Cyclic-Permutation Invariant Networks for Modeling Periodic Sequential Data: Application to Variable Star Classification

Keming Zhang*
Department of Astronomy
University of California at Berkeley
Berkeley, CA 94720
kemingz@berkeley.edu

Joshua S. Bloom
Department of Astronomy
University of California at Berkeley
Berkeley, CA 94720
joshbloom@berkeley.edu

Abstract

Current recurrent and convolutional neural network (RNN; CNN) architectures are sub-optimal for modeling periodic sequences because of their acyclic topology. Here, we present cyclic-permutation invariant networks, a new type of CNN for which the invariance to cyclic-permutations, or phase shifts, is guaranteed by representing data in phase-space and performing convolutions in polar coordinates, which we implement by means of "Symmetry Padding." Across three different datasets of periodic variable star light-curves (time-series of brightness) [1, 2, 3], we show that two particular implementations of the cyclic-permutation invariant network: the iTCN and the iResNet, consistently outperform both non-invariant versions of the same network, as well as previous state-of-the-art approaches including Random Forests. This methodology is first presented in [4] and is applicable to a wide range of science domains where periodic data abounds due to physical symmetries.

1 Introduction

Strictly periodic data is common in the physical sciences where periodicity occurs both spatially and temporally. Neural networks for which invariances arising from periodicity are explicitly considered either in the input features or the loss function, have been previously applied to particle physics [5] and atomic physics/chemistry [6]. In this study, we focus on periodic time series arising from variable stars, which is important to many sub-fields of astronomy, including the study of the cosmic distance ladder, galactic archaeology, and stellar evolution. Although different variability mechanisms underlie these objects, their light curves — time series of brightness versus time — are strictly periodic, with periods spanning from minutes to years. Until recently, random forest (RF) classification of hand-crafted features of periodic variable star light curves had been the state-of-the-art for variable star classification ([7]; [8]). Recently, RNNs have seen competitive performance compared to the RF techniques for periodic variable star classification ([9]; [10]; [11]). Although works including [9] took advantage of periodicity by period-folding the multi-cycle time series into a single cycle, i.e., from time space into phase space, none of them took advantage of the fact that classifications should be invariant to phase shifts, or formally cyclic permutations. Since the initial-phase of the period-folded light curve is completely determined by the phase at the first measurement, it is extrinsic to the physical source and thus irrelevant for classification.

In this work, we present cyclic-permutation invariant convolutional networks in which periodicity of the input data is directly embedded in the network architecture. The technique to embed

*Correspondence to: kemingz@berkeley.edu

Third Workshop on Machine Learning and the Physical Sciences (NeurIPS 2020), Vancouver, Canada.

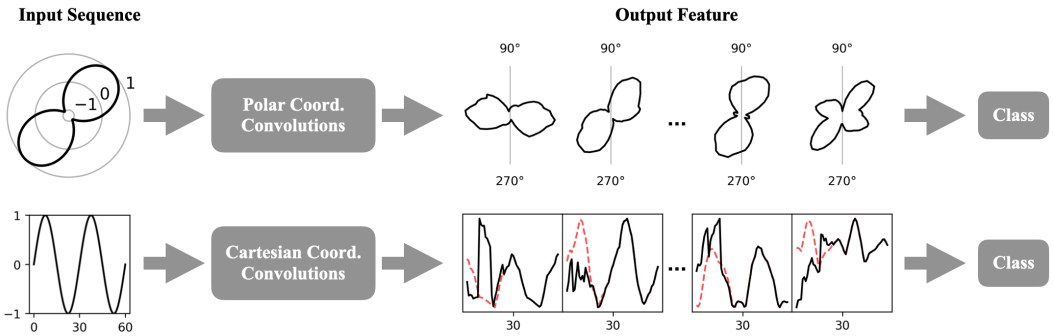


Figure 1: Schematic illustration of the effect of polar coordinate convolutions in preserving cyclic-permutation invariance. The input and output sequences are shown in polar coordinates for iTCN, and in Cartesian coordinates for TCN. The input sequence is a sine curve with two full oscillations in both cases. In the upper diagram, 1-D feature maps of the periodic input remains periodic; rotational symmetry is preserved. These periodic feature maps are also shown in Cartesian coordinates of the lower plots in red dashed lines for comparison. As demonstrated by the discrepancy, feature maps are distorted for the first full oscillation in the non-invariant network, which is shown in solid black lines.

cyclic-permutation invariance is generically applicable to all convolutional networks. We describe specific implementations with 1D residual convolutional networks (ResNets), and with dilated 1-D convolutions that have recently been shown to achieve state-of-the-art for a variety of sequence modeling tasks [12]. To facilitate applications of the cyclic-permutation invariant networks, we are releasing code at `https://github.com/kmzzhang/periodicnetwork`.

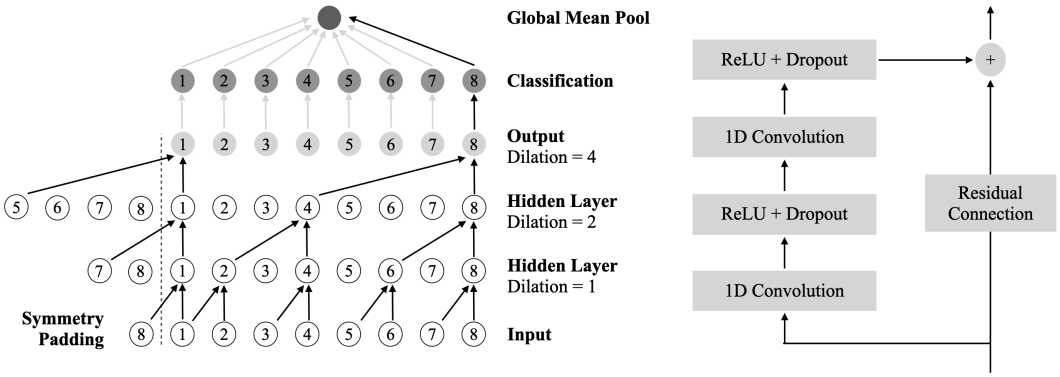
2 Method

We begin by emphasizing that cyclic-permutation invariant networks do not refer to any particular network architecture, but to any neural network satisfying the following condition. Given input sequence $x \in \mathbb{R}^{\mathrm{N}}$, network $f : x \to y$ is invariant to cyclic-permutation if

$$\forall i \in [1, N], f(x_{1:N}) = f(\mathrm{concat}(x_{i:N}, x_{1:i})) \tag{1}$$

Under the cyclic-permutation invariant network framework, the multi-cycle periodic time series is first period-folded into a single cycle by transforming from temporal space $(\vec{t}, \vec{m})$ into phase space $(\vec{\phi}, \vec{m})$: $\vec{\phi} = \vec{t} \mod p$, where m_i is the magnitude measurement at phase ϕ_i and p, the period, is determined with periodogram analysis [13, 14] and used as an auxiliary input. Then, rather than the typical Cartesian coordinates, convolutions are performed in polar coordinates such that the period-folded sequence is essentially wrapped in a "closed ring" (Figure 1: Input Sequence) whereby invariance to phase shifts, or "rotation," is naturally guaranteed, rendering the initial phase (ϕ_0) irrelevant. Indeed, since ϕ_0 is completely determined by the phase at the first observation, it is extrinsic to the physical source and thus should not affect the classification. We also note that such invariance, formally cyclic-permutation invariance, is forbidden in RNNs because of its acyclic topology. Polar coordinate convolution is implemented in CNNs by replacing zero-padding of length (`kernel` size $-$ 1) in ordinary Cartesian CNNs with "Symmetry Padding," which pads the input or hidden sequence not with zeros, but with the sequence itself (Figure 2a).

Based on the above framework, we present two particular implementations of the cyclic-permutation invariant network: the invariant Temporal Convolutional Network (iTCN) and the invariant Residual Convolutional Network (iResNet). The iTCN is based on the Temporal Convolutional Network (TCN; [12]), and is composed of "residual blocks" (Figure 2b) of 1D dilated convolutions (Figure 2a), where the input to each "residual block' is concatenated to the output, creating a "gradient highway" for back-propagation, thus allowing for improved network optimization. Dilated convolutions refer to convolutions where the convolution `kernel` is applied over a region larger than the `kernel` size by skipping input values with a step of 2^{n-1} for the n-th layer. This dilation allows the network to achieve an exponential increase in the receptive field — the extent of input data accessible with respect to a particular output neuron — with network depth. The receptive field is calculated as



(a) iTCN **(b)** Residual Block

Figure 2: (a) Simplified illustration of the cyclic-permutation invariant Temporal Convolutional Network (iTCN). Numbers refer to the ordering of the period-folded sequence. Dilated convolutions are represented by arrows where the dilation factor is indicated to the right of each layer. Gray arrows in the final two layers represent operations which are present only in the iTCN not the TCN. The classification layer consists of two convolutions of `kernel` size 1. (b) The residual block, which is the actual hidden layer used in the iTCN. Residual connections are to be replaced with $k = 1$ convolutions when consecutive layers have different hidden dimensions.

$L = (K-1) \times \sum_{n=1}^{D} 2 \times 2^{n-1} = (K-1) \times (2^{D+1} - 2)$, where K is the kernel size, D the number of layers, 2^{n-1} the dilation factor for the nth layer, and the additional factor of 2 due to the fact that each residual block consists of two dilated convolutions. Network depth is required to be large enough for the receptive field to be larger than the input sequence length, such that each feature vector in the output layer has complete information over the input sequence. Simultaneous predictions are then made for every initial phase of the input sequence (Figure 2a: "classification" layer), which are averaged with a `global mean pooling` layer as input to the `softmax` function for classification. By averaging predictions from all possible initial phases, the invariant network makes more robust predictions, as compared to non-invariant CNNs and RNNs, which can only predict for one particular initial phase with one network forward pass.

To visualise the effects of cyclic-permutation invariance on modelling periodic sequences, we compare output sequences produced by the iTCN and the TCN in Figure 1. We create an iTCN and a TCN with the same weights and the same receptive field of 30 at a network depth of 4 with kernel size 2. The input sequence is a length-60 sine function with two full oscillations (0 to 4π). As seen in the figure, while output feature maps produced by the iTCN remain symmetrical in polar coordinates, the first half of the output sequence produced by the TCN is distorted by zero-padding, thus degrading the fidelity of output feature maps.

The second implementation, the iResNet, is also composed of stacks of "residual blocks," but is different from the iTCN in that the exponential receptive-field increase is achieved through `max pooling` layers, instead of dilated convolutions. A `max pooling` layer of `kernel` size 2 and `stride` 2, which combines every two adjacent feature vectors into one by selecting the maximum value, is added after every "residual block" to distil information extracted. After every such operation, the temporal dimension is reduced by half, while the number of hidden dimension is doubled until a specified upper limit. Unlike the iTCN, the feature vectors in the iResNet output layer do not have a one-to-one correspondence to the input sequence because the temporal dimension of the output feature map is reduced by a factor of 2^{D-1}, where D is the network depth. Because of this discreteness of featurisation, the iResNet is only invariant to phase shifts of 2^{D-1} steps (when the input sequence length is divisible by the same factor). Nevertheless, the iResNet potentially benefit from data augmentation of the input-sequence initial phase, as could non-invariant networks.

3 Results

We first perform an ablation study where the iTCN/iResNet is compared to the TCN/ResNet to identify the gains made solely by cyclic-permutation invariance. Input data is given in phase-space in

Model	MACHO	OGLE-III	ASAS-SN
iTCN	**92.7%** $\pm 0.43\%$	**93.7%** $\pm 0.09\%$	**94.5%** $\pm 0.14\%$
TCN	$92.0\% \pm 0.35\%$	$92.9\% \pm 0.11\%$	$93.0\% \pm 0.20\%$
$-^1$	$(-0.58\%^{+0.05\%}_{-0.17\%})$	$(-0.75\%^{+0.04\%}_{-0.02\%})$	$(-1.52\%^{+0.12\%}_{-0.09\%})$
iResNet*	$92.6\% \pm 0.45\%$	**93.7%** $\pm 0.09\%$	$93.9\% \pm 0.14\%$
ResNet	$92.1\% \pm 0.34\%$	$93.4\% \pm 0.11\%$	$93.2\% \pm 0.22\%$
$-^1$	$(-0.48\%^{+0.05\%}_{-0.13\%})$	$(-0.25\%^{+0.03\%}_{-0.02\%})$	$(-0.64\%^{+0.01\%}_{-0.07\%})$
GRU	$92.3\% \pm 0.37\%$	$92.8\% \pm 0.19\%$	$93.6\% \pm 0.42\%$
$-^2$	$(-0.34\%^{+0.05\%}_{-0.02\%})$	$(-0.86\%^{+0.17\%}_{-0.13\%})$	$(-0.71\%^{+0.07\%}_{-0.12\%})$
LSTM	$91.7\% \pm 0.53\%$	$92.6\% \pm 0.61\%$	$93.5\% \pm 0.23\%$
$-^2$	$(-0.93\%^{+0.45\%}_{-0.16\%})$	$(-0.85\%^{+0.17\%}_{-0.48\%})$	$(-1.00\%^{+0.11\%}_{-0.06\%})$

[1] Absolute difference compared to the invariant version of the same network.

[2] Absolute difference compared to the best performing network.

* Semi-invariant due to use of discrete max-pooling layers.

Table 1: **Ablation study test accuracies demonstrating gains afforded by cyclic-permutation invariance**. Test accuracies are the mean values for 8 different data splits. The network with the top accuracy for each dataset is shown in bold. We apply randomised, stratified 60/20/20 train/validation/test splits for each dataset. To properly account for dataset boot-strapping noise—accuracy variations due to the particular choices of data splits—the same random splits are used to test every network, whose accuracies are compared pairwise in splits. Median test accuracy differences of the different data partitions are shown in parentheses with the uncertainty interval corresponding to 1-σ range of test accuracy differences calculated pair-wise for the same random partitions of data (see footnote above). The standard deviation of the accuracy differences, rather than the accuracy themselves, should then serve as the basis for comparison. Within each split, each full light-curve is divided into sequences of length 200 in temporal order and transformed into phase-space with respect to the period provided in the catalogs. Compared to random sampling, subdividing by temporal order preserves the irregular samplings which resemble how data is accumulated. Each segment is then individually normalised (zero-mean and unit variance) while measurement times are rescaled by the period into phase ([0, 1]). The measurement phase intervals $\Delta\vec{\phi}$ are fed together with the rescaled light curve as inputs to the network. The mean and standard deviation of each light curve segment, along with $\log(p)$, are concatenated to the network output layer as auxiliary inputs. We perform extensive hyperparameter optimisation for each pair of network and dataset (see Appendix A).

all cases as the advantages compared to time-space have been demonstrated in previous studies [9]. RNN baselines of GRU [15] and LSTM [16] are also included as additional baseline methods for comparison. The networks are trained on fixed-length light-curve segments ($L = 200$) from three periodic variable star datasets (see Appendix B): the All-Sky Automated Survey for Supernovae (ASAS-SN; [17, 2]), the Massive Compact Halo Object (MACHO; [3]), and the Optical Gravitational Lensing Experiment (OGLE-III; [1]). Details of the ablation study are provided in the Table 1 caption. As shown in Table 1, the improvements of the iTCN and the iResNet from their respective non-invariant baselines, as well as the RNNs, are significant by more than 5-σ in most cases, demonstrating the advantages of cyclic-permutation invariance. The improvements in classification accuracies correspond to reductions in overall error rates by between 4% to 22%, depending upon the non-invariant baseline and the dataset.

We then compare the iTCN/iResNet with published results in the literature, which include time-space RNNs and Random Forests. Becker et al. [11] presents OGLE-III classification results with their time-space GRU and a Random Forest baseline [18]. As seen in Table 2, the cyclic-permutation invariant networks outperform both results. The invariant network accuracies are significantly higher for most classes, reducing error rates by as much as 69% for the minority classes. We find this result to be critically important, as the hard-to-classify minority classes are tend to be the least well-understood and often are the most interesting to identify for further study. Additionally, Naul et al. [9] published Random Forest benchmark accuracies for the MACHO dataset of 90.50% with the Richards et al. features [7] and 88.98% with the Kim/Bailer-Jones features [8], although these benchmarks would have lower accuracies if preformed strictly within our data partitioning scheme (see Appendix B).

Class	iTCN	iResNet	GRU[1]	RF[2]
Cep	**98.3%**±0.3%	**98.4%**±0.7%	72%	97%
RRab	**99.7%**±0.1%	**99.7%**±0.4%	85%	99%
RRc	**99.0%**±0.2%	**99.1%**±0.1%	30%	98%
Dsct	**97.6%**±0.8%	**97.8%**±0.6%	72%	93%
EC	**87.9%**±0.9%	**87.8%**±0.7%	54%	79%
ED	**95.0%**±0.3%	**94.8%**±0.4%	93%	92%
ESD	68.7%±1.0%	**70.7%**±0.9%	24%	61%
Mira	**97.1%**±0.6%	**96.8%**±0.3%	92%	**97%**
SRV	**96.0%**±0.4%	**95.9%**±0.2%	93%	82%
OSARG	93.2%±0.4%	93.4%±0.2%	90%	**97%**
Mean	**93.4%**	**93.3%**	70.5%	89.5%

[1]Time-space GRU with grouping moving window. [11]

[2]Random Forest with 59 features generated from the light curve data. [11, 18]

Table 2: **Test accuracies for OGLE-III full-length light curves compared to classifications from the literature**. The Becker et al. time-space GRU work groups each full OGLE-III light curve with a moving window of size 50 and stride 25, whereby the effective sequence length is reduced by a factor of 25, thus alleviating the vanishing gradient problem. To facilitate this comparison, we have used the same OGLE-III data selection as their work (see Methods). Since a $L > 300$ requirement has been applied to their OGLE-III data selection, we trained the iTCN/iResNet on $L = 300$ segments, and average classifications on $L = 300$ segments for each full light curve during testing. For all but one subclass, the cyclic-permutation invariant networks outperform previous results. Similar to Table 1, we note that uncertainties are dominated by the bootstrapping noise arising from randomized data partitioning, and as such, are only upper limits to uncertainties in the accuracy differences for each class.

4 Conclusion

We have presented a summary of cyclic-permutation invariant networks [4], which explicitly embed temporal periodicity in CNN architectures. Such networks achieve state-of-the-art performance in variable star classification, and is applicable to periodic data in a wide range of science domains.

Broader impact

The development of robust automatic classification systems in the next generation public astronomical surveys enables researchers with less access to expensive computational facilities to more easily acquire structured data. It also allows a broader participation in observational science which is previously restricted to those with (usually exclusive) access to observing resources.

Acknowledgments

K.Z. and J.S.B are supported by a Gordon and Betty Moore Foundation Data-Driven Discovery grant. K.Z. thanks the LSSTC Data Science Fellowship Program, which is funded by LSSTC, NSF Cybertraining Grant 1829740, the Brinson Foundation, and the Moore Foundation; his participation in the program has benefited this work. This work is supported by the AWS Cloud Credits for Research program.

Appendix

A Neural Network Hyper-Parameter Optimization

We search for optimal hyper-parameters independently for each network and for each dataset with the validation set in a fixed train/validation/test split. For all networks, among possible combinations of input features: phase interval ($\Delta\vec{\phi}$), magnitude ($\vec{m}$), magnitude change ($\Delta\vec{m}$), and gradient ($\Delta\vec{m}/\Delta\vec{\phi}$), we find the combination of $(\Delta\vec{\phi}, \vec{m})$ to yield the highest validation accuracy. For iTCN/TCN, we perform a hyper-parameter search over network depth (6, 7), hidden dimension (12,

24, 48), dropout (0, 0.15, 0.25), and kernel size (iTCN/TCN: 2, 3, 5). For iResNet/ResNet, we perform a grid search over initial hidden dimension (16, 32), maximum hidden dimension (32, 64), network depth (4, 5, 6), and kernel size (3, 5, 7). For GRU/LSTM, we search over network depth (2, 3), hidden dimension (12, 24, 48), and dropout rate (0, 0.15, 0.25). We find that a dropout rate of 0.15 works best for both GRUs and LSTMs across all three datasets, while no dropout works best for all other networks.

All networks are trained with the ADAM optimiser [19] with initial learning rates of 0.005, which are scheduled to decrease by a factor of 0.1 when training loss does not decrease by 10% for 5 epochs. Models are saved at the best validation accuracy for testing.

B Data

B.1 All-Sky Automated Survey for Supernovae (ASAS-SN) Data

The ASAS-SN dataset consists of 282,795 light curves from eight classes of variable stars: 288 W Virginis ($p > 8$ day), 102 W Virginis ($p < 8$ day), 941 Classical Cepheids, 297 Classical Cepheids (Symmetrical), 1,631 Delta Scuti, 25,314 Detached Eclipsing Binaries, 12,601 Beta Lyrae, 43,151 W Ursae Majoris-type, 2,149 High Amplitude Delta Scuti, 9,623 Delta Scuti, 14046 Rotational Variables, 26,956 RR Lyrae type A/B, 7,469 RR Lyrae type C, 364 RR Lyrae type D, and 137,847 Semi-regular Variables. The class label of each variable star is classified by [20] and only those with class probability greater than 99% are used. The maximum number of full light curve per class is capped at 20,000 to reduce the number of light curves of the dominant classes. Finally, segmenting into $L = 200$ chunks results in 106,005 fixed-length light curves.

B.2 Massive Compact Halo Object (MACHO) Project data

The MACHO dataset consists of 21,470 red band light curves from eight classes of variable stars: 7,403 RR Lyrae AB, 6,833 Eclipsing Binary, 3,049 Long-Period Variable Wood (sub-classes A–D were combined into a single super-class), 1,765 RR Lyrae C, 1,185 Cepheid Fundamental, 683 Cepheid First Overtone, 315 RR Lyrae E, and 237 RR Lyrae/GB Blend. Segmenting into $L = 200$ chunks has resulted in 80,668 fixed-length light curves. Although we have use the same MACHO dataset as [9], our results are not directly comparable because [9] preformed randomised train/test split on the $L = 200$ segmented light curves, which have caused different versions of the same light curve to exist in both training and test split, resulting in information leakage and thus a higher accuracy.

B.3 Optical Gravitational Lensing Experiment: OGLE-III

The OGLE-III dataset is identical to that used in [11], except for the selection of OSARGs. The OGLE-III data consists of 357,748 light curves from ten classes of variable stars: 6862 Eclipsing Contact Binaries, 21503 Eclipsing Detached Binaries, 9475 Eclipsing Semi-detached Binaries, 6090 Miras, 234,932 Small Amplitude Red Giants (OSARG), 25943 RR Lyrae type A/B, 7990 RR Lyrae type C, 34835 Semi-regular Variables, 7836 Classical Cepheids, 2822 Delta Scuti. Of the 234,932 OSARGs, 40,000 random ones are selected. Absent of a fixed random seed in their relevant code section, we have not been able to procure their exact selection, although the number selected is large enough for the difference to be small. Finally, segmenting into $L = 200$ chunks results in 540,457 fixed-length light-curves.

References

[1] A. Udalski. The Optical Gravitational Lensing Experiment. Real Time Data Analysis Systems in the OGLE-III Survey. *Acta Astronomica*, 53:291–305, December 2003. ISSN 0001-5237. URL http://adsabs.harvard.edu/abs/2003AcA....53..291U.

[2] T. Jayasinghe, C. S. Kochanek, K. Z. Stanek, B. J. Shappee, T. W.-S. Holoien, Toda A. Thompson, J. L. Prieto, Subo Dong, M. Pawlak, J. V. Shields, G. Pojmanski, S. Otero, C. A. Britt, and D. Will. The ASAS-SN catalogue of variable stars I: The Serendipitous Survey. *Monthly Notices of the Royal Astronomical Society*, 477(3):3145–3163, July 2018. ISSN 0035-8711. doi: 10.1093/mnras/sty838. URL https://academic.oup.com/mnras/article/477/3/3145/4961151.

[3] C. Alcock, R. A. Allsman, T. S. Axelrod, D. P. Bennett, K. H. Cook, K. C. Freeman, K. Griest, S. L. Marshall, B. A. Peterson, M. R. Pratt, P. J. Quinn, A. W. Rodgers, C. W. Stubbs, S. Sutherland, and D. L. Welch. The MACHO Project LMC Variable Star Inventory.II.LMC RR Lyrae Stars- Pulsational Characteristics and Indications of a Global Youth of the LMC. *The Astronomical Journal*, 111:1146, March 1996. ISSN 0004-6256. doi: 10.1086/117859. URL http://adsabs.harvard.edu/abs/1996AJ....111.1146A.

[4] Keming Zhang and Joshua S. Bloom. Classification of Periodic Variable Stars with Novel Cyclic-Permutation Invariant Neural Networks. *arXiv:2011.01243 [astro-ph, physics:physics]*, November 2020. URL http://arxiv.org/abs/2011.01243. arXiv: 2011.01243.

[5] Steffen Maeland and Inga Strümke. Deep Learning with Periodic Features and Applications in Particle Physics. In Hien Nguyen, editor, *Statistics and Data Science*, Communications in Computer and Information Science, pages 140–147, Singapore, 2019. Springer. ISBN 9789811519604. doi: 10.1007/978-981-15-1960-4_10.

[6] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller. SchNet – A deep learning architecture for molecules and materials. *The Journal of Chemical Physics*, 148(24):241722, March 2018. ISSN 0021-9606. doi: 10.1063/1.5019779. URL https://aip.scitation.org/doi/full/10.1063/1.5019779.

[7] Joseph W. Richards, Dan L. Starr, Nathaniel R. Butler, Joshua S. Bloom, John M. Brewer, Arien Crellin-Quick, Justin Higgins, Rachel Kennedy, and Maxime Rischard. On Machine-learned Classification of Variable Stars with Sparse and Noisy Time-series Data. *The Astrophysical Journal*, 733:10, May 2011. ISSN 0004-637X. doi: 10.1088/0004-637X/733/1/10. URL https://iopscience.iop.org/article/10.1088/0004-637X/733/1/10.

[8] Dae-Won Kim and Coryn A. L. Bailer-Jones. A package for the automated classification of periodic variable stars. *Astronomy & Astrophysics*, 587:A18, March 2016. ISSN 0004-6361, 1432-0746. doi: 10.1051/0004-6361/201527188. URL https://www.aanda.org/articles/aa/abs/2016/03/aa27188-15/aa27188-15.html.

[9] Brett Naul, Joshua S. Bloom, Fernando Pérez, and Stéfan van der Walt. A recurrent neural network for classification of unevenly sampled variable stars. *Nature Astronomy*, 2(2):151–155, February 2018. ISSN 2397-3366. doi: 10.1038/s41550-017-0321-z. URL http://arxiv.org/abs/1711.10609. arXiv: 1711.10609.

[10] Benny T.-H. Tsang and William C. Schultz. Deep Neural Network Classifier for Variable Stars with Novelty Detection Capability. *The Astrophysical Journal*, 877(2):L14, May 2019. ISSN 2041-8213. doi: 10.3847/2041-8213/ab212c. URL http://arxiv.org/abs/1905.05767. arXiv: 1905.05767.

[11] I. Becker, K. Pichara, M. Catelan, P. Protopapas, C. Aguirre, and F. Nikzat. Scalable End-to-End Recurrent Neural Network for Variable Star Classification. *Monthly Notices of the Royal Astronomical Society*, February 2020. doi: 10.1093/mnras/staa350. URL https://academic.oup.com/mnras/advance-article/doi/10.1093/mnras/staa350/5728517.

[12] Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. *arXiv:1803.01271 [cs]*, April 2018. URL http://arxiv.org/abs/1803.01271. arXiv: 1803.01271.

[13] N. R. Lomb. Least-squares frequency analysis of unequally spaced data. *Astrophysics and Space Science*, 39(2):447–462, February 1976. ISSN 0004-640X, 1572-946X. doi: 10.1007/BF00648343. URL http://link.springer.com/10.1007/BF00648343.

[14] J. D. Scargle. Studies in astronomical time series analysis. II - Statistical aspects of spectral analysis of unevenly spaced data. *The Astrophysical Journal*, 263:835, December 1982. ISSN 0004-637X, 1538-4357. doi: 10.1086/160554. URL http://adsabs.harvard.edu/doi/10.1086/160554.

[15] Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1724–1734, Doha, Qatar, October 2014. Association for Computational Linguistics. doi: 10.3115/v1/D14-1179. URL https://www.aclweb.org/anthology/D14-1179.

[16] Sepp Hochreiter and Jürgen Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9 (8):1735–1780, November 1997. ISSN 0899-7667. doi: 10.1162/neco.1997.9.8.1735. URL https://www.mitpressjournals.org/doi/10.1162/neco.1997.9.8.1735.

[17] B. J. Shappee, J. L. Prieto, D. Grupe, C. S. Kochanek, K. Z. Stanek, G. De Rosa, S. Mathur, Y. Zu, B. M. Peterson, R. W. Pogge, S. Komossa, M. Im, J. Jencson, T. W.-S. Holoien, U. Basu, J. F. Beacom, D. M. Szczygieł, J. Brimacombe, S. Adams, A. Campillay, C. Choi, C. Contreras, M. Dietrich, M. Dubberley, M. Elphick, S. Foale, M. Giustini, C. Gonzalez, E. Hawkins, D. A. Howell, E. Y. Hsiao, M. Koss, K. M. Leighly, N. Morrell, D. Mudd, D. Mullins, J. M. Nugent, J. Parrent, M. M. Phillips, G. Pojmanski, W. Rosing, R. Ross, D. Sand, D. M. Terndrup, S. Valenti, Z. Walker, and Y. Yoon. The Man behind the Curtain: X-Rays Drive the UV through NIR Variability in the 2013 Active Galactic Nucleus Outburst in NGC 2617. *The Astrophysical Journal*, 788:48, June 2014. ISSN 0004-637X. doi: 10.1088/0004-637X/788/1/48. URL https://iopscience.iop.org/article/10.1088/0004-637X/788/1/48.

[18] Isadora Nun, Pavlos Protopapas, Brandon Sim, Ming Zhu, Rahul Dave, Nicolas Castro, and Karim Pichara. FATS: Feature Analysis for Time Series. *arXiv:1506.00010 [astro-ph]*, August 2015. URL http://arxiv.org/abs/1506.00010. arXiv: 1506.00010.

[19] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980 [cs]*, December 2014. URL http://arxiv.org/abs/1412.6980. arXiv: 1412.6980.

[20] T. Jayasinghe, K. Z. Stanek, C. S. Kochanek, B. J. Shappee, T. W.-S. Holoien, Todd A. Thompson, J. L. Prieto, Subo Dong, M. Pawlak, O. Pejcha, J. V. Shields, G. Pojmanski, S. Otero, C. A. Britt, and D. Will. The ASAS-SN catalogue of variable stars – II. Uniform classification of 412 000 known variables. *Monthly Notices of the Royal Astronomical Society*, 486(2):1907–1943, June 2019. ISSN 0035-8711. doi: 10.1093/mnras/stz844. URL https://academic.oup.com/mnras/article/486/2/1907/5420510.