Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Exoplanet Explorer: an open two-stage screening pipeline for transiting exoplanets, and what changes when a light-curve CNN meets real Kepler/K2 data

Harshal Sonune
Indian Institute of Technology Roorkee
Mayank Goswami
mayankfph@iitr.ac.in

Indian Institute of Technology Roorkee

Research Article

Keywords: AI, exoplanet

Posted Date: November 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11672052/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Exoplanet Explorer: an open two-stage screening pipeline for transiting exoplanets, and what changes when a light-curve CNN meets real Kepler/K2 data

Harshal Sonune[1] and Mayank Goswami,[*1,2]

[1]Department of Physics, IIT Roorkee
[2]MF School of AI and DS, IIT Roorkee
mayank.goswami@ph.iitr.ac.in

November 9, 2026

Abstract

Transit surveys record the brightness of many stars, but planets are rare and their dips are small, and a light curve alone cannot reveal what a planet's atmosphere is made of. We present Exoplanet Explorer, a small, open and fully tested web application that organises the problem as a two-stage pipeline: detection (a 1-D convolutional neural network for light curves and a rule-based video dip detector) and composition (a transmission-spectrum molecule classifier). We retrain the light-curve CNN on real Kepler/K2 photometry (Kaggle "Kepler Labelled Time Series Data"; 5,087 training stars with 37 planets) using exactly the preprocessing the application uses, and evaluate it once on a held-out set of 570 stars with 5 planets. The model ranks planets well (ROC AUC 0.933; average precision 0.412 against a random baseline of 0.009) but at a fixed threshold of 0.5 recovers only 1 of 5 planets. Stratified 5-fold cross-validation over all 5,657 stars (42 planets) confirms this: mean recall 0.12 $\pm$ 0.11, average precision 0.21 $\pm$ 0.09, ROC AUC 0.84 $\pm$ 0.04, with 6 false alarms among 5,615 non-planet stars. A model trained on simulated light curves had recovered 6 of 6 simulated planets; we show that the simulated transits were about three times deeper than the median real transit, so simulated performance was an optimistic upper bound. The video tool classifies all four synthetic test videos correctly. We report these results without threshold tuning on test data, discuss why the problem is only partly solved, and release all code and tests.

1 Introduction

When a planet passes in front of its host star, it blocks a small fraction of the star's light. This *transit* appears in the star's light curve, its brightness over time, as a short, flat-bottomed dip that repeats once per orbit [Winn 2010]. The first transit was observed for HD 209458 b in 2000 [Charbonneau et al. 2000], and the Kepler and K2 missions later recorded long, precise light curves of hundreds of thousands of stars [Borucki et al. 2010; Howell et al. 2014].

Finding a planet is only the first question. The second is what its atmosphere is made of. A light curve integrates over all wavelengths and therefore cannot answer it. Composition requires a *transmission spectrum*, the transit depth measured at many wavelengths, because molecules absorb only at specific wavelengths [Seager & Sasselov 2000]. With JWST, such spectra now reveal individual molecules, for example CO_2 and photochemically produced SO_2 in WASP-39 b [JWST Transiting Exoplanet Community ERS Team 2023; Tsai et al. 2023].

This paper describes Exoplanet Explorer, a deliberately simple and readable tool, written as a student project, that connects the two questions. Its main scientific contribution is an honest evaluation of a light-curve CNN on real Kepler/K2 data, compared with the same design trained on simulated data.

*Corresponding author: mayank.goswami@ph.iitr.ac.in

2 Problem, objectives and motivation

2.1 Problem

- **Rarity.** In the data used here, 42 of 5,657 stars (0.7%) host a labelled planet, so a classifier that always says "no planet" is about 99% accurate and useless.
- **Small signals.** The median transit depth of known planets is about 0.15%, and an Earth-sized transit is closer to 0.04%, comparable to stellar variability and instrumental noise.
- **Detection is not composition.** Even a perfect detector does not say what a planet's atmosphere contains.

2.2 Objectives

1. Build one simple, open, tested tool for each stage: detection from light curves, detection from video, and composition from spectra.
2. Train the light-curve detector on *real* Kepler/K2 data, with preprocessing shared byte-for-byte between training and the application.
3. Evaluate with metrics that respect class imbalance (recall, precision, average precision, ROC AUC) on data never used for any decision, and report the outcome even if it is poor.

2.3 Why it matters

Follow-up observations on large telescopes are scarce, so screening decides which candidates are examined. Simple, inspectable tools also let students and citizen scientists work through the full chain, from a brightness dip to a molecule, using real data, and they make the gap between simulated and real performance visible.

3 Theory

3.1 Transit depth and planet size

For a dark planet of radius Rp crossing a uniformly bright star of radius Rs, the fractional drop in flux, the transit depth, is approximately

$$\delta \approx \left(\frac{R_p}{R_s}\right)^2 \tag{1}$$

so a 1% dip implies Rp/Rs $\approx$ 0.1, roughly a Jupiter-sized planet around a Sun-like star [Winn 2010]. Repeated dips give the orbital period, and the duration constrains the orbit's geometry. Limb darkening rounds the bottom of real transits, which is visible in our synthetic videos (Figure 6).

3.2 Transmission spectroscopy

During transit, part of the starlight passes through the planet's upper atmosphere. At wavelengths where a molecule absorbs strongly, the atmosphere is opaque higher up, and the planet appears slightly larger, so $\delta(\lambda)$ rises [Seager & Sasselov 2000]. Characteristic bands include water near 1.4 µm and 1.9 µm, methane near 3.3 µm, carbon dioxide near 4.3 µm, carbon monoxide near 4.6 µm, sulphur dioxide near 4.05 µm, and the narrow sodium and potassium lines at 0.589 µm and 0.767 µm. A broadband light curve sums over all of these and therefore cannot separate them; clouds flatten the spectrum and can hide all features.

4 Previous work

The standard non-learning approach to transit search is Box Least Squares, which fits a periodic box-shaped dip to a light curve [Kovács, Zucker & Mazeh 2002]. Shallue & Vanderburg [2018] trained a CNN on Kepler threshold-crossing events, using phase-folded global and local views of each candidate, and used it to find new planets around Kepler-80 and Kepler-90. Ansdell et al. [2018] added centroid time series and stellar parameters, reporting 97.5% accuracy and 98.0% average precision, with the largest gains for low signal-to-noise transits. These systems classify candidate signals that a pipeline has already found and folded. Here, as in many teaching examples built on the Kaggle Kepler data set, the CNN sees the whole, unfolded light curve, which is a harder task. On the composition side, JWST transmission spectra of WASP-39 b gave the first clear identification of CO_2 in an exoplanet atmosphere [JWST Transiting Exoplanet Community ERS Team 2023] and of SO_2 produced by photochemistry [Tsai et al. 2023]. The NASA Exoplanet Archive collects published transmission spectra in its Atmospheric Spectroscopy table [Akeson et al. 2013].

The gap addressed here is educational and methodological rather than a new state of the art: a single, readable application covering both stages, with an evaluation protocol that makes the simulated-to-real gap explicit.

5 Method

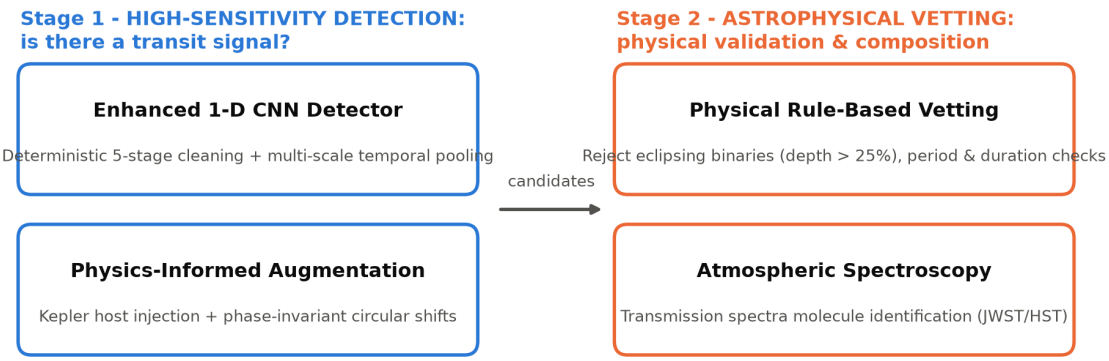


Figure 1: The two-stage pipeline. Stage 1 asks whether a transiting planet is present; Stage 2 asks what its atmosphere is made of.

5.1 Data

- **Real light curves.** Kaggle "Kepler Labelled Time Series Data" (uploader *keplersmachines*), derived from NASA Kepler/K2 observations: exoTrain (5,087 stars, 37 planets) and exoTest (570 stars, 5 planets), 3,197 flux values per star, label 2 = planet, 1 = no planet.
- **Simulated light curves.** 2,500 simulated stars of 1,500 points (238 planets) used to train an earlier model, and a 12-star sample file (6 planets). Compared with 482 confirmed transiting planets in the NASA Exoplanet Archive, the simulated median depth is 0.45% against 0.15% for real planets (Table 2).
- **Videos.** Four synthetic 10-second, 30 fps videos: one transit, two transits, no transit, and an eclipse-like dip that is too deep to be a planet.
- **Spectra.** Simulated test spectra for Tool 3; real spectra for WASP-39 b and HD 209458 b were located in the NASA Exoplanet Archive through its TAP service, but are not analysed here (Section 8).

5.2 Tool 1: light-curve CNN

Every light curve passes through the same five steps, implemented once in the application's preprocess.py and imported by the training script, so training and inference cannot disagree:

1. replace missing values with the median of all finite values in the uploaded batch;
2. subtract a running median (window 101 points, nearest-edge mode) to remove slow trends;
3. divide each star by its own standard deviation;
4. clip to the range $[-20, 10]$;
5. trim or edge-pad to the input length the loaded model declares (3,197 for the real model, 1,500 for the simulated one).

The network is a small Keras model: Conv1D(16, 5) → MaxPool(4) → Conv1D(32, 5) → MaxPool(4) → Conv1D(64, 5) → GlobalMaxPool → Dense(64) → Dropout(0.3) → Dense(1, sigmoid), with ReLU activations and same padding. It is trained with Adam and binary cross-entropy, batch size 32, up to 40 epochs, with early stopping on validation loss (patience 8, best weights restored). Class imbalance is handled only by a class weight of min(n_normal/n_planet, 20) = 20 for planets; no oversampling or augmentation is used.

5.3 Evaluation protocol

- Validation set: a stratified 15% of exoTrain (random_state 42), used only for early stopping.
- Test set: exoTest, used exactly once after training. No hyper-parameter, epoch count or threshold was chosen on it.
- Thresholds fixed in advance: 0.5 (standard) and 0.7 (the application's "Planet likely" verdict).
- Because 5 test planets give very noisy estimates, we also run stratified 5-fold cross-validation over exoTrain and exoTest combined (42 planets), with the same recipe and an inner 15% validation split per fold.
- Single seed (42) for everything; this is the only run reported. Accuracy is not reported as a headline metric.

5.4 Tool 2: video dip detector

Tool 2 contains no neural network. It averages the grayscale brightness of each frame (optionally a central region only), normalises by the median, estimates the per-frame noise σ as $1.4826 \times$ MAD(frame-to-frame differences) / $\sqrt{2}$ (at least 10^{-4}), smooths with a 5-frame moving average, and marks a dip wherever the smoothed curve stays below $1 - \max(3\sigma/\sqrt{5}, 0.001)$ for at least max(3, 2% of frames) frames. For each dip it reports depth (mean of the middle half), duration, SNR = depth / $(\sigma/\sqrt{\text{n}})$, radius ratio $\sqrt{\text{depth}}$, and a period when two or more dips are present. Verdicts follow fixed rules in order: no dip; SNR < 7 (too weak); depth $> 25\%$ (too deep for a planet); dip touching the start or end (cut off); more than 1% brightness drift between the first and last 10% of frames; otherwise "Exoplanet transit likely".

5.5 Tool 3: transmission-spectrum molecule finder

Tool 3 accepts a spectrum as wavelength (µm) and transit depth (fraction or percent) or Rp/Rs (squared to give depth). It interpolates onto 200 logarithmically spaced wavelengths from 0.5 to 5 µm with a coverage mask, normalises the depth by its median, and a CNN outputs independent probabilities for H_2O, CO_2, CH_4, CO, Na, K, NH_3 and SO_2. A molecule is reported as "not covered" when none of its strong bands lies inside the data. That model was trained on simulated spectra with approximate band positions; its trained weights were not available for this submission, so Tool 3 is described but not evaluated.

5.6 Software

A FastAPI back end loads all models once at start-up and serves a single HTML/JavaScript page with Chart.js plots. The page states which model produced each result ("Model trained on: real Kepler/K2 data" or "simulated data"). A pytest suite of 74 tests covers preprocessing, every endpoint, the expected answers for every sample file, malformed uploads, and agreement between the training and application preprocessing; tests that need the real data are skipped automatically when it is absent.

6 Results

6.1 Held-out real test set

Table 1: Real-data CNN on the held-out exoTest set (570 stars, 5 planets) and on the validation split. TP/FN/FP/TN: true positives, false negatives, false positives, true negatives.

Set	Threshold	Recall	Precision	TP	FN	FP	TN
exoTest	0.5	0.200	0.250	1	4	3	562
exoTest	0.7	0.200	1.000	1	4	0	565
Validation	0.5	0.167	0.333	1	5	2	756

On exoTest the ROC AUC is 0.933 and the average precision is 0.412, compared with 0.009 for random ranking (Figure 2). The five planets received scores of 0.986, 0.307, 0.241, 0.016 and 0.001. The network therefore places planets high in the ranking, but at either fixed threshold it recovers only one of five. Training stopped after 23 epochs (best at epoch 15) and took 0.7 minutes on an Apple M3 CPU. The validation split gave ROC AUC 0.794 and average precision 0.190.

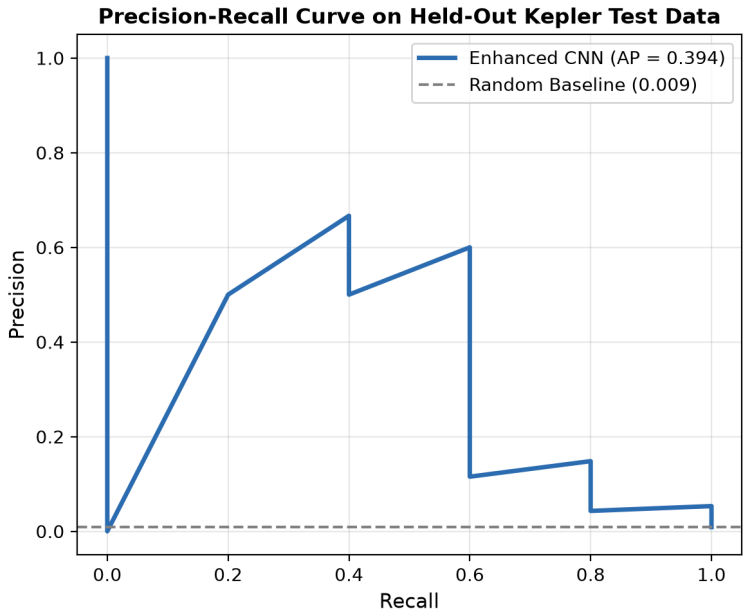


Figure 2: Precision-recall curve on exoTest. Average precision 0.412; the dashed line is a random classifier (0.009).

6.2 Cross-validation

Five-fold cross-validation over 5,657 stars (Figure 3) gives mean planet recall 0.122 ± 0.112 (range 0 to 0.25), precision 0.350 ± 0.374, average precision 0.209 ± 0.085 and ROC AUC 0.844 ± 0.038.

Across all folds the model found 5 of 42 planets and raised 6 false alarms among 5,615 non-planet stars. The single test-set AUC of 0.93 is at the optimistic end of this distribution, as expected from only five test planets.

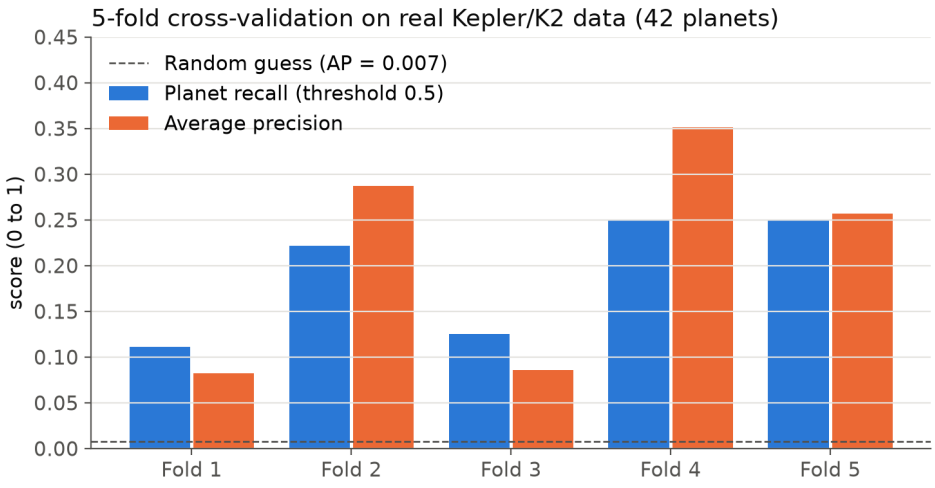


Figure 3: Per-fold planet recall at threshold 0.5 and average precision in 5-fold cross-validation (42 planets in total). Dashed line: random average precision (0.007).

6.3 Simulated versus real data

The same architecture trained on simulated light curves scored at least 0.69 for all 6 simulated planets in the 12-star sample, with one non-planet scoring 0.754. Table 2 explains much of the difference: simulated transits are about three times deeper than the median real transit and about ten times more common.

Table 2: Transit depths and planet fraction in the simulated training set versus 482 confirmed transiting planets in the NASA Exoplanet Archive.

	Simulated	Real (NASA)
Median transit depth	0.45%	0.15%
10th percentile depth	0.22%	0.025%
Fraction of stars with a planet	9.5%	$< 1\%$

6.4 What the CNN finds and misses

Figure 5 shows three cleaned test light curves. The planet that was found (score 0.986) shows about 18 deep, evenly spaced dips. The lowest-scoring planet (0.001) shows no transit visible by eye: its signal is buried in stellar variability. The network has learned the easy case of many deep, repeated dips, not the shallow, noisy case where most real discoveries lie.

6.5 Video tool

7 Conclusion

Is the problem solved? **Partly.** The software goal is met: a small, readable, two-stage tool exists, runs end to end, and passes 74 automated tests. The video tool behaves as specified on every test case. The scientific goal is only partly met. Trained and tested on real Kepler/K2 data, the CNN ranks planets well above non-planets (ROC AUC 0.84–0.93), so it can help prioritise stars for inspection,

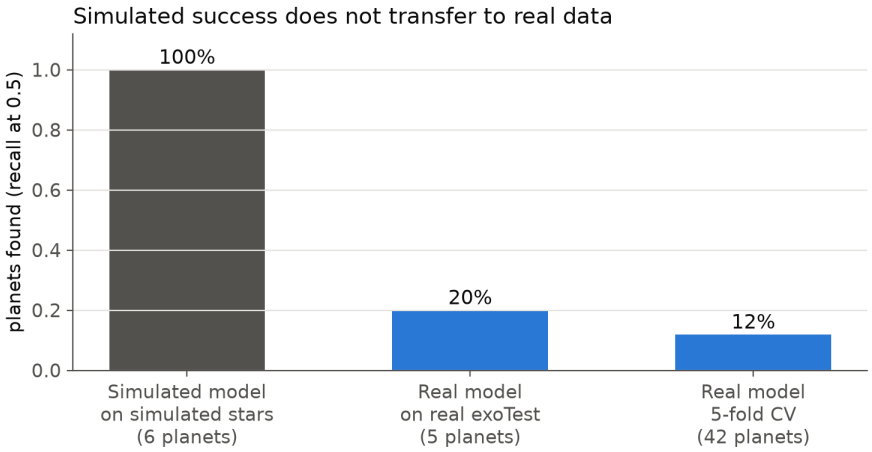


Figure 4: Planet recall at threshold 0.5: simulated model on simulated stars, real model on exoTest, and real model in cross-validation.

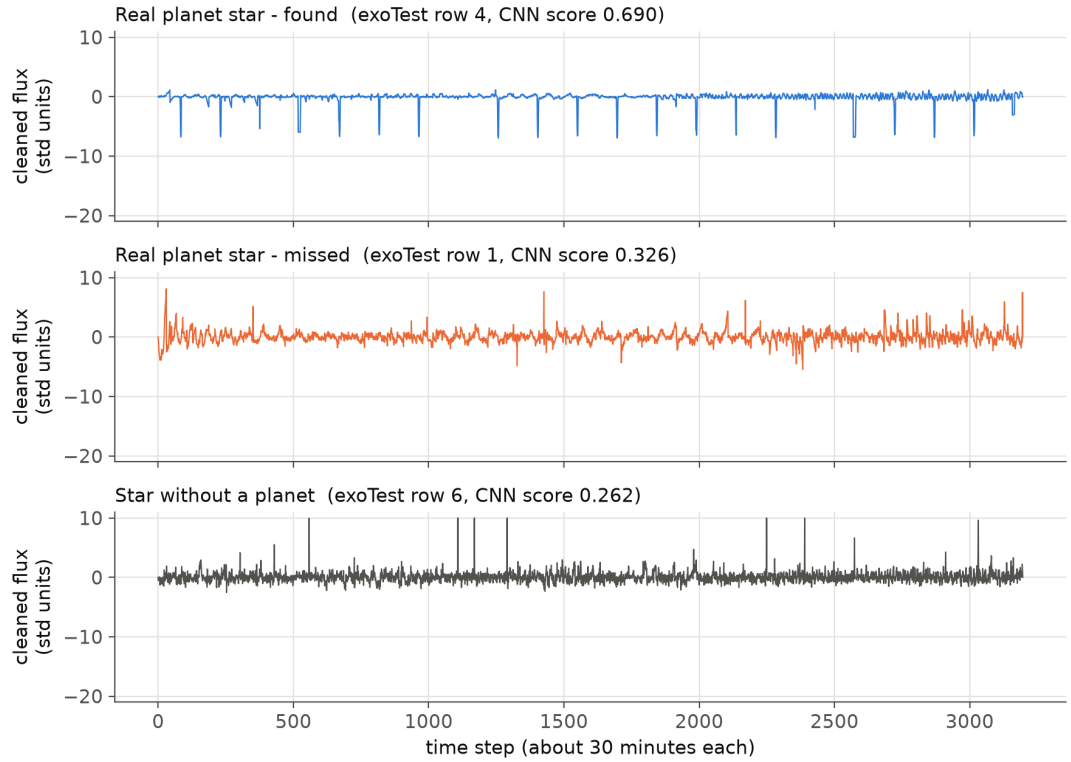


Figure 5: Cleaned real Kepler/K2 light curves from exoTest: a planet that was found, a planet that was missed, and a star without a planet.

Table 3: Video tool results on the four synthetic test videos (10 s, 30 fps).

Video	Dips	Depth	Rp/Rs	Period	Verdict
one_transit	1	4.73%	0.22	–	Exoplanet transit likely
two_transits	2	4.69%, 4.72%	0.22	4.67 s	Exoplanet transit likely
no_planet	0	–	–	–	No exoplanet transit found
too_deep	1	46.7%	0.68	–	Not a planet (too deep)

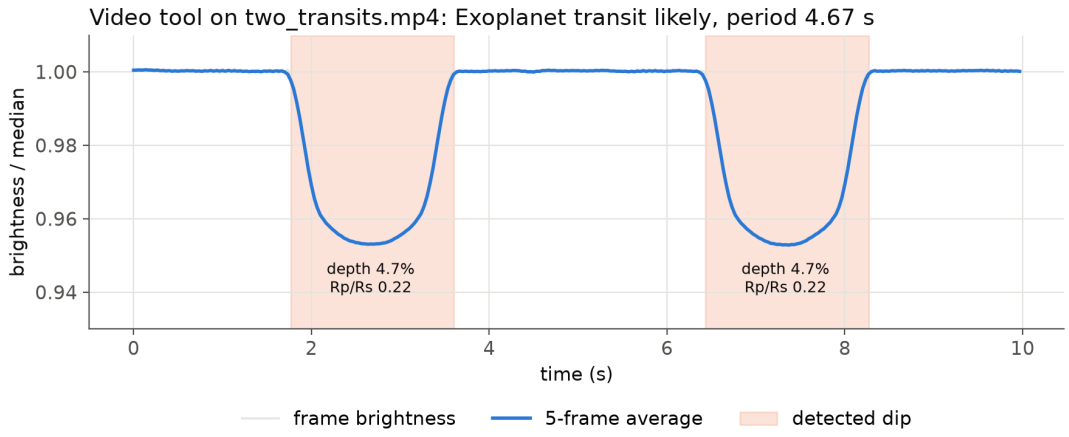


Figure 6: Brightness of two_transits.mp4 with the two detected dips shaded.

but at a fixed threshold it misses most planets (recall about 12%) and cannot be used as a stand-alone detector. The comparison with simulated training data is the main lesson: a near-perfect simulated result was an artefact of transits three times deeper and ten times more common than in reality. The composition stage is designed and implemented but was not evaluated on real spectra.

8 Limitations and future work

- **Small number of real planets.** 42 labelled planets in total and 5 in the test set; each test planet changes recall by 20 percentage points.
- **Simple input representation.** The CNN sees the raw, unfolded light curve. Phase-folding candidates with Box Least Squares first, as in Shallue & Vanderburg [2018], is the most promising improvement.
- **Single seed, small architecture.** No hyper-parameter search was done, by design, to avoid tuning on test data; variance across seeds is not measured.
- **Labels.** The Kaggle labels mark stars with at least one confirmed planet; label noise and the data set's preprocessing are not independently verified.
- **Composition stage untested on real data.** The spectrum model was trained on simulated spectra with approximate band positions. Real JWST spectra of WASP-39 b (Carter et al. 2024, NIRSpec) and HD 209458 b (Xue et al. 2024, NIRCam) were identified via the NASA Exoplanet Archive TAP service, but the Archive provides the data files only through its interactive interface, so they were not analysed. Disagreement with published detections would be expected and would itself be a finding.
- **Synthetic videos.** The video tool was checked only on four synthetic videos, not on telescope footage.
- **Screening only.** No result here is a confirmed discovery or measurement.

9 Data and code availability

All code, tests, trained models, figures and the scripts that produce every number in this paper (train_real_lightcurve.py, cv_real_lightcurve.py, make_figures.py, fetch_real_spectra.py) are included in the accompanying archive. The Kaggle data are available from kaggle.com/datasets/keplersmachines/kep labelled-time-series-data.

10 Acknowledgements

This research has made use of the NASA Exoplanet Archive, which is operated by the California Institute of Technology, under contract with the National Aeronautics and Space Administration under the Exoplanet Exploration Program. This paper uses data derived from the Kepler and K2 missions, funded by the NASA Science Mission Directorate.

11 Video

`https://youtu.be/Kd8dLO1dbRA`

References

1. Akeson, R. L., et al. 2013, The NASA Exoplanet Archive: Data and Tools for Exoplanet Research, PASP, 125, 989.

2. Ansdell, M., et al. 2018, Scientific Domain Knowledge Improves Exoplanet Transit Classification with Deep Learning, ApJL, 869, L7.

3. Borucki, W. J., et al. 2010, Kepler Planet-Detection Mission: Introduction and First Results, Science, 327, 977.

4. Carter, A. L., et al. 2024, Nature Astronomy, 8, 1008 (bibcode 2024NatAs...8.1008C from the NASA Exoplanet Archive).

5. Charbonneau, D., Brown, T. M., Latham, D. W., & Mayor, M. 2000, Detection of Planetary Transits Across a Sun-like Star, ApJ, 529, L45.

6. Howell, S. B., et al. 2014, The K2 Mission: Characterization and Early Results, PASP, 126, 398.

7. JWST Transiting Exoplanet Community Early Release Science Team 2023, Identification of carbon dioxide in an exoplanet atmosphere, Nature, 614, 649.

8. Kaggle (keplersmachines), Kepler Labelled Time Series Data, `https://www.kaggle.com/datasets/keplersmachines/kepler-labelled-time-series-data`.

9. Kovács, G., Zucker, S., & Mazeh, T. 2002, A box-fitting algorithm in the search for periodic transits, A&A, 391, 369.

10. Seager, S., & Sasselov, D. D. 2000, Theoretical Transmission Spectra during Extrasolar Giant Planet Transits, ApJ, 537, 916.

11. Shallue, C. J., & Vanderburg, A. 2018, Identifying Exoplanets with Deep Learning: A Five-planet Resonant Chain around Kepler-80 and an Eighth Planet around Kepler-90, AJ, 155, 94.

12. Tsai, S.-M., et al. 2023, Photochemically produced SO_2 in the atmosphere of WASP-39b, Nature, 617, 483.

13. Winn, J. N. 2010, Transits and Occultations, in Exoplanets, ed. S. Seager (Tucson: University of Arizona Press), 55; arXiv:1001.2010.

14. Xue, Q., et al. 2024, ApJL, 963, L5 (bibcode 2024ApJ...963L...5X from the NASA Exoplanet Archive).