Contrib. Astron. Obs. Skalnaté Pleso **54/2**, 167 – 170, (2024)
https://doi.org/10.31577/caosp.2024.54.2.167

Deep-learning classification of eclipsing binaries

Š. Parimucha[1], P. Gajdoš[1,2], Y. Markus[1,3] and V. Kudak[4]

[1] *Institute of Physics, Faculty of Science, Pavol Jozef Šafárik University, 040 01 Košice, Slovakia*

[2] *Astronomical Institute of the Czech Academy of Sciences 251 65 Ondřejov, The Czech Republic*

[3] *International Center of Astronomical, Medical, and Ecological Research, NASU, Kyiv, Ukraine*

[4] *Laboratory of Space Researches, Uzhhorod National University, Uzhhorod, Ukraine*

Received: October 31, 2023; Accepted: December 18, 2023

Abstract. We present a deep-learning model for the classification of eclipsing binaries. Our classifier provides a tool for the categorization of light curves of eclipsing binaries into four classes: detached systems with and without spots, and over-contact systems with and without spots. The classifier was trained on 200 000 synthetic light curves created using ELISa code. We randomly selected 100 light curves from the GAIA catalogue, which were fitted for evaluation purposes, and their morphologies were determined. We tested several classifiers and found that the best-performing classifier combined a Long Short-Term Memory (LSTM) layer and two one-dimensional convolutional neural networks. The precision from the evaluation set was 97% compared with the predicted precision of 94 % for the validation of synthetic data. Our classifier is more likely to successfully process data from subsequent large observational surveys.

Key words: Stars: binaries: eclipsing – deep-learning

1. Introduction

Eclipsing binaries (EBs) are a well-known group of variable stars in which light variations are caused by the mutual obscuration of stars with respect to the observer. It produces typical light curves where valuable information about the physical properties of the stars within the system is coded, like the sizes and shapes of the stars, mass ratio, relative temperatures, and the inclination of the orbital plane. Moreover, the light curves of many EBs show irregularities caused by spot(s) on one or both components (Hilditch, 2001; Prša, 2018).

EBs can be classified in two different ways. The first is the morphological division into three classes (Algol, β Lyr, and W UMa) based only on the shape of the light curve. The second is based on the amount of Roche lobe filling in the binary systems (Wilson, 1994; Kallrath & Milone, 2009), where three classes are listed (detached, semi-detached, and over-contact).

From a geometrical point of view, the semi-detached system is, in principle, detached; both components can be described by a convex surface (Čokina et al., 2021). Moreover, to model such a system, we must know two potentials (Kallrath & Milone, 2009). Therefore, we divided the EBs for machine-learning purposes into four groups: 0 – over-contact without spots, 1 – over-contact with spots, 2 – detached without spots and 3 – detached with spots.

This approach will allow us to classify possibly all EBs, which were found in large surveys e.g. GAIA or KEPLER will also be found in prepared surveys, such as Vera Rubin (LSST). Sorting into these groups will enable the use of different approaches to determine other parameters (physical and geometrical) of EBs using machine-learning approaches.

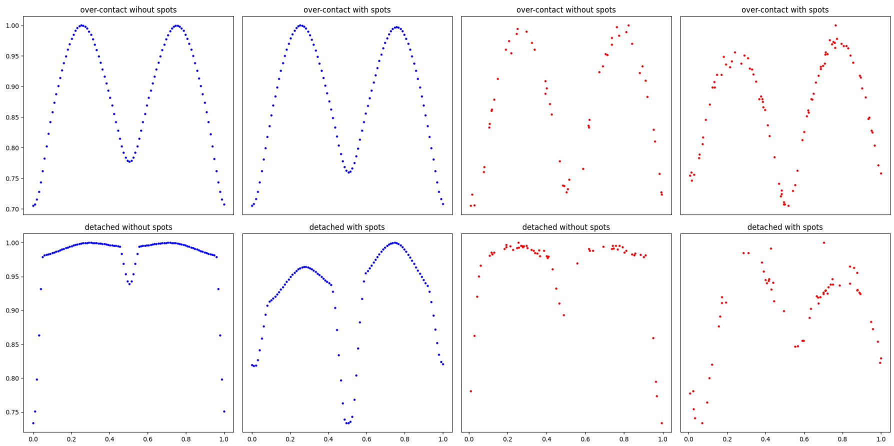

Figure 1. Examples of synthetic (blue) and GAIA (red) light curves from different groups.

2. Training and evaluation dataset

Deep-learning classification requires a large labelled training dataset, where all groups of objects have equal representations. In the analysis of EBs, we can create such datasets using software packages dedicated to EB modelling. In our study, the training dataset was created using the ELISa code (Čokina et al., 2021). For each group, we simulated 50 000 light curves created from the parameters covering a wide range of physically correct values for a specific group. Each light curve was represented by 100 data points of normalized flux phased to $< 0, 1 >$ interval.

To evaluate our model, we randomly selected 100 EB light curves published in the GAIA-DR3 catalogue (Mowlavi et al., 2023). All these light curves were fitted using the ELISa code to determine the basic parameters of the systems and eventually spot(s) on the components, as well as to determine their morphology. Examples of synthetic and GAIA light curves are shown in Fig. 1.

3. Deep-learning model and its performance

We tested several classification models and found that the best performance was achieved by combining two one-dimensional convolutional neural networks (CNN) and a Long Short-Term Memory (LSTM) layer. We used Adam optimizer and sparse categorical cross-entropy loss function (Chollet et al., 2015).

The training of our model was performed for 10 epochs, during which the loss and accuracy were determined. We randomly selected 20% of the training light curves for the validation dataset. Using this, the predicted precision was calculated and a confusion matrix was created. The predicted precision of the validation data is 94%. A detailed inspection of the confusion matrix (Fig. 2) reveals that our model misclassified (on the level of approximately 10%) the spotted light curves for both the detached and contact systems. This is probably because small spots cause only small changes in the light curve, and the model is unable to recognize the changes. To solve this issue, a new, more complicated model should be trained on a much larger dataset with a better coverage of spot parameters.

We used an evaluation dataset with GAIA light curves to test our model using real data. The precision of the model is 97%. One overcontact system without spots was misclassified as a detached system with spots and two detached systems without spots were misclassified as detached systems with spots. This is probably because of the relatively poor data quality of the GAIA light curves (outliers and noise).

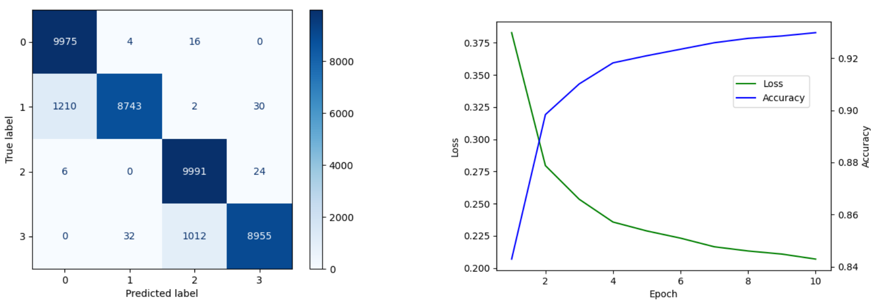

Figure 2. Confusion matrix (left) and performance of the model depending on epoch (right).

To achieve better results, it is necessary to process observational data better and train the model on more diverse synthetic data (wider range of system parameters and spots, different levels of noise, and/or adding outliers to synthetic data). Nevertheless, our classifier is now more likely to be applicable to GAIA data and promising for data analysis from large observational surveys, such as the Vera Rubin Telescope (LSST).

Acknowledgements. This research was supported by the Slovak Research and Development Agency under Contract No. APVV-20-0148. Y.M. was funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under project No. 09I03-03-V01-00011. V.K. was supported by Ukrainian National Grant 0122U000937.

References

Chollet, F. et al. 2015, Keras, https://github.com/fchollet/keras

Hilditch, R. W. 2001, *An Introduction to Close Binary Stars* (Cambridge University Press)

Kallrath, J. & Milone, E. 2009, *Eclipsing Binary Stars: Modeling and Analysis* (New York: Springer-Verlag)

Mowlavi, N., Holl, B., Lecoeur-Taïbi, I., et al., Gaia Data Release 3. The first Gaia catalogue of eclipsing-binary candidates. 2023, *Astronomy and Astrophysics*, **674**, A16, DOI: 10.1051/0004-6361/202245330

Prša, A. 2018, *Modeling and Analysis of Eclipsing Binary Stars; The theory and design principles of PHOEBE* (IOP Publishing)

Čokina, M., Fedurco, M., & Parimucha, Š., ELISa: A new tool for fast modelling of eclipsing binaries. 2021, *Astronomy and Astrophysics*, **652**, A156, DOI: 10.1051/0004-6361/202039171

Wilson, R. E., Binary Star Light-Curve Models. 1994, *Publications of the ASP*, **106**, 921, DOI: 10.1086/133464