Planetary Markers in Stellar Spectra: Jupiter-host Star Classification

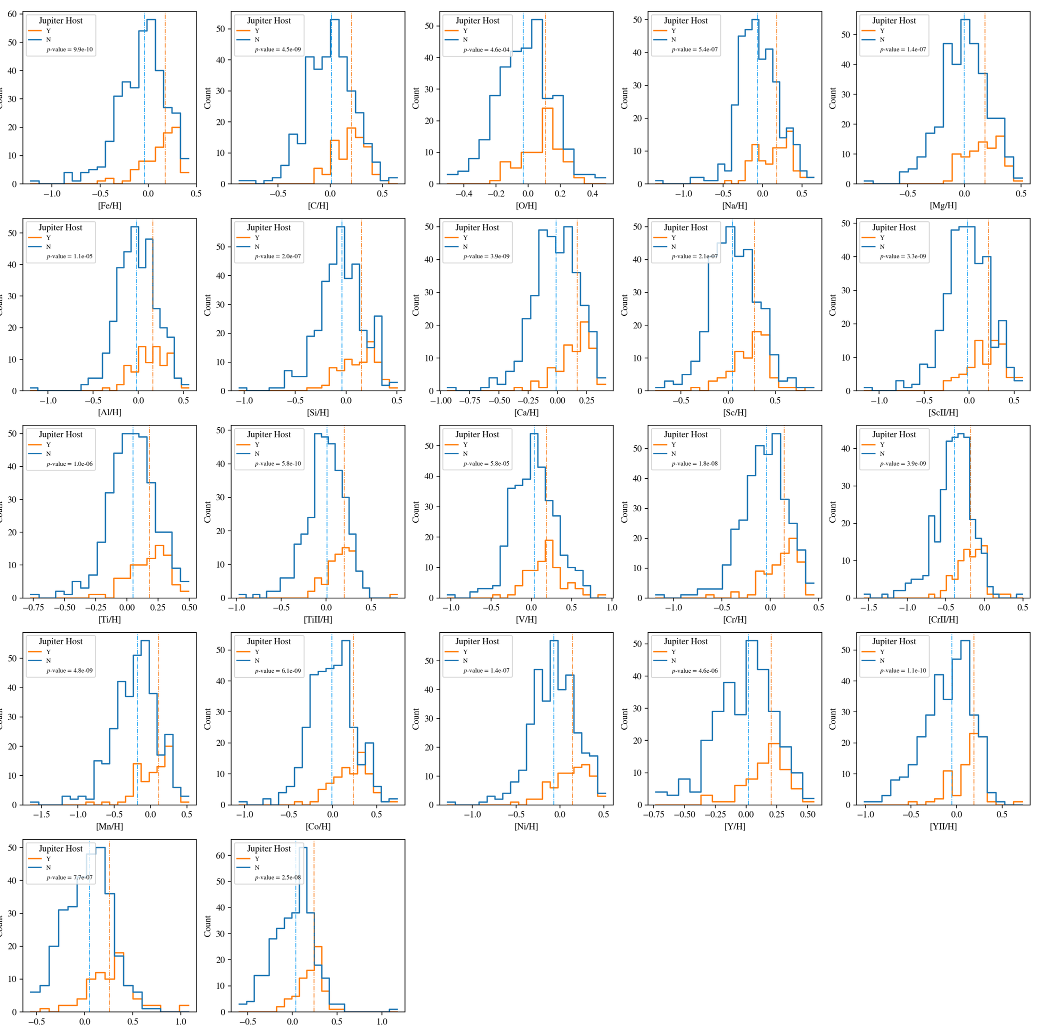
Miguel Andrea Zammit, Kristian Zarb Adami

Institute of Space Sciences and Astronomy (ISSA), University of Malta, Malta

Decades of observational & theoretical research has explored the relation between a star's chemical, physical and galactic properties with the presence of orbiting planetary companions. Certain sources suggest that observed correlations are indicators of the environment of the system's protoplanetary disc, and subsequently its proclivity to facilitate planetary formation. This project aims to use the predictive power of machine learning to develop a classifier that uses spectral data of labelled target stars, to learn to model subtle discriminating markers and predict a binary class (Jupiter host or non-host) for every instance. Two approaches were highlighted: The first method was to use raw high-resolution stellar spectra as inputs, in order to preserve any inherent information within the spectrum. The second method was to use homogeneous elemental abundance data curated from a preexisting catalog, and implement a system capable of separating the planetary hosts from comparison stars based solely on the abundance levels of certain elements. To determine whether using raw high-resolution stellar spectra leads to consistent learning and generalisation, several convolutional neural networks, particularly for their strengths in image classification, were implemented in a stacked architecture. Every CNN model was assigned a particular spectral range to collectively cover the entire spectrum, the results of which were then fed into a meta learner to aggregate their votes. Multi-objective optimization will be used to train a model using the elemental abundance feature data while imputing incomplete records in an online approach. Cross-examination of both approaches will then be conducted.

Aims & Motivation

It has been a coveted goal for the field to map out the distribution of planet occurrence with all manner of stellar & orbit parameters [6, 10]. In particular for Jupiter-class exoplanets, there has been several corroborated work reinforcing the hypothesis that certain stellar properties of a planetary-host star yield a higher probability for gas-giant formation in its primordial protoplanetary disk [4, 1, 2, 5].



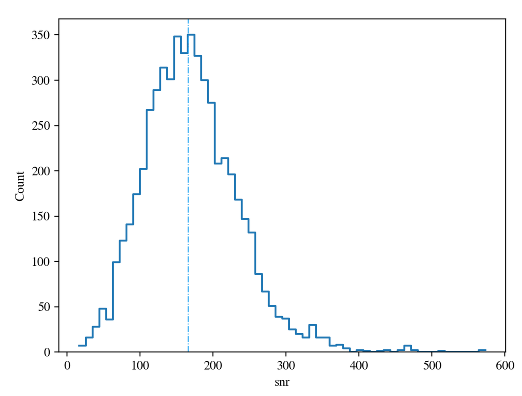
In this work, we apply ML techniques to develop a classifier capable of discriminating between Jupiter hosts and comparison stars solely based on the spectral data collected from the planetary system's host star. The focus was placed on gas giant exoplanets due to the higher overall confidence in the significance of the primordial link between stellar chemistry and core accretion gas-giant formation. Furthermore, with current technological capabilities, gas giant detection has lower margins of error and higher chances for detection. Thus, any labelling of such a dataset of stellar hosts and non-hosts can be done with greater confidence. We decided to approach the problem through two different input designs.

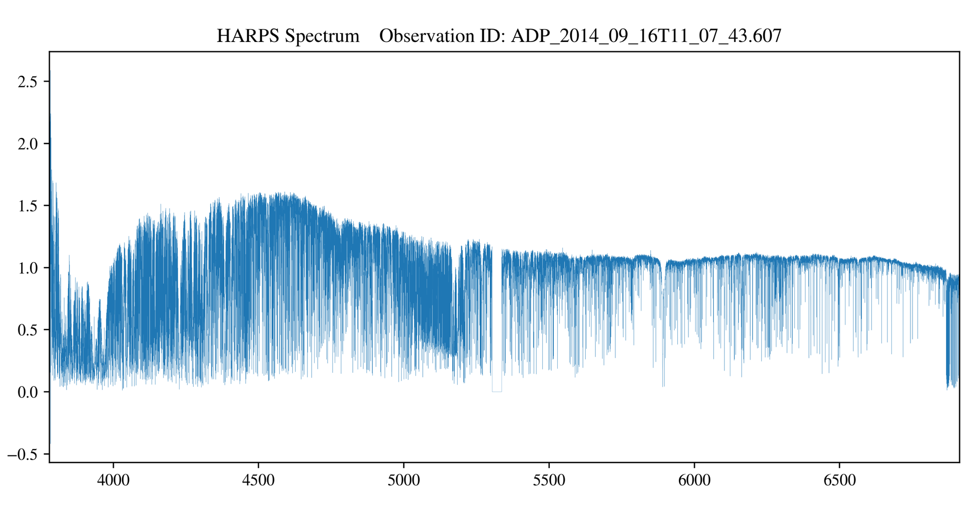
The first method was to use raw high-resolution spectra as direct inputs, and implement a stacked CNN classification architecture to assign a classification probability on whether that spectrum's star is a Jovian host. A particular draw to such a design comes from the fact that if it truly leads to a stable, accurate and precise classifier, such a system could hypothetically serve as a preliminary check for exoplanet detection.

The second technique which we will apply focuses on the use of elemental abundance data and stellar parameters as input features for the host and comparison stars in the dataset. Partially motivated by the work found in [3], applying this method to the same stars within the spectral dataset will allow for cross-examination of both input designs.

Dataset

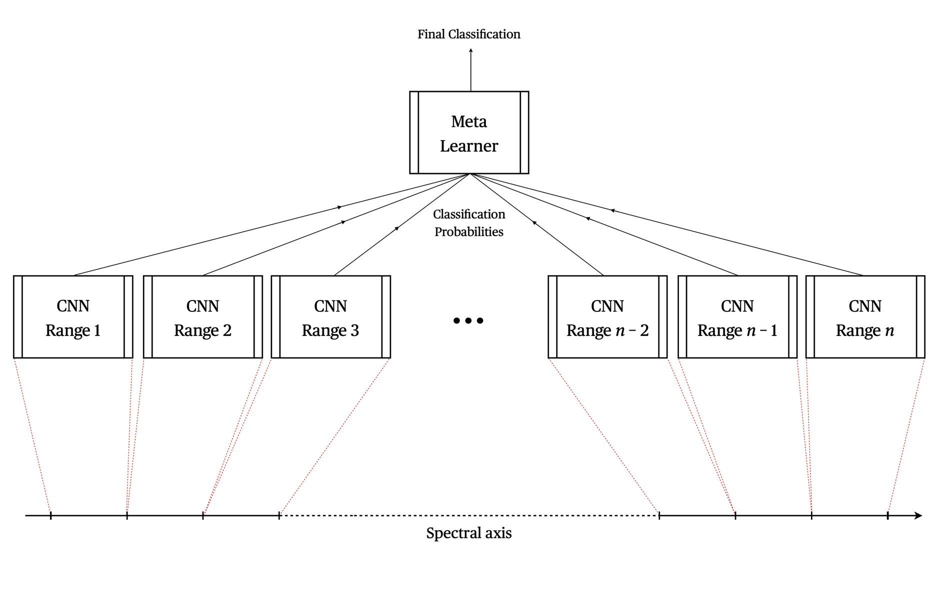
A spectral dataset was compiled using high-resolution optical spectra observed with either the HARPS (High Accuracy Radial velocity Planet Searcher) [9] or FEROS (Fiber-fed Extended Range Optical Spectrograph) [7] spectrographs mounted on the European Southern Observatory (ESO) La Silla 3.6m telescope and the ESO/MPI 2.2m telescope respectively. An elemental abundance dataset was compiled of the same target stars from entries in the Hypatia Catalog, an unbiased dataset of spectroscopic abundance data from 233 different literature sources for a total of 9,982 stars in the solar neighbourhood and 80 elements & species. We restrict the number of elemental features to 22 elements linked by previous work to correlations with gas-giant core accretion.





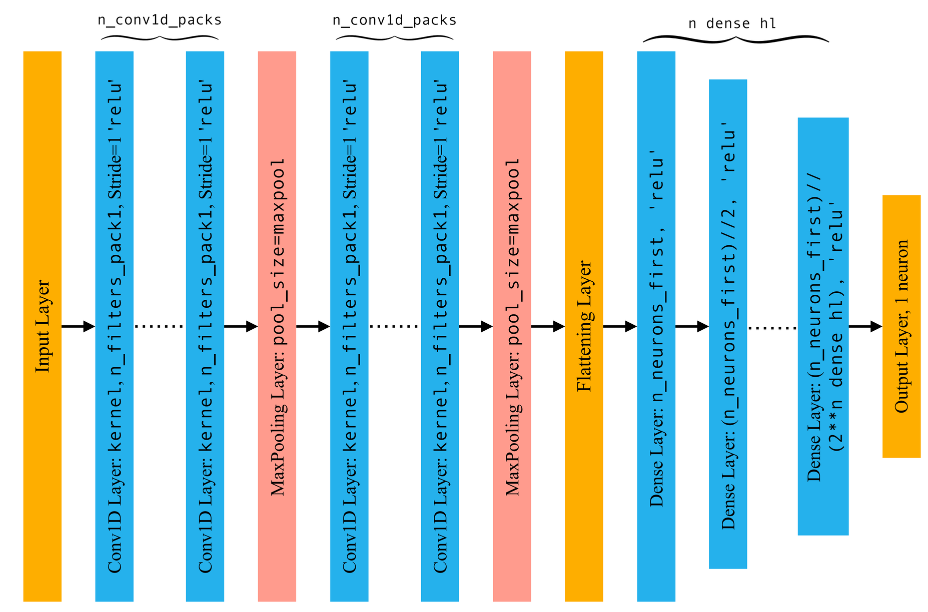
A strict methodology was put in place while selecting spectra, to ensure that the dataset is clean and absent of any major sources. Multiple spectra for single targets were allowed, with a maximum of 25 spectra of each target star. As this work is partly focused on generalising a model capable of detecting Jupiters from spectroscopic observations, the use of several spectra of one source can be deemed acceptable in favour of generating a trainable dataset. Once the selection process was completed, a dataset of 5,417 instances of 434 stars was compiled and ready to be prepared for use. It was made sure that labels to the training data are confidently assigned. As the class distinction selection for this work is whether or not the host star in question has a Jovian-class companion, it was important that prior observations and analysis were taken into account.

The final step in preparing the spectral dataset for training was to normalize the data and process it into a model-readable format. It was ensured that the spectral range & resolution of each data instance is invariant to the facility used to collect them. Thus, both FEROS and HARPS spectra were truncated and spectrally binned such that all spectra comprised of 104,350 datapoints across the same spectral range (∼3781–6912Å). We then generate a fit for the spectrum's continuum to then use for normalization.

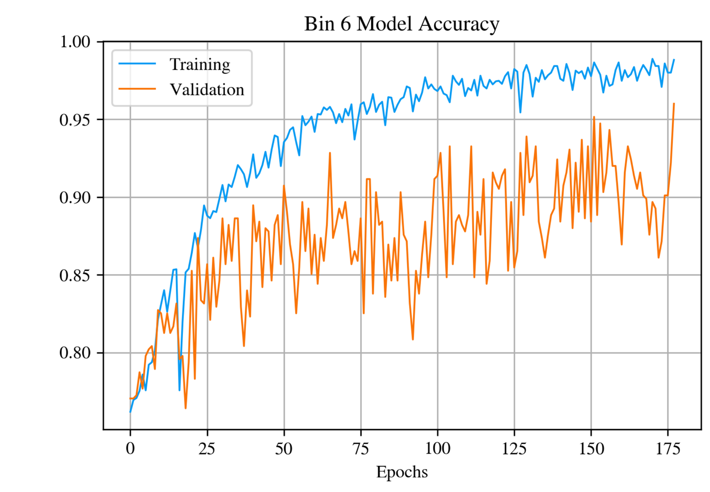


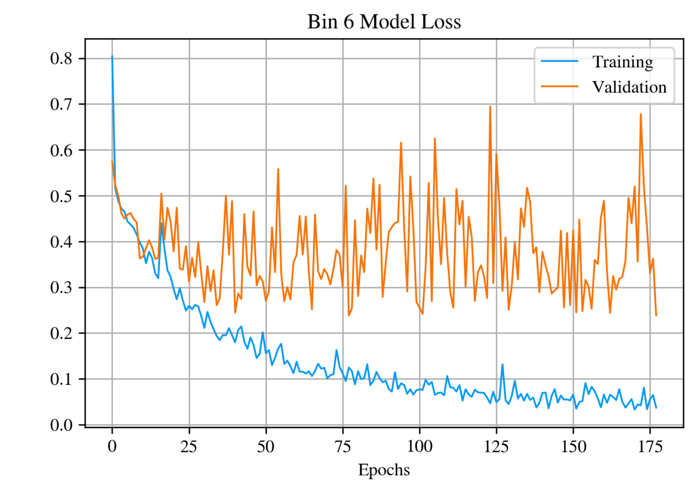
Spectral Classifier Design

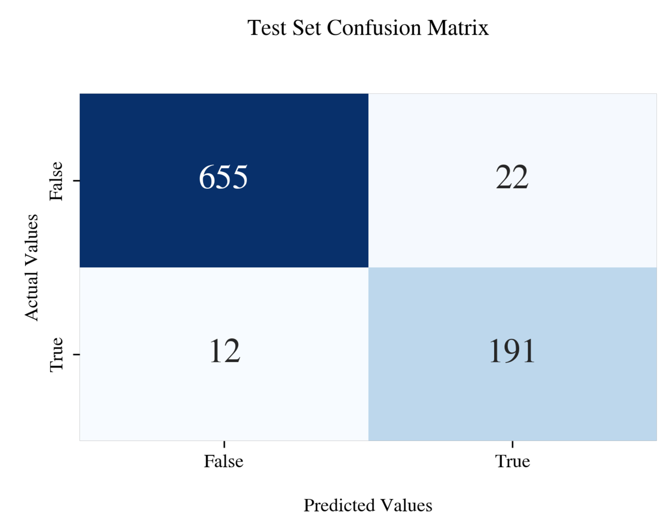
The spectral classifier follows a stacked ensemble classifier design, with the first-level voters all implementing a conservatively deep 1D-CNN design. Each spectrum was binned into 25 disjoint subsets, such that the data instances for each spectral bin were used to train their respective classifier and validate the model's hyperparameter settings for that particular spectral bin. Each spectral bin is fed to the respective first level CNN classifier, which in turn then assigns a classification probability score to be used as an input feature by the meta learner. The meta learner then aggregates these probability scores (from all first level classifier which demonstrate any form of generalisation) to predict a final classification for the data instance.



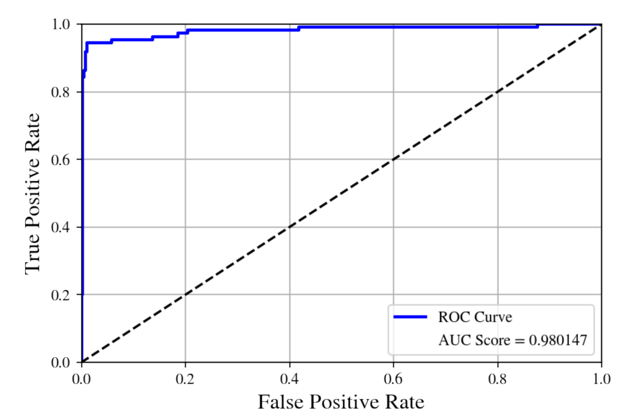
Results







Performance Metric	Validation Set (%)	Test Set (%)
Accuracy	97.05	96.14
Precision	92.79	89.67
Recall	94.50	94.09
F_1 score	93.64	91.83



Abundance Classifier

The next step will be to use the elemental abundance dataset to train a classifier capable of discriminating between Jovian hosts and comparison stars. As the dataset has several incomplete records, a multi-objective optimization model as proposed by [8] will be implemented. This will allow for online imputation of missing features within the dataset during model selection. Once the model is trained & tested, cross-examination with the spectral classifier's performance will be conducted.

Preliminary Conclusions

We applied a spectral dataset, compiled alongside a corresponding elemental abundance dataset, in order to train a classifier capable of discriminating between stars with Jupiter planets and those without. The spectral classifier shows strong generalisation in its' performance on the validation and test set, albeit without stable minimization of the loss function. The next step will be to observe how it compares to an abundance data approach and explore ways to improve the model architecture. We will then test the model on a holdout set from a third facility in order to check for telescope invariability.

References

[1] Xue-Ning Bai and James M Stone. Dynamics of solids in the midplane of protoplanetary disks: Implications for planetesimal formation. *The Astrophysical Journal*, 722(2):1437, 2010.

[2] Barbara Ercolano and CJ Clarke. Metallicity, planet formation and disc lifetimes. *Monthly Notices of the Royal Astronomical Society*, 402(4):2735–2743, 2010.

[3] Natalie R Hinkel, Cayman Unterborn, Stephen R Kane, Garrett Somers, and Richard Galvez. A recommendation algorithm to predict giant exoplanet host stars using stellar elemental abundances. *The Astrophysical Journal*, 880(1):49, 2019.

[4] Anders Johansen, Andrew Youdin, and Mordecai-Mark Mac Low. Particle clumping and planetesimal formation depend strongly on metallicity. *The Astrophysical Journal Letters*, 704(2):L75, 2009.

[5] Jarrett L Johnson and Hui Li. The first planets: The critical metallicity for planet formation. *The Astrophysical Journal*, 751(2):81, 2012.

[6] John Asher Johnson, Kimberly M Aller, Andrew W Howard, and Justin R Crepp. Giant planet occurrence in the stellar mass-metallicity plane. *Publications of the Astronomical Society of the Pacific*, 122(894):905, 2010.

[7] A Kaufer, O Stahl, S Tubbesing, P Nørregaard, G Avila, P Francois, L Pasquini, and A Pizzella. Commissioning feros, the new high-resolution spectrograph at la-silla. *The messenger*, 95:8–12, 1999.

[8] Hadi A Khorshidi, Michael Kirley, and Uwe Aickelin. Machine learning with incomplete datasets using multi-objective optimization models. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2020.

[9] M Mayor, F Pepe, D Queloz, F Bouchy, G Rupprecht, G Lo Curto, G Avila, W Benz, J-L Bertaux, X Bonfils, et al. Setting new standards with harps. *The Messenger*, 114:20–24, 2003.

[10] Michael Perryman, Joel Hartman, Gáspár Á Bakos, and Lennart Lindegren. Astrometric exoplanet detection with gaia. *The Astrophysical Journal*, 797(1):14, 2014.

L-Università ta' Malta
Institute of Space Sciences & Astronomy