Searching Exoplanet Transits through Machine Learning Techniques

Li-Chin Yeh[1] and Ing-Guey Jiang[2]

[1]Institute of Computational and Modeling Science, National Tsing Hua University, Hsin-Chu, Taiwan (e-mail: lichinyeh@mx.nthu.edu.tw)
[2]Department of Physics & Institute of Astronomy, National Tsing Hua University, Hsin-Chu, Taiwan (e-mail: jiang@phys.nthu.edu.tw)

Abstract

Due to the availability of high-performance computers and the data release from big-scale sky survey projects, the machine learning has been employed in various fields in astronomy. We have developed a procedure of machine learning to search for exoplanet transits from several survey data. Our procedure employs the convolutional neural network (CNN) of deep learning techniques. Both 1D-CNN and 2D-CNN with different training processes are used and compared. The details of the procedure and the results will be presented.

Method

- **Data Sources**

Data from The Transiting Exoplanet Survey Satellite (Tess STScI | Mikulski Archive for Space Telescopes (MAST) Portal)

- **Data preparation :** (from Yeh & Jiang [4])

(1) Clustering and Standardization (2) Removing Outliers
(3) Data Folding (4) Interpolation for Imputation

- **Generating data set:**

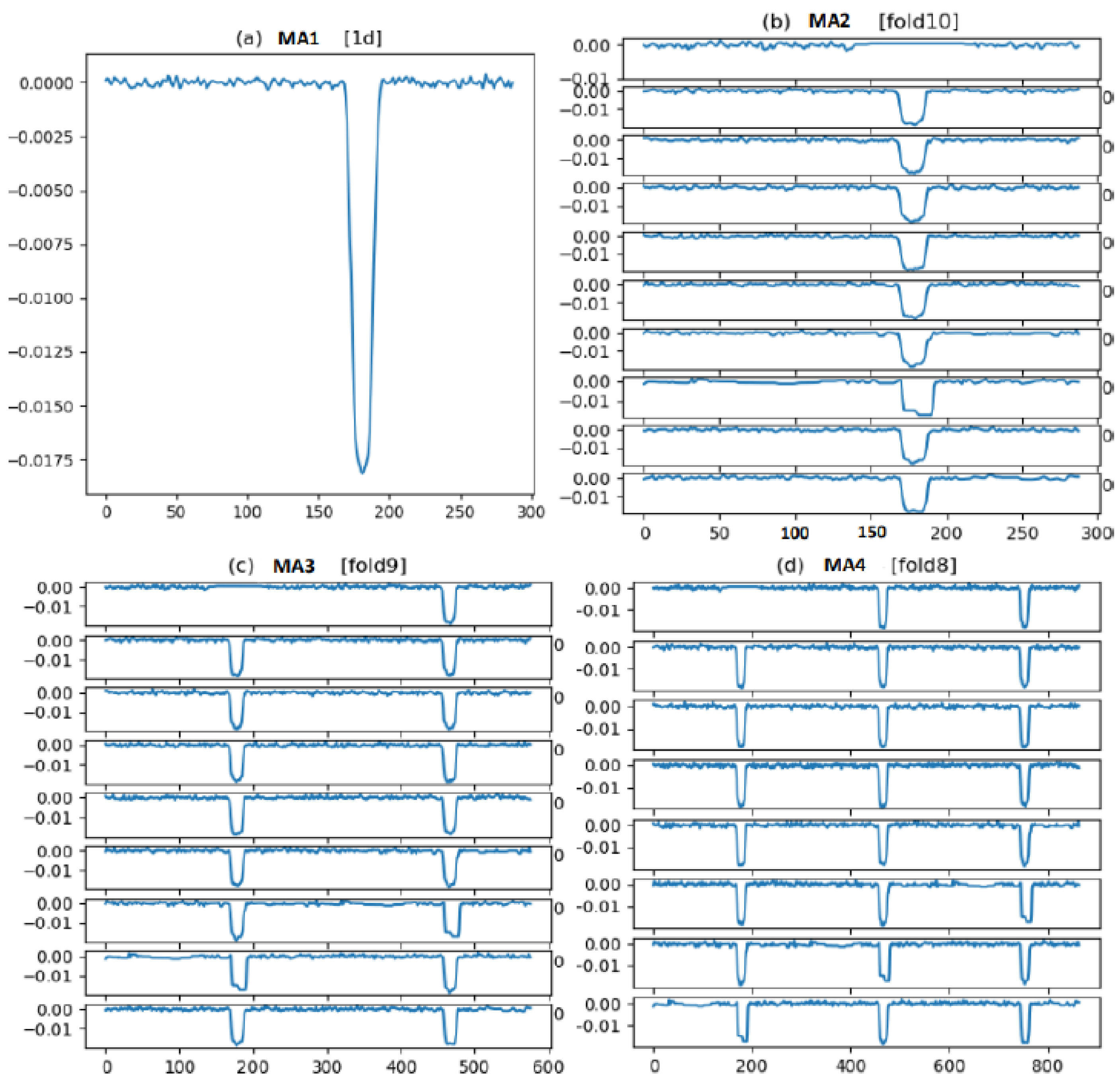
(1) construct the noisy background of light curves (called non-signal light curves) from the Tess data set for the training data set.
(2) The transit signal data from Mandel Agol [2] model by using different parameters into the original light curve

- **CNN Models Training Process:**

(1) testing data: 80% of light curves ;validation data: 10% of the light curves are for validation data;testing data on the remaining light curves.
(2) The stopping condition :the difference of five successive accuracies was less than 0.001.

CNN Models

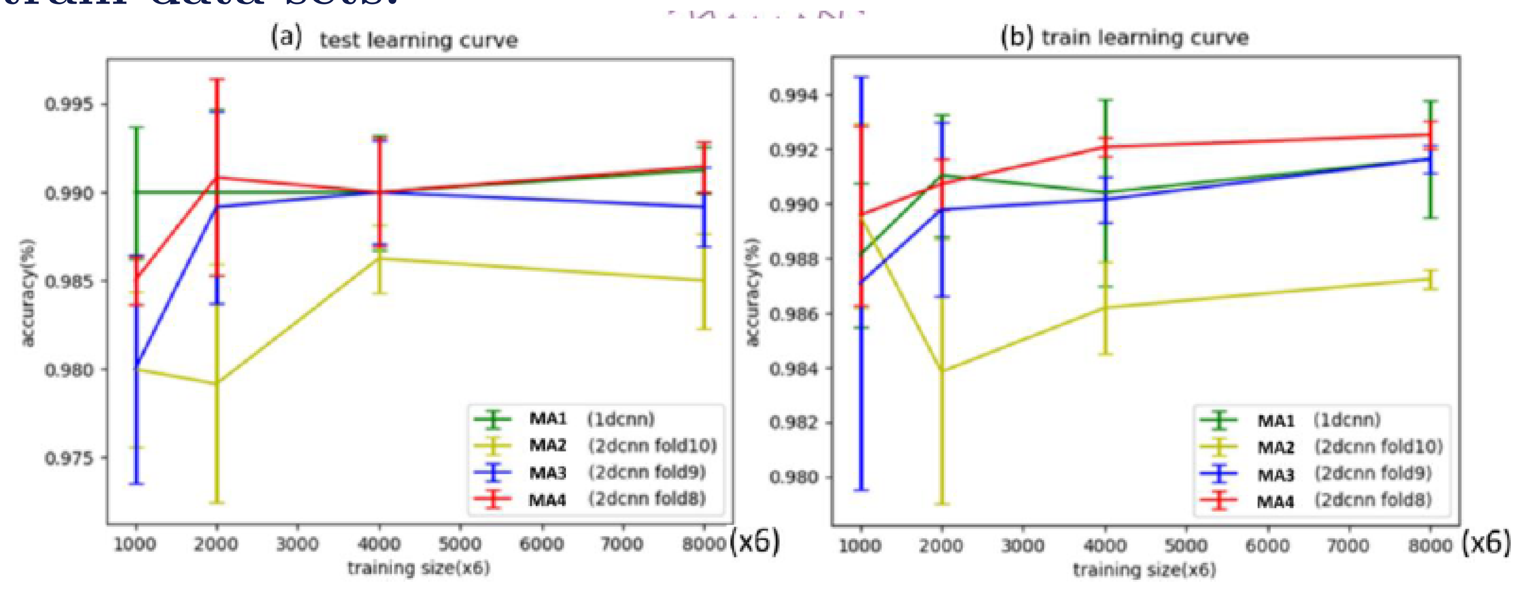
- **From Chintarungruangchai and Jiang [1], we construct the following CNN models :**



- **(a) 1D-folding CNN (MA) (b) 2D-folding CNN-1 (MA1) (c) 2D-folding CNN-2 (MA2) (d) 2D-folding CNN-3 (MA3)**

Results

- **Compare the superiority for these four models for test and train data sets:**



- **2D-folding CNN-3 (MA3) is the best model**

Possible exoplanet candidates

- **The error estimation:**

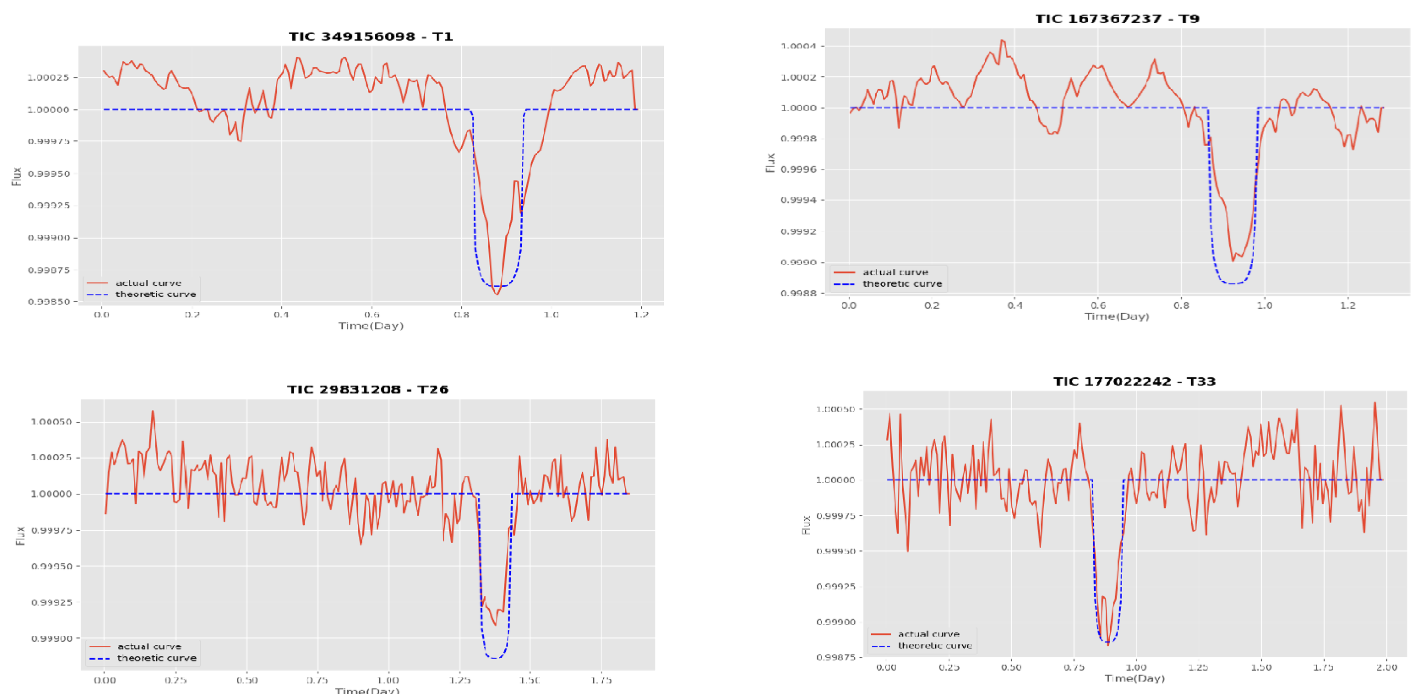
$$Er = \sqrt{\frac{\sum_{i=1}^{n}(T_i - D_i)^2}{n}},$$

where T_i represents the theoretical value and D_i represents the actual value.

- There are four possible candidates in the following Table :

ID	TIC Number	Period P	The error E_r
1	TIC 349156098	1.1944	2.79×10^{-4}
2	TIC 167367237	1.2903	1.9×10^{-4}
3	TIC 29831208	1.836	1.96×10^{-4}
4	TIC 177022242	1.98	2.15×10^{-4}

- There are four light curves in the following:(from Han,Wei-His[3])



Conclusions

- **2D-folding CNN-3 (MA3) is the best CNN model.**
- **Four possible exoplanet candidates are found from Tess data set.**

References

[1] Chintarungruangchai, P., & Jiang, I.-G. (2019). Publications of the Astronomical Society of the Pacific, 131(1000)
[2] Mandel, K., Agol, E. (2002). The Astrophysical Journal, 580(2), L171.
[3] Han,Wei-His, NTHU master thesis.
[4] Yeh, L.-C., Jiang, I.-G. (2020). Publications of the Astronomical Society of the Pacific, 133(1019), 014401.