ExoAID

Artificial Intelligence for Exoplanet Detection

Asier Abreu[1]
PhD Supervisors : Jorge-Lillo Box [2] , Ana Maria Perez [2]

[1]ATG Science & Engineering S.L. Asier.Abreu@ext.esa.int

[2]CAB-INTA jorge.lillo@cab.inta-csic.es , CAB-INTA aperez@cab.inta-csic.es

UCM , December 2023

Overview

Detection : Statistics

The current census of exoplanets

- To date (2023) $\sim$ 5000 confirmed exoplanets.
- This sample has selection biases introduced by the dominant detection techniques.
- Transit favours close orbiting large planets and only edge-on detection.
- Radial Velocity maps massive exoplanets, close-in to the host star.

Detection Statistics Breakdown

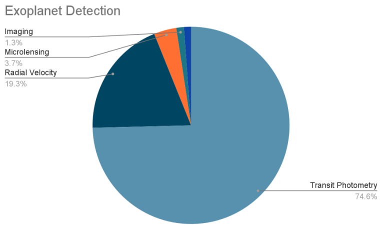


Figure: (from Author) Current breakdown of exoplanets confirmed by detection method. Data Source : http://exoplanetarchive.ipac.caltech.edu

Detection : Limitations

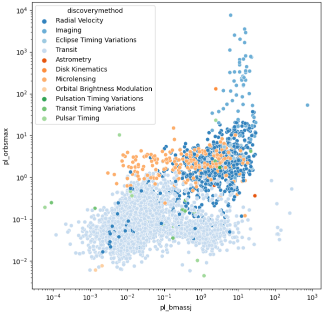


Figure: (from Author) Orbit Size vs Mass for all confirmed exoplanets (2023) color coded by detection method.Data Source: http://exoplanetarchive.ipac.caltech.edu

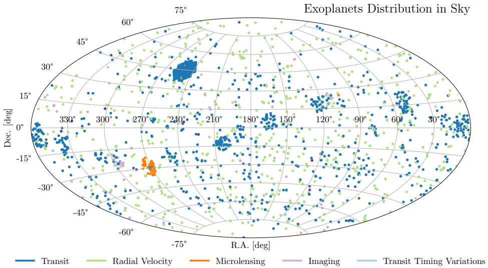


Figure: (from Author) Sky position of stars hosting confirmed exoplanets (2023) Data Source: http://exoplanetarchive.ipac.caltech.edu

Exoplanet Detection at scale

Thesis Goal and Approach:

- Improve the current statistics and remove selection biases.
- Analyze and combine large available astronomical datasets using in Deep Learning.
- Detect exoplanets at scale over the full sky and characterize their physical properties.
- Revise current theories of planet formation in view of the results.

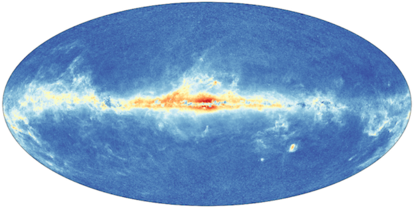
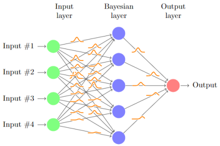


Figure: Credit : ESA/Gaia/DPAC, CC BY-SA 3.0 IGO.



Astrometric Exoplanet Detection

The technique: A companion induces a displacement in the star's position causing it to "wobble" around the system c.o.m.

The challenge: The magnitude of this distortion is so small ($\sim \mu$as) that so far we only have 2 astrometrically detected exoplanets.

The approach: Apply artificial intelligence to the Gaia catalog for systematic search of companion signature in stars.

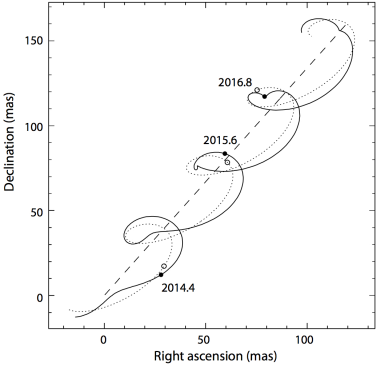


Figure: Credit: Perryman M. Schematic of the path on the sky of a $1M_{\odot}$ star at a $d = 50pc$, with a proper motion of 50 milli-arcsec per year, and orbited by a planet of $15M_J$, $e = 0.2$, and semi-major axis $a = 0.6AU$. The straight dashed line shows the system's barycentric motion viewed from the solar system barycenter.

Astrometric Signature

Distorsion Magnitude:

$$\alpha = \left(\frac{M_p}{M_*}\right)\left(\frac{a_p}{d}\right) [arcsec]$$

where:

M_p = exoplanet mass [$M_\odot$]
M_* = host star mass [$M_\odot$]
a_p = exoplanet orbit semi-major axis [AU]
d = distance to the host start [pc]

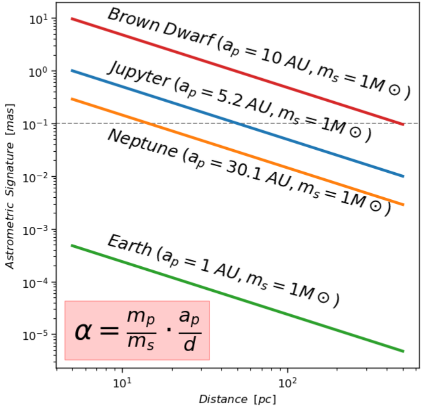


Figure: (from Author) Astrometric signature α for example (Solar like) exoplanets as a function of distance in pc. The gray dashed line shows the Gaia astrometric precision 100 μas

Use Gaia to detect Exoplanets!

Gaia Mission:
ESA's Gaia mission provides ultra precise ($\sim 100\mu$as) **all sky** measurements of $(\alpha, \delta, \mu_\alpha, \mu_\delta, \varpi)$ for 1.8 billion sources!.

	# sources in Gaia DR3
Total number of sources	**1,811,709,771**
Number of 5-parameter sources	585,416,709
Number of 6-parameter sources	882,328,109
Number of 2-parameter sources	343,964,953
Sources with mean G magnitude	1,806,254,432
Sources with mean GBP-band photometry	1,542,033,472
Sources with mean GRP-band photometry	1,554,997,939

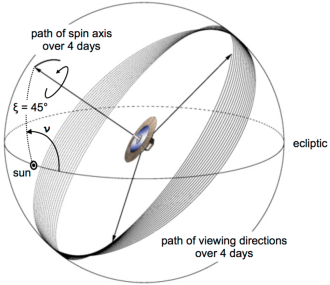
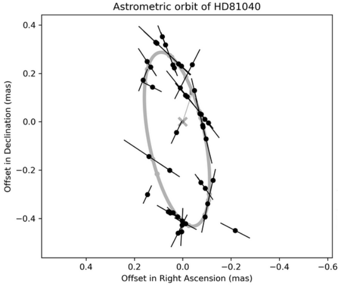
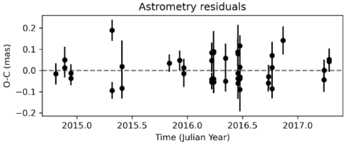


Figure: Credit : ESA/Gaia/DPAC

Estimations: Gaia should be able to detect $\sim$ 20000 large-mass (1-15 M_J) long period exoplanets ([1]) out to $d = 500pc$ using astrometric exoplanet detection.

Gaia Observations





Credit : ESA/Gaia/DPAC

Astrometric observations:
But Gaia does not measure (α), it measures the 1D displacement $(\Delta\eta)$ of the star's position along the scan direction (many times).

Astrometric model (5p):

$$\Delta\eta = [\Delta\alpha^* + \mu_{\alpha^*} t] \sin\theta + [\Delta\delta + \mu_\delta t] \cos\theta + \varpi f_\varpi$$

Astrometric fit: We can perform a least squares fit to minimize the sum of squared residuals :

$$R_i = \Delta\eta_i^{observed} - \Delta\eta_i^{computed}$$

Artificial Intelligence

Basic hypothesis: The presence of companion should reflect in the astrometric fit quality. We can use fit quality statistics to induce the presence of a companion.

$$\chi^2 = \frac{1}{\nu}\sum_{i=1}^{N}\frac{R_i^2}{\sigma_i^2}$$

where:

$R_i = \Delta\eta_i$ Residuals
$\sigma_i = \Delta\eta_i$ measurement uncertainty
$\nu = N_{obs} - 5$

Methodology: Generate deep learning models capable of predicting the probability of a star in the Gaia catalog to hold a substellar companion.

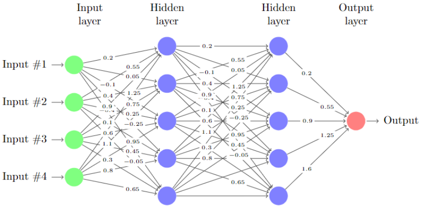


Figure: (from Author) Illustrative Deep Neural Network Architecture showing the weights of each neuron.

Artificial Intelligence : Supervised Learning

Approach: Use a labeled dataset to train neural networks.

Attributes (Fit statistics)

$$x_1 = ruwe$$
$$x_2 = astrometric\_chi2\_al$$
$$x_3 = astrometric\_excess\_noise$$
$$x_n = \ldots$$

Labels

1 = star with Gaia detectable companion
0 = star with NO Gaia detectable companion

Output:

$$p_0(\bar{x}) = \mathrm{P}(c = 0|x_1, x_2, \ldots, x_n)$$
$$p_1(\bar{x}) = \mathrm{P}(c = 1|x_1, x_2, \ldots, x_n)$$

But ... No negative labels!: There is no catalog of stars with *NO* companion.

Generate synthetic data: We simulate 1D Gaia-like observations for sources with and without a companion and perform a least-squares fit.

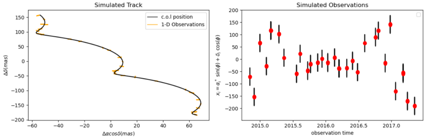


Figure: (from Author) Simulated Gaia-like observations.

Synthetic Dataset Characteristics

Simulations:

- **1:** Neptunes, Jupyters and Brown-dwarfs in close orbits (0<P<10 yr) around $1M_\odot$ star.
- **2:** Neptunes, Jupyters and Brown-dwarfs in close orbits (10<P<100 yr) around $1M_\odot$ star.
- **3:** Neptunes, Jupyters and Brown-dwarfs in close orbits (P>100 yr) around $1M_\odot$ star.

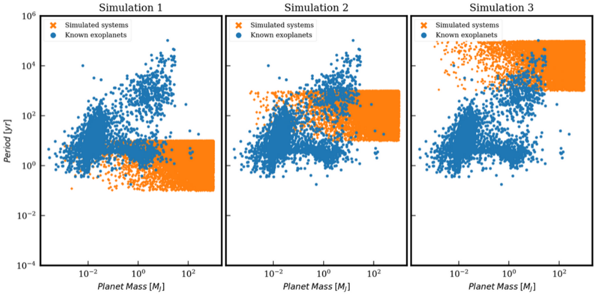
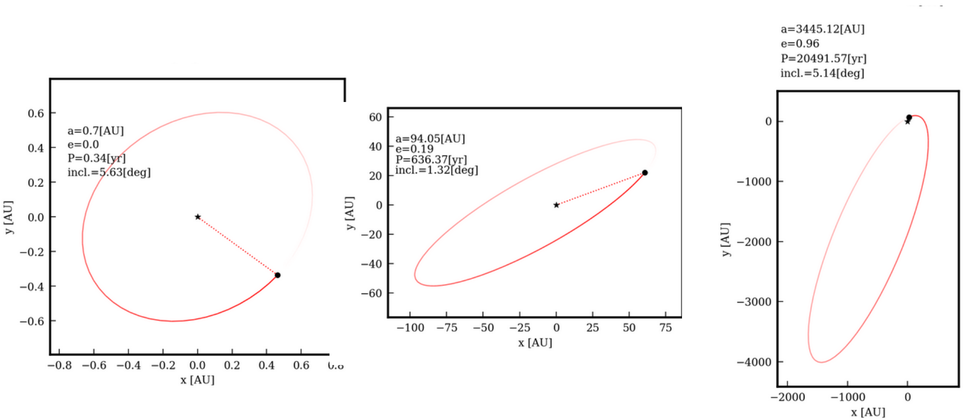


Figure: (from Author) Period vs Mass for simulated systems (orange) and known exoplanets (blue).

Synthetic Dataset Examples



Artificial Intelligence : Attribute Behaviour

How do the attributes behave?: We verify how the astrometric fit quality parameters (attributes) behave in the simulated dataset

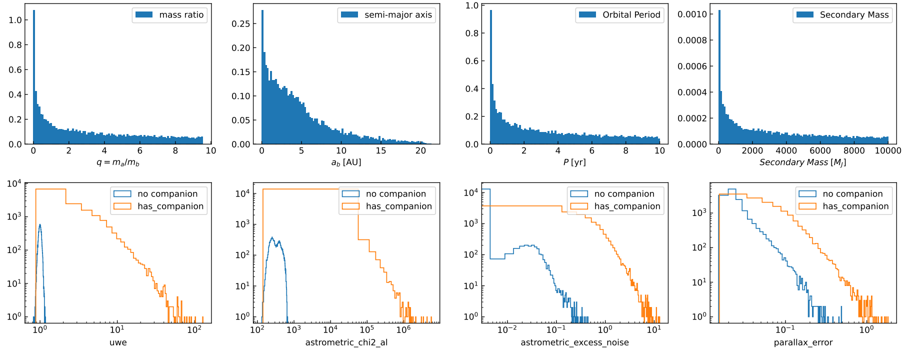


Figure: (from Author) (Top row) Orbital parameter distributions of simulated binaries, (Bottom row) Astrometric fit statistics distributions color coded by learning flag , simulated companion (orange) and those with no companion (blue).

Artificial Intelligence : Attribute Importance

Which are the most important attributes for learning?: How significant are the astrometric fit quality statistics in determining the target?

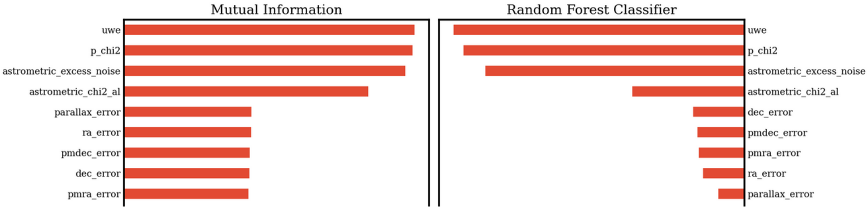


Figure: (from Author) Attributes relevance for determining the target order by descending importance estimated using 2 independent methods , information retrieval theory and random forest classification.

AI Model : Architectures and Metrics

Deep Learning Architectures:
We use deterministic and probabilistic neural network architecture.

1. (Deterministic) Deep Neural Network architecture for companion detection.
2. (Probabilistic) Deep probabilistic Neural Network for orbital parameters characterisation.

The later to incorporate the ability to estimate the uncertainty on orbital parameter estimation.

Quality Metrics:
Classification metrics:

1. F1-score
2. Log Loss

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN}$$

$$L = \frac{1}{N} \sum_{i=1}^{N-1} y_i \cdot log(p_i) + (1 - y_i) \cdot log(1 - p_i)$$

$F1 - score$: trade-off between precision and recall. In this domain it is important that we correctly classify stars that do host exoplanets as such (precision), and that we capture as many of those as possible (recall).

Results: Internal Calibration (1)

Deep Neural Network (Results) : Synthetic Dataset provides internal calibration.

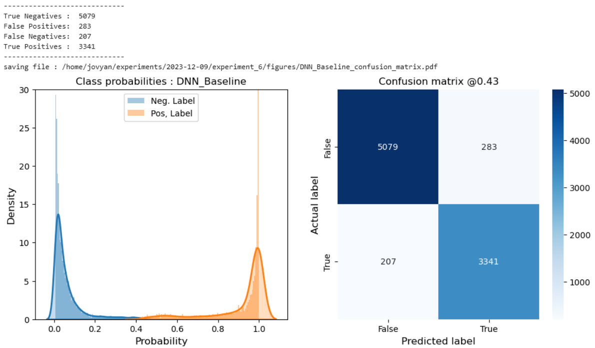


Figure: (from Author) Learned probability distribution of the target class (Left). Confusion matrix over the synthetic dataset (right).

Deep Neural Network (Architecture) : Model and layers description

Model: "model"

Layer (type)	Output Shape	Param #
input (InputLayer)	[(None, 7)]	0
batch_normalization (BatchNormalization)	(None, 7)	28
dense (Dense)	(None, 56)	448
dense_1 (Dense)	(None, 56)	3192
dense_2 (Dense)	(None, 56)	3192
dense_3 (Dense)	(None, 56)	3192
dropout (Dropout)	(None, 56)	0
output (Dense)	(None, 1)	57

Total params: 10109 (39.49 KB)
Trainable params: 10095 (39.43 KB)
Non-trainable params: 14 (56.00 Byte)

Figure: (from Author) Deep learning architecture (Tensorflow).

Results: Internal Calibration (2)

Deep Neural Network (Detection) : Results on Synthetic Dataset provide internal calibration (how good does the AI model perform on controlled dataset).

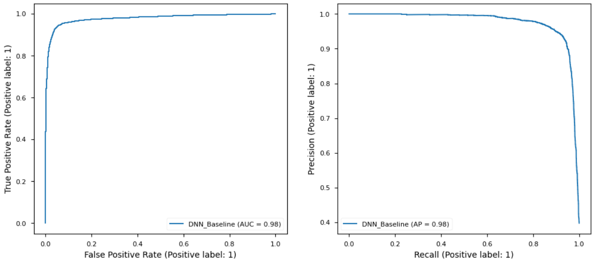


Figure: (from Author) AUC Curve (Left). Precision-Recall curve (right).

	0
model	DNN_Baseline
f1_score	0.932
log_loss	0.166

Figure: (from Author) Deep Neural Model Performance.

Results: External Calibration (1)

- External Calibrator : List of NSS Astrometric Binaries
- Data source: Gaia Non-Single-Stars from Gaia DR3 release.
- Model Performance : only ∼50% of binaries detected as such (too few!)

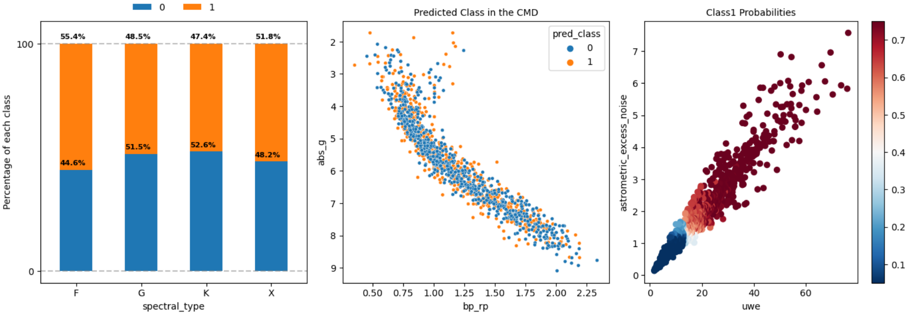


Figure: (from Author) Normalized fraction of detected objects per spectral type (Left). Color-magnitude distribution obf objects per target class (middle). Predicted probabilities of belonging to class 1 as a function of uwe and excess noise features (right).

Results: External Calibration (2)

- External Calibrator : List of NASA Confirmed Exoplanets
- Data source: NASA Exoplanets cross-matched with Gaia DR3.
- AI model performance: ∼50% of exop. detected as such (too many!)

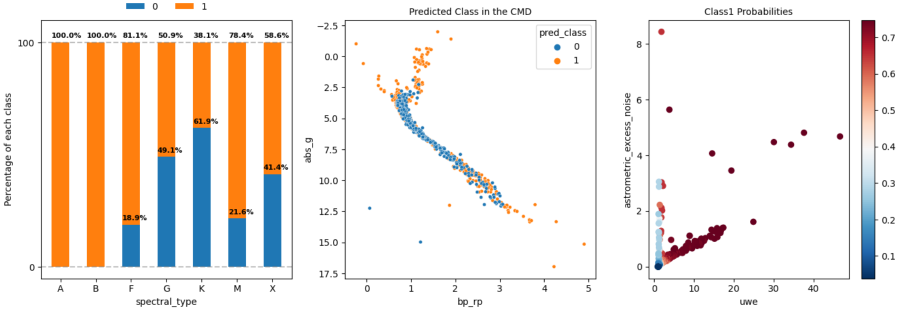


Figure: (from Author) Normalized fraction of detected objects per spectral type (Left). Color-magnitude distribution obf objects per target class (middle). Predicted probabilities of belonging to class 1 as a function of uwe and excess noise features (right).

Results: External Calibration (3)

- External Calibrator : List of Gaia Exoplanets candidates
- Data source: Gaia Non-Single-Star exoplanets candidate list
- AI Model Performance : ∼30% of candidates detected as such (too few!)

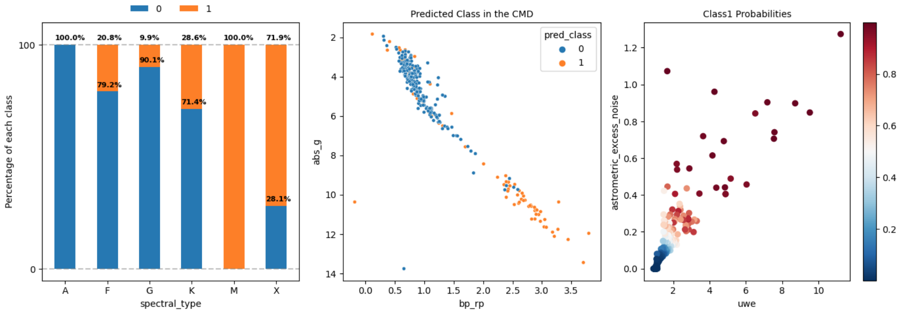


Figure: (from Author) Normalized fraction of detected objects per spectral type (Left). Color-magnitude distribution obf objects per target class (middle). Predicted probabilities of belonging to class 1 as a function of uwe and excess noise features (right).

Look-Ahead: Planned Activities

Planned work - short term:

1. Improve **external calibration** e.g. using further Gaia Non Single-Stars results
2. Improve **external calibration** to extend the coverage of the Mass-Period parameter space.
3. Apply AI model to unseen data from the Gaia DR3 catalog.

Planned work - long term:

1. Propose follow up observations on candidate exoplanet list.
2. Work on orbital parameter $(i, e, a, \Omega, \omega, \nu)$ recovery with Bayesian Neural Networks.
3. Incorporation other data types (radial velocity) for binary parameters (mass) contraining.

Thanks ! ... Q/A

Michael Perryman, Joel Hartman, Gáspár Bakos, Lennart Lindegren
Astrometric exoplanet detection with Gaia
The Astrophysical Journal (2014).

Erik A. Petigura E. A., Howard A. W. and Marcy G. W.
Prevalence of Earth-size planets orbiting Sun-like stars
PNAS 110 19273 (2013).

Schmidhuber, Jürgen
Deep Learning in Neural Networks: An Overview
Neural Networks (2015).