Contrib. Astron. Obs. Skalnaté Pleso **55/3**, 366 – 369, (2025)
https://doi.org/10.31577/caosp.2025.55.3.366

EBOP MAVEN

A machine learning model for predicting eclipsing binary light curve fitting parameters

S. Overall and J. Southworth

Astrophysics Group, Keele University, Staffordshire, ST5 5BG, UK
(E-mail: s.p.overall@keele.ac.uk)

Received: November 16, 2024; Accepted: January 31, 2025

Abstract. Detached eclipsing binary stars (dEBs) are a key source of data on fundamental stellar parameters. While there is a vast source of candidate systems in the light curve databases of survey missions such as *Kepler* and TESS, published catalogues of well-characterised systems fall short of reflecting this abundance. We seek to improve the efficiency of efforts to process these data with the development of a machine learning model to inspect dEB light curves and predict the input parameters for subsequent formal analysis by the JKTEBOP code.

Key words: binaries: eclipsing – methods: data analysis – methods: machine-learning

1. Introduction

Detached eclipsing binary systems are a key source of stellar parameters used in the development and testing of models of stellar evolution. By combining photometric light curve and spectroscopic radial velocity (RV) observations it is possible to measure the components' physical characteristics with great precision (Andersen, 1991; Torres et al., 2010).

Machine learning (ML) approaches are becoming widely used in many areas of science and technology, often as a means to gain new insights into existing datasets or as a way of addressing the vast volumes of data produced by the latest scientific programmes. The key feature of the machine learning approach is that a given program learns to carry out its task through direct experience with the data rather than through a predefined algorithm.

2. Developing the model

EBOP MAVEN is a CNN (Convolutional Neural Network) model implemented in python with the TensorFlow (Abadi et al., 2015) and Keras (Chollet et al., 2015) libraries. CNN models are a type of supervised machine learning model, one that requires training before use, which are widely used in computer vision scenarios. They typically consist of one or more convolutional layers which with training learn the convolution kernels best able to extract the characteristic features within the input data from which a subsequent neural network can make predictions.

We have developed a regression model which views the 1-D time-series data of a phase-folded dEB light curve to predict the value of six parameters used as the input for subsequent formal analysis with the JKTEBOP code (Southworth, 2008). The predicted parameters are the sum $(r_A + r_B)$ and ratio $(k \equiv r_B/r_A)$ of the stars' fractional radii, the central surface brightness ratio (J), the orbital eccentricity through the Poincaré elements ($e\cos\omega$ and $e\sin\omega$), and the orbital inclination through the primary impact parameter (b_P).

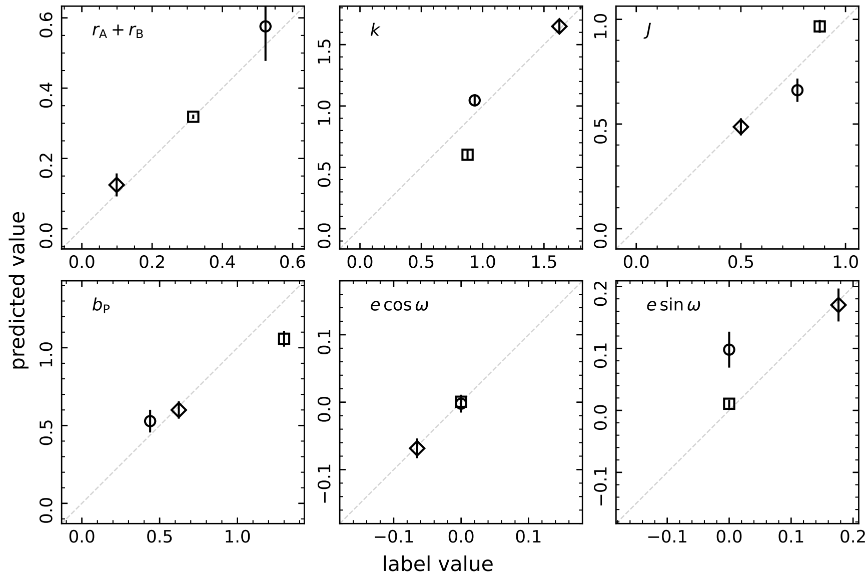
The model was trained on a fully synthetic dataset of 250,000 labelled phase-folded light curves, split 80:20 between training and validation instances. The distribution of the training data was chosen to cover the parameter space supported by JKTEBOP. While training, augmentation steps added random noise and shifts to the data. Dropout layers are included in the model as a regularisation measure to combat overfitting.

3. Testing and results

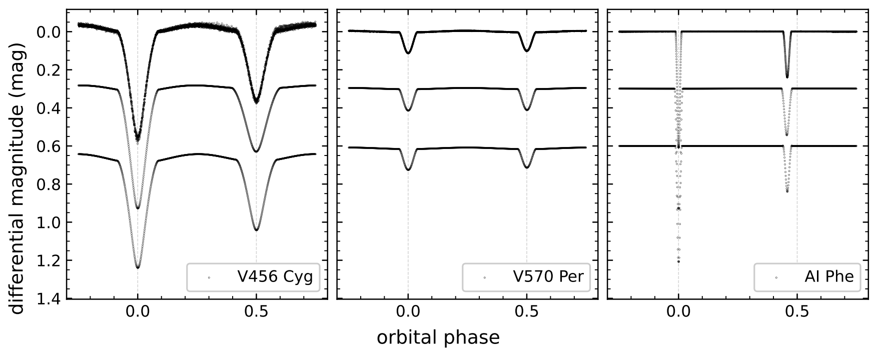
The test dataset consists of synthetic instances representing a population of physically plausible dEB systems. Stellar parameters were taken from MIST (Choi et al., 2016) stellar models with population demographics based on the product of the initial mass function from Chabrier (2003) and the binarity function from Wells & Prša (2021). The treatment of apparent brightness, noise and limb darkening was based on the TESS mission.

A second test dataset of real dEB systems was assembled based on membership of the DEBCat (Southworth, 2015) catalogue of well-characterised detached systems and the availability of TESS timeseries photometry. The test labels were derived from the published works giving the most recent characterisation of each system. A processing pipeline was implemented to prepare and fit light curves from these systems using JKTEBOP, and a set of control fits were made with input parameters based on the test labels. The final set of systems represent a range of targets where it was possible to achieve a stable control fit with results comparable to those published.

The EBOP MAVEN model was used to predict the fitting parameters for the real dEB systems. With the predictions used as the input parameters for the fitting step of the processing pipeline, while all other settings and processes were carried over from the control fits, 22 of the 23 systems yielded a final fitted characterisation in agreement with the equivalent control fit. The exception was AI Phe (see Fig. 1), however this is an especially challenging target where the secondary star is significantly larger than the primary ($k \simeq 1.6$).



(a) The predictions made for V456 Cyg (circles), V570 Per (squares) and AI Phe (diamonds) plotted against the corresponding label values



(b) The phase-folded light curves of V456 Cyg, V570 Per and AI Phe with corresponding generated light curves based on the predicted input parameters (shifted by 0.3 mag) and the fitted characterisation with JKTEBOP (shifted by 0.6 mag)

Figure 1.: The EBOP MAVEN predictions for a subset of the test dataset of real dEB systems below which are the corresponding phase-folded light curves of the input feature, and generated light curves of both the predicted and final fit

4. Conclusion and next steps

We have demonstrated that it is possible to train a machine learning model to predict input parameters for dEB light curve fitting with the JKTEBOP code. When tested with TESS photometry from real systems we achieved a good fit in 22 of 23 targets.

Development of the model is ongoing and a full article is being prepared for publication. The code to build, train and test the version of the model discussed here is publicly available on GitHub[1]. The EBOP MAVEN model will be incorporated into a processing pipeline with which a catalogue of dEB systems will be assembled. This is intended to provide sufficient data for the selection of suitable targets for further observation and formal characterisation.

Acknowledgements. We gratefully acknowledge the support of the Science and Technologies Facilities Council (STFC) in the form of a PhD studentship.

References

Abadi, M., Agarwal, A., Barham, P., et al. 2015, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, https://www.tensorflow.org/

Andersen, J., Accurate masses and radii of normal stars. 1991, *Astronomy and Astrophysics Reviews*, **3**, 91, DOI:10.1007/BF00873538

Chabrier, G., Galactic Stellar and Substellar Initial Mass Function. 2003, *Publications of the ASP*, **115**, 763, DOI:10.1086/376392

Choi, J., Dotter, A., Conroy, C., et al., Mesa Isochrones and Stellar Tracks (MIST). I. Solar-scaled Models. 2016, *Astrophysical Journal*, **823**, 102, DOI:10.3847/0004-637X/823/2/102

Chollet, F. et al. 2015, Keras, https://keras.io

Southworth, J., Homogeneous studies of transiting extrasolar planets - I. Light-curve analyses. 2008, *Monthly Notices of the RAS*, **386**, 1644, DOI:10.1111/j.1365-2966.2008.13145.x

Southworth, J., DEBCat: A Catalog of Detached Eclipsing Binary Stars. 2015, in Astronomical Society of the Pacific Conference Series, Vol. **496**, *Living Together: Planets, Host Stars and Binaries*, ed. S. M. Rucinski, G. Torres, & M. Zejda, 164

Torres, G., Andersen, J., & Giménez, A., Accurate masses and radii of normal stars: modern results and applications. 2010, *Astronomy and Astrophysics Reviews*, **18**, 67, DOI:10.1007/s00159-009-0025-1

Wells, M. A. & Prša, A., Building and Calibrating the Binary Star Population Using Kepler Data. 2021, *Astrophys. J., Suppl. Ser.*, **253**, 32, DOI:10.3847/1538-4365/abd5ba

[1] https://github.com/SteveOv/ebop_maven/tree/kopal2024