A WORLD AWAY

HUNTING FOR EXOPLANETS WITH AI

TEAM RED SPOT

Data from NASA missions like Kepler, K2, and TESS have led to thousands of exoplanet discoveries, but most were identified manually. With AI/ML, we can automatically analyze these large datasets to detect exoplanets more efficiently and accurately. Our challenge is to build an AI/ML model using NASA's open datasets to classify and identify new exoplanets.

EXOPLANETARY EXPLORATION TIMELINE

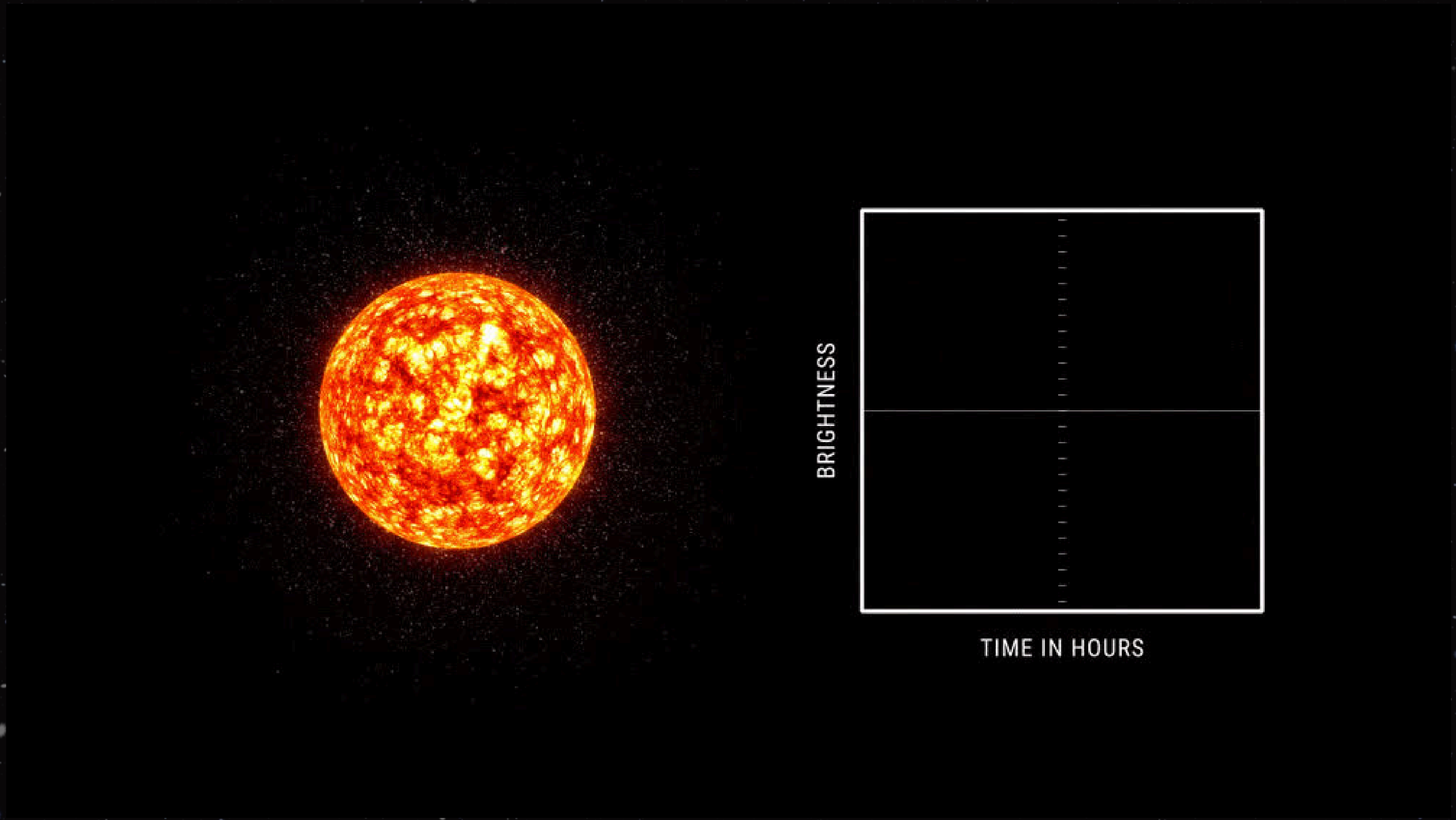
2009 – Kepler Launch
- First dedicated space telescope using the transit method.
- Collected data for nearly a decade.
- Thousands of exoplanet candidates identified (mostly manually).

2014 – K2 Mission
- Successor to Kepler, using the same hardware but a new survey path.
- Continued manual exoplanet identification with improved coverage.

2018 – TESS Launch
- Ongoing mission, scanning almost the entire sky.
- Similar transit-based detection, delivering massive amounts of new data.

Present
- Public datasets from Kepler, K2, and TESS available through NASA.
- Each entry includes features like orbital period, transit duration, planetary radius, etc.
- Growing use of AI/ML to automate identification → promising results, but much of the work is still done manually.

RED SPOT



Light curves will help to detect transit through change in Brightness whereas radial velocity measures star wobble

A DECADE OF DISCOVERY: FROM KEPLER'S KOI PIPELINES TO AI-POWERED EXOPLANET VETTING

- Classical Transit Detection (Kepler Era)

Jenkins (2010), Mullally (2015), Thompson (2018) used TPS to detect periodic dips, relying on manual vetting.

⚠️ Limitations: High false positives, noise sensitivity, poor detection of small/long-period planets.

- Automated Feature-Based ML

Gradient-Boosted Trees + TSFRESH features improved speed and interpretability, achieving AUC ≈ 0.94 (Kepler), ≈ 0.81 (TESS).

⚠️ Limitations: Class imbalance, lower precision, heavy manual feature engineering.

- Deep Learning Revolution — ExoMiner

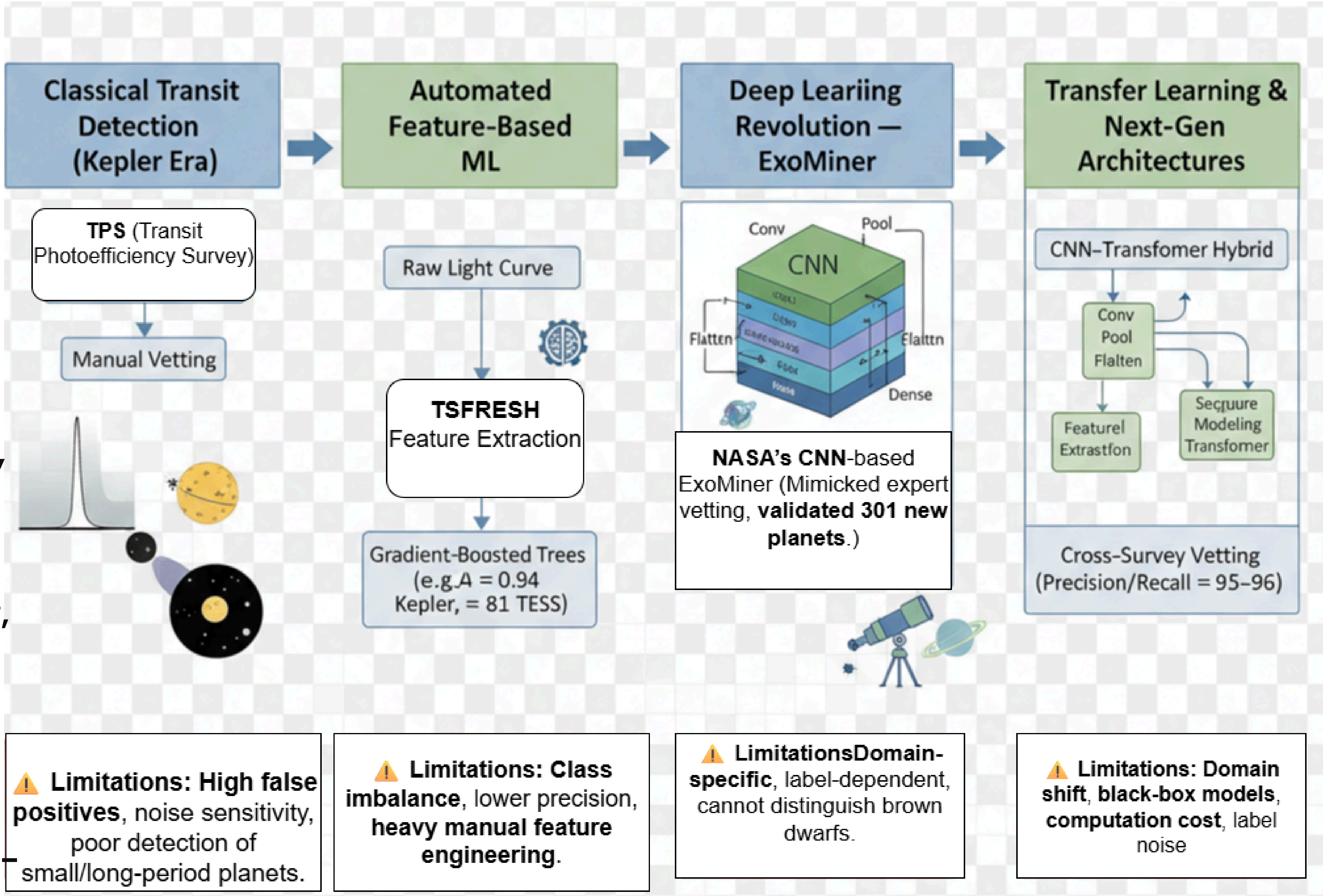
NASA's CNN-based ExoMiner mimicked expert vetting, validating 301 new planets.

⚠️ Limitations: Domain-specific, label-dependent, cannot distinguish brown dwarfs.

- Transfer Learning & Next-Gen Architectures

ExoMiner++ + CNN–Transformer hybrids enabled cross-survey vetting with Precision/Recall ≈ 0.95–0.96.

⚠️ Limitations: Domain shift, black-box models, computation cost, label noise.



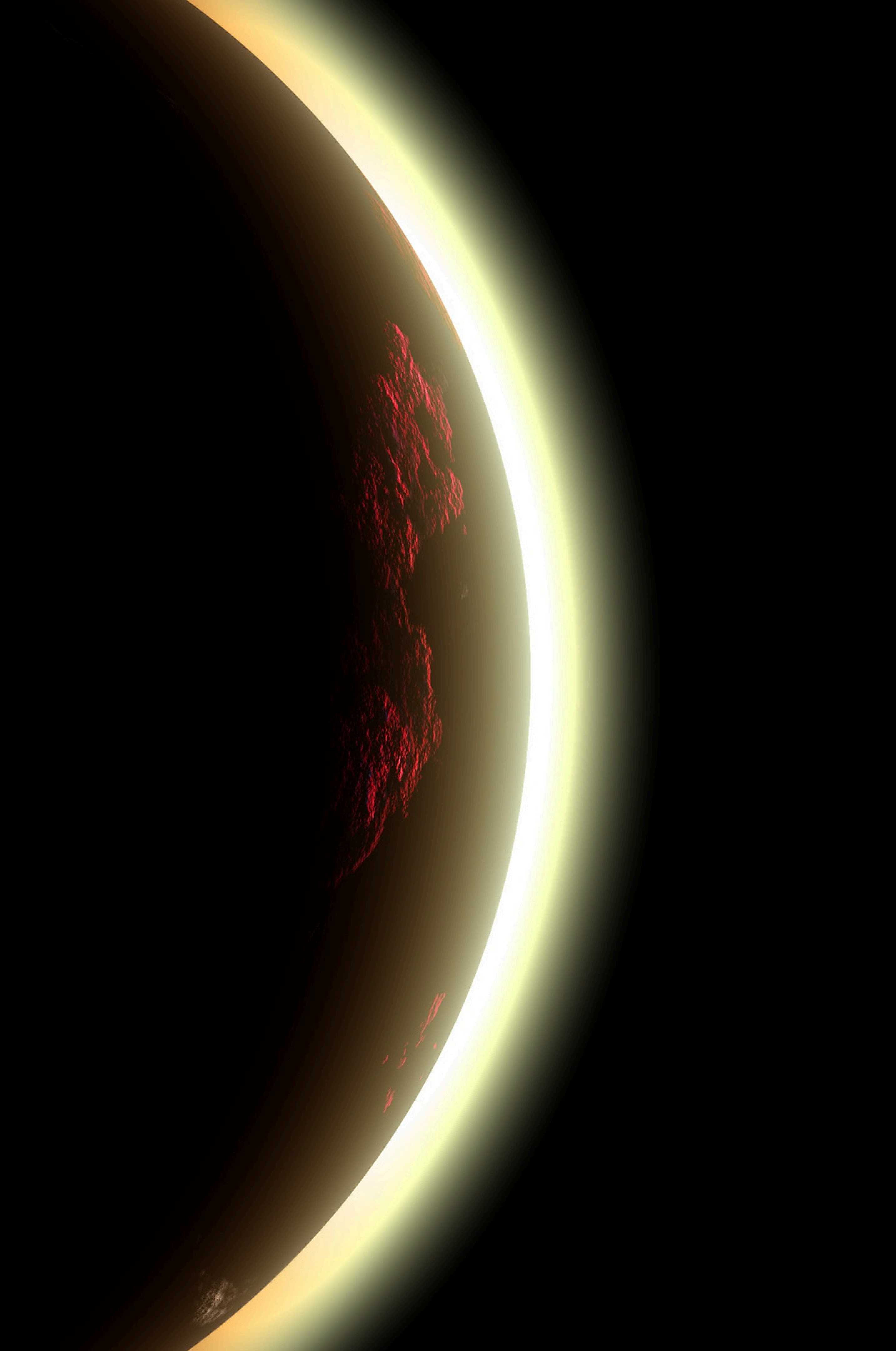
LITERATURE SURVEY

01 Robovetter : Transit-shape metric + Centroid / difference-image analysis + Ephemeris matching + Flag Logics + Monte Carlo

02 Multiplicity Boost of Transit Signal Classifiers : given an existing transit-signal vetter (classifier), improve its performance using multiplicity information.

03 ExoMiner++: Enhanced Transit Classification and a New Vetting Catalog for 2-Minute TESS Data

04 Astronet : (Shallue & Vanderburg, 2017) applied a deep CNN with global and local light-curve views to Kepler data, achieving ~96% accuracy and discovering Kepler-90i, though limited by generalization, interpretability, and dependence on Kepler preprocessing.



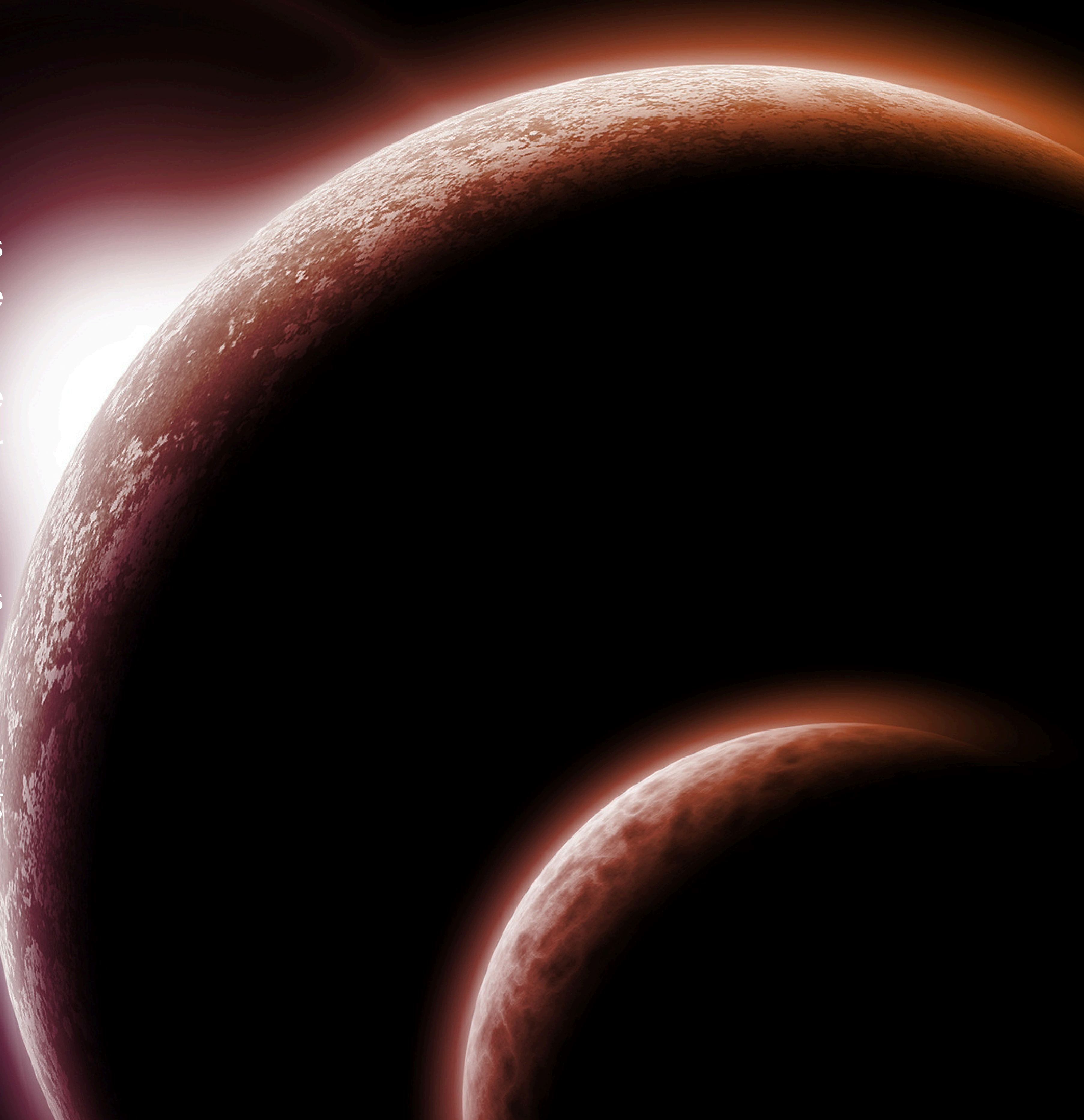
RED SPOT

OUR APPROACH

- Current SOTA Approaches, heavily rely on Robovetter and it’s Monte Carlo simulations to fight with any bias coming with rule based logics.
- ExoMiner++ and Multiplicity Boost of Transit Signal are the state of the art frameworks as well as neural networks for classifications.

These approaches are difficult for explainability which makes us come to a conclusion, are these verifiable?

For this we come with approaches heavily relying on Classical ML Techniques like Naive Bayes and Random Forest in a Ensembling method.



THANK YOU!