclusively sinusoidal light curves could bias our FAP against Eclipsing Binaries and other more complex light curves.

Fig. 3 shows an example of each of the forms of light curves generated with 1000 measurements with an amplitude of 1 mag. These equations aim to roughly (but not exactly or comprehensively) model the common types of pulsators and binary light curves that are seen (Molnar et al. 2022). Type 1 is a distorted sinusoid which is a fairly standard form for synthetic light curves (Cincotta, Mendez & Nunez 1995; Huijse et al. 2012). Types 2 and 5 are eclipsing binary-like light curves (i.e. more than one turning point per period). Type 3 is used to mimic the common identifying feature of a contact binary system (Kirk et al. 2016) and Type 4 is a simple sinusoid. An important reason for using multiple shapes to train the network is to remove as much of a dependency on light-curve shape as possible. This is similar to the methodology for contrastive learning. By showing the network multiple different shapes of a periodic signal we aim to remove any biases related to its shape.

The method by which the synthetic data are created also means that the light-curve parameters are drawn from the distributions shown in Fig. 2. The periods used are randomly selected from a uniform distribution between 0.1 and half the length of the light curve [period $\sim \mathbb{U}(0.1, \sim 1500)$]. We note that the period (or number of

cycles) used for the synthetic light curves is largely inconsequential to how it is perceived by the RNN. The RNN is only shown the phase fold of the light curve and so there is little difference between otherwise identical light curves with different periods. This is also the case for the total time range of the light curve. Provided at least one cycle is captured, the number of measurements and signal-to-noise are the limiting factors. This is a potential caveat for this method as a low FAP could be assigned for a light curve with only one cycle, which is not sufficient for the actual identification of periodicity. We recommend only trusting the FAP from this method if the period is less than half of the length of the light curve (i.e. at least two cycles are captured).

$$\texttt{Type 1.}\ m(t) = 0.5\sin\left(\frac{2\pi t}{P}\right) - B_1\sin\left(\frac{4\pi t}{P}\right)$$
$$- B_2\sin\left(\frac{6\pi t}{P}\right) \tag{1}$$

$$\texttt{Types 2 \& 5.}\ m(t)$$
$$= 1 \pm \left(A_1\sin\left(\frac{2\pi t}{P}\right)^2 + A_2\sin\left(\frac{\pi t}{P}\right)^2\right) \tag{2}$$

$$\texttt{Type 3.}\ m(t) = \left|\sin\left(\frac{2\pi t}{P}\right)\right| \tag{3}$$

$$\texttt{Type 4.}\ m(t) = \sin\left(\frac{2\pi t}{P}\right). \tag{4}$$

A periodic signal is added to the source light curve, and the photometric error is derived using a KNN search of a data set containing information about the photometric uncertainty of 1 000 000 data points from the VIRAC data base. This data set is utilized to identify the 100 nearest neighbours, from which the mean and standard deviation are computed.

Each data point in the light curve has its photometry (m) and photometric error (m_{err}) drawn from a Gaussian constructed of these 100 nearest neighbours.

3.3 Synthetic training data: aperiodic light curves

We employ two methods to generate aperiodic light curves: a real or synthetic periodic variable has its photometric order randomly shuffled. The time data are left unmodified to conserve the observing cadence of the original survey. We effectively create a light curve of random noise with the survey's observing pattern conserved. This method also removes any other correlated effects, such as photometric uncertainty, that may be present in real aperiodic light curves. One caveat present is that by destroying correlated effects, the neural network could differentiate between the aperiodic and periodic synthetic light curves with greater ease. The second method of aperiodic synthetic light curve generation involves taking a known non-variable star (identified with a Stetson index <0.1) and re-sampling the photometric points with a larger scatter. For each measurement a Gaussian is constructed with $\mu = m_i$ and $\sigma \geq 3 \times m_{i,\text{err}}$, where $m_{i,\text{err}}$ is the measurement error. The light curve is then re-scaled to ensure a realistic amplitude.

This method retains as much temporally correlated, but non-periodic, information as possible compared with the random shuffle method. An example of this is with astronomical seeing, which can vary on long time-scales, affecting multiple measurements. With VVV (and subsequent catalogue VIRAC 2-β) data we have instances where bad seeing causes DoPHOT to systematically underestimate flux in crowded fields. Such a case could appear as a non-periodic signal in the light curve. In less crowded fields, poor weather will
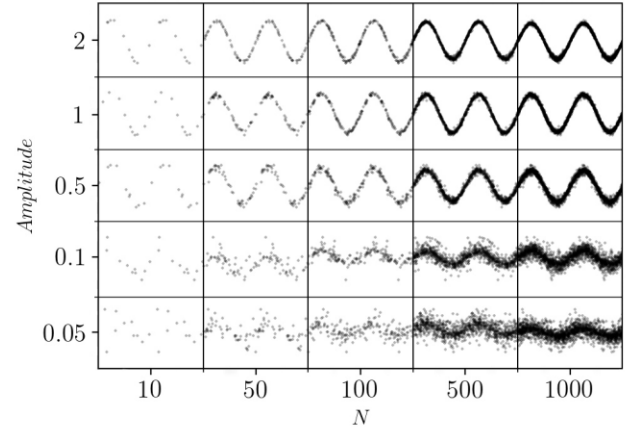

Figure 4. Examples of a synthetic sinusoidal light curve varying through the number of data points in the light curve on the *x*-axis and the amplitude of the light curve in the *y*-axis. The median magnitude error for each point was 0.1.

increase the uncertainty at times, creating correlated uncertainty which may occasionally lead to a spurious aperiodic signal. This is of particular note as the neural network is never shown the photometric uncertainty. This method of inflating measurement error will weaken but not fully destroy these correlated effects.

The random shuffle method enables training with non-Gaussian aperiodic signals. Due to the limitations of these methods, it is beneficial to also have real training data. The synthetic data has the advantage of volume with the certainty of aperiodicity. This allows us to construct a training data set large enough to train an RNN.

3.4 Test data

We generate three data sets to test our classifier. A real data set was constructed by manually classifying 8000 previously unseen real light curves taken from the same VVV survey. These 8000 sources were identified from the same ZTF and ASAS-SN periodic catalogues that were used in training. Each light curve has a $A/\bar{\sigma} > 2$, where 'A' is the amplitude calculated as the difference between the 1 per cent and 99 per cent percentile after sigma clipping and '$\bar{\sigma}$' is calculated as the median value of the magnitude error. The manual classification of the real light curves involved selecting phase-folded periodic variables by eye. This was independently repeated multiple times by three astronomers to ensure reliability. All of the astronomers agreed on classification. Any ambiguous light curves were removed from the set. Two synthetic data sets were also constructed via the method described in Section 3. The data set 'Variable N' was generated as 80 000 identical synthetic light curves with only the number of data points per light curve varied ($10 < N < 600$). A median signal-to-noise ratio (SNR) ($A/\bar{\sigma}$) of 10 was generated for each of these. The data set 'Variable SNR' was generated as 80 000 identical synthetic light curves with only the signal-to-noise ratio varied. For each light curve in the Variable SNR data set, there were 200 data points used. Fig. 4 exemplifies both 'Variable N' and 'Variable SNR' on the *x*- and *y*-axes, respectively. The four types of synthetic variables used were evenly split for both of the synthetic data sets.
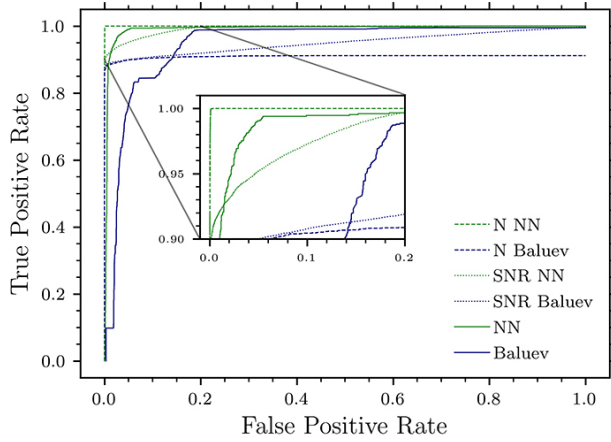
Figure 5. Showing the ROC curve for the neural network and Baluev methods as a binary classifier. *Solid line* shows real data set classified by eye. *Dashed line* shows synthetic data set where the number of measurements was varied (Fig. 8). *Dotted line* shows a synthetic data set where the SNR was varied (Fig. 9).

Table 1. Showing the AUC for each data set and method.

Source	NN FAP	Baluev
Real	0.99 193	0.95 245
Variable SNR	0.99 808	0.97 843
Variable N	0.99 703	0.97 393

4. EXPERIMENTAL RESULTS FROM RNN

To quantify the performance of the NN FAP we can test its ability as a binary classifier and compare it with the commonly used Baluev method. We use the generalized Lomb–Scargle periodogram along with its associated FAP as described by Zechmeister & Kürster (2009) for our calculations of the Baluev FAP. The Baluev FAP typically lies in a range between unity and 10^{-200} and so the y-axis of the Baluev FAP plots have been shown as both linear and logarithmic scaling.

4.1 Performance measurements

The receiver operator characteristic (ROC) curve is used to measure the capability of a binary classifier as the threshold for classification is varied. An idealized binary classifier will have a threshold at which the *sensitivity* and *specificity* are equal to 1.

Fig. 5 shows the true positive rate (otherwise known as the sensitivity) versus the false positive rate (otherwise known as specificity). Equation (5) shows more clearly how sensitivity and specificity are defined:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad \text{Specificity} = \frac{TN}{TN + FP} \tag{5}$$

where *TP* and *TN* are true positive and true negative, respectively, and *FP* and *FN* are false positive and false negative, respectively.

The area under the ROC curve (AUC) can be calculated as an evaluation metric for a binary classifier. Table 1 shows the AUC for each tested data set. This shows that the NN FAP method has a larger AUC for each data set than the Baluev method. This indicates that the NN FAP method performs better in each test. However, the AUC metric does not tell the whole story, as discussed below.
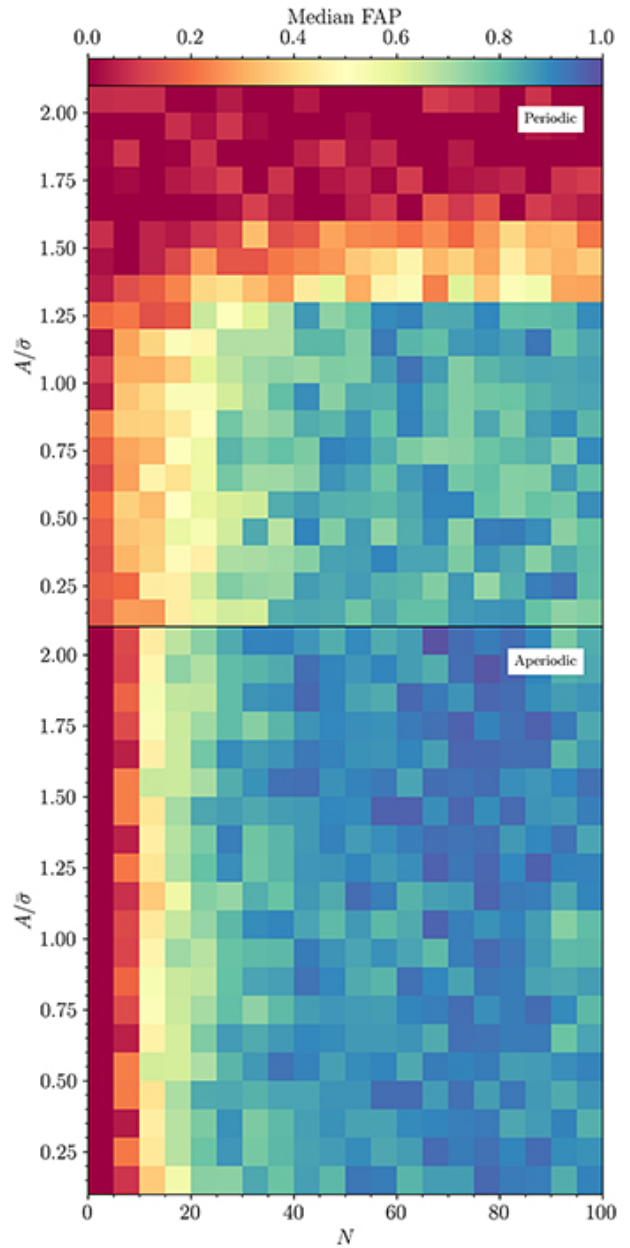
Figure 6. Showing the FAP calculated for synthetic light curves as a function of the number of data points 'N' in the x-axis and $A/\bar{\sigma}$ in the y-axis.

Fig. 6 shows the median NN FAP as a function of N and $A/\bar{\sigma}$. This was calculated for 80 000 synthetic sinusoidal periodic light curves (i.e. generated with equation 4). The calculations were performed with a range of $3 < N < 100$ and $0.1 < A/\bar{\sigma} < 2.1$. We can see that this method appears reliable provided $A/\bar{\sigma} > 1.5$ and $N > 40$. We note that a median value does not reveal occasional failures and we suggest a limit of $N > 50$ for greater reliability, based on the results in Section 4.2 and Fig. 8. A small amplitude with respect to the uncertainty is likely to give false negatives whereas a small number of measurements is likely to give false positives.

We also randomly selected 1000 eclipsing binary stars from the VIVACE catalogue (Molnar et al. 2022). This catalogue was generated from the same VVV data that this model was trained on. All light curves were independently verified as eclipsing binary for
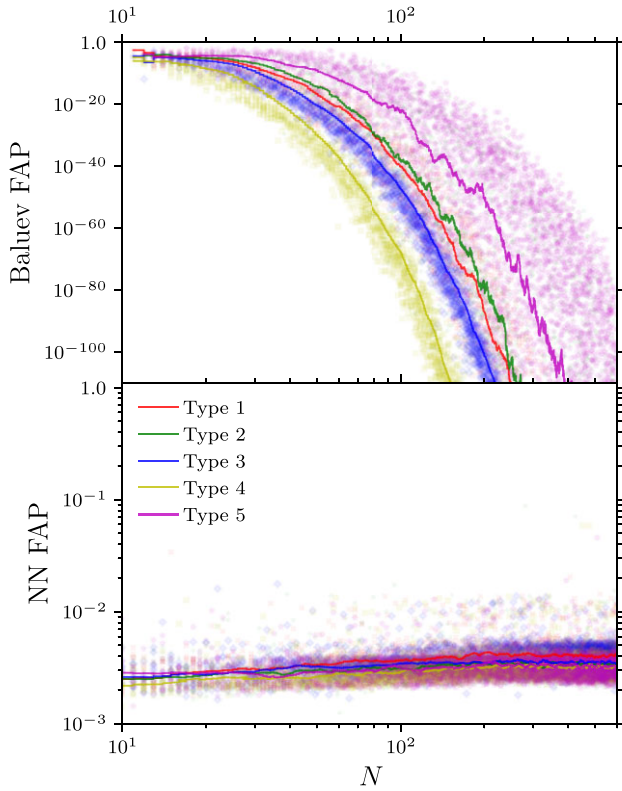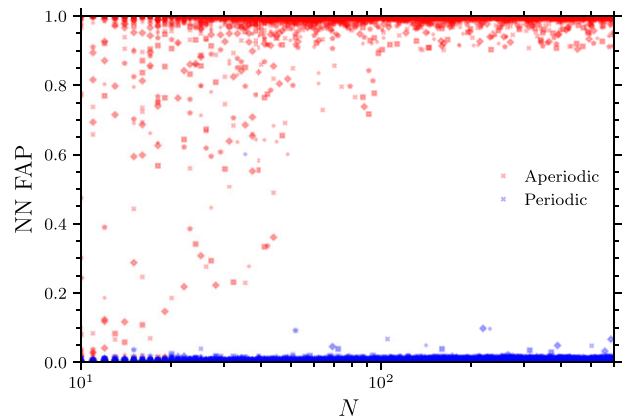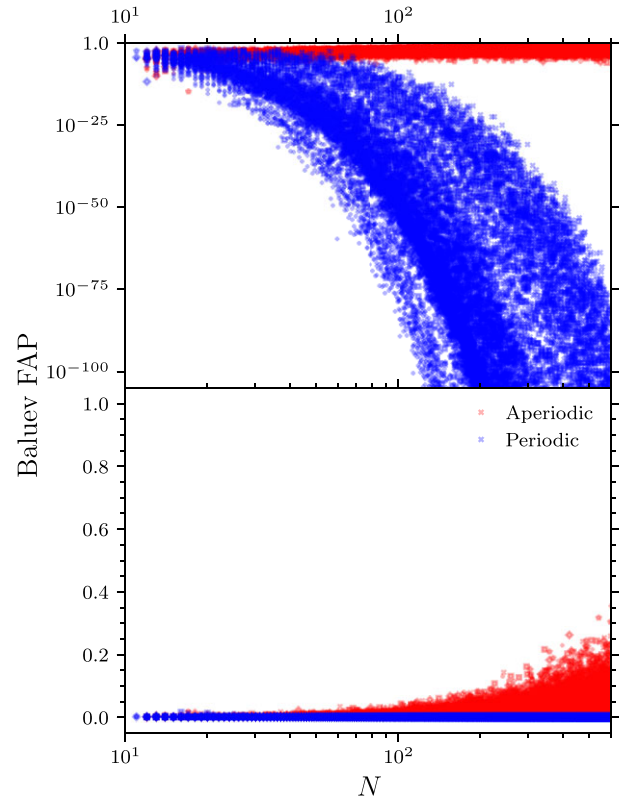
Figure 7. Both FAPs versus the number of measurements in the synthetic periodic light curve described in Section 3.4. The colours and markers correlate to those shown in Fig. 3. Top panel: Baluev FAP versus *N*. Bottom panel: NN FAP versus *N*.

this test. We construct a periodogram with both Lomb–Scargle and PDM and choose whichever period produced the lowest NN FAP. We find that 997 of the 1000 were identified as periodic with a FAP < 0.1. The three light curves which failed to be identified each featured an FAP > 0.6. In each of these three light curves there was a significantly shorter transit time paired with $N < 60$ measurements. The identification of the correct periodicity can be an issue when a light curve can look periodic when phase folded at multiple different periods. If an eclipsing binary features a similar size and shape for each eclipse then the NN FAP can erroneously be assigned half the true period as the two dips in the light curve are likely to be undifferentiable in the phase fold. This can be problematic for equal mass eclipsing binary systems.

4.2 FAP versus *N*

The number of measurements used to constitute a light curve can vary by orders of magnitude dependent on the survey. Surveys such as *Kepler* and *TESS* feature highly sampled light curves which should not pose an issue to any FAP technique. However, this is not always the case and many surveys feature light curves with fewer than 100 measurements. Fig. 7 shows how the Baluev FAP and the NN FAP vary as a function of the number of measurements '*N*' for the synthetic light curve described in Section 3.4. The NN FAP does not produce any significant number of false negatives as the number of measurements decreases to 10. The Baluev FAP has a clear trend as a function of *N* and starts to increase to a problematic range of values as *N* approaches ~50 measurements. It can also be seen that the Baluev FAP has a dependency on the shape of the light curve with

Figure 8. Both FAPs versus the number of measurements in the synthetic light curves described in Section 3.4. The red points show the FAP assigned to aperiodic light curves and the blue shows periodic light curves. It can be seen that the Baluev FAP is more likely to assign false negatives whereas the NN FAP is more likely to assign false positives. The Baluev FAP rarely exceeds 0.1 and never approaches unity. Top panel: Baluev FAP versus *N*. Bottom panel: NN FAP versus *N*. Each light curve here featured $A/\bar{\sigma} = 10$. The marker shapes correspond to those shown in Fig. 3 (i.e. a cross represents 'Type 1' and a plus 'Type 2'...).

more sinusoidal light curves assigned a lower FAP compared with more complex light curves such as eclipsing binary shapes. This is an issue as it can lead to incorrect conclusions on the demographics of variable stars.

Again using the synthetic light curves described in Section 3.4, in Fig. 8 (bottom panel) we show that the NN FAP sometimes falls to low values for aperiodic light curves as $N < 50$, potentially leading

to false positive classifications. These false positives arise as any variable light curve with a small number of points will more easily represent a periodic light curve at a given phase fold. Caution should be taken with this method when searching for periodic variables with fewer than 50 measurements. By contrast, the Baluev FAP does not suffer from this problem but Fig. 8 (top panel) shows that it is more likely to assign false negatives to periodic light curves within the same range.

It is not possible to define a threshold for either method which we can use to perfectly separate the periodic and aperiodic light curves. Such a threshold must be set by the user depending on preference regarding completeness and purity. The periodic light curves shown in Fig. 8 have a maximum NN FAP of 0.791 but the minimum NN FAP for aperiodic light curves is 0.01. The Baluev FAP has a maximum value of 0.015 for periodic light curves but a minimum value of 1.197×10^{-15} for the aperiodic light curves. The Baluev FAP values for periodic and aperiodic light curves overlap despite never approaching 1. The median Baluev FAP for the aperiodic light curves when $N \leq 100$ is 0.0012 and when $N \leq 50$ it is 0.0005. Using the widely adopted criterion for the Baluev FAP of $\log_{10}(FAP) < -2$ (Herbst et al. 2000; Koeltzsch et al. 2009; Chen et al. 2020a; Botan et al. 2021) yields misidentification of only four of the synthetic aperiodic light curves plotted in Fig. 8 as periodic, while incorrectly categorising 13 849 (46.4 per cent) aperiodic stars as periodic. This indicates that a lower threshold is more suitable for our synthetic light curves. In Molnar et al. (2022), a Baluev FAP selection of $\log_{10}(FAP) < -10$ was used to define a reliable but incomplete set of VVV light curves for training. If we were to use that cut for these data we would misidentify 1152 (3.86 per cent) periodic light curves as aperiodic and 502 (1.68 per cent) aperiodic light curves as periodic. If we use an NN FAP of 0.15 we misidentify 1 periodic light curve as aperiodic and 400 (1.34 per cent) aperiodic light curves as periodic.

False positives will arise, or not, depending on the FAP threshold value that is adopted. The NN FAP method performs very well in the AUC test for Variable N (see Table 1) because, even where aperiodic light curves have a low FAP, the periodic light curves have even lower FAP values. This allows the binary classifier to be successful, in principle, if the threshold FAP could be ideally selected. However, in practice, this will rarely be possible.

4.3 FAP versus amplitude

The signal-to-noise of a light curve is a common source of erroneous periodicity classification. Periodic variable stars can host a range of amplitudes depending on the source of variability. As such, it is not uncommon to investigate variable stars whose variability is similar to, or below, the photometric uncertainty.

It can be seen in Fig. 9 that both the NN and Baluev FAP feature a dependency on $A/\bar{\sigma}$. Both the NN FAP and the Baluev FAP suffer from false negative rates as $A/\bar{\sigma} \rightarrow 1.5$. Again, it can be seen that the NN FAP does not suffer from a structure-dependent FAP, unlike the Baluev FAP. This is not surprising as the Lomb–Scargle method, to which the Baluev FAP is applied, is effectively a sinusoidal fitting method and hence will feature such structure-based dependencies. In Fig. 10 it is shown that neither of the methods appear to suffer from false positives as $A/\bar{\sigma} \rightarrow 1.5$.

The median Baluev FAP for the periodic light curves when $A/\bar{\sigma} \leq 1.25$ is 0.011 and 0.020 for aperiodic sources. The NN FAP at the same $A/\bar{\sigma}$ has a median value of 0.959 for periodic sources and 0.998 for aperiodic sources. Both methods feature a significant level of confusion at such a low $A/\bar{\sigma}$ but they do so at different absolute values, the Baluev FAP rarely features values larger than 0.1. If we
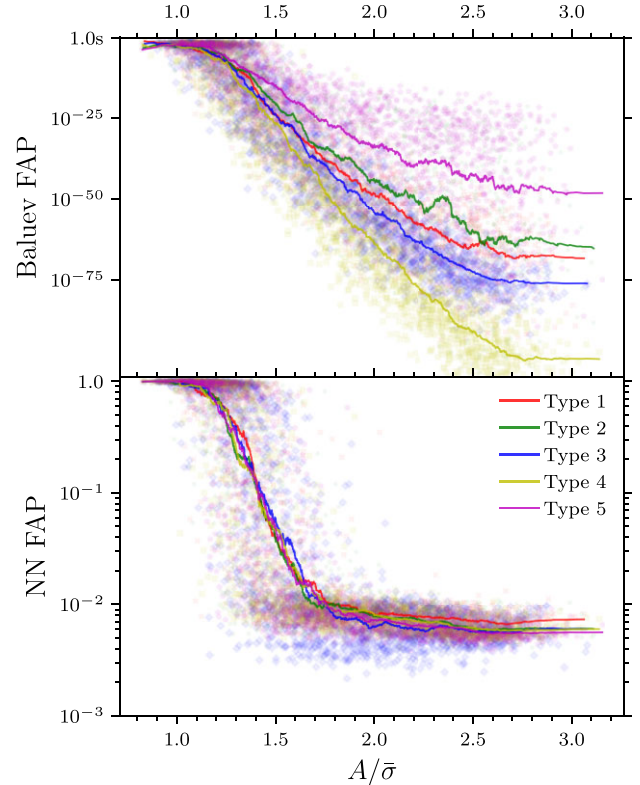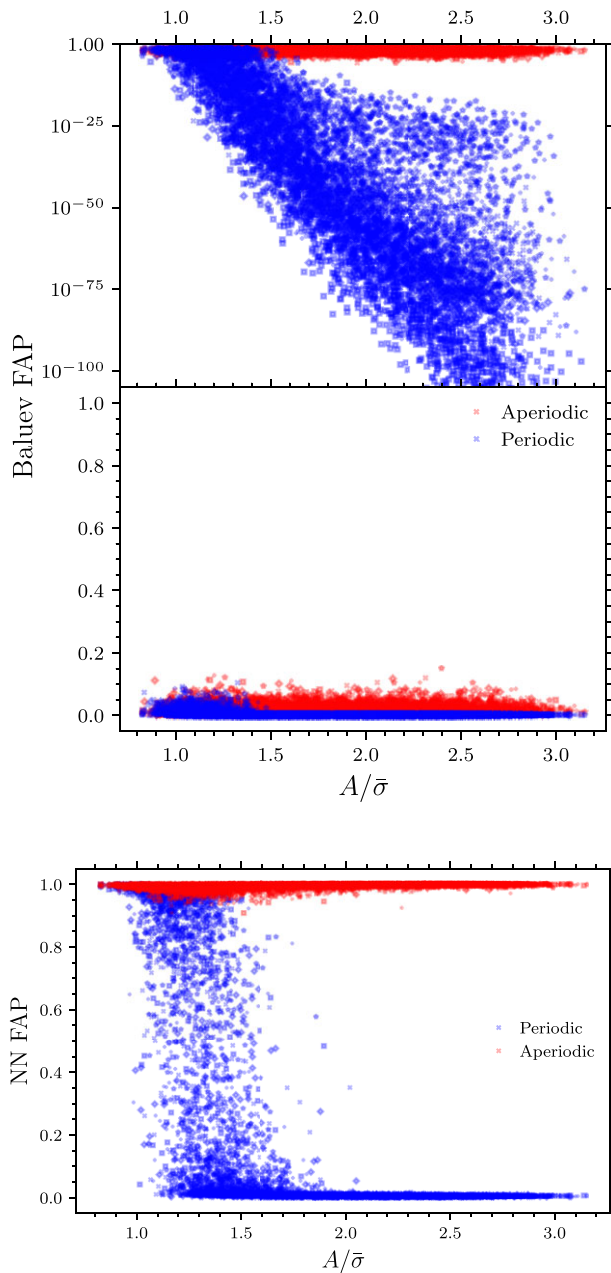

Figure 9. Both FAPs versus the amplitude of a synthetic periodic light curve divided by the median average of the photometric uncertainty. The colours and markers are the same as that shown in Fig. 3. Top panel: Baluev FAP versus $A/\bar{\sigma}$. Bottom panel: NN FAP versus $A/\bar{\sigma}$.

use the same value of $\log_{10}(FAP) < -10$ from Molnar et al. (2022) for the Baluev FAP we misidentify 2028 (3.86 per cent) periodic light curves as aperiodic. If we use a NN FAP of 0.15 we misidentify 1996 (3.79 per cent) periodic light curve as aperiodic. Neither method misidentifies any aperiodic sources as periodic sources with this selection.

5. TESTING WITH OTHER SURVEYS

Our proposed method of calculating an FAP is universal and independent of the method of period detection. We can also show that the NN FAP method can be applied to data that are not drawn from the same distribution as the training data. Fig. 11 shows periodic and aperiodic variable stars in the CRTS (Drake et al. 2009). Fig. 12 shows them for ZTF. Fig. 13 shows them for *Kepler*. Fig. 14 shows periodic variables in OGLE (Udalski, Szymański & Szymański 2015) data. Each of the subplots in these figures show the assigned NN FAP.

The top-left panel of Fig. 14 shows the light curves in two filters for the source 'OGLE-BLG-ECL-124368' which appears much more clearly periodic in 'I' than 'V'. The NN FAP reflects this, showing a higher FAP for the 'V' band data. The 'V' band data does still show a poorly sampled transit at the phase-folded period, hence the NN FAP is above 0.5 but below 0.9. The model used to identify these variables was trained as described in Section 3 with VVV light curves. The periodicity of each of these stars was identified by choosing the period which produced the lowest NN FAP extracted from a PDM periodogram. For both the CRTS and ZTF light curves the Baluev FAP was sufficient for differentiating between aperiodic and periodic variable stars. For three of the *Kepler* light curves the Lomb–Scargle

Figure 10. Both FAPs versus the amplitude of synthetic periodic and aperiodic light curves divided by their median average of the photometric uncertainty. Both methods show how their reliability begins to fail at $A/\bar{\sigma} \approx 1.5$. Top panel: Baluev FAP versus $A/\bar{\sigma}$. Bottom panel: NN FAP versus $A/\bar{\sigma}$. 200 data points were used for these light curves. The marker shapes correspond to those shown in Fig. 3.

periodogram incorrectly assigned half of the period with a low Baluev FAP. One of the *Kepler* light curves was not identified as periodic by the Baluev FAP (bottom-left panel of periodic variables in Fig. 13). Only one of the OGLE light curves was correctly identified as periodic in both 'V' and 'I' by the Baluev FAP (top-right panel in Fig. 14) although with a notably different FAP of 9×10^{-60} in 'V' and 1×10^{-235} in 'I'. Each of the other OGLE light curves were either incorrectly given half of the true period or given a Baluev FAP indicative of aperiodic variability. The Lomb–Scargle periodogram
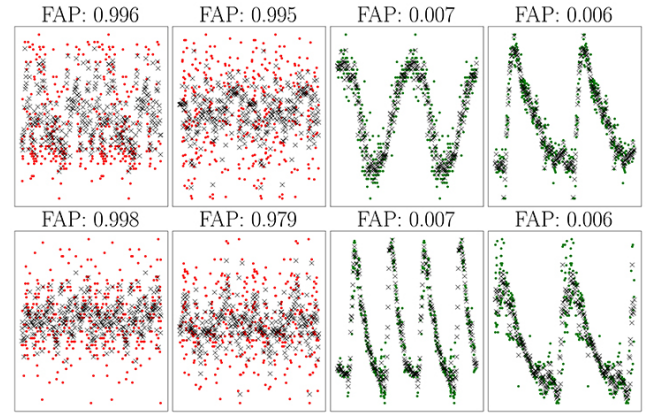
Figure 11. Randomly selected examples of identified aperiodic (red, left) and periodic (green, right) variable stars found in the CRTS survey. Each subplot displays the assigned NN FAP as its title. The green and red points represent the raw magnitude as a function of phase for the periodic and aperiodic light curves, respectively. The black points represent the KNN interpolated fit to the raw light curve. The Baluev FAP for each of the aperiodic sources was above 2×10^{-5} and the periodic sources were all below 1×10^{-60}.
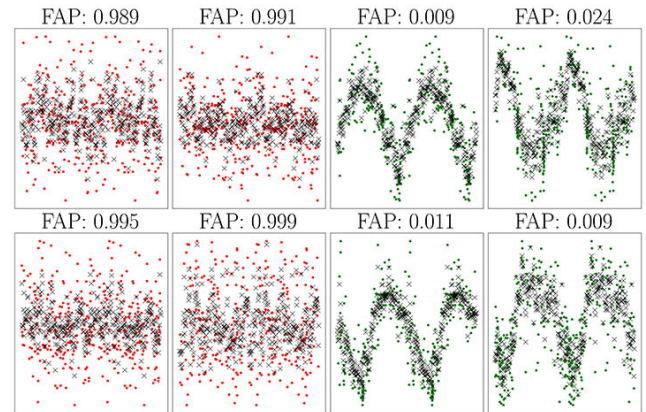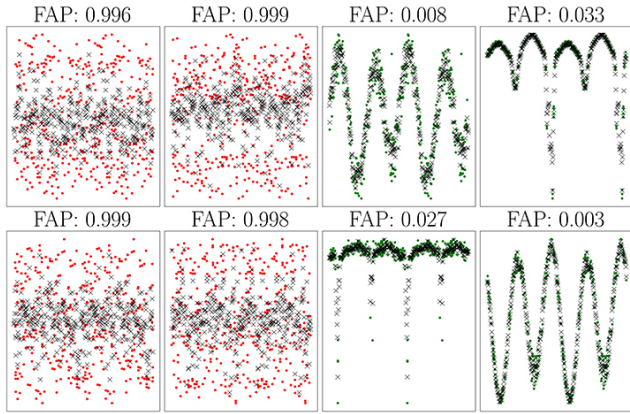

Figure 12. A random sample of identified aperiodic (red, left) and periodic (green, right) variable stars found in the ZTF survey. With each subplot showing the assigned NN FAP as its title. The Baluev FAP for each of the aperiodic sources was above 2×10^{-14} and the periodic sources were all below 1×10^{-51}.

also correctly identified the 'I' band period of the bottom-left panel with a Baluev FAP of 9.51×10^{-141} but failed to extract the correct period for 'V' band. The Lomb–Scargle periodogram and Baluev FAP predominantly struggled with more complex eclipsing binary-shaped light curves.

6. FAP PERIODOGRAM

The NN FAP method presented above can be seen as something analogous to a neural network version of the PDM method so we can try to use it as such, i.e. for the construction of a periodogram rather than FAP calculation. We can calculate an FAP for a set of trial periods and the period which returns the lowest FAP should be the correct period. This has the added benefit of generating a periodogram on a universal scale and thus the FAP is given along with the periodogram. Currently, this approach is limited by its computationally demanding nature. Future developments in computing paired with this method being modified for periodogram

Figure 13. A random sample of identified aperiodic (red, left) and periodic (green, right) variable stars found in the *Kepler* survey. Each subplot shows the assigned NN FAP as its title. The Baluev FAP for each of the aperiodic sources was above 2×10^{-4}. The bottom left periodic variable has a Baluev FAP of 2.746×10^{-6}. The other periodic sources were all below 1×10^{-43} but each had an incorrect period of half the true period.
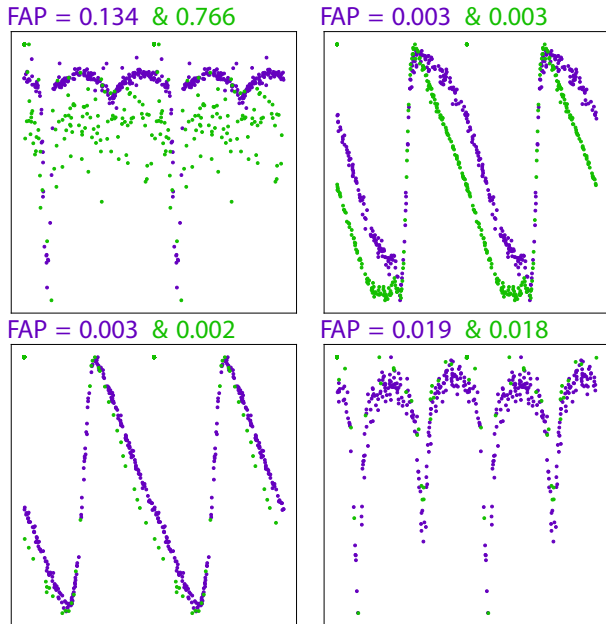
Figure 14. Periodic variables from the OGLE selection of variable stars. The green points represent the light curve of the star in the 'V' filter and the purple represent the 'I' filter. Three of the stars are identified as periodic in both V and I filters. The top-left panel (OGLE-BLG-ECL-124368) was not identified as clearly periodic at any period in 'V' and a higher NN FAP was given (although below that found for the aperiodics in Figs 11–13). From top-left to bottom-right panel, the Baluev FAPS are 0.967 for 'V' and 0.9 for the incorrect period in 'I', 9×10^{-60} in 'V', and 1×10^{-235} in 'I', 0.016 for the incorrect period in 'V' and 9.51×10^{-141} in 'I', 0.160 in 'V', and 6.9×10^{-65} for the incorrect period in 'I'.

construction purposes will make this work more practical. Fig. 15 shows the periodogram constructed for a synthetic light curve (of Type 5, equation 2) with 200 points, an SNR of 2 and a period of 296.4 d. This periodogram took 23 min to construct and correctly extracted the correct period (inference was run on 64 CPU cores). This compares with the 0.2 s it took for the PDM method to construct the same periodogram and achieve the same results (without an FAP).
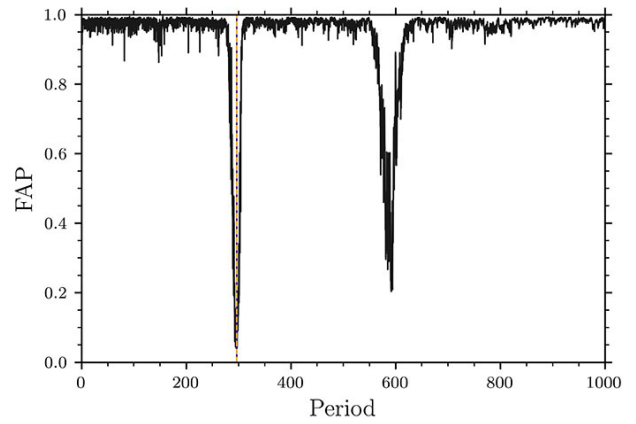
Figure 15. A periodogram constructed from the NN FAP method. The periodogram took 23 min on 64 cores to compute. The correct minima is identified despite the binary-like construction of the periodogram.

Both the NN periodogram and the PDM periodogram suffered from aliasing at multiples of the true period but both also correctly assigned the true period the largest peak value.

7. CONCLUSIONS

We have shown that utilizing the flexibility afforded by neural networks allows a more robust analysis of light curves. Using synthetic and real data, RNNs can be trained to produce a reliable and universal measure of periodicity. A study of the parameter space (namely the signal-to-noise ratio and temporal density) demonstrated how and when this method fails in comparison with the commonly used Baluev method. This method remains reliable where $N > 50$ with $A/\bar{\sigma} > 10$ or $A/\bar{\sigma} > 1.5$ with $N \geq 200$. As we analyse the phase-folded light curve and not the periodogram, the NN FAP is independent of the tools used to construct the periodogram. This method is more analogous to a universally scaled PDM and so the network is effectively analysing the structure of the phase-folded light curve. This has further implications for a possible method of period detection that were explored in Section 6.

Fig. 5 and Table 1 have shown how this method outperforms the Baluev method for both synthetic and real data. Given a data set for candidate periodic variable stars, this method will provide a more complete search for periodicity, at the expense of occasionally generating more false positives for small N.

We highlight that the most challenging aspect of this method is the data preparation which is outlined in Section 3. Care must be given to how the training data are constructed and prepared. This method is provided both with the ability to retrain on different data sets as well as pre-trained with the data described above. We expect the method to be fully functional in its pre-trained state within the parameters outlined in this paper. Conversely, this is not the case for the network's architecture which was shown by ablative testing to be relatively inconsequential to the performance.

DATA AVAILABILITY

The code used in this project is available at github.com/nialljmiller/NN_FAP. The model and weights are hosted at www.kaggle.com/datasets/niallmiller/nn-fap. The model, weights, and code can also be found at https://www.nialljmiller.com/projects/FAP/FAP.html. The

training data used will be available upon any reasonable request to the first author.

REFERENCES

Baluev R. V., 2008, MNRAS, 385, 1279
Baluev R. V., 2009, MNRAS, 395, 1541
Bellm E. C. et al., 2019, PASP, 131, 018002
Borucki W. J. et al., 2003, in Fridlund M., Henning T., Lacoste H., eds, ESA Special Publication, Vol. 539, Towards Other Earths: DARWIN/TPF and the Search for Extrasolar Terrestrial Planets. ESA Publications Division, Noordwijk, p. 69
Botan E., Saito R. K., Minniti D., Kanaan A., Contreras Ramos R., Ferreira T. S., Gramajo L. V., Navarro M. G., 2021, MNRAS, 504, 654
Burhanudin U. F. et al., 2021, MNRAS, 505, 4345
Chen T., Kornblith S., Norouzi M., Hinton G., 2020, preprint (arXiv:2002.05709)
Chen X., Wang S., Deng L., de Grijs R., Yang M., Tian H., 2020b, ApJS, 249, 18
Cho K., van Merrienboer B., Gulcehre C., Bahdanau D., Bougares F., Schwenk H., Bengio Y., 2014, preprint (arXiv:1406.1078)
Chollet F. et al., 2015, Keras. GitHub. Available at: https://github.com/fchollet/keras
Cincotta P. M., Mendez M., Nunez J. A., 1995, ApJ, 449, 231
Drake A. J. et al., 2009, ApJ, 696, 870
Fix E., Hodges J. L., 1951, USAF School of Aviation Medicine, 1-2
Gaia Collaboration, 2021, A&A, 649, A1
Graham M. J., Drake A. J., Djorgovski S. G., Mahabal A. A., Donalek C., Duan V., Maker A., 2013, MNRAS, 434, 3423
Herbst W., Rhode K. L., Hillenbrand L. A., Curran G., 2000, AJ, 119, 261
Hochreiter S., Schmidhuber J., 1997, Neural Comput., 9, 1735
Huijse P., Estevez P. A., Protopapas P., Zegers P., Principe J. C., 2012, IEEE Trans. Signal Process., 60, 5135
Irwin J., Aigrain S., Bouvier J., Hebb L., Hodgkin S., Irwin M., Moraux E., 2009, MNRAS, 392, 1456
Ivezić Ž. et al., 2019, ApJ, 873, 111
Kingma D. P., Ba J., 2014, preprint (arXiv:1412.6980)
Kirk B. et al., 2016, AJ, 151, 68
Koeltzsch A. et al., 2009, Astron. Nachr., 330, 482
Lomb N. R., 1976, Astrophys. Space Sci., 39, 447
Mainzer A. et al., 2014, ApJ, 792, 30
Minniti D. et al., 2010, New Astron., 15, 433
Molnar T. A., Sanders J. L., Smith L. C., Belokurov V., Lucas P., Minniti D., 2022, MNRAS, 509, 2566
Neumann J. V., 1941, Ann. Math. Stat., 12, 367
Park W. et al., 2021, ApJ, 920, 132
Pawlak M. et al., 2019, MNRAS, 487, 5932
Pedregosa F. et al., 2011, J. Mach. Learn. Res., 12, 2825
Plavchan P., Jura M., Kirkpatrick J. D., Cutri R. M., Gallagher S. C., 2008, ApJS, 175, 191
Ricker G. R. et al., 2015, J. Astron. Tel. Instr. Syst., 1, 014003
Saito R. K. et al., 2012, A&A, 537, A107
Scargle J. D., 1982, ApJ, 263, 835
Schechter P. L., Mateo M., Saha A., 1993, PASP, 105, 1342
Smith L. C. et al., 2018, MNRAS, 474, 1826
Sokolovsky K. V. et al., 2017, MNRAS, 464, 274
Stellingwerf R. F., 1978, ApJ, 224, 953
Stetson P. B., 1996, PASP, 108, 851
Templeton M. R., Mattei J. A., Willson L. A., 2005, AJ, 130, 776
Udalski A., Szymański M. K., Szymański G., 2015, Acta Astron., 65, 1
Wright E. L. et al., 2010, AJ, 140, 1868
Zechmeister M., Kürster M., 2009, A&A, 496, 577
Zhang R., Zou Q., 2018, J. Phys.: Conf. Ser., 1061, 012012

SUPPORTING INFORMATION

Supplementary data are available at *RASTAI* online.

FAP_Paper(3).zip

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

This paper has been typeset from a TEX/LATEX file prepared by the author.