Classification of light curves of eclipsing binaries using deep-learning models

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

[1] Institute of Physics, Faculty of Science, Pavol Jozef Šárik University, Košice, Slovakia, stefan.parimucha@upjs.sk
[2] Astronomical Institute of the Czech Academy of Sciences, Ondřejov, The Czech Republic
[3] Astronomical Institute of Slovak Academy of Sciences, Taranská Lomnica, Slovakia
[4] Laboratory of Space Researches, Uzhhorod National University, Uzhhorod, Ukraine

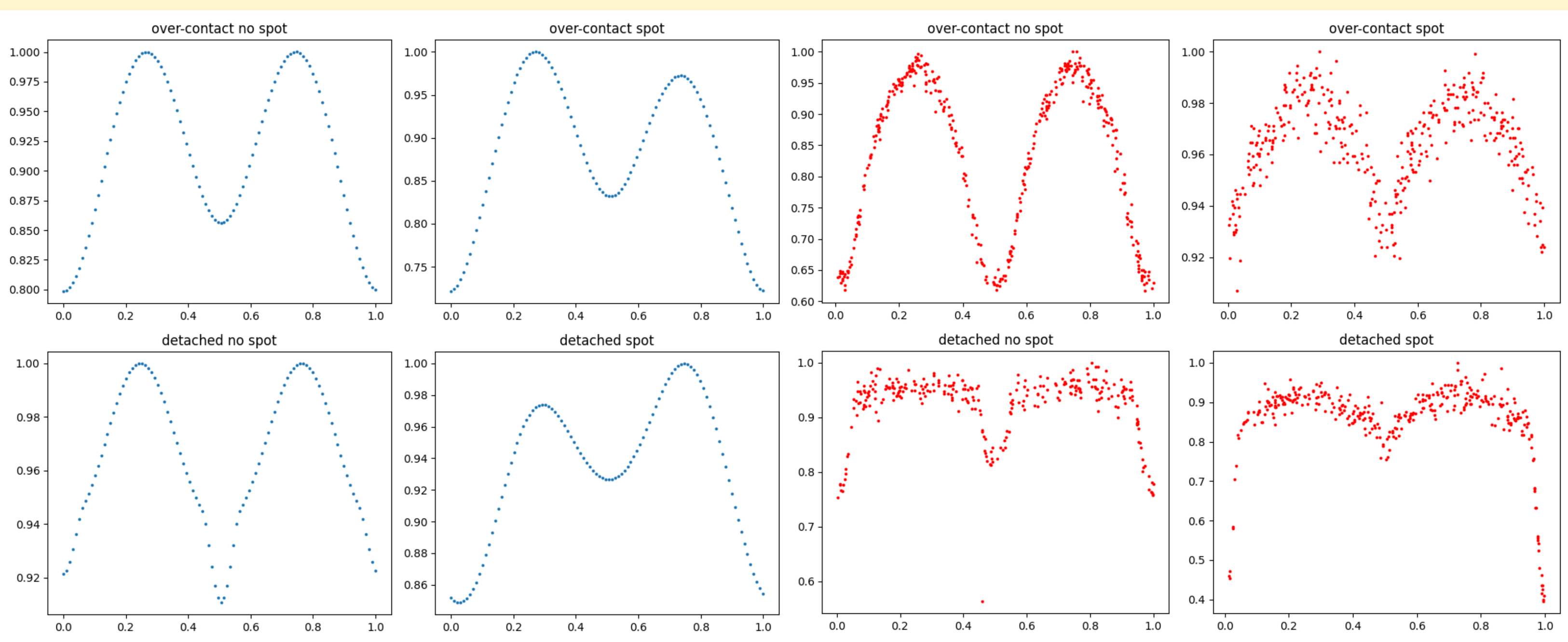
Eclipsing binaries (EBs) – binary stars, where mutual eclipses of components cause typical light curves. From LCs we can determine many parameters of the system and stars – their sizes and shapes, potentials, photometric mass ratio, relative temperatures, inclination of the orbital plane, presence of irregularities caused by spot(s) and/or pulsations

For deep-learning classification we divided EBs into 4 groups (Parimucha et al., 2024), overcontact and detached with and without spots Using this approach allows us to classify possibly all EBs, which were found in large surveys e.g. GAIA, KEPLER or TESS and also will be found in prepared surveys like Vera Rubin Observatory (LSST).

Sorting EBs into those groups is important for another study of these systems – we can use different machine learning approaches to determine other basic parameters

TRAINING AND EVALUATION DATASET

- training dataset was created with ELISa code (Čokina et al, 2021)
- light curves were simulated for overcontact and detached binaries from parameters covering a wide range of physically correct values of stars, system and spot parameters for a specific group, together with the orbital period. Altogether we created more than 2,5 million LCs.
- for evaluation dataset OGLE Eclipsing and Binary Stars Catalog (Soszynski et al., 2016; Pawlak et al., 2016) was used. EBs from bulge, LMC and SMC are classified to contact C and non contact systems NC
- light-curves in I passband were folded by period from catalog, binned to 100 points and normalized to maximum flux



Simulated LC   OGLE LCs

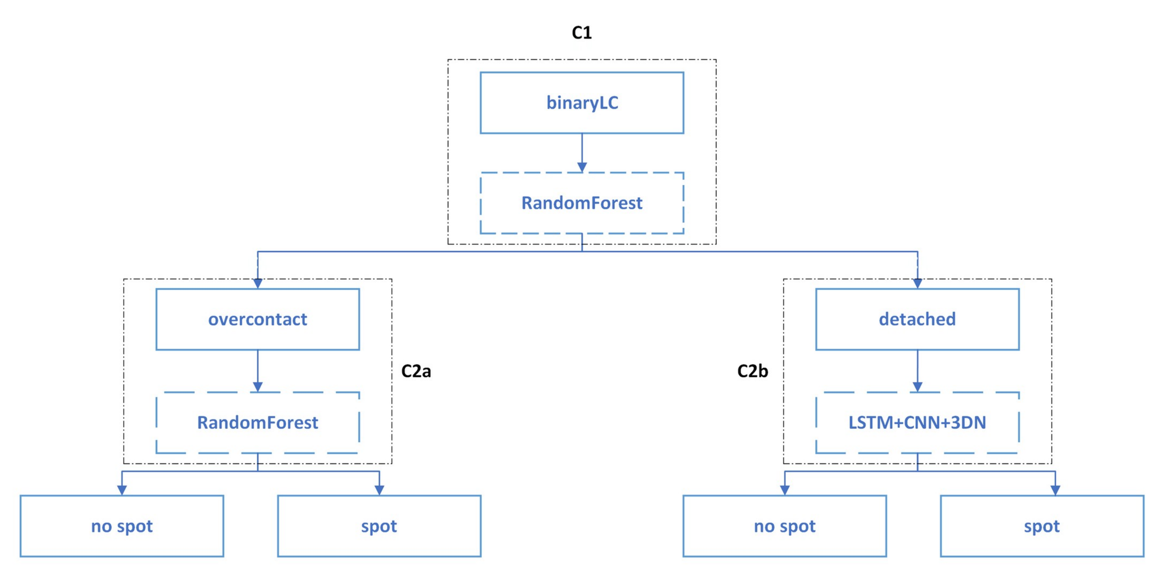
DEEP LEARNING MODEL AND ITS PERFORMANCE

- our classification scheme allows us to use **hierarchical model** – different models can be used for different groups – better overall performance and faster training
- for each group we randomly selected 100 000 LCs from training dataset and it was expanded by gaussian noise augmentation. 20% of them were selected for validation
- several different models and their combinations were tested – the best performance was achieved for the following configuration

1st level (C1)- categorize the LCs into two groups: overcontact and detached
RandomForest Classifier

2nd level - two independent binary classifiers:

C2a – overcontact with and with no spot - **RandomForest Classifier**
C2b – detached with and with no spot - **LSTM+CNN+3DN**



Level	Model	Precision
C1	RandomForest	97.8%
C2a	RandomForest	95.1%
C2b	LSTM+CNN+3DN	92.7%

APPLICATION TO OGLE ECLIPSING BINARIES

Classification		SMC	LMC	Bulge	Σ
OGLE	C	124	1613	84365	86102
	NC	7814	36533	329033	373380
Model	over	643	1691	184512	186846
	detached	7295	36455	228866	272616
over	no spot	637	1568	172022	174227
	spot	6	45	12490	12541
detached	no spot	7211	35633	182611	225455
	spot	84	792	46255	47131

- our model classifies all C binaries from OGLE as overcontact, except for two in SMC, seven in LMC, and 29 in the bulge.
- a significantly large number of NC binaries were classified as overcontact – probably misclassification in OGLE
- detection of spots on OGLE data from LMC and SMC with our model is unusable. In bulge, we classified a much higher number of spotted binaries, but this number is also underestimated; detection of spotted binaries strongly depends on the quality of observed LC – very faint stars with large scatter in data
- our findings suggest that our classifier is well-suited for processing data from future large observational surveys

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 and grant UPJŠ VVGS 2024-3096 V.K. was supported by Ukrainian national grant 0122U000937.

SLOVAK RESEARCH AND DEVELOPMENT AGENCY
PLÁN OBNOVY
Financované Európskou úniou NextGenerationEU