

The Classification of Periodic Light Curves from non-survey optimized observational data through Automated Extraction of Phase-based Visual Features

Paul R. McWhirter

Computing and Mathematical Sciences
Liverpool John Moores University
Liverpool, UK.
P.R.McWhirter@2014.ljmu.ac.uk

Iain A. Steele

Astrophysics Research Institute
Liverpool John Moores University
Liverpool, UK.
I.A.Steele@ljmu.ac.uk

Dhiya Al-Jumeily

Computing and Mathematical Sciences
Liverpool John Moores University
Liverpool, UK.
D.Aljumeily@ljmu.ac.uk

Abir Hussain

Computing and Mathematical Sciences
Liverpool John Moores University
Liverpool, UK.
A.Hussain@ljmu.ac.uk

Marley M. B. R. Vellasco

Dept. of Electrical Engineering
Pontifical Catholic University of Rio de Janeiro
Rio de Janeiro, RJ - Brazil
marley@ele.puc-rio.br

Abstract— We present Random Forest, Support Vector Machine and Feedforward Neural Network models to classify 2519 variable star light curves. These light curves are generated from a reduction of non-survey optimized observational images gathered by wide-field cameras mounted on the Liverpool Telescope. We extract 16 features found to be highly informative in previous studies and achieve an area under the curve of 0.8495 using a feedforward neural network with 50 hidden neurons trained with stratified 10-fold cross-validation with 3 repeats. We propose using an automated visual feature extraction technique by transforming bin-averaged phase-folded light curves into image based representations. This eliminates much of the noise and the missing phase data, due to sampling defects, should have a less destructive effect on these shape features as they still remain at least partially present. There is also no need for feature engineering as the learning algorithms can learn shape features directly from the light curves. We produced a set of scaled images based on a threshold of data points in each pixel. Training on the same feedforward network, we achieve an area under the curve of 0.6348. By introducing the Period and Amplitude as features into this dataset therefore giving meaning to the dimensions of the image we show this improves to 0.7952. Our current models lack translational-invariance and the method may be better suited to specific sub-classification problems common in the variable object hierarchical multi-class problem.

Keywords—Data analysis; Machine Learning; Light Curve Classification; Variable Stars; Visual feature extraction

I. INTRODUCTION

Time-domain Astronomy is an active field of discovery driven by recent technological advancements in observation, storage and data processing. Recent years have allowed for extended sky surveys such as the Sloan Digital Sky Survey (SDSS) [1] and the Gaia satellite which is mapping billions of stars allowing their fundamental parameters to be determined [2]. In the next few years this capability will grow through the deployment of even more powerful surveys such as the Large Synoptic Survey Telescope (LSST) generating approximately 20 TB of raw data every night [3]. These surveys are capable of regularly gathering data on wide regions of the sky. The regularity of this sampling is defined by the survey cadence. Each survey is optimized to a given cadence describing the approximate duration between observations of the same area of sky.

Despite the immense timescales involved in many Astrophysical processes, a number of variable phenomena occur within more human-comprehensible time scales. Few of these events are as well-studied as variable stars. These stars are at a volatile stage in their evolution resulting in perceived brightness fluctuations due to physical processes in their atmospheres. These stars allow for the study of stellar evolution and galactic structure [4]. Additionally, certain types of variable stars such as RR Lyraes or Delta Cepheids exhibit

specific luminosity-period relationships that allow the determination of their distances from Earth [5, 6]. Other important variable light sources are eclipsing binary systems. These systems exhibit periodic brightness changes from binary stars eclipsing each other due to the orbital plane of the system having a low inclination relative to the Earth [7].

It is of great importance to reliably identify and monitor these objects and large wide-field sky surveys are an ideal method of accomplishing this task. The quantity of data from these surveys makes this a daunting exercise. Fortunately, the field of machine learning has provided techniques that can be developed for the automated classification of light sources. There have been a number of studies investigating the production of both general purpose and more specific learned classification models through the extraction of useful features from the data of known variable objects.

Debosccher et al. developed a method of fitting harmonic models to light curves using multiple periods identified by a Lomb-Scargle Periodogram. These models were used to extract Fourier-based features for the production of learned classification models [8]. These features were extended into a set of 30 plus descriptive properties for light curves by adding non-periodic features by Richards et al [9]. These features were then processed into 16 highly informative features for general purpose variable star classification in the form of the Upsilon software package by Kim et al. [10]. Nun et al. have collated these feature extraction methods into the Feature Analysis for Time Series package [11]. Pichara et al. have also proposed using meta-classification allowing the use of multiple high-performance, specific classification models named experts in general purpose classification tasks [12]. Puegert et al. have also proposed the extraction of shape-defining features from phase-folded light curves through the coefficients of fitted chains of polynomial models. These features were then used to classify eclipsing binary light curves using learned models built using a feedforward neural network [7].

These studies focused on datasets comprised of well-sampled fixed-cadence light curves. How would these methods perform on a different style of light curves derived from wide-field observations without a guaranteed cadence? These observations can have a significantly uneven distribution in time for individual light sources [13]. An example of one of these datasets is the observations produced by the Small Telescopes Installed at the Liverpool Telescope (STILT) [14]. The cameras are mounted to the frame of the Liverpool Telescope aiming co-parallel with the main telescope’s field of view capturing a ten second exposure every minute whilst the telescope is in operation. They have no control over the position of the telescope and therefore no ability to influence their observational cadence. These observations from March 2009 to March 2012 have been processed into a dataset containing over 27 million individual light sources.

In this paper we introduce the initial results and problems from the application of the methods from these previous studies to the STILT observations and propose a method of automatically extracting shape-based features from the phase-folded light curves through the use of multiple learning algorithms trained to recognize visual features mirroring the

methodology employed manually by astronomers. In the future additional topologies will be introduced to further power this feature extraction and allow for the classification of super and sub-classes in a large hierarchical multiclass problem.

The rest of this paper is structured as follows. In Section 2, a selection of variable star light curves, generated from the STILT observations through coordinate comparison with the American Association of Variable Star Observers (AAVSO) variable star index catalogue is presented. Our method of analysing this light curve dataset is also introduced in this section. Section 3 presents the results of classifiers trained using a number of algorithms on features from previous studies as well as from our newly proposed. In Section 4 the study is summarised and we propose the direction of future work.

II. DATASET AND METHOD

The Small Telescopes Installed at the Liverpool Telescope (STILT) dataset is a wide field object SQL database. It contains 1.24 billion separate object observations of 27.74 million independent stellar objects. It was generated through the pre-processing of observational images gathered by the STILT instruments [14]. This database contains light curves for many objects, including many of unknown classification. Reliable class information is required for a subset of objects in the database in order to test classification methods on these light curves. The optimal method to extract this class information is through a comparison between the STILT data to a variable star catalogue. The American Association of Variable Star Observers (AAVSO) operates one of the largest and best-updated catalogues of nearby bright variable stars in the world, The AAVSO International Variable Star Index. This catalogue does not contain any of the AAVSO gathered light curves but it does contain data on 373,565 known variable stars including their name, coordinates in right ascension and declination and their currently identified class. The coordinates of these variable stars were matched to objects in the STILT database with a tolerance of 3.6 arc seconds (seemingly sufficient to avoid detection collisions between nearby stars) and a minimum of 100 individual observations. This resulted in the production of 12461 variable stars of various types. Five variable star super-classes were selected which describe a large number of known periodic variable object types. They were all well-represented in this dataset leaving 2519 corresponding objects. Table 1 demonstrates the class by class breakdown of this dataset.

TABLE I
2519 OBJECT DATASET

Class	Dataset Statistics		
	Type	Acronym	Count
1	Delta Cepheid Variables	DCEP	132
2	Delta Scuti Variables	DSCT	499
3	Eclipsing Binaries	EB	1409
4	Long Period Variables	LPV	365
5	RR Lyrae Variables	RR	114

Following the selection of these 2519 variable light curves, the performance of the features used in previous high-

performance general purpose classifiers was established. The 16 features used by Kim et al. [10] were chosen for this operation as they had been shown to be capable of reliably separating super-classes as well as achieving respectable inter-class accuracy. These features are shown in Table 2.

TABLE II
KIM ET AL. 16 VARIABILITY FEATURES

Feature	Description	Reference
Period	Period derived by the Lomb-Scargle Periodogram	Kim et al. 2014
ψ^η	η of a phase-folded light curve	Kim et al. 2014
ψ^{CS}	Cumulative sum index of a phase-folded light curve	Kim et al. 2014
R_{21}	2 nd to 1 st amplitude ratio	Kim et al. 2014
R_{31}	3 rd to 1 st amplitude ratio	Kim et al. 2014
Φ_{21}	Difference between 2 nd and 1 st phase	Kim et al. 2014
Φ_{31}	Difference between 3 rd and 1 st phase	Kim et al. 2014
γ_1	Skewness	Kim et al. 2014
γ_2	Kurtosis	Kim et al. 2014
K	Stetson K index	Kim et al. 2014
Q_{3-1}	Difference between 3 rd and 1 st quartiles	Kim et al. 2014
A	A ratio of magnitudes brighter or fainter than the average	Kim et al. 2016
H_1	Amplitude from Fourier decomposition	Kim et al. 2016
W	Shapiro-Wilk normality test	Kim et al. 2016
m_{p10}	10 th percentile of slopes of a phase-folded light curve	Long et al. 2012
m_{p90}	90 th percentile of slopes of a phase-folded light curve	Long et al. 2012

The Lomb-Scargle Periodogram [15] utilised in this method operated over a linear frequency grid from the reciprocal of the total observation time of a light curve up to 20 cycles per day. The interval between candidate frequencies is shown in equation 1 where t_{max} and t_{min} are the last and first observation times respectfully.

$$f_{step} = \frac{0.25}{t_{max} - t_{min}} \quad (1)$$

When calculated using the Lomb-Scargle Periodogram using light curves from the STILT dataset, the Period feature appears to have a correct match rate of under 5% relative to the AAVSO reference period (which is treated as the ground truth in this study) for many of the classes. As the period, calculated by the Lomb-Scargle Periodogram, is the basis in which phase-folded light curves are generated, this inaccuracy heavily pollutes an additional 9 features. This is over half the number of features used in this analysis. As for the non-folded features, the distribution of these features amongst the classes appears to centre at or near their expected means. However, the range is much greater than expected increasing the overlap between classes. This is likely a result of the larger-than-usual noise threshold in the STILT data [13]. Classifiers trained using these polluted features resulted in models of accuracies only

slightly better than the no-information accuracy, the expected result of a completely randomly trained model. This is primarily due to the Long Period Variable class as it exhibits long period sinusoidal variations with a high amplitude signal. Therefore, for the following analysis, the periodogram-derived period was replaced with the AAVSO reference period purifying the features shown in figure 1.

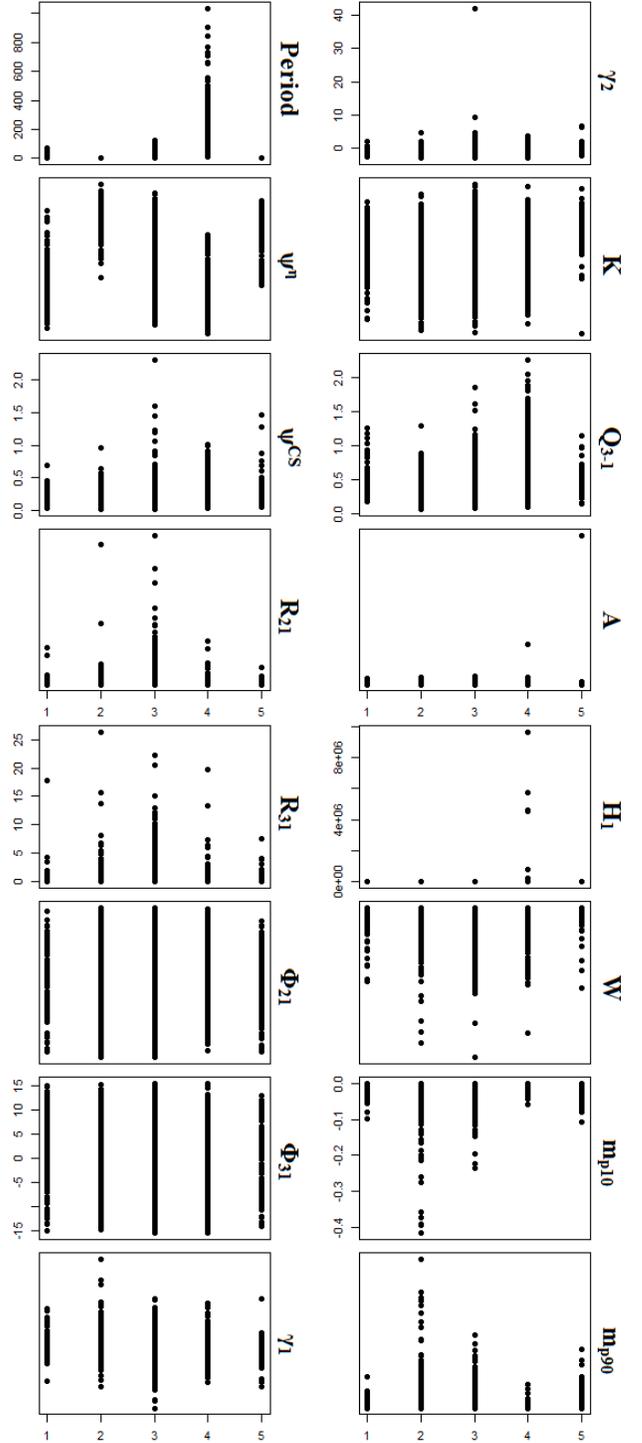


Fig. 1. Plot of each of the 16 features against the five super-classes in the order shown in table 1. Many features appear to poorly differentiate the super-classes with light curves from the STILT dataset.

These 16 features have been shown to be fully capable of training useful classifiers in previous work however they required the fitting of Fourier models to the light curves. These Fourier models are used to generate features such as the amplitude ratios, the phase information and the amplitude of the first harmonic. In the method of Kim et al., utilised in this work, a five harmonic Fourier model has been fit to the light curves [10]. These models are very versatile but can suffer from three serious drawbacks on our dataset. Firstly, the sums of sinusoids used to assemble Fourier models fit some classes of astrophysical signal poorly such as the sharp dips associated with eclipsing binary stars [9]. Secondly, the more harmonics used in a Fourier model, the more complex the signal it can fit. Unfortunately this can also result in overfitting the light curve data causing the noise to have an unwanted contribution to any resulting features. Finally, as these Fourier models are being fit in the time dimension, the poorly sampled regions can cause the fit model to deviate outside of expected ranges again resulting in a non-signal contribution to the Fourier coefficients. The cadence concerns raised in the STILT dataset can cause this to become a considerable source of poor results.

It is important for the analysis method to address the above dangers as the extracted features are important. We propose, as the classes we are attempting to train models on are highly periodic, to transform the representation of the light curves into an epoch-folded representation and extract features. These features describe the shape of brightness changes through the dominant periodic variation by ‘folding’ all the gathered data points into one waveform. This is very useful in astronomy due to the limitations in gathering data. In fact, this is one of the most powerful techniques in eliminating sampling issues as long as the light curve does have a dominant period [7]. For non-periodic variable objects in astronomy, such as transient light sources, other approaches must be considered. In the case of many periodic variable objects, the shapes of the light curves in these phase-folded representations carry significant information about the class of the light source. Figure 2 shows an example of three of the light curves in this dataset, a Mira-type Long Period Variable, an Algol-type eclipsing binary and a Delta Cepheid. These light curves have been folded at the AAVSO period of the associated objects. Therefore, a light curve must clearly demonstrate these shape features in order for the 16 Kim et al. features (and many more) to extract enough of this information from any noise.

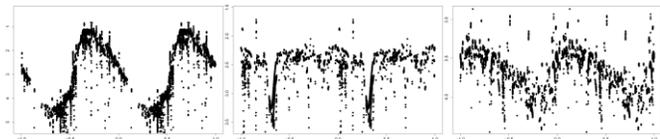


Fig. 2. STILT dataset folded Light curves of the star Mira (Mira class) with a 332 day period, Algol (Algol-type eclipsing binary) with a 2.86 day period and Eta Aquilae (Delta Cepheid class) with a period of 7.18 days. The shape of each light curve is distinctive to the associated class.

In this form, the poorly sampled regions due to cadence limitations in the original observations can be removed with the restriction of the dominant period must be a fraction of the complete observed time. The 16 features of the previous research does contain a number of features extracted from the phase-folded light curve however the Fourier model is not performed in this representation. We propose to replace the

shape features from the Fourier model with new shape features automatically extracted from the folded light curves through applying machine learning algorithms directly to a visualised image-based representation of this folded light curve. In essence allowing the learning of class-specific shapes.

The light curves in figure 2 are fairly typical of the better sampled light curves from the STILT database yet they do exhibit potential issues. Firstly, there are a lot of points with significant noise. This is possibly instrumental in nature but is much more likely due to a number of simplifications made to reduce the computational load of the pre-processing pipeline. This noise is likely to be the cause of both the larger range on the non-periodic features as well as the cause of the very low period match rate from the Lomb-Scargle Periodogram. Secondly, whilst the examples in figure 2 are well sampled across the whole phase space, there are other light curves that lack this due to the highly variable cadence of the STILT observations. This means that important shape features may only be partly present and not to the level required for the extracted features in previous studies.

Yet, despite these obvious limitations, human astronomers can still look at these light curves and recognise the main shape patterns. Therefore it seems reasonable to conclude that even in the more poorly sampled, noisy STILT light curves, there are still features that have not yet been extracted which are being gathered for manual classification. Ideally the models used to fit the light curves should attempt to parameterise the shape of the actual variable object classes rather than some predefined or abstract form. This can be done by determining the specific form of different astrophysical signals directly from the astrophysics driving the variability of these object types. This is quite an undertaking further complicated by a lack of consensus about the dominant physical processes shaping these variabilities in many classes. Therefore, it would seem to be more appropriate to have a model that can identify the patterns in the shape of the light curves without requiring an underlying physically produced model. This can be accomplished through a learning process applied to visualized examples of light curves. Over the last decade, neural networks have been developed into platforms for visual reasoning [16]. The ImageNet classification is a good example, a large dataset of images collected into 1000 classes. Respectable classification accuracy has been found through the use of deep networks with convolutional layers for visual feature extraction [16]. We attempt to replicate these visual feature extraction layers through the construction of hidden layers tuned to find visual features. As this is just the initial investigation, convolutional layers have not yet been utilised and this does result in limitations discussed in later sections.

We first phase-folded the light curves for each of the STILT dataset light sources. This task required a candidate period. As the Lomb-Scargle Periodogram is performing poorly on our data, we instead used the AAVSO period as we had when the 16 features used in Kim et al. were produced. This phase space exists from a phase of 0 to 1 with the brightest data point defined as 0.25. Outliers were also eliminated from the light curves by defining the brightness range of the folded light curve from the mean brightness as the amplitude of the light curve. This amplitude is defined as the

difference between the median of the maximum 5% of data points and the median of the minimum 5% of data points divided by two. In order to emphasise shapes present at the edge of the folded light curve (as shapes can be split as phase values over 1 loop around to 0), the folded light curve was duplicated operating over a new phase space of -1 to 1. In order to reduce the noise we applied a bin-averaging process to the folded light curve data. The phase-space was binned into 100 phase-bins each bin having a phase range of 0.02. All observed data points in each phase bin are mean averaged and retained. Empty phase bins are removed. The bin-averaged phase-folded light curves were then used to generate pixelated images. This has a number of important uses. First we can guarantee an identical number of inputs into our neural network regardless of the sampling of the light curve. Second, it can be minimized to a level which optimizes for computational cost. For this task we decided to transform the light curves into 100x20 pixel images giving 2000 input ‘feature’ pixels. Each light curve produced a magnitude-scaled image of the amplitude of the light curve centred on its weighted mean.

III. RESULTS

Previous studies found their features performed best when trained using the random forest algorithm. Therefore, in this study we make use of Random Forest models with a number of different parameters, a Feedforward Neural Network with a single hidden layer of appropriate size and a Support Vector Machine with a linear kernel as the radial basis function kernel was unable to extract usable information from the features resulting in all predictions being assigned to the dominant class in the dataset, the eclipsing binaries. All the results were obtained through training on a 3.4 GHz Intel Core i7-3770 processor with 16 GB of memory. RStudio was used as the running environment. The STILT data was stored on a separate 1 TB hard drive within a MySQL database.

The 2519 light curves are evaluated through a process of stratified 10-fold cross-validation with 3 repeats. In this procedure the dataset is split into ten sections whilst maintaining the ratio of each class in the subsets relative to the whole dataset. Each subset is then used as a validation set for models trained using the other nine subsets. This validation involves the prediction of the classes of the validation set light curves followed by the computation of the Area under the Curve (AUC) statistic from the computed multi-class Receiver operating characteristic (ROC) curve. This validation is performed ten times for each data subset and mean averaged to produce the validation statistic for that repeat. This process is repeated three times with differing random seed values and again mean averaged to produce the final validation statistic. As the AUC is expected to vary around a mean value due to slight difference in the quality of the light curves being used for training and validation in each cycle, this procedure is hoped to be sufficient to eliminate any potential variation in the trained models performance.

The 10-fold cross-validation was first applied to the full 16 features dataset determined for each of the 2519 light curves. The random forest model was tuned with a hyper-parameter that defines the number of predictors sampled for splitting at each node. The best performance was obtained with this

parameter set to 4 features. The Neural Network was trained using backpropagation on a single hidden layer feedforward neural network with 16 input neurons, 50 neurons in the hidden layer and 5 neurons in the output layer using a softmax classifier. The Hyperbolic-Tangent function was used for non-linearity and complexity control was introduced through a momentum term valued at 0.9. All neurons are initialised with a uniform random number between 0 and 0.07. The learning rate was set at 0.005. The network was trained using backpropagation for 600 iterations. The Support Vector Machine was tuned using a grid based search for the best performing cost value which was found to be 32 for this evaluation. The results of this evaluation are shown in table 3.

TABLE III
AUC OF THE THREE ALGORITHMS ON THE 16 KIM ET AL. FEATURES

Model	Mean AUC
Random Forest Mtry parameter = 4	0.84195528
Support Vector Machine Linear Kernel, Cost = 32	0.80900461
Feedforward Neural Network 16-50-5	0.84946005

Additionally, due to neural networks often performing better with scaled data, the 16 features were scaled to a mean of 0 and a standard deviation of 1 and all three machine learning methods validated on this scaled dataset with results shown in table 4. The hyper-parameters were re-tested but did not change from the previous validation.

TABLE IV
AUC OF THE THREE ALGORITHMS ON THE SCALED 16 KIM ET AL. FEATURES

Model	Mean AUC
Random Forest Mtry parameter = 4	0.84195528
Support Vector Machine Linear Kernel, Cost = 32	0.80897909
Feedforward Neural Network 16-50-5	0.76962864

Surprisingly, the scaled dataset resulted in a small drop in performance for the Support Vector Machine and Neural Network. The results show that our 50 hidden layer feedforward neural network achieved the best performance on the STILT light curves. The random forest model had a similar performance, a result expected by previous studies [9] with the no information rate being an AUC statistic value of 0.5.

By using the probabilities predicted by the random forest algorithm for each of the light curves to determine five binary one verses all ROC curves, a form of multi-class ROC curve can be plotted. These curves are a measure of a class’s true positive rate against the false positive rate with the ideal classifier maximising the true positive rate whilst minimising the false positive rate. Therefore, the better performing a class, the closer it will deviate towards the top left corner from the random-state as a straight line with slope 1 shown by the dotted black line in the figures. Figure 3 shows the ROC curve

generated by one of the validation 16 feature random forest models with a multiclass AUC of 0.8102. Each line is related to a one-vs-many prediction on a specific class given by the line colour in the legend. This is performed by assigning a class label of 1 to the appropriate class and a label of 0 to all other classes.

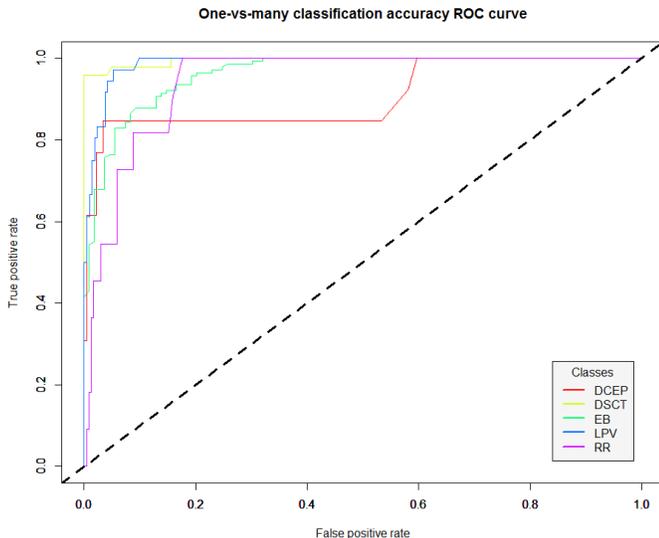


Fig. 3. ROC curve for the 16 feature trained random forest model. The Long Period Variable (LPV) and Delta Scuti (DSCT) classes exhibit the best performance.

These results are reasonable considering the cadence limitations of this survey with all the classes detectable with greater than 80% retrieval rate at a cost of at worst a 10% false retrieval rate. The Delta Scuti and Long Period Variable classes achieve the best AUC with values of 0.9959 and 0.9877 respectively. This is likely a result of the two classes exhibiting clear periodic and amplitude features. Long Period Variables tend to have highly sinusoidal variations with periods of the order of years and amplitudes of multiple magnitudes whereas Delta Scuti variables have periods of only a few hours and variations on the order of a tenth of a magnitude. This conclusion can be reinforced through obtaining the feature importance of the random forest model defined by the Mean Decrease GINI statistic. Table 5 demonstrates this importance statistic for the model used to generate the ROC curve in figure 3. This shows that the Period, variability in the folded light curve and the amplitude of the Fourier model are dominant.

Overall, the most important features are the period and the amplitude of the primary harmonic of the Fourier model. This amplitude is superior to the range of magnitude for a light curve as the range can be prone to noisy observations. The Fourier model does have the disadvantage of poor fits due to sampling as discussed previously however the models have still selected it as a strong feature in these 2519 light curves. Additional features of interest are the Kurtosis and slope gradient features m_{p10} and m_{p90} . These features are strong at identifying light curves with sharp peaks or dips in brightness which is a feature commonly associated with a large number of eclipsing binary light curves. Finally, the feature ψ^1 shows how strongly aligned the data points are for a given period. This can be useful for Long Period Variables due to secondary periods.

TABLE V
MEAN DECREASE GINI COEFFICIENTS FOR THE 16 KIM ET AL. FEATURES

Feature	Mean Decrease GINI
Period	2019.2553583
ψ^1	549.73631507
ψ^{CS}	192.76109940
R_{21}	209.11562924
R_{31}	136.55228001
Φ_{21}	69.433471117
Φ_{31}	71.555076485
γ_1	97.304429700
γ_2	107.25834895
K	142.53335421
Q_{3-1}	150.26993955
A	81.838996126
H_1	413.76605741
W	88.665784819
m_{p10}	425.10419581
m_{p90}	320.10894756

Like the previous 16 feature models, the images produced from the bin-averaged phase folded light curves are used to produce a new validation dataset. The same machine learning algorithms from the previous validation were applied to this dataset generated by the new method. The primary difference from the previous models was there were now 2000 input units where each one is the value of a specific pixel from a concatenated 100x20 image representation vector, -0.5 for an off (black) pixel and +0.5 for an on (white) pixel.

The random forest model was tuned with the number of predictors sampled for splitting at each node hyper-parameter valued at 4 like the previous validation. The Neural Network was trained using backpropagation on a single hidden layer feedforward neural network with 2000 input neurons, 200 neurons in the hidden layer and 5 neurons in the output layer using a softmax classifier. The number of neurons in the hidden layer was increased in order to model more complex patterns expected to be present in the input features. Potentially more might be required but this was limited by the available resources. The Hyperbolic-Tangent function was used for non-linearity and complexity control was introduced through a momentum term valued at 0.9. All neurons are initialised with a uniform random number between 0 and 0.07. The learning rate was set at 0.005. The network was trained using backpropagation for 600 iterations. The Support Vector Machine was tuned using a grid based search for the best performing cost value which was found to be 1 for this evaluation. The results of this evaluation are shown in table 6.

TABLE VI
AUC OF THE THREE ALGORITHMS ON THE VISUAL FEATURES

Model	Mean AUC
Random Forest Mtry parameter = 4	0.63483958
Support Vector Machine Linear Kernel, Cost = 1	0.58861276
Feedforward Neural Network 2000-200-5	0.61050239

Whilst the AUC results are notably inferior to the previous results from Period and Amplitude features from the previous study, the result does show that features were automatically extracted by the machine learning algorithms and used to train to recognise visual shapes for use in a classification task. It is also worth noting that this approach may prove better at discriminating between two similar subclasses than on an overall superclass problem. This network is also extremely limited in the visual features it can extract. For example, despite attempts to position certain magnitude features at specific phases, noise quite often causes these features to be placed at slightly different phases. This results in the requirement of any visual feature layer to implement translation invariance. This can be accomplished by neural networks using convolutional layers [16] but this has not been implemented in these models, which is a big limitation. Figure 4 shows the ROC curves from one of the image representation random forest models with a multi-class AUC of 0.6386.

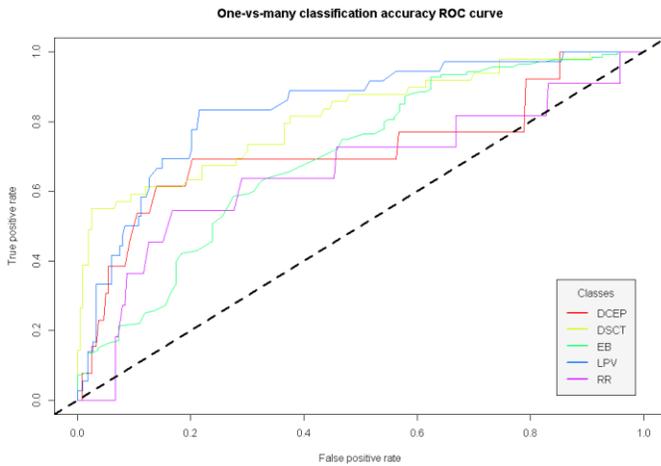


Fig. 4. ROC curve for the 0.6386 AUC image representation model. Global performance is poorer than the 16 features models with the best resolved classes remaining the Long Period Variables and the Delta Scuti variables.

The image representation performance can be augmented through the recognition that significant information is lost through the lack of scaling in the images. The Delta Cepheid and Long Period Variable folded waveforms can look very similar until the realisation is made that the amplitude of the Long Period Variables is significantly larger. Therefore we included two features that describe the two axes of the images. As the horizontal direction shows the phase of the folded light curve, the period describes the length of time this phase covers. As for the vertical direction, this is by definition the amplitude of the light curve as defined above. Including these two features along with the 2000 input pixel values produced the AUC cross-validation results displayed in table 7.

TABLE VII
AUC OF THE THREE ALGORITHMS ON THE VISUAL FEATURES W/ PER AND AMP

Model	Mean AUC
Random Forest Mtry parameter = 4	0.66047666
Support Vector Machine Linear Kernel, Cost = 1	0.76388919
Feedforward Neural Network 2000-200-5	0.79524852

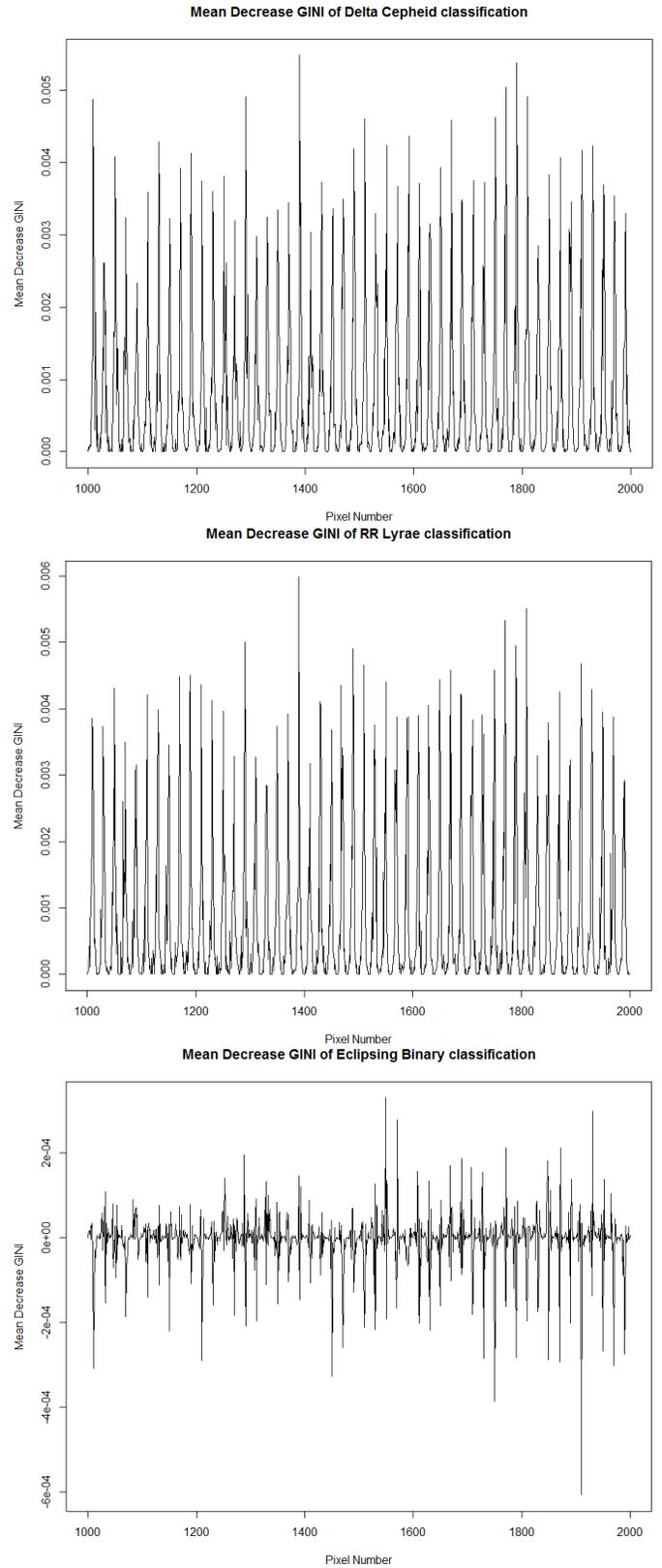


Fig. 5. The Mean Decrease GINI feature importance against pixel numbers 1000 to 2000 (defined as the phase regime from 0 to 1) for three of the main super-classes. The Delta Cepheid variables at the top, RR Lyrae variables at the middle and Eclipsing Binaries in the bottom plot. Each pixel individually has a low weighting but together can communicate important class knowledge.

Figure 5 demonstrates the importance of the individual image pixels for the classification task on three of the super-classes using their mean decrease GINI coefficients. Whilst each individual pixel has a minimal weighting, together they can identify interesting structures. All three plots show a clear rhythm which corresponds to 20 pixels. This is the importance varying from low weight at the exterior of the image to higher weighting close to the centre which is where important signal structures are expected to be found. In the Delta Cepheid and RR Lyrae classes we can see a clear secondary structure where the importance rises to peaks at pixels near the phase of 0.25 and 0.75 where major peaks and dips are expected given we set the max phase to occur near 0.25 when the light curves were epoch-folded. Finally, the Eclipsing Binary plot shows the weighting is a fraction of that from the other two classes. It appears clear that our new proposed method was struggling to resolve usable detail from the eclipsing binary light curves. Whilst the reason for this is unclear it may be a result of the observational cadence of these objects. If the characteristic dip in the light curve due to the transit event is not sufficiently observed the resulting folded light curve will exhibit reduced amplitude and without this characteristic dip feature except for possibly a gap near the expected light curve dip but it would be very unlikely for this gap to occur at the same phase location.

IV. SUMMARY AND FURTHER WORK

In this paper we presented our work on using visual features for light curve classification. We built a number of models based on features from Kim et al. and our own visual approach using stratified 10-fold cross-validation on 2519 STILT variable light curves from five object super-classes. We showed these features contained important information when the light curves were noisy and poorly sampled. Our method initially struggled to compete with features engineered from previous studies attaining a best AUC of 0.6348 compared to the 16 Kim et al. features with a best AUC of 0.8495 until we introduced the period and amplitude features into the training phase. These features give context to the two dimensions on the image representations allowing for an improvement in the best AUC to 0.7952. These strengths were offset by limitations in the machine learning algorithms we used when applied to image based representations especially in the lack of translational and scale invariance. This caused test light curves of a well-known class but with a different phase alignment to be misclassified. We plan to implement convolutional layers in our feedforward neural network topology to introduce these invariances [16].

There are also a number of hyper parameters that haven't been fully investigated such as the optimal pixel 'resolution' for the light curve images. The horizontal pixels contain the majority of the sampling noise and the vertical pixels carry a lot of the magnitude noise which is heavily instrumental and data-reduction limited. A superior method of determining the candidate period to phase-fold the light curve at must also be determined or the resulting image carries no useful information on the class of the light curve. These efforts will improve light curve classification and potentially redefine the limitations of survey cadence required for scientific analysis.

ACKNOWLEDGMENT

The Liverpool Telescope is operated on the island of La Palma by Liverpool John Moores University in the Spanish Observatorio del Roque de los Muchachos of the Instituto de Astrofísica de Canarias with financial support from the UK Science and Technology Facilities Council [17]. We acknowledge with thanks the variable star observations from the AAVSO International Database contributed by observers worldwide and used in this research.

REFERENCES

- [1] D. G. York, J. Adelman, et al., "The Sloan Digital Sky Survey: Technical Summary," *The Astronomical Journal*, vol. 120, 3, pp. 1579–2000, 2000.
- [2] M. A. C. Perryman, K. S. de Boer, G. Gilmore, et al., "GAIA: Composition, formation and evolution of the Galaxy," *Astronomy and Astrophysics*, vol. 369, pp. 339–363, 2001.
- [3] Z. Ivezić, J. A. Tyson, B. Abel, et al., "LSST: from science drivers to reference design and anticipated data products," *ArXiv e-prints* [arXiv:0805.2366], 2008.
- [4] L. Eyer, N. Mowlavi, "Variable stars across the observational hr diagram," *Journal of Physics: Conference Series*, vol. 118, 1, 012010, 2008.
- [5] A. K. Vivas, R. Zinn, P. Andrews, et al., "The QUEST RR Lyrae Survey: Confirmation of the Clump at 50 kiloparsecs and Other Overdensities in the Outer Halo," *The Astrophysical Journal*, vol. 554, 1, pp. L33–L36, 2001.
- [6] T. R. Bedding, A. A. Zuckerman, "HIPPARCOS Period-Luminosity relations for Mira and semiregular variables," *The Astrophysical Journal*, vol. 506, pp. 47–50, 1998.
- [7] 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*, vol. 148, 2, article id. 31, 16, 2014.
- [8] J. Debosscher, L. M. Sarro, et al., "Automated supervised classification of variable stars I. Methodology," *Astronomy and Astrophysics*, vol. 475, pp. 1159–1183, 2007.
- [9] J. W. Richards, D. L. Starr, et al., "On Machine-Learned Classification of Variable Stars with Sparse and Noisy Time-Series Data," *The Astrophysics Journal*, vol. 733, 1, pp. 10–32, 2011.
- [10] D.-W. Kim and C. A. L. Bailer-Jones, "A package for the automated classification of periodic variable stars," *Astronomy and Astrophysics*, vol. 587, A18, 2016.
- [11] I. Nun, P. Protopapas, et al., "FATS: Feature Analysis for Time Series," *ArXiv e-prints* [arXiv:1506.00010], 2015.
- [12] K. Pichara, P. Protopapas and D. León, "Meta-Classification for Variable Stars," *The Astrophysical Journal*, vol. 819, 1, 2016.
- [13] P. R. McWhirter, S. Wright, I. A. Steele, D. Al-Jumeily, A. Hussain, P. Fergus, "A Dynamic, Modular Intelligent-Agent Framework for Astronomical Light Curve Analysis and Classification," *Lecture Notes in Computer Science*, vol. 9771, pp. 820–831, 2016.
- [14] N. R. Mawson, I. A. Steele and R. J. Smith, "STILT: System design and performance," *Astronomische Nachrichten*, vol. 334, 7, pp. 729–737, 2013.
- [15] J. D. Scargle, "Studies in Astronomical Time Series Analysis. II. Statistical aspects of spectral analysis of unevenly spaced data," *The Astrophysical Journal*, vol. 263, pp. 835–853, 1982.
- [16] A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, 2012.
- [17] I. A. Steele, R. J. Smith, et al., "The Liverpool Telescope: performance and first results," *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, 2004.