# Automated Identification of Exoplanets with Machine Learning

Author: Vicenç Fernández Fernández, vfernafe18@alumnes.ub.edu
*Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.*

Advisor: Daniel del Ser Badia, danieldelser@ub.edu

**Abstract:** The detection of exoplanets is a rapidly evolving field, increasingly supported by advances in Machine Learning. In this work, we explore the capabilities of the AstroNet deep learning algorithm when applied to the light curves preprocessed by the TFAW algorithm. The goal is to classify Threshold Crossing Events (TCEs) and identify new potential exoplanet candidates.

We first validate the performance of the model on a subset of previously confirmed exoplanets, showing that the algorithm successfully recovers the expected high prediction scores. Subsequently, we analyze a visually selected subset of 478 candidates from the TFAW survey with assigned priority levels, using the model output to propose priority reclassifications based on objective criteria. Finally, we apply the model to a dataset of 65.970 K2 light curves, identifying 3.800 previously unreported candidates.

Our results demonstrate that AstroNet, when combined with TFAW, is a powerful tool for automatic exoplanet candidate classification. However, we also emphasize that such models are not definitive, and complementary validation methods remain essential to confirm the planetary nature of any new transiting candidate.
**Keywords:** Exoplanets, Planetary Systems, Photometry, Data Analysis, Machine Learning

## I. INTRODUCTION

The discovery and study of exoplanets has become one of the most dynamic and impactful fields in modern astrophysics. Since the first confirmed detection in the early 1990s, thousands of exoplanets have been identified. These discoveries have revolutionized our understanding of planetary formation, migration, and habitability. With each new detection, researchers gain further insight into the frequency of Earth-like planets and the potential for life elsewhere in the universe.

Following the success of the original Kepler mission, NASA launched the K2 mission [1] as an extended campaign after the failure of two of Kepler's reaction wheels. K2 made use of solar radiation pressure to maintain pointing stability, allowing observations of various regions along the ecliptic plane. Despite reduced stability, K2 maintained remarkable photometric precision and, when combined with advanced data processing algorithms, remained highly effective in detecting exoplanets.

To further exploit the potential of K2, especially in the search for previously undetected exoplanet candidates, new methodologies were required. In this context, the TFAW survey [2] emerges. Its current goal is to search for exoplanet candidates that may have been missed by previous studies, by further enhancing the photometric precision of EVEREST 2.0-corrected light curves. The survey combines TFAW, a new wavelet-based detrending and denoising algorithm developed by [3], in conjunction with the EVEREST 2.0 pipeline [4] and the Transit Least Squares (TLS) search algorithm [5]. As demonstrated in [6], TFAW achieves superior photometric precision and improved planet characterization compared to other detrending techniques applied to K2 data.

In recent years, machine learning (ML) techniques have been increasingly adopted to automate analysis of light curve data in exoplanetary science. From decision trees and random forests to deep neural networks, ML has enabled significant advances in vetting transit-like signals and classifying Threshold Crossing Events (TCEs). Notably, convolutional neural networks (CNNs) have proven especially effective due to their capacity to extract features from time-series data analogous to image analysis. Projects such as Autovetter [7] and Robovetter [8] have demonstrated the power of supervised learning to reduce human bias and improve classification performance across large datasets.

In this work, we apply a deep learning model based on the AstroNet CNN [9] architecture to classify K2 TCEs. Unlike previous applications that relied on Kepler pipeline products, we preprocess the K2 light curves using the TFAW algorithm to enhance signal quality. The aim of this project is to assess whether the use of TFAW-processed input can improve the performance of neural network classifiers in distinguishing real exoplanetary transits from false positives.

## II. METHODOLOGY

The AstroNet algorithm follows the typical three-phase structure common to most Machine Learning workflows: training, evaluation, and prediction. In this section, we detail how this algorithm has been adapted and applied to the specific context of our study.

### A. CREATING OUR TRAINING SET

To train the AstroNet model for exoplanet classification, it is essential to convert the raw data into a format compatible with TensorFlow. For this purpose, we employ the TFRecord format, a binary format widely used in TensorFlow to efficiently store and read large amounts of data. TFRecord files consist of sharded files that contain serialized tf.Example protocol buffers. Each tf.Example represents a single Threshold Crossing Event (TCE), encoded with a set of input features derived from the light curves. These features include the rowid (an integer identifying the row in the TCE table), the Kepler ID of the target star, the TCE number associated with that star, the Autovetter training set label which classifies each TCE as a planet candidate (PC), astrophysical false positive (AFP) or non-transiting phenomenon (NTP), and the period of the detected event in days.

In this context, each serialized example includes two key representations of the phase-folded light curve: **global_view**, a vector of length 2001 that captures the entire orbital phase of the signal, allowing the model to learn from the full structure and periodicity of the transit; and **local_view**, a shorter vector of length 201 centered on the transit event itself, offering a high-resolution view of the transit shape and depth. These two complementary views allow the model to leverage both the overall periodic behavior and the fine transit details, improving classification performance.

To create our training set, we first compiled a list of 17.221 Threshold Crossing Events (TCEs). These labeled TCEs, which form the basis of our dataset, can be downloaded from the NASA Exoplanet Archive as the DR24 TCE Table in CSV format. The events are characterized by the following parameters: `tce_planet_num`, `tce_period`, `tce_time0bk`, `tce_duration`, `av_training_set`, `campaign`, `tce_depth`, and `tce_impact`. Additionally, we extracted K2 light curves and preprocessed them using the TFAW method to enhance the quality and reliability of the input data. This dataset is divided into 80% for training, 10% for testing, and 10% for model evaluation.

The model architecture is based on a one-dimensional convolutional neural network (CNN), a type of deep learning model particularly effective for analyzing spatially structured input data, such as light curves. This design assumes that relevant features within the light curves are locally distributed and that the model's predictions should be invariant to small translations in the input signal. To capture these patterns, the model applies convolutional filters followed by max pooling operations, which progressively reduce the dimensionality while preserving essential information. In cases where both global and local representations of the light curve are available, these are processed through separate convolutional branches before being merged into a set of fully connected layers. This dual-path approach allows the network to integrate information across multiple temporal scales, improving its ability to detect transiting exoplanet signals.

### B. MODEL EVALUATION

To evaluate our model's classification performance, we considered several key metrics commonly used in machine learning. **Precision** measures the proportion of predicted planets that are actually true planets, indicating the reliability of positive predictions. **Recall** assesses the model's ability to recover real planets from the dataset, representing the fraction of actual planets that the model successfully identifies. **Accuracy** reflects the overall performance, quantifying the proportion of all signals, both planetary and non-planetary, that are correctly classified. Lastly, the **AUC** (Area Under the Receiver Operating Characteristic Curve) evaluates the model's ability to rank a true planet above a false positive, providing a global measure of separability between classes. While precision, recall, and accuracy are sensitive to the choice of classification threshold, AUC remains threshold-independent and provides a global measure of separability.

Table I summarizes the main evaluation results obtained in training step 625. These results confirm that the model performs remarkably well in distinguishing true exoplanets from false positives. The high AUC score of 0.992 indicates excellent discriminative capacity across classification thresholds. The accuracy of 98,7% reflects the model's overall robustness, with 1.696 of 1.718 examples correctly classified. Despite a slightly lower recall of 78,8%, which suggests some true planets are missed, the precision of 86,7% demonstrates that most predicted planets are indeed real. The confusion matrix also highlights this balance: the model correctly identified 1644 negatives and 52 positives, with 8 false positives and 14 false negatives. These results suggest that the model is correct for automated identification of exoplanets.

TABLE I: Evaluation metrics at training step 625.

| Metric | Accuracy | Precision | Recall | AUC |
|---|---|---|---|---|
| Value | 0.987 | 0.867 | 0.788 | 0.992 |

### C. MAKING PREDICTIONS

Once the model has been trained and evaluated, generating predictions for new Threshold Crossing Events (TCEs) requires only a few input parameters: the orbital period, transit duration, transit epoch ($t_0$), and the Kepler ID of the target star. Given this information, the model returns both a classification prediction indicating whether the signal is likely to be a planet and the corresponding local and global view inputs, which can be used for visual inspection.

Of course, this automated classification is just an initial step in the broader process of exoplanet discovery and validation. A model prediction alone is not sufficient to confirm or rule out the planetary nature of a signal. Proper validation involves extensive follow-up analysis by expert astronomers, including statistical vetting and complementary observational evidence [9]. For a detailed overview of this validation procedure, refer to Sections 6.3 and 6.4.

## III. RESULTS

### A. VALIDATION WITH PUBLISHED DATA

To validate the performance of the implemented algorithm, we conducted an initial test applying it to a sample of confirmed exoplanets previously published in the literature [2]. This control data set serves as a reference point for assessing whether the model is capable of correctly identifying known transit signals. By comparing the predictions generated by the algorithm with the expected outcomes for these well-characterized planetary systems, we aim to establish a baseline for its classification accuracy and robustness.

TABLE II: Prediction values for a subset of TFAW targets.

| EPIC | Pred. | Disp. | EPIC | Pred. | Disp. |
|---|---|---|---|---|---|
| 210768568 | 0.924 | VP | 246078343 | 0.974 | VP |
| 246220667 | 0.960 | VP | | | |
| 210418253 | 0.818 | PC | 210706310 | 0.921 | PC |
| 210708830 | 0.847 | PC | 210945680 | 0.249 | PC |
| 210967369 | 0.947 | PC | 211436876 | 0.808 | PC |
| 218701083 | 0.896 | PC | 247874191 | 0.964 | PC |
| 247223703 | 0.906 | PC | 247744801 | 0.820 | PC |
| 247560727 | 0.923 | PC/CC | | | |
| 205979483 | 0.971 | FP/CC | 220471100 | 0.963 | FP/CC |
| 246022853 | 0.943 | FP/CC | 246163416 | 0.733 | FP/CC |
| 220356827 | 0.660 | FP | 246048459 | 0.925 | FP |
| 211705502 | 0.922 | FP | 211572480 | 0.893 | FP |

As seen in Table II, all validated planets (VPs) in the sample show consistently high prediction scores, all well above the 0.9 level. This demonstrates the robustness of the algorithm for identifying true positives.

Almost all planet candidates (PCs) also receive high scores, such as EPIC 210706310 (0.921), 210708830 (0.847), 210967369 (0.947) 247874191 (0.964), 247223703 (0.906), and 247744801 (0.820). These results reinforce the validity of the 0.8 threshold, as they reflect confident detections consistent with their planetary disposition.

A notable exception is EPIC 210945680, which is classified as a planet candidate but only receives a low prediction score of 0.248. Upon inspection, this object failed the centroid test, a method that detects shifts in the star's apparent position during transit to identify false positives caused by background sources, which likely led the algorithm to classify it with low confidence. It remains listed as a candidate in our sample due to its prior identification by Zink et al. (2021) [10], illustrating the importance of additional context and human vetting in this kind of cases.

True false positives (FP) such as EPIC 220356827 (0.660) and 246163416 (0.733) receive scores below the threshold, which aligns with their classification. This demonstrates the model's ability to down-rank likely non-planetary signals.

Some objects labeled as FP/CC (False Positive or Contaminated Candidate), such as EPIC 205979483 (0.971), 220471100 (0.963), and 246022853 (0.943), display high prediction values. This suggests that the algorithm may respond to planetary-like signals even in systems that are flagged as potentially contaminated. These cases highlight the need to combine Machine Learning predictions with complementary validation metrics, including high-resolution imaging and astrometry.

For example, two FPs, EPIC 211705502 (0.922) and 211572480 (0.893), obtain relatively high prediction scores. In both cases, follow-up observations with speckle imaging have revealed the presence of nearby contaminating stars. In addition, during the vetting process, the authors used Gaia eDR3 data to search for nearby stars and flag possible binary systems by evaluating astrometry indicators such as `GOF_AL`, `D`, and `RUWE` (see Section 2.5 in [2] for details), to rule out the candidates' classification as likely exoplanet candidates.

Finally, EPIC 246048459 (0.925), although having a high prediction score, failed the vetting process in [2] and should be treated with caution.

Based on the previous results, we adopt a threshold of 0.8 for accepting predictions as indicative of potential planet candidates. In this way, we can increase the chances of detecting new transiting signals of planetary nature without introducing too many false positives. The algorithm performs well on validated planets and most high-quality candidates, while also correctly down-ranking likely false positives. Exceptions appear to be linked to contamination or marginal data quality, reinforcing the need for a combined approach that leverages both Machine Learning and traditional vetting methods.

### B. APPLICATION TO TFAW SURVEY CANDIDATES

Another important analysis was conducted on the TFAW-survey dataset, which consists of 478 light curves visually selected as potential exoplanet candidates. Each target in this set was initially assigned a priority level based on visual inspection and human vetting criteria. These priorities are divided into three groups: priority 1 corresponds to low-priority candidates that show transit-like signals but fail some of the vetting criteria; priority 2 includes candidates with clear transit-like signals that require additional vetting; and priority 3 consists of candidates with very clear transit-like signals, presumed to be non-binary in nature, or those belonging to multiplanetary systems. To further refine this classification and assess the effectiveness of our model, we ran the trained algorithm on all the targets in this dataset. The resulting prediction scores were then used to update the original priority flags following specific thresholds. In particular, candidates initially labeled as priority 2 were downgraded to priority 1 if the model prediction was lower than 0.8, while priority 1 candidates were either promoted to priority 2 if their score exceeded 0.8 or demoted to priority 0 otherwise. This process allowed for a more objective reassessment of the significance of the candidate, integrating machine learning-based inference with expert-driven classification.

The three histograms 1 represent the normalized distribution of the prediction values returned by the algorithm for targets classified with priority 1, 2, and 3, respectively. These results provide insights into the

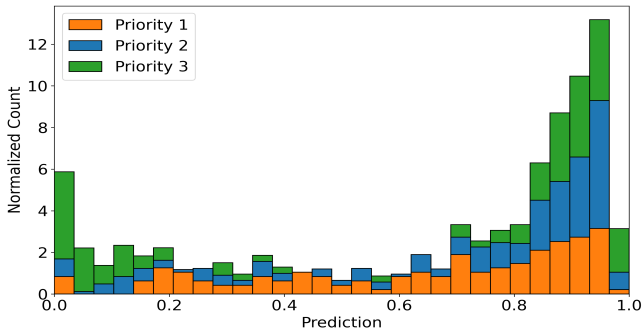


FIG. 1: Normalized histogram of Predictions of priority 1, 2 and 3 TFAW dataset.

behaviour of the model under different levels of initial human-assigned priority. The most interesting behaviour is seen in the Priority 3 targets, which correspond to multi-planetary systems identified in the TFAW visual analysis. The histogram for this group is highly polarized, with a large number of predictions clustered close to 0 or near 1. This bimodal behaviour is a consequence of the model architecture: the algorithm tends to correctly detect the dominant transit signal (usually corresponding to the deepest transit), while failing to detect additional planets in the same light curve, or mistaking them with stellar binary systems. A representative example of this behaviour is found in EPIC 251351134, which appears twice in the dataset with two distinct prediction values. One instance is assigned a high prediction score of 0.948, corresponding to the dominant transit signal in the light curve, while the other receives a much lower score of 0.105, likely associated with a secondary, weaker transit event. This discrepancy illustrates the model's sensitivity to the most prominent signal in multiplanet systems, while often overlooking additional, subtler features within the same dataset.

This limitation is intrinsic to the current algorithm configuration, which was trained assuming a single-transit input. To improve performance on multiplanet systems, a masking strategy should be implemented to isolate individual signals prior to classification.

TABLE III: New priority classification for TFAW survey candidates.

| | Initial Priority | New Priority |
|---|---|---|
| Priority 1 | 139 | 100 |
| Priority 2 | 242 | 201 |
| Priority 3 | 97 | 97 |

### C. APPLICATION TO ALL K2 TFAW LIGHT CURVES

In addition to the validation test, we performed a large-scale analysis on the full set of available TFAW-corrected K2 light curves, excluding those already used during the training phase and those associated with the TFAW-survey candidates (see Section B). The final dataset consisted of 65.970 unique light curves. Each of these was processed through the trained model to evaluate the presence of transit-like signals. From this analysis, the algorithm identified 3.980 targets with prediction scores that exceeded the established threshold of 0.8, indicating a high probability of being genuine exoplanet candidates.

Of the 3.980 light curves with prediction scores above the 0.8 threshold, 180 correspond to candidates previously identified in other surveys but not yet included in the NASA Exoplanet Archive [11]. These systems have been considered documented in the literature and are therefore excluded from our new candidate sample. As a result, we retain a final sample of 3.800 2.

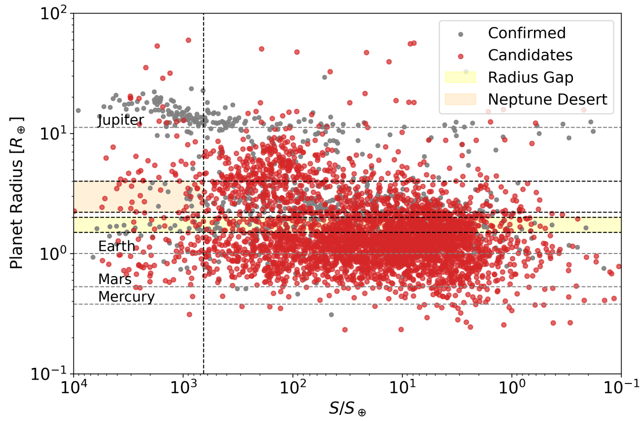


FIG. 2: Planet radius as a function of the stellar insolation for our planet candidates (red) sample versus the distribution of confirmed planets from NASA Exoplanet Archive.

Although not the main objective of this work, we provide a brief description of our new candidate sample. In Figures 2 and 3, we observe that our candidates follow a distribution similar to that of previously confirmed exoplanets. However, a few candidates exhibit radii more than five times larger than Jupiter's, suggesting they may be false positives such as stellar binaries or contaminated candidates rather than true exoplanets. We also identify some Mercury-like planets, which are relatively rare among confirmed exoplanets [12]. We find several candidates located within two regions of interest for exoplanetary science: the Radius Gap and the Neptune desert. The Radius Gap [13] is a region between 1.5 and $2R_\oplus$ where there is a significant deficit of planets, separating smaller super-Earths from larger sub-Neptunes; this gap is linked to planetary formation and atmospheric loss processes. The Neptune desert [14] refers to the scarcity of Neptune-sized planets (2-4 $R_\oplus$) that receive very high stellar irradiation, where intense photoevaporation likely strips away their atmospheres, leaving few planets detectable in this region. Furthermore, we highlight the detection of several Ultra Short Period (USP) planets, defined as having orbital periods shorter than one day [15]. Often called hot Earths or lava worlds, these planets have day-side surface temperatures exceeding the melting point of most rock-forming minerals. USPs play a crucial role in theories of planetary formation and evolution, as their extreme proximity to their host stars challenges existing models.

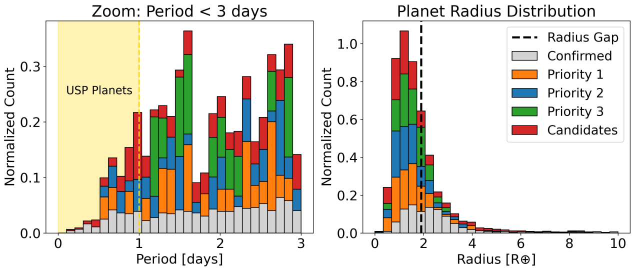


FIG. 3: Summary histograms were we highlight our candidates contribution to the USP and Radius Gap sample.

As an example of one of our newly detected candidates, in Figure 4, we present the top-scoring prediction (EPIC 220438985) from our set of new candidates, with a confidence of 0.984. This example illustrates a typical output of our algorithm. Visual inspection suggests a transit shape consistent with that of a planetary signal. In addition, it passes the Gaia eDR3 astrometric checks, with values of `GOF_AL` = 0.5407, `D` = 0.5407, and `RUWE` = 1.019, indicating no strong evidence of contamination from unresolved companions. The candidate exhibits an orbital period of 6.75 days and an estimated radius of 0.52 Earth radius. However, this and the other candidates still require further study and thorough vetting to confirm their planetar nature, as noted in Section 3.A.

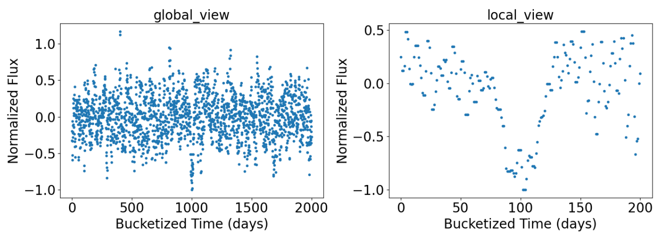


FIG. 4: EPIC 220438985 light curves output from AstroNet CNN. Left: global view; right: local view.

## IV. CONCLUSIONS

The results of this study demonstrate the effectiveness of applying a Machine Learning model to TFAW-processed light curves. Initially, we established a prediction threshold of 0.8, based on previously published data. This threshold allowed us to refine the selection process by reclassifying the initial priority groups and discarding several visually selected candidates from the TFAW sample that did not meet the confidence criteria. From the original set of 478 targets, we keep 398, which were given updated priority groups as follows: 100 classified as priority 1, 201 as priority 2, and 97 as priority 3.

After this refinement step, we scaled up the process and applied the trained model to a dataset consisting of 65.970 unique light curves. This led to the identification of approximately 3.800 potential planetary candidates, which represent a promising dataset for future vetting and analysis.

As a future work, we plan to study and vet the TFAW candidates with updated priorities. On the other hand, we aim to upgrade the algorithm to identify multiplanetary candidates, such as those currently labeled as priority 3, which may be better classified/detectable through masking techniques. Additionally, we intend to thoroughly analyze the entire newly identified subset of potential candidates by running the full TFAW vetting procedure, assigning priority values, and conducting follow-up observations (e.g., speckle imaging, adaptive optics) for those new candidates with high priority flags. We could also improve our Machine Learning algorithm by implementing Gaia DR3 data, specifically `RUWE`, `GOF_AL`, and `D` parameters, to enhance false positive detection. Other future work will include applying the prediction procedure to other surveys such as Kepler and TESS, of course creating new test and validation sets.

### Acknowledgments

I acknowledge Dr. Daniel del Ser for his exceptional support throughout the development of this work.

[1] S. B. Howell, C. Sobeck, et al., PASP **126**, 398 (2014).
[2] D. del Ser, O. Fors, et al., MNRAS **518**, 669 (2023).
[3] D. del Ser, O. Fors, et al., A&A **619**, A86 (2018).
[4] R. Luger, E. Kruse, et al., AJ **156**, 99 (2018).
[5] M. Hippke and R. Heller, A&A **623**, A39 (2019).
[6] D. del Ser and O. Fors, MNRAS **498**, 2778 (2020).
[7] Kepler Project, NASA IPAC DataSet, NEA23 (2021).
[8] J. L. Coughlin, Kepler Science Document KSCI-19114-001, id. 1 (2017).
[9] C. J. Shallue and A. Vanderburg, AJ **155**, 94 (2018).
[10] J. K. Zink et al., AJ **162**, 259 (2021).
[11] Caltech, *NASA Exoplanet Archive*, `https://exoplanetarchive.ipac.caltech.edu/` (2025).
[12] Barros, S. C. C. et al., A&A **665**, A154 (2022).
[13] B. J. Fulton, E. A. Petigura, et al., AJ **154**, 109 (2017).
[14] M. S. Lundkvist, H. Kjeldsen, et al., Nat. Commun **7**, 11201 (2016).
[15] E. R. Adams, B. Jackson, et al., AJ **152**, 47 (2016).

# Identificació Automàtica d'Exoplanetes amb *Machine Learning*

Author: Vicenç Fernández Fernández, vfernafe18@alumnes.ub.edu
*Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.*

Advisor: Daniel del Ser Badia, danieldelser@ub.edu

**Resum:** La detecció d'exoplanetes és un camp en ràpida evolució, cada cop més recolzat pels avenços en l'àmbit del *Machine Learning*. En aquest treball, explorem les capacitats de l'algoritme de *deep learning* AstroNet quan s'aplica a corbes de llum preprocesades amb l'algoritme TFAW. L'objectiu és classificar els esdeveniments de creuament i identificar nous candidats a exoplaneta.

En primer lloc, validem el rendiment del model sobre un subconjunt d'exoplanetes prèviament confirmats, mostrant que l'algoritme recupera amb èxit puntuacions de predicció elevades, tal com s'esperava. Posteriorment, analitzem un subconjunt seleccionat visualment de 478 candidats de l'estudi TFAW amb nivells de prioritat assignats, utilitzant la sortida del model per proposar reclassificacions de prioritat basades en criteris objectius. Finalment, apliquem el model a un conjunt de dades de 65.970 corbes de llum de la missió K2, identificant 3.800 candidats anteriorment no reportats.

Els nostres resultats demostren que AstroNet, combinat amb TFAW, és una eina potent per a la classificació automàtica de candidats a exoplaneta. Tot i això, també remarquem que aquests models no són definitius, i que els mètodes de validació complementaris continuen essent essencials per confirmar la naturalesa planetària de qualsevol nou candidat en trànsit.

**Paraules clau:** Exoplanetes, Sistemes Planetaris, Fotometria, Anàlisi de Dades, Aprenentatge Automàtic
**ODSs:** Educació de qualitat

**Objectius de Desenvolupament Sostenible (ODSs o SDGs)**

| | | | |
|---|---|---|---|
| 1. Fi de la es desigualtats | | 10. Reducció de les desigualtats | |
| 2. Fam zero | | 11. Ciutats i comunitats sostenibles | |
| 3. Salut i benestar | | 12. Consum i producció responsables | |
| 4. Educació de qualitat | X | 13. Acció climàtica | |
| 5. Igualtat de gènere | | 14. Vida submarina | |
| 6. Aigua neta i sanejament | | 15. Vida terrestre | |
| 7. Energia neta i sostenible | | 16. Pau, justícia i institucions sòlides | |
| 8. Treball digne i creixement econòmic | | 17. Aliança pels objectius | |
| 9. Indústria, innovació, infraestructures | | | |