Classifying Stellar Variability in the V and g bands with the All-Sky Automated Survey for SuperNovae

A Senior Research Thesis

Presented in Partial Fulfillment of the Requirements for graduation *with research distinction* in Astronomy in the undergraduate colleges of The Ohio State University

by

Collin Christy

The Ohio State University
April 2021

Project Adviser: Dr. Krzysztof Stanek, Department of Astronomy

Abstract

The All-Sky Automated Survey for SuperNovae (ASAS-SN) is a wide-field photometric survey that monitors the entire night sky every night. Recently, it has begun observing in the g-band with an improved cadence ($\lesssim 24$ hours) and reduced diurnal aliasing compared to their survey in the V-band, allowing for up to 100 million Milky Way stars to be observed. Many candidates for variable stars have been recovered which require classification based on their time-domain photometric modulations. ASAS-SN currently uses a machine learning classifier that utilized Fourier transforms to classify variable candidates in their data. However, machine learning classifiers are prone to incorrectly classifying rare phenomena and objects that are the result of systematic errors in detection. Here, I present additional methods of classification including *Citizen ASAS-SN* and a different approach to machine learning classification. *Citizen ASAS-SN* is a citizen science project hosted on the Zooniverse platform in which volunteers are presented with ASAS-SN g-band light curves of variable star candidates. The classification workflow allows volunteers to classify these sources, while also allowing for the identification of unique variable stars for additional follow-up. We aim to improve existing and future variable star classifications in our time series catalog by implementing data from Citizen ASAS-SN and updating our machine learning classifiers with promising results shown in this thesis.

Contents

1 Introduction

Recent advancements in observing technologies have led Astronomy into the era of big data where large astronomical surveys are the main source of discovery for astronomical objects and the accumulation of observational data (Mickaelian, 2015). Some surveys go a step further and regularly monitor the entire night sky by utilizing a global network of telescopes; in recent years, the All-Sky Automated Survey for SuperNovae (ASAS-SN) has done exactly this. The primary focus of ASAS-SN was to observe bright supernovae (Holoien et al., 2016); however, it also observes many other transient and variable sources in the process.

1.1 The All-Sky Automated Survey for SuperNovae

The All-Sky Automated Survey for SuperNovae is a wide-field photometric survey that monitors the entire night sky using 20 telescopes located in both the Northern and Southern hemispheres (Shappee et al., 2014; Kochanek et al., 2017; Jayasinghe et al., 2018). The field of view of an ASAS-SN camera is 4.5 deg^2, the pixel scale is 80 and the FWHM is typically $\sim$2 pixels. ASAS-SN uses image subtraction (Alard & Lupton, 1998; Alard, 2000) for the detection of transients and to generate light curves. Recently, we have been using ASAS-SN data to study bright variable stars (see, for e.g., Jayasinghe et al. 2018, 2019b,c,d, 2020a,b,c). In the initial V-band catalog, $\sim$60 million stars were classified through machine learning techniques, resulting in a catalog of $\sim$426,000 variables, of which $\sim$220,000 were new discoveries (Jayasinghe et al., 2020d).

1.2 From V-band to g-band

Since 2018, ASAS-SN has shifted its survey to the g-band with a limiting magnitude of $g \lesssim 18.5$ mag, allowing for up to 100 million Milky Way stars to be observed and characterized. When compared to the V-band data, the new g-band data is especially promising for the discovery and characterization of variable stars because of its improved cadence ($\lesssim 24$ hours in the g-band vs. $\sim 2-3$ days in the V-band) and reduced diurnal aliasing due to the longitudinal spread of the ASAS-SN units.

1.3 Motivation

As I stated earlier, machine learning techniques were utilized to classify variable candidates in ASAS-SN's initial V-band surveys. However, some of these candidates are simply the result of systematic errors from detection or false positives due to signal contamination. Such instances are prone to be incorrectly classified as real variable phenomena. When training new classifiers, previously catalogued variables are used as training sets. If the training set contains "junk" data, the classifier will suffer in performance. In this thesis, I will outline our efforts to improve existing classifications and discuss our work on constructing a new machine learning classifier to be used on the g-band data.

2 Citizen ASAS-SN

ASAS-SN's citizen science project, *Citizen ASAS-SN*, primarily focuses on the classification of g-band light curves to identify both classical and anomalous variables. In particular, the identification and follow-up study of unusual variable stars is of interest as these systems often lead to new astrophysical insights. ASAS-SN has

serendipitously found a number of examples, including the most extreme heartbeat star found to date (Jayasinghe et al., 2019a), and a Tabby's star analogue (dubbed Zachy's star, Way et al. 2019).

As our input source catalog for this project, we used the `refcat2` catalog (Tonry et al., 2018). The g-band light curves were extracted as described in Jayasinghe et al. (2018) using image subtraction (Alard & Lupton, 1998; Alard, 2000) and aperture photometry on the subtracted images with a 2 pixel radius aperture. We corrected the zero point offsets between the different cameras as described in Jayasinghe et al. (2018) and calculated periodograms using the Generalized Lomb-Scargle (GLS, Zechmeister & Kürster 2009; Scargle 1982) periodogram. Candidate variable sources were identified with simple cuts in light curve and periodogram statistics as in Jayasinghe et al. (2019c). Volunteers working on Citizen ASAS-SN are shown images with light curves phased by both the best GLS period and twice the best GLS period along with the observed light curve (Figure 1).

2.1 Citizen Science

While machine learning methods are efficient at analyzing known behaviors, the scarcity of rare phenomena makes it difficult for them to account for new or extreme cases (Alhammady & Ramamohanarao, 2004). One alternative approach is crowd-sourcing through citizen science, enabling volunteers worldwide to participate in the scientific process. In recent years, the Zooniverse[1] has hosted many successful citizen science projects that often lead to serendipitous discoveries (Trouille et al., 2019). Citizen scientists working with the *Planet Hunters* project (Boyajian et al., 2016) led to the discovery of KIC8462852 (also known as Tabby's star), which is

[1]Zooniverse:https://www.zooniverse.org/

an F-type main-sequence star subject to continued follow-up (Sacco et al., 2018; Thomas et al., 2019; Hitchcock et al., 2020). The *Galaxy Zoo* project used citizen science to classify galaxies based on their visual morphology and has produced over 60 publications, millions of classifications, and multiple spin-off projects (Masters & Galaxy Zoo Team, 2020). One of the most unusual objects found was an emission nebula near the spiral galaxy IC 2497 known as "Hanny's Voorwerp", discovered by Dutch schoolteacher Hanny van Arkel (Józsa et al., 2009).

2.2 Classification Workflow

As a first foray into citizen science, our goal was to address several simple but common problems in the classification process. Most variable stars can be classified based on their light curves, however ambiguities may arise for certain variable types (such as RRC RR Lyrae systems and EW eclipsing binaries). For eclipsing binaries, the best period returned by the GLS periodogram is often 1/2 of the orbital period, which is why the light curve phased with twice the best GLS period is shown (see Figure 1). When phased with the correct orbital period, the light curves of eclipsing binaries show a distinct separation of the primary and secondary eclipses allowing for their accurate classification. The observed light curve is useful for identifying long-period variables such as Miras, and evolving variables like rotating spotted stars.

We designed our project workflow to be easy to navigate and accessible to a wide array of volunteers. Because we expect no prior knowledge of variable star classification, we first present users with a tutorial that details the classification process and briefly summarizes the science involved. Volunteers also have access to a field guide that describes common variable stars and their light curves. Our

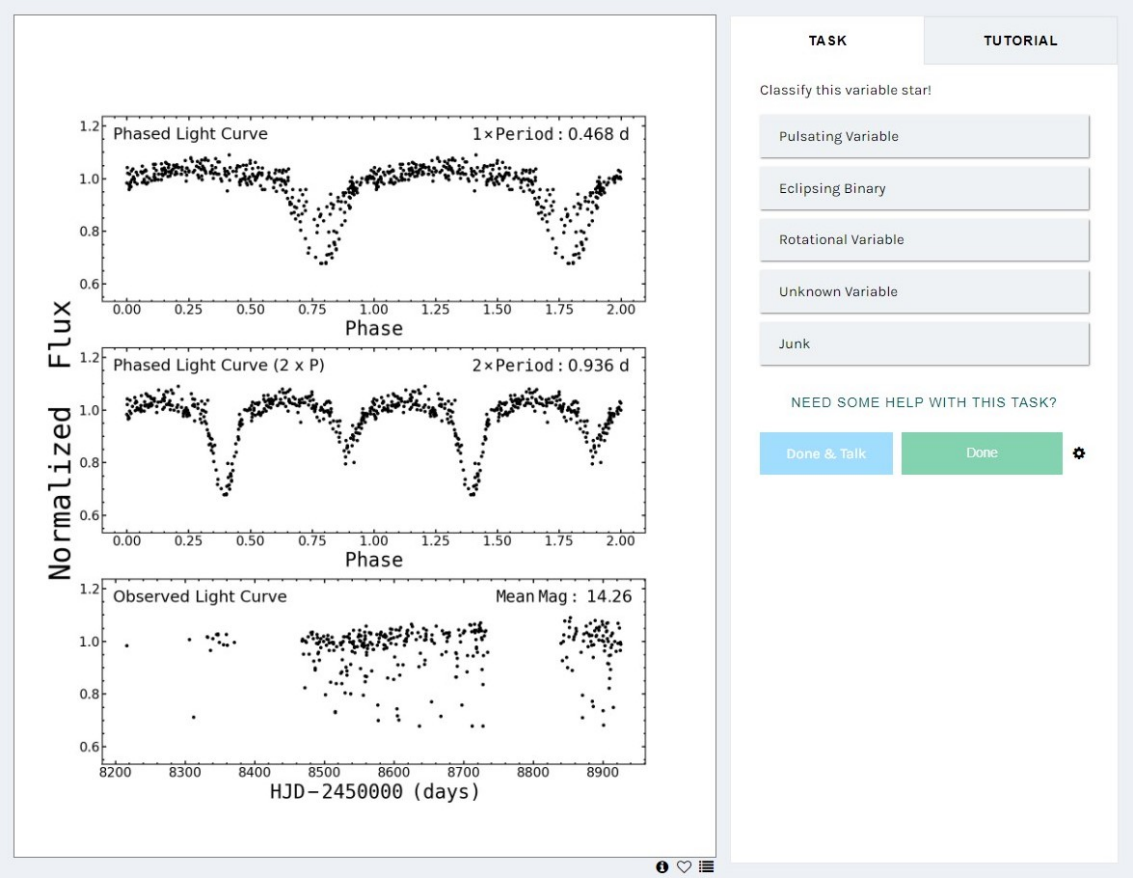

Figure 1: The left panel shows the data presented to the users. The top and middle panels show the light curve phased at the best GLS period and twice that period, while the bottom panel shows the actual light curve. The classification options appear to the right of the light curves.

workflow tasks users to determine the correct basic classification, selecting between three broad classes (Pulsating Variables, Eclipsing Binaries, Rotational Variables), choosing the option "Unknown Variable" for ambiguous cases, or flagging the light as "Junk". As users get started, we present them with a variety of "gold-standard" (GS) candidates that have been classified by the science team. These GS variables provide the user with feedback on their classifications to train them in the process. As users make more classifications, GS variables become less frequent and the user

begins to classify new light curves.

2.3 Citizen ASAS-SN Beta and Beyond

In preparation for launch, we submitted Citizen ASAS-SN to the Zooniverse team for their internal review. During this phase, 58 volunteers from Zooniverse tested the workflows and provided feedback on their experience. Many users suggested that our draft tutorial was too complicated and that they would prefer more examples in the field guide, leading to a final update for these features. During this stage, users classified a set of 200 new candidates and $\sim$ 400 GS variables. Figure 2 also shows the resulting confusion matrix; the "voted" class represents the final majority vote that a candidate received while the "true class" indicates the correct classification. When we analyzed the classifications, we found that users could reliably distinguish between the three main types. Individuals correctly identified 86% of pulsating variables in our testing sample, along with 70% of eclipsing binaries, 65% of rotational variables, and 89% of variables labeled as "junk" by our team. For the non-GS variables, the classification accuracy for pulsating variables and eclipsing binaries remained high; however, rotational variable classifications were systematically less accurate, probably because their appearance can vary widely leading to inconsistent classifications (e.g.,Thiemann et al. 2021). We also found that $\sim$ 20% of our 200 test variables were classified as junk. Based on the results, we also removed the second task for eclipsing binaries which asked users to determine if period doubling was occurring. Users were quick to label binaries as having the original GLS period (1 x Period), when in fact, most of the binaries in our sample had twice this period. We suspect that this question is better addressed in a workflow dedicated solely to eclipsing binaries.

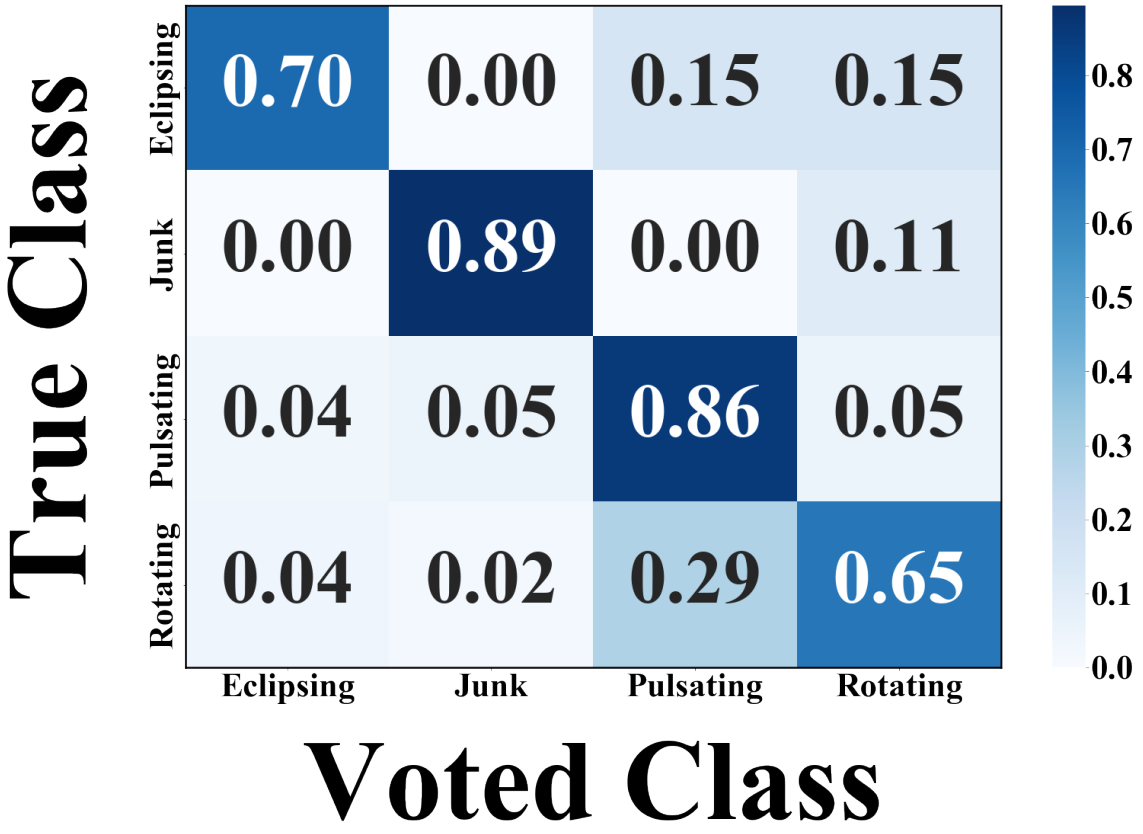

Figure 2: The normalized confusion matrix derived from the internal review with the user-selected classifications on the horizontal axis and the correct classifications on the vertical axis.

Citizen ASAS-SN was released for public use on January 5th, 2021, and has since accrued over 1,800 volunteers and ∼400,000 classifications. We launched the project with an initial subject set containing 40,640 variable candidates detected in a portion of the southern sky (from -60 to -90 deg in DEJ2000). Of these classifications, we found many interesting variable stars which will be discussed in more detail in Section 4.1. Note that a variable candidate stops being show to users once it has reached its retirement limit. For our workflow presented at launch, we set the retirement limit at 10 votes. If the number of "Junk" votes reaches 5, then the

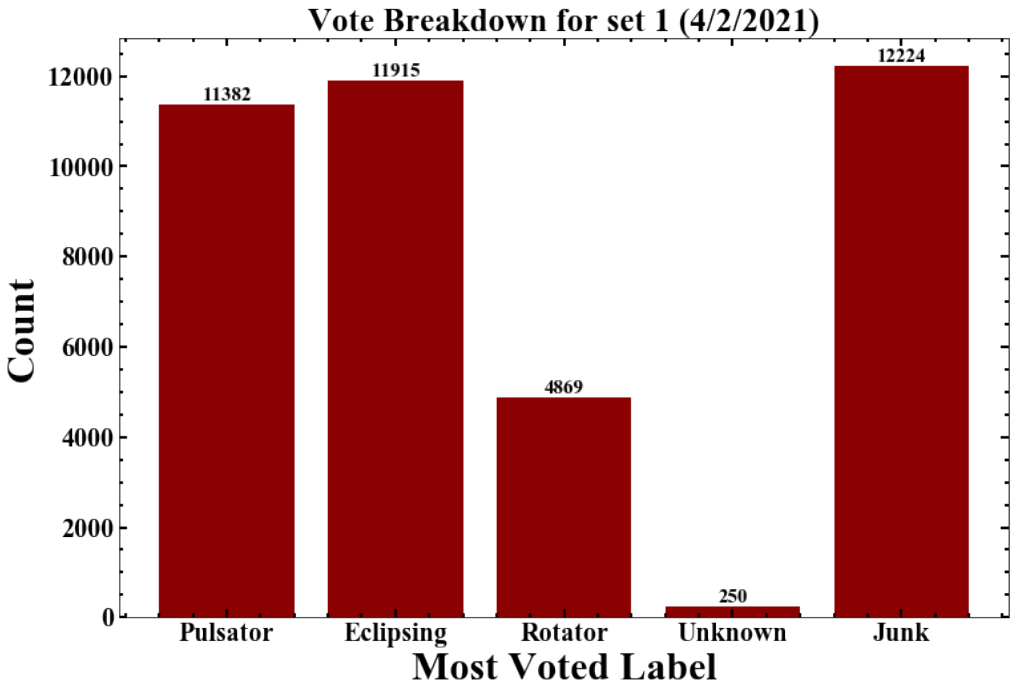

Figure 3: Breakdown of the frequency for the most voted on class; from Citizen ASAS-SN's first subject set

candidate is retired early with the label "Junk". A breakdown of the most voted class for each retired variable is shown in Figure 3. The most notable take away from these results are that nearly one quarter (∼12,000) of the first subject set appeared to be false positive signals (i.e., junk). When training future machine learning classifiers, we can remove such variables from the training set to improve the classifiers performance.

Out of the 40,640 candidates in our initial subject set, 15,368 had been previously classified in the AAVSO VSX catalog (Watson et al., 2006). After the first set was fully classified, we found that our volunteers were able to recover 99.0% of these variables with ∼ 150 marked as being junk. The resulting confusion matrix of these classifications are shown in Figure 4; this matrix uses that same formulation as in

Figure 4: The normalized confusion matrix derived from the first round of completed classifications. User-selected classifications are on the horizontal axis and the correct classifications are on the vertical axis.

the beta review. Overall, the results are quite similar to those found in the beta review with pulsating variables and eclipsing binaries being the most identifiable. Another metric that we can look at from our classifications is how correlated our users votes were for each variable type. To do this, I computed a parameter called "Classification Strength", which describes the ratio between the most voted classification type and the total number of votes. This metric would evaluate to 1.0 if all user classifications agree for a particular variable candidate. In Figure 5, I show the distribution of classification strengths for each variable class used in our work-

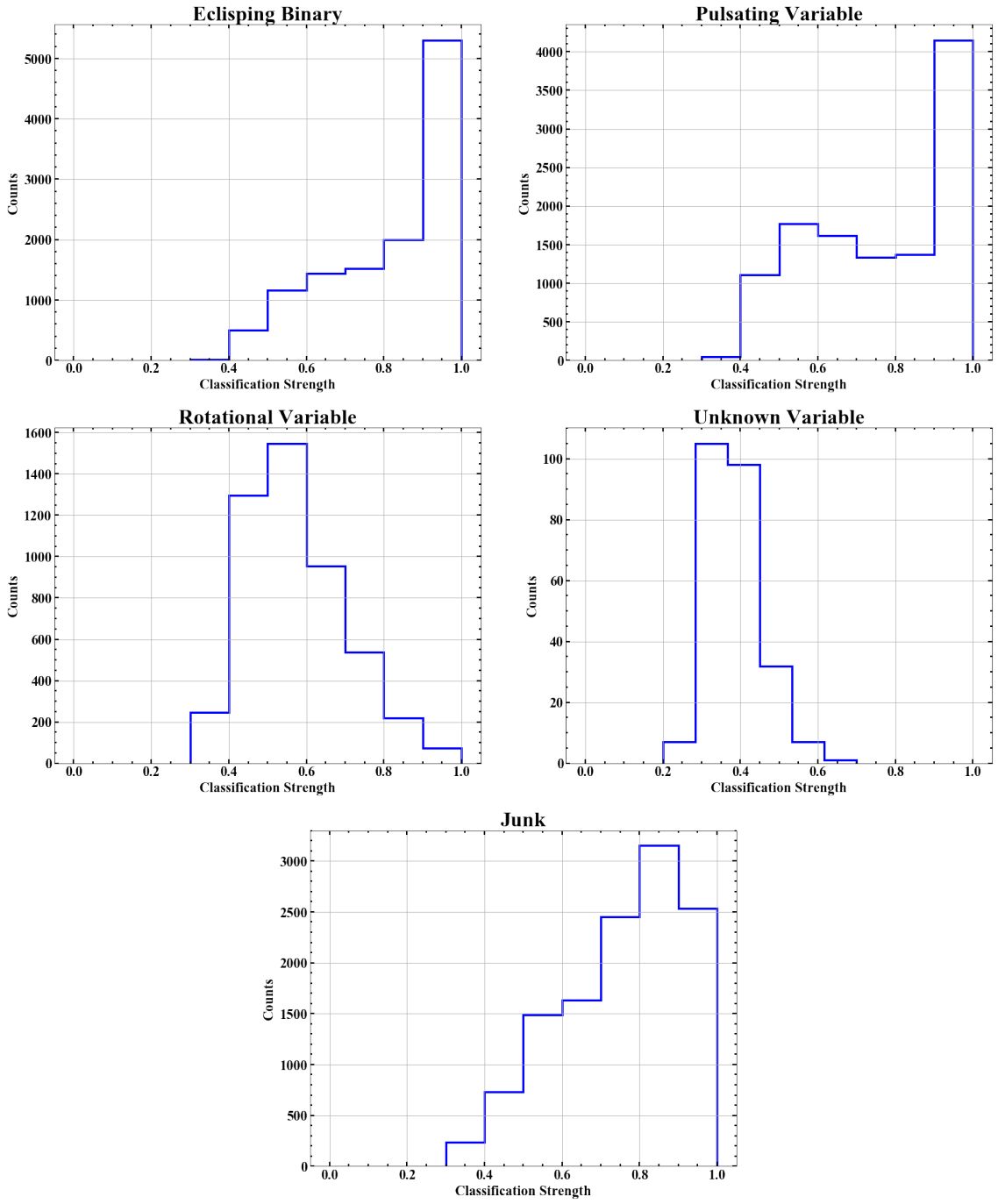

Figure 5: Distribution of classifications strength for each variable class used in *Citizen ASAS-SN* periodic variables workflow.

flow. This plot indicates that our users are likely to agree on variables most voted as Eclipsing Binary, Pulsating Variable, and Junk. For candidates most voted as Rotational Variable and Unknown Variable, there are more disagreements between users.

As we progress with the project, we plan to introduce other workflows, including classifications of irregular variables and higher-order classifications of specific variable types. For the variables that were studied using the V-band data, we will compare the classifications made by Zooniverse volunteers to those obtained through our machine learning classifiers (e.g., Jayasinghe et al. 2019b), and update the ASAS-SN variable stars database[2] with improved classifications as needed.

3 Machine Learning

In recent years, ASAS-SN has discovered millions of variable candidates; classifying all of them by hand would be nearly impossible. Therefore, the ASAS-SN team employed the used of machine learning to aid in their classification efforts. Machine learning is an incredibly powerful tool that allows researchers to analyze large and complex datasets. In the context of variable star classification, an ideal machine learning classifier can quickly assess a particular candidate and accurately identify its variability type. Of course, such an ideal is very unlikely, so we are very interested in improving the quality and efficiency of such classifiers.

[2]ASAS-SN variable stars database:https://asas-sn.osu.edu/variables

3.1 V-band Classifier

The machine learning classifier used on ASAS-SN's V-band time series data is extensively described in Jayasinghe et al. (2019). The first classifier they used was a random forest model from `scikit-learn` (Pedregosa et al., 2018). This initial classifer was trained to distinguish between broad variable types. A Fourier series was fit to the variables in the training set as described by,

$$M(\phi) = m_0 + \sum_{i=1}^{N} (a_i sin(2\pi i\phi) + b_i cos(2\pi i\phi)) \tag{1}$$

where m_0 is the median magnitude of each candidate, N is the order of the Fourier fit, and ϕ is the phase at a given epoch (Jayasinghe et al., 2019). The RF classifier was then trained using the Fourier fit parameters and various stellar statistics including pulsation period and color measurements. While this classifier was extremely accurate at identifying common variable types, it often mislabeled rare phenomena and objects contaminated with systematic errors.

3.2 g-band Classifier

My approach to designing a machine learning classifier was to mimic the methodology used in our citizen science project, *Citizen ASAS-SN*. In particular, I wanted to build a classifier that would correctly classify variable stars found in ASAS-SN's new g–band data based on their folded light curves. From these images, I trained a random forest classifier to correctly distinguish between variable types.

Before training the classifier, I first had to prepare the folded light curves. To do this, I first had to construct light curves from ASAS-SN's photometric data, then convert the image data to a NumPy arrays (see Figure 6). Afterwards, I

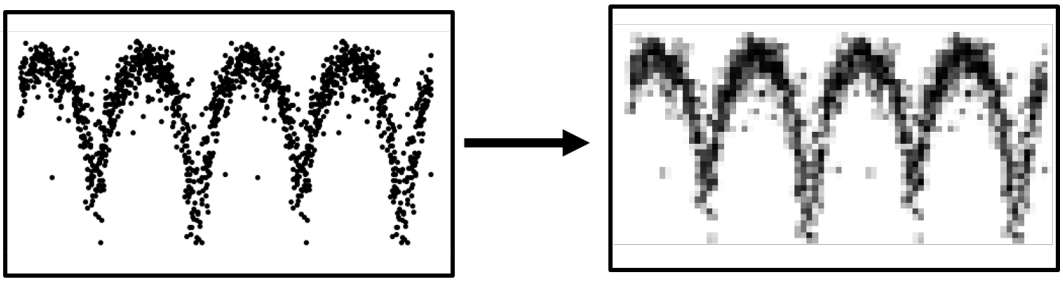

Figure 6: (Left) Folded light curve of a contact binary star system. (Right) The same light curve converted to NumPy array of pixel data (100 x 50)

constructed a random forest classifier from `scikit-learn` with initial parameters: `n_estimators = 100`, `max_depth = 5`, `random_state = 42`, `criterion = 'gini'`. For my initial pass through, I trained the RF classifier to identify $\sim$ 8,700 light curves for the following variable types: RR Lyrae (RRAB), Cepheid (DCEP), Detached Eclipsing Binary (EA), Contact Binary (EW), Mira (M), and Semi – Regular (SR). This classifier was trained on 80% of the full data, allocating the remaining 20% to be used as the testing set. After training, the over all accuracy was found to be,

Accuracy: 0.8092783505154639

Here, accuracy is defined as the percentage of light curves that were identified as their true classification. The resulting confusion matrix from this iteration is shown in Figure 8. From this, we can see that the model is only classifying variables as EA, EW, or RRAB types. This occurred because these types of variables accounted for $\sim$ 93 % of the stars in the training data. Over-representation of a certain class or classes in a training set can cause a machine classifier to develop a bias towards those outcomes. In this case, the RF classifier became highly sensitive towards the EA,EW, and RRAB labels because they were the most prevalent types encountered by the algorithm during its training. These problems, over-representation, sampling

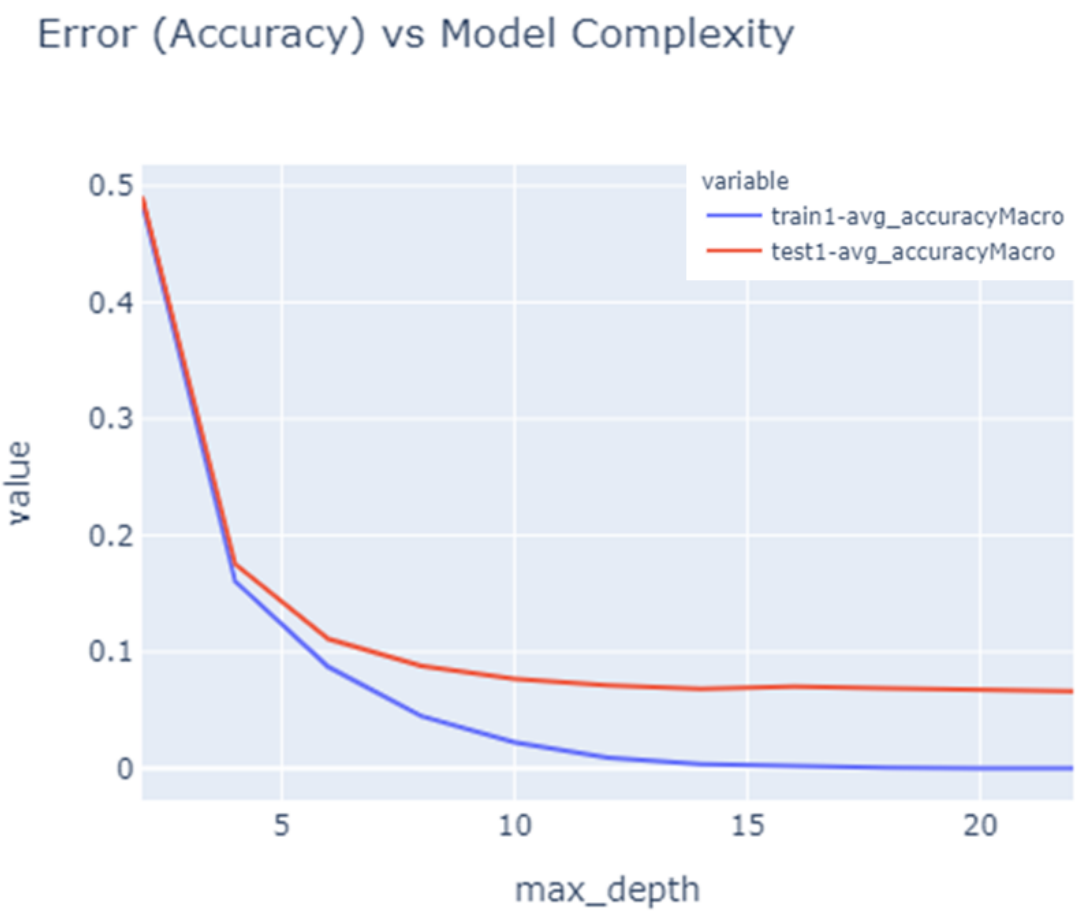

Figure 7: This plot displays model error v.s. complexity for the simplified model using broad variable classes. Here, the maximum depth of the RF model is used a proxy for complexity.

issues, etc., were present in the ASAS-SN V-band classifier (although not nearly as severe). Machine classifiers often have difficulty describing the behavior of rare objects; this fact is what motivated the ASAS-SN team to employ the use of citizen science, which ultimately led to the formation of *Citizen ASAS-SN*.

To combat the sampling issue, I constructed another model that would consider the underrepresented variables (DCEP, M, SR) as a new class called "Pulsators" along with RRAB types. Making this adjustment ensures that the classifications present in the training set are now more evenly distributed. To optimize this model

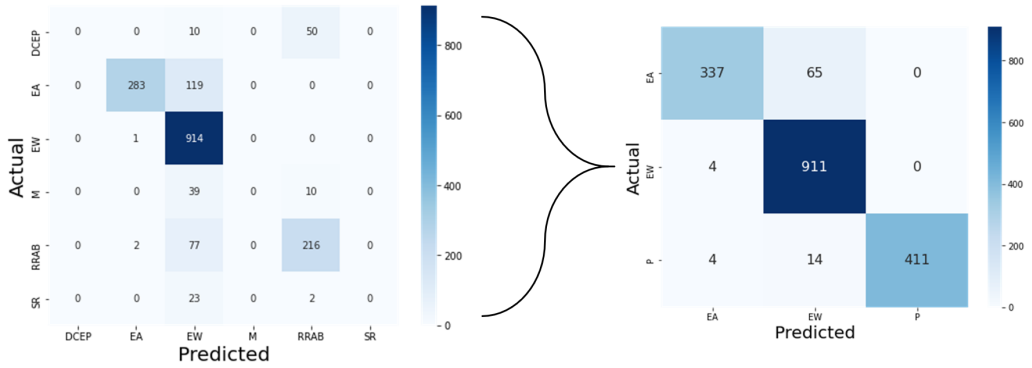

Figure 8: (Left) The classification matrix for my initial RF model. (Right) The confusion matrix for my revised RF model with an optimized max depth.

further, we can analyze how the error in the model changes with complexity. Here, complexity is linked to the `max_depth` parameter of a RF classifier. This parameter describes the number of decision trees that the RF classifier will use. Having a shallow depth in a random forest will likely cause the classifier to suffer in performance accuracy, while a large depth can cause the classifier to over-fit the data. The error trend shown in Figure 7 suggests that an optimal depth to use for the RF model is 10. After retraining the RF classifier with this updated parameter, the model accuracy increases to,

Accuracy: 0.9501718213058419

The confusion matrix for the optimized model (Figure 8) indicates that the classifier can now distinguish between the desired variable types. In practice, there are hundreds of variable types; therefore, a more complete model would have to consider other factors (pulsation amplitude, periods, etc.) in order to classify the complete catalog of variable types. That being said, this is a good metric to determine the broad variable type of unknown transient stars.

3.3 New g-band Classifier

We want the new g-band classifier to be an improvement from their previous models. Now that *Citizen ASAS-SN* is launched, we can begin incorporating human classifications into the automated classification pipeline that is in place currently. In simple terms, the previous V-band classifier identified variables primarily based on a Fourier fit to their light curves, while my g-band classifier identified variables based on the morphology of the image data containing their light curves. Ideally, we would like to reconcile these methods into a model that incorporates both of these techniques. Previously, ASAS-SN has only used a random forest model to classify their light curves. Other models such as recurrent neural networks (RNN) are highly effective at performing predictive tasks related to image classification (Shallue et al., 2019). A particular type of RNN known as a Long Short Term Memory network (LSTM) is especially useful in predictive classification because it is capable of retaining long term trends in a dataset. Because of this, an LSTM inspired model may be more equipped to handle input features containing Fourier fit parameters and image data compared to our previous models.

4 V-band v.s. g-band

ASAS-SN's shift to surveying in the g-band allows us to examine variability features for any particular star in both the V and g bands. Most variable stars exhibit correlated variability patterns in the different wavelengths. Using the newer observations in g allows us to refine our detection for new variables, refine existing period measurements, and distinguish microvariability trends from noise (Süveges et al., 2012). In this section, I will examine some interesting light curves flagged by users

on *Citizen ASAS-SN* and analyze their variability in both bands.

4.1 Hybrid Systems

Stars that exhibit multiple pulsation behaviors are of great interest because these systems can act as stellar laboratories (Thiemann et al., 2021). An example of such a star is J193359.68-680634.9, which began generating discussion on the *Citizen ASAS-SN* talk forum. This system displayed a large regular fluctuation in brightness during its pulsation cycle which suggests that their may be some competing variability trends. This star was previously classified in the V-band as a rotational variable (ROT) by the ASAS-SN variable stars database with a classification probability of 0.889. The light curves for this system in the V and g bands are shown in Figure 9. The observed light curve in V displays prominent amplitude modulation likely due to large spots on their surface; this supports the rotational variable classification. In the g-band, modulation in the observed light curve is less distinct. The phased light curve in g clearly displays an eclipse-like dip which suggests that this star is in an eclipsing binary system. The period returned by the GLS periodogram did not parse out the eclipsing behavior in V; instead, the phased light curve using this period displayed noisy variability.

What is particularly interesting about J193359.68-680634.9, is that the phased light curve in g is very stable, something unusual in systems with stellar spots. However, RS Canum Venaticorum (RS CVn) type variables can explain this strange behavior. RS CVn systems are detached binaries with active chromospheres and nearly synchronized orbits (Hall, 1976). RS CVn stars that display large variation in brightness imply the presence of enormous spots on their surfaces which can cover up to 50% of the visible disc (Berdyugina, 2005). The synchronized orbits of

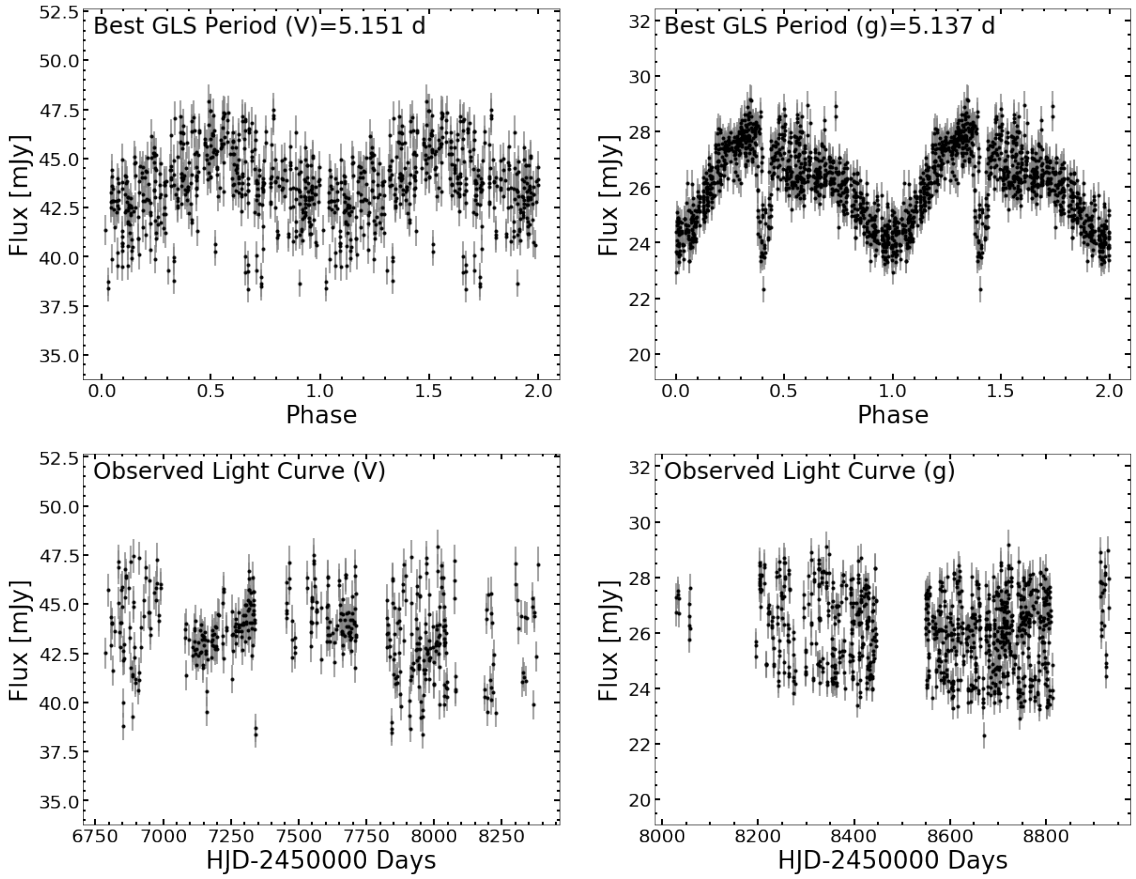

Figure 9: Phased and observed light curves for J193359.68-680634.9; V-band (Left), g-band (Right).

RS CVn types possibly explain why the eclipsing and spotting variability are both visible in the phased light curve of this system.

A particularly interesting, and rare, possibility in stellar evolution is for a pulsating variable to be in an eclipsing binary. These systems are powerful tools because they allow researchers to derive the fundamental stellar parameters and probe the internal structure of stars (Kahraman Aliçavuş et al., 2017). The system J090020.74-644127.9, again flagged by users on *Citizen-ASAS-SN*, showed possible signs of being this type of hybrid system. The phased light curve presented on the project (shown

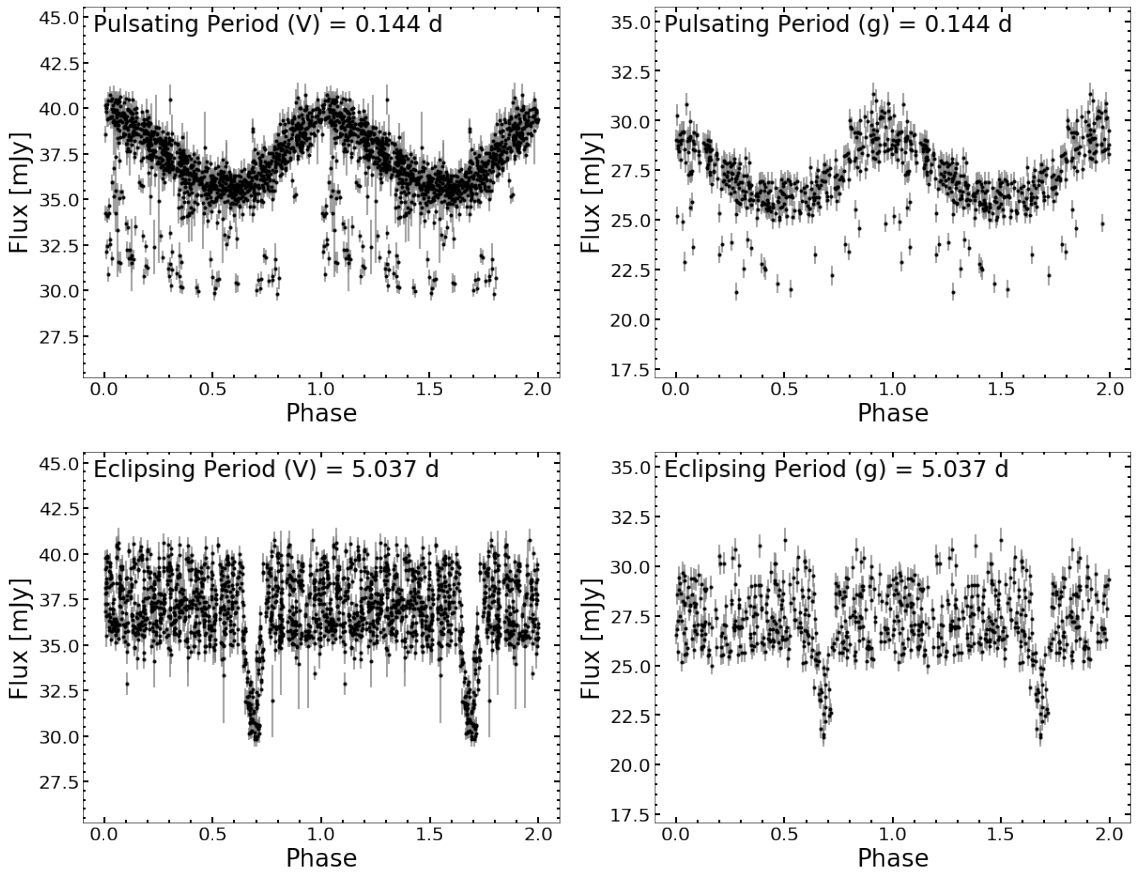

Figure 10: Phased light curves for J090020.74-644127.9 using the pulsating and orbital periods; V-band (Left), g-band (Right). The light curve in the upper right was shown to volunteers on *Citizen ASAS-SN*.

in Figure 10) displayed pulsating behavior with a period of 0.144 days and recurring outliers. Further investigation found that the outliers were the result of a primary eclipse with a period of 5.037 days. The phased light curves using each period show similar trends between the V and g bands. The pulsation behavior shows signs of being a High Amplitude δ Scuti (HADS) type, while the eclipsing pattern looks like an Algol binary (EA). Discovering systems containing HADS in EA type binaries are of great interest due to their potential for followup study. Examining these systems in great detail is important for expanding our theoretical knowledge of pulsating

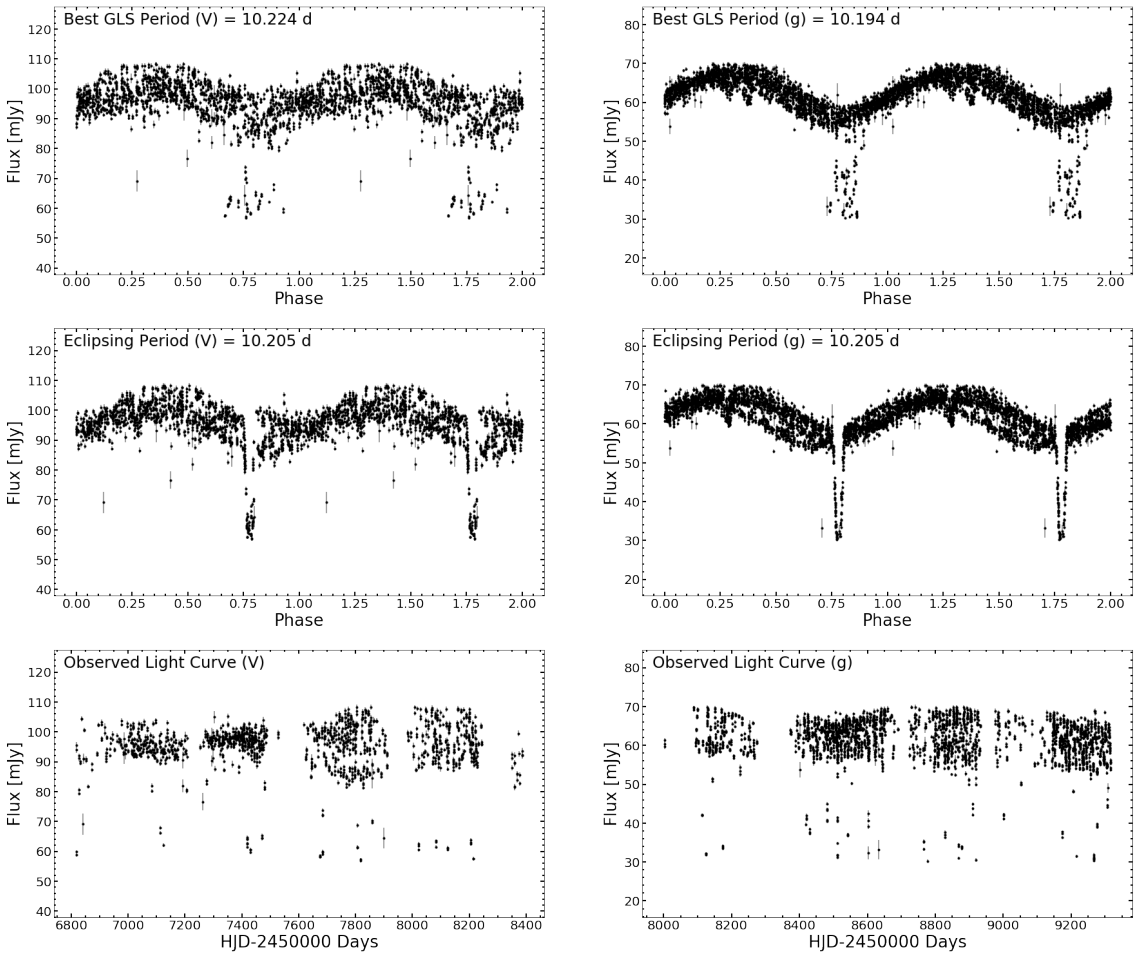

Figure 11: Phased light curves for J085305.34-824360.0 using the rotational and orbital periods; V-band (Left), g-band (Right). The light curve in the upper right was shown to volunteers on *Citizen ASAS-SN*

stars and understanding the internal structures and evolution of stars (Kahraman Aliçavuş et al., 2017).

Another interesting object found by users on *Citizen ASAS-SN* was J085305.34-824360.0 (see Figure 11). The light curve for this variable displayed a strong sinusoidal variation with recurring outliers at each minima. The observed light curve in both bands show distinct amplitude modulation, likely due to spotting on the surface. The regular nature of these outliers indicated that this system might be

an eclipsing binary. Further investigation showed that this star had been previously classified in the V-band as an eclipsing binary (EB type) by the ASAS-SN variable stars database with a classification probability of 0.962 and eclipsing period of 20.4 days. After running a GLS periodogram on the g-band data, I found the eclipsing period to be 1/2 of this value at $\sim$ 10.2 days. This period is very close to the best GLS period presented to our users on *Citizen ASAS-SN*, which suggests that this system is a synchronized binary with active spotting.

5 The ASAS-SN Variable Star Atlas

Variable stars are useful astrophysical tools that have been used to study various aspects of stellar evolution and galactic structure. They also allow astronomers to probe the distance to nearby star clusters and galaxies. Due to the importance of these stellar objects, we plan to utilize the years worth of variable star data in our time series catalog to help inform those in are interested in the field. This led to the formation of the ASAS-SN Variable Star Atlas, a new branch of the ASAS-SN website [3] aimed to educate professional and amateur astronomers about the study and classification of variable stars.

5.1 Variability Tree

Stars can exhibit variability due to various reasons and observable factors. In the past, variable stars were classified in accordance to a prototype star; as the availability of stellar data increased, so did the number of subcategories and higher-order classifications. To help clarify the often confusing landscape of variable star taxon-

[3] ASAS-SN website:https://asas-sn.osu.edu

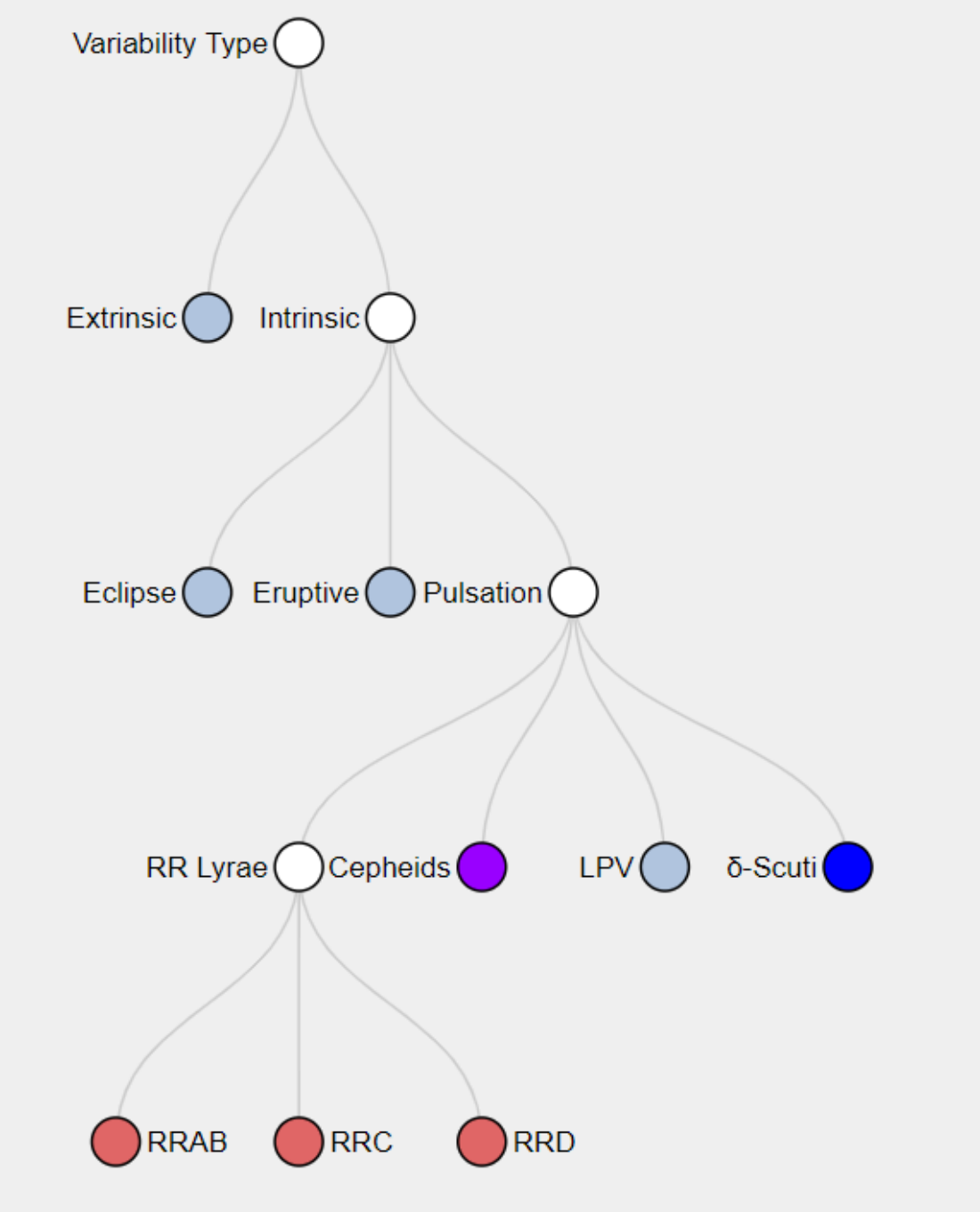

Figure 12: The variability tree in the ASAS-SN Variable Star Atlas. In this case, the path outline above documents the locations of the RR Lyrae subclasses.

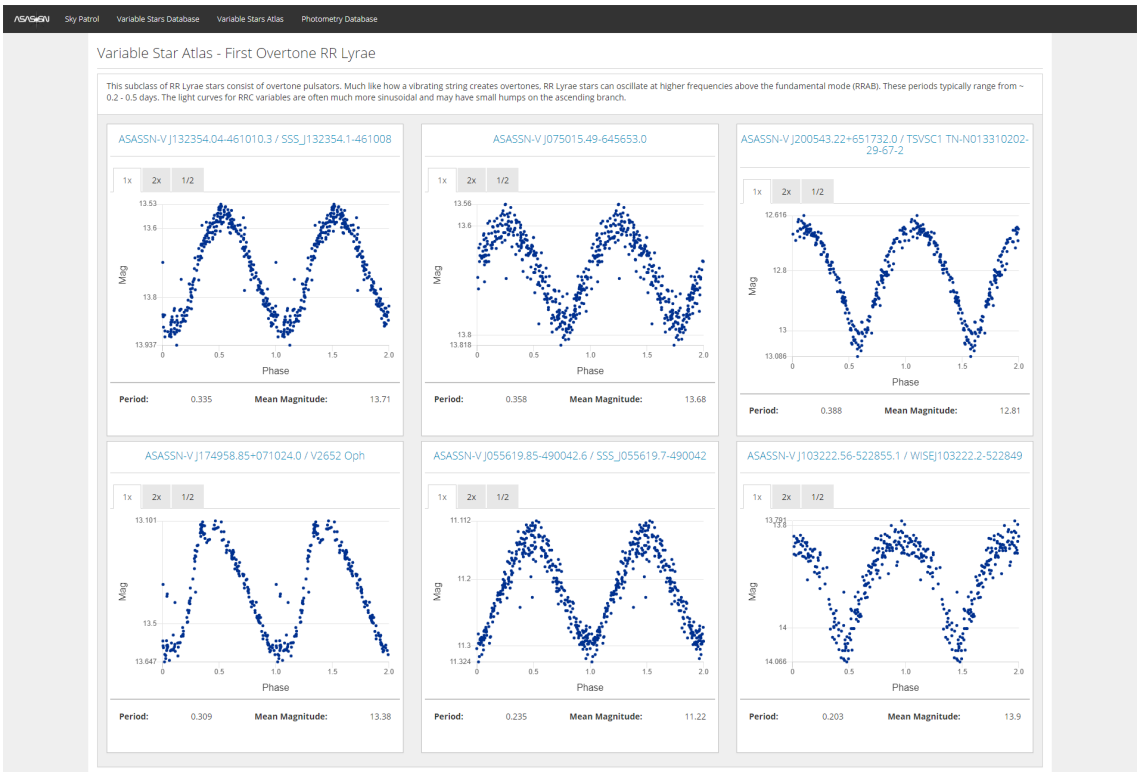

Figure 13: The First Overtone RR Lyrae (RRC) location page; ASAS-SN Variable Star Atlas.

omy, we decided to create a visual classification guide in the form of a variability tree diagram. Variable stars can largely be categorized into variability caused by intrinsic and extrinsic factors. Intrinsic variables encompass stars that vary in brightness due to physical properties; pulsating stars like RR Lyrae and Cepheids are considered this type. Conversely, extrinsic variables are stars that exhibit variability due to external properties, such as eclipsing binaries. This distinction is the first node on the variability tree (see Figure 12). While seemingly broad, this categorization can be quite insightful to those who are new to variable stars. From this, the tree narrows down the variability type even further. In particular, Figure 12, shows a path Intrinsic $\rightarrow$ Pulsation $\rightarrow$ RR Lyrae. Upon clicking the RR Lyrae node, the three

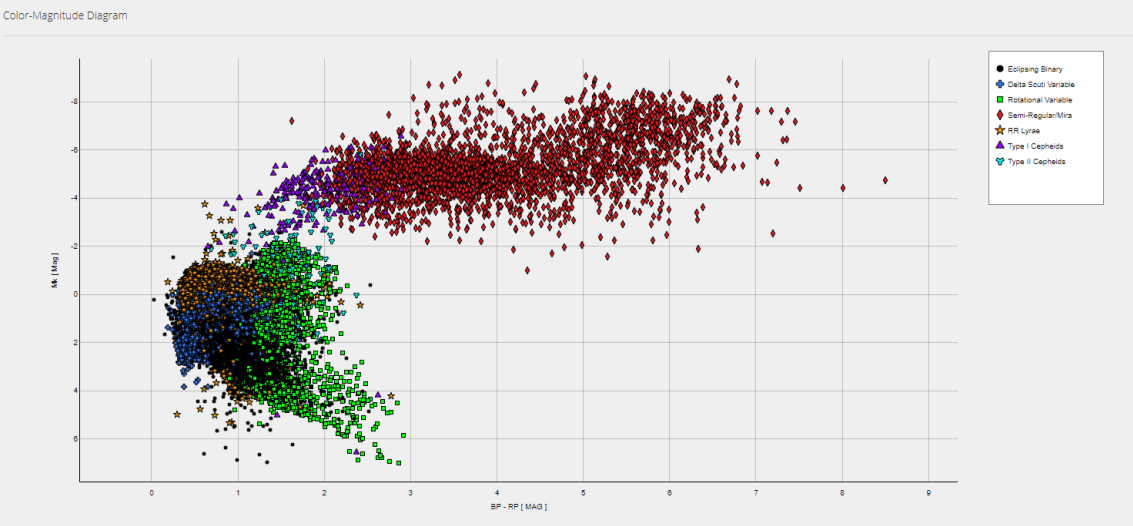

Figure 14: Color-Magnitude Diagram; ASAS-SN Variable Star Atlas

standard subtypes of this class, RRAB, RRC, and RRD, are shown. From here, the user can click on any of these nodes are see what their respective light curves look like. For example, Figure 13 shows the atlas destination for First Overtone RR Lyrae (RRC) types. Here, a short description of the variable class and notable light curve features are shown along with example light curves.

5.2 Visualizations

It is often enough to understand the visual morphology of light curves of a given subclass. However, understanding how other variables relate to one another can offer just as much insight. Visualizations are a power tool in any learning environment; in the ASAS-SN Variable Star Atlas, we will also provide additional materials in the form of helpful visuals. The first of which is a color-magnitude diagram (see Figure 14). Similar to a Hertzsprung–Russell diagram, a color-magnitude diagram (CMD) shows the relationship between magnitude (brightness) and color (temperature) for

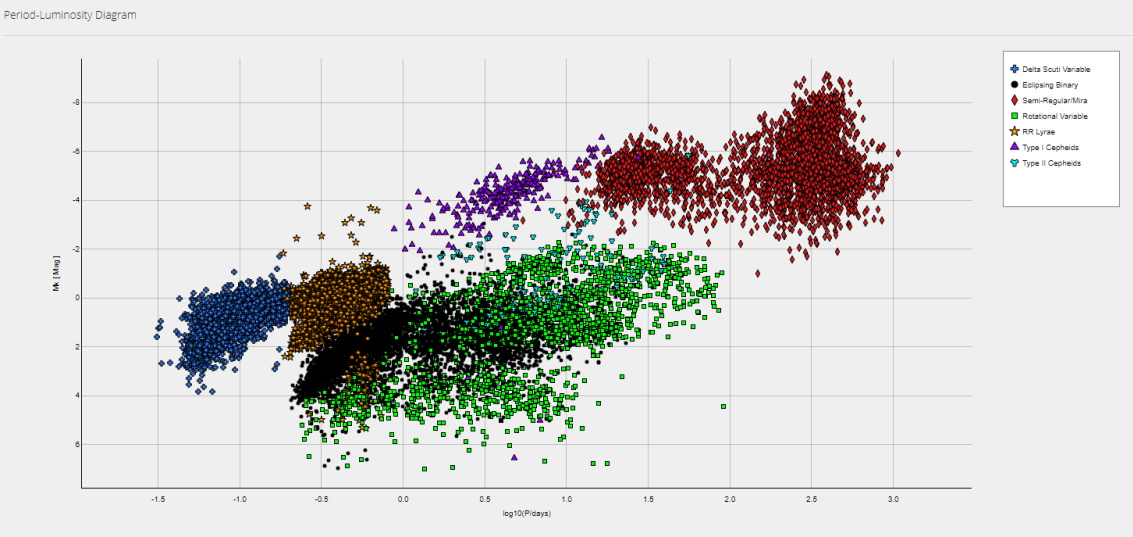

Figure 15: Period-Luminosity Diagram; ASAS-SN Variable Star Atlas

a given stellar population. In stellar astronomy, it is often helpful to see where a star lies on a color-magnitude diagram to understand where they are at in their evolution. In this diagram various types of variable stars occupy different regions in the CMD.

Another helpful visual is a period-luminosity plot. Many types of stellar variability are governed by a distinct period-luminosity relationship; Type I v.s. Type II Cepheids for example. Because of this, the Atlas was designed to also include a plot that illustrates how the brightness of different variable groups change with their period (see Figure 15).

6 Acknowledgements

I would like to thank Prof. Stanek, Tharindu Jayasinghe, and the rest of the ASAS-SN team for their guidance and support through my work on Citizen ASAS-SN and with our new g-band data. I would also like to thank Prof. Hughes for adequately

teaching me techniques related to machine learning. Lastly, I would like to thank Andrew Schneider for his incredible work on the ASAS-SN Variable Star Atlas.

References

Alard, C. 2000, , 144, 363, doi: 10.1051/aas:2000214

Alard, C., & Lupton, R. H. 1998, , 503, 325, doi: 10.1086/305984

Alhammady, H., & Ramamohanarao, K. 2004, in Fourth IEEE International Conference on Data Mining (ICDM'04), 315–318

Berdyugina, S. V. 2005, Living Reviews in Solar Physics, 2, 8, doi: 10.12942/lrsp-2005-8

Boyajian, T. S., LaCourse, D. M., Rappaport, S. A., et al. 2016, , 457, 3988, doi: 10.1093/mnras/stw218

Hall, D. S. 1976, The RS CVn Binaries and Binaries with Similar PROPERTIÉS, ed. W. S. Fitch, Vol. 60, 287, doi: 10.1007/978-94-010-1175-4_15

Hitchcock, J., Fossey, S. J., & Savini, G. 2020, arXiv e-prints, arXiv:2005.06569. https://arxiv.org/abs/2005.06569

Holoien, T.-S., Stanek, K. Z., Kochanek, C. S., et al. 2016, Monthly Notices of the Royal Astronomical Society, 464, 2672–2686, doi: 10.1093/mnras/stw2273

Jayasinghe, T., Stanek, K. Z., Kochanek, C. S., et al. 2019a, , 489, 4705, doi: 10.1093/mnras/stz2460

Jayasinghe, T., Kochanek, C. S., Stanek, K. Z., et al. 2018, , 477, 3145, doi: 10.1093/mnras/sty838

Jayasinghe, T., Stanek, K. Z., Kochanek, C. S., et al. 2019b, , 486, 1907, doi: 10.1093/mnras/stz844

—. 2019c, , 485, 961, doi: 10.1093/mnras/stz444

—. 2019d, Monthly Notices of the Royal Astronomical Society, 491, 13, doi: 10.1093/mnras/stz2711

Jayasinghe, T., Stanek, K. Z., Kochanek, C. S., et al. 2019, Monthly Notices of the Royal Astronomical Society, 486, 1907, doi: 10.1093/mnras/stz844

Jayasinghe, T., Stanek, K. Z., Kochanek, C. S., et al. 2020a, , 493, 4045, doi: 10.1093/mnras/staa518

—. 2020b, , 493, 4186, doi: 10.1093/mnras/staa499

Jayasinghe, T., Kochanek, C. S., Stanek, K. Z., et al. 2020c, arXiv e-prints, arXiv:2006.10057. https://arxiv.org/abs/2006.10057

—. 2020d, arXiv e-prints, arXiv:2006.10057. https://arxiv.org/abs/2006.10057

Józsa, G. I. G., Garrett, M. A., Oosterloo, T. A., et al. 2009, , 500, L33, doi: 10.1051/0004-6361/200912402

Kahraman Aliçavuş, F., Soydugan, E., Smalley, B., & Kubát, J. 2017, , 470, 915, doi: 10.1093/mnras/stx1241

Kochanek, C. S., Shappee, B. J., Stanek, K. Z., et al. 2017, , 129, 104502, doi: 10.1088/1538-3873/aa80d9

Masters, K. L., & Galaxy Zoo Team. 2020, in IAU Symposium, Vol. 353, Galactic Dynamics in the Era of Large Surveys, ed. M. Valluri & J. A. Sellwood, 205–212, doi: 10.1017/S1743921319008615

Mickaelian, A. M. 2015, Astronomical Surveys and Big Data. https://arxiv.org/abs/1511.07322

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2018, Scikit-learn: Machine Learning in Python. https://arxiv.org/abs/1201.0490

Sacco, G., Ngo, L. D., & Modolo, J. 2018, Journal of the American Association of Variable Star Observers (JAAVSO), 46, 14. https://arxiv.org/abs/1710.01081

Scargle, J. D. 1982, , 263, 835, doi: 10.1086/160554

Shallue, C. J., Lee, J., Antognini, J., et al. 2019, Measuring the Effects of Data Parallelism on Neural Network Training. https://arxiv.org/abs/1811.03600

Shappee, B. J., Prieto, J. L., Grupe, D., et al. 2014, , 788, 48, doi: 10.1088/0004-637X/788/1/48

Süveges, M., Bartholdi, P., Becker, A., et al. 2012, Springer Series in Astrostatistics, 253–261, doi: 10.1007/978-1-4614-3323-1_27

Thiemann, H. B., Norton, A. J., Dickinson, H. J., McMaster, A., & Kolb, U. C. 2021, Monthly Notices of the Royal Astronomical Society, doi: 10.1093/mnras/stab140

Thomas, S., Schuler, S., Fernanda Martinez, C., Cunha, K., & Smith, V. 2019, in American Astronomical Society Meeting Abstracts, Vol. 233, American Astronomical Society Meeting Abstracts #233, 247.30

Tonry, J. L., Denneau, L., Flewelling, H., et al. 2018, , 867, 105, doi: 10.3847/1538-4357/aae386

Trouille, L., Lintott, C. J., & Fortson, L. F. 2019, Proceedings of the National Academy of Sciences, 116, 1902, doi: 10.1073/pnas.1807190116

Watson, C. L., Henden, A. A., & Price, A. 2006, Society for Astronomical Sciences Annual Symposium, 25, 47

Way, Z., Stanek, K. Z., Kochanek, C. S., et al. 2019, The Astronomer's Telegram, 13346, 1

Zechmeister, M., & Kürster, M. 2009, , 496, 577, doi: 10.1051/0004-6361:200811296