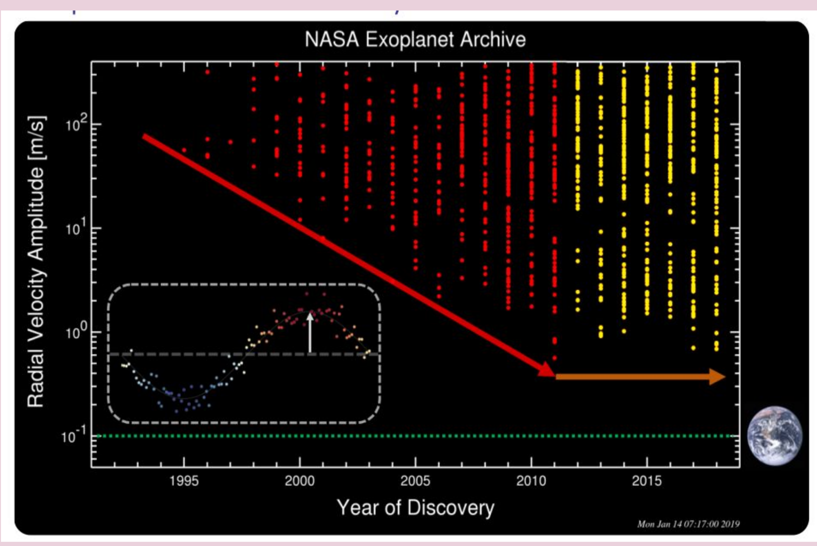
Radial Velocity Planet Detection with Vision Transformers

Anoop Gavankar[1], Tanish Mittal[2], Joe Ninan[1], Shravan Hanasoge[1]
[1]Tata Institute of Fundamental Research, Mumbai, [2]Birla Institute of Technology and Science, Pilani

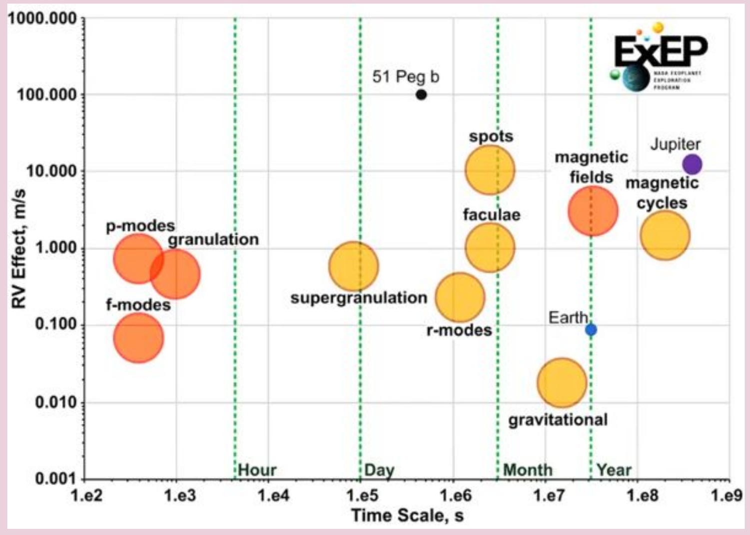
Introduction

Radial Velocity Method for Exoplanet Detection:

- Measuring Periodic Doppler shift in the parent star's spectrum.
- Current precision **constrained** by Stellar jitter of parent star
- Stellar Jitter: main source of RV noise below **1ms^{-1}**
- Stellar jitter characterization and removal : key to measure "Extreme Precision Radial Velocities" (EPRVs) accurately.
- Traditional methods like FF' and GPs : **empirical** in nature.
- Machine learning : can potentially utilize all spectral data.



Radial Velocity method : Detection threshold evolution*



Sources of Stellar Jitter**

*NASA Exoplanet Archive

**NASA EPRV Working Group Report

OBJECTIVES

- To **disentangle** Keplerian planetary RV signal from solar jitter, for NEID solar data, using Machine Learning
- To **extract** synthetic Keplerian RV signal with semi-amplitude < 1ms^{-1}
- To **apply** this technique for extraction of Keplerian orbital parameters like period, amplitude, eccentricity etc.
- To **extrapolate** this technique for application on stellar spectra.

DATA

NEID spectrograph solar feed data:

- High precision RV measurements : precision well below 1ms^{-1}
- 380-930 nm high resolution(~117000) spectral data(Dec '20 - Jun '22).
- A Cross Correlation Function(CCF) generation pipeline converts spectral data to a CCF : weighted sum of spectral lines.
- A single spectrum is converted to **10 CCFs** based on line depth of averaged spectral lines: appended with activity sensitive lines to form a **CCCF vector**.
- Since line depth **approximates** the photospheric depth of spectral line formation, it helps **prevent** significant averaging of photospheric fluid flows.
- An Image generation pipeline converts multiple such randomly selected CCCF vectors into an **image**: time dependent Keplerian Doppler shift inserted in every CCF vector.

NEID DATA PROCESSING AND MACHINE LEARNING ARCHITECTURE

"Clear sky" spectra → Normalization → Spectral line selection and segregation → CCF vector generation → Appending activity lines

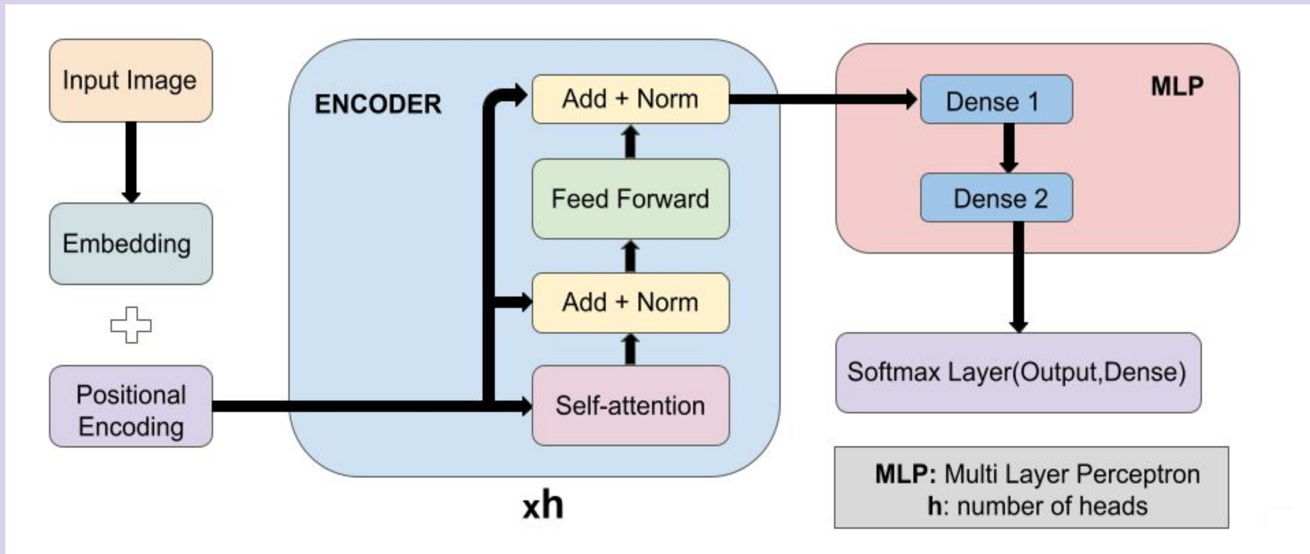
CCCF generation pipeline

100 CCF vectors: random selection → Stacking CCFs: "100x1120" image → Row wise timestamp assignment → Keplerian signal injection

Image generation pipeline

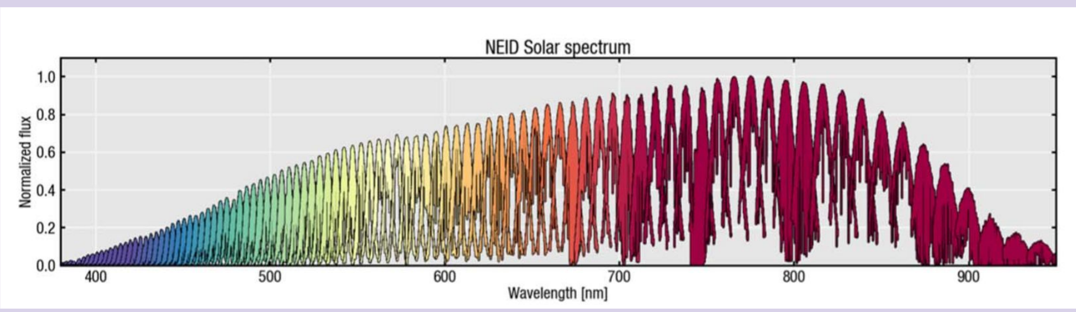
The Keplerian signals are sampled for one-planet-Sun systems, with period P, and semi-amplitude K:

- Input log(P) and K values : mapped to 10 and 5 uniform bins respectively.
- Sampling of P: **log-uniform** distribution with 12-365 day range
- Sampling of K : **uniform** distribution with 0.05-3 ms^{-1} range
- Output : probability arrays of length 10 and 5 for P and K respectively.
- Standard ML models : **not designed to handle** aperiodically sampled data.
- Vision Transformers : allow the variable observation times of our data to be **encoded**.

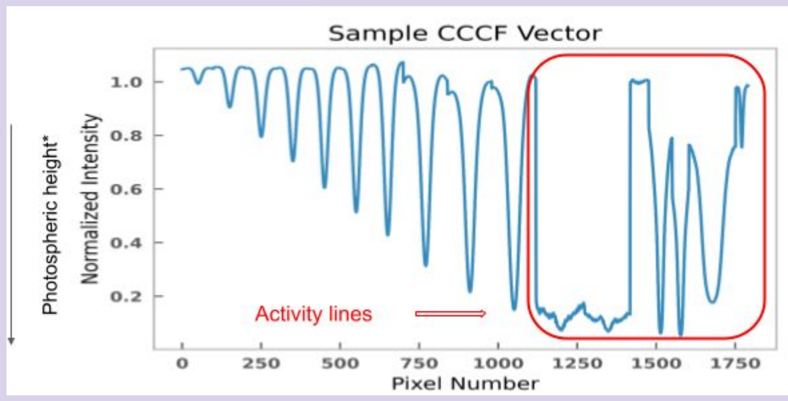


ViT Architecture

Raw NEID data processed to generate images with Keplerian signal



Raw data[1]



CCF vector

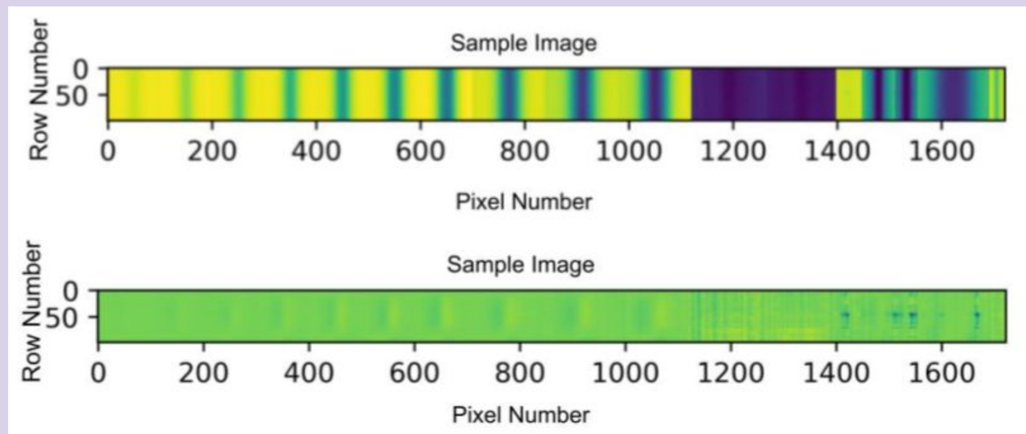


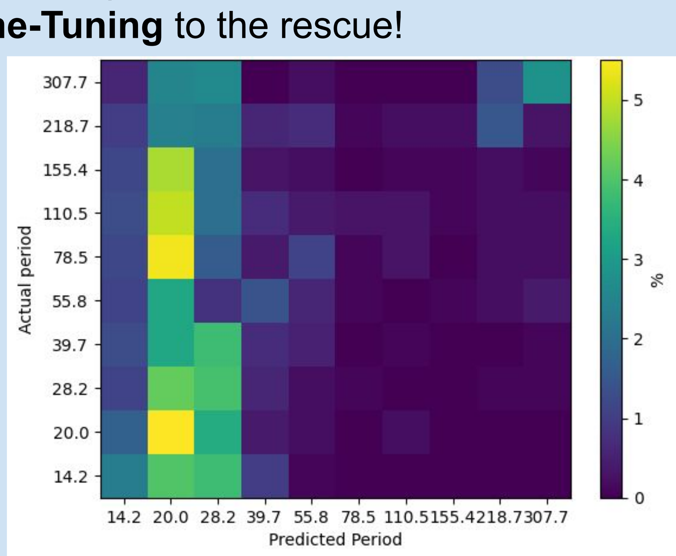
Image with Keplerian signal

SUMMARY OF RESULTS

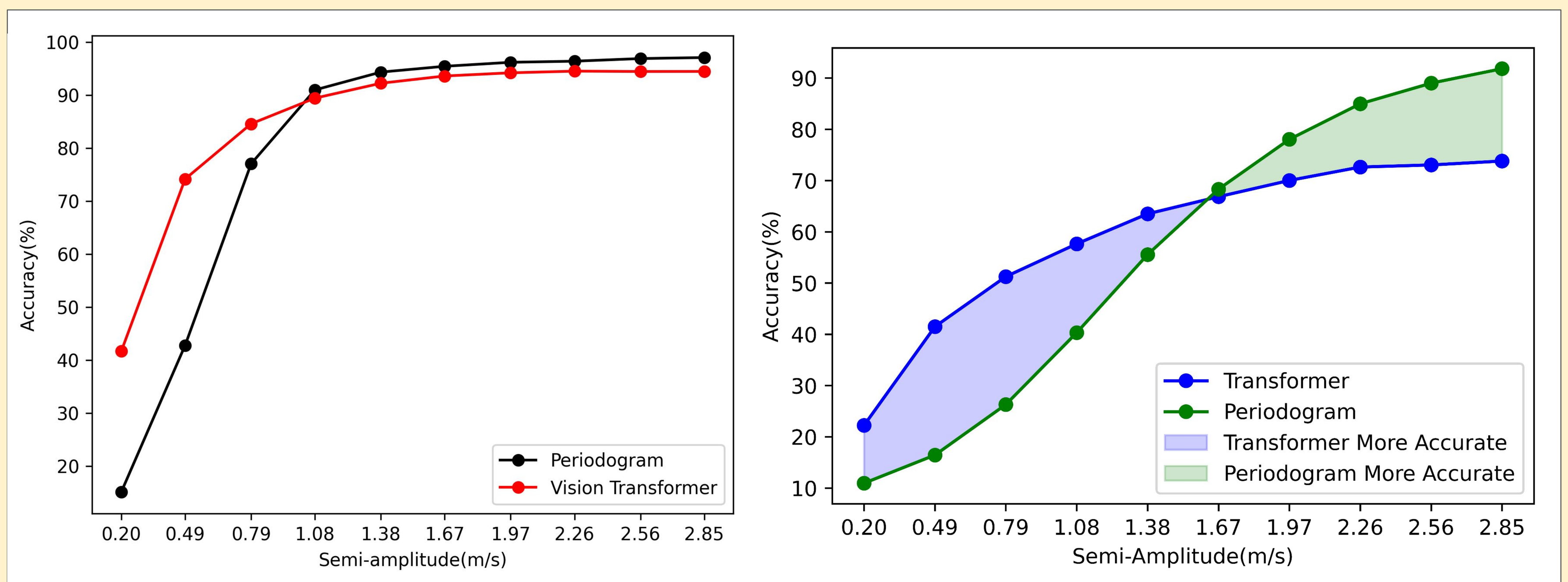
- **Vision Transformer :** Predicts period **P** and semi-amplitude **K** as orbital parameters
- Shuffled data: P accuracy : **86%,** K accuracy : **76%** for their respective 10 and 5 class classifications.
- Ordered Data: Solar rotation **interference** in Period Prediction
- **Fine-tuning**: Decorrelates Solar rotation from period prediction for ordered data.
- **Performance Trend**: Accuracy in prediction of P of Keplerian signal increases with increasing semi-amplitude K
- **Comparison**: Direct comparison hints at model outperformance over Periodogram (K ~< 1.7ms^{-1})

Rotational Interference!

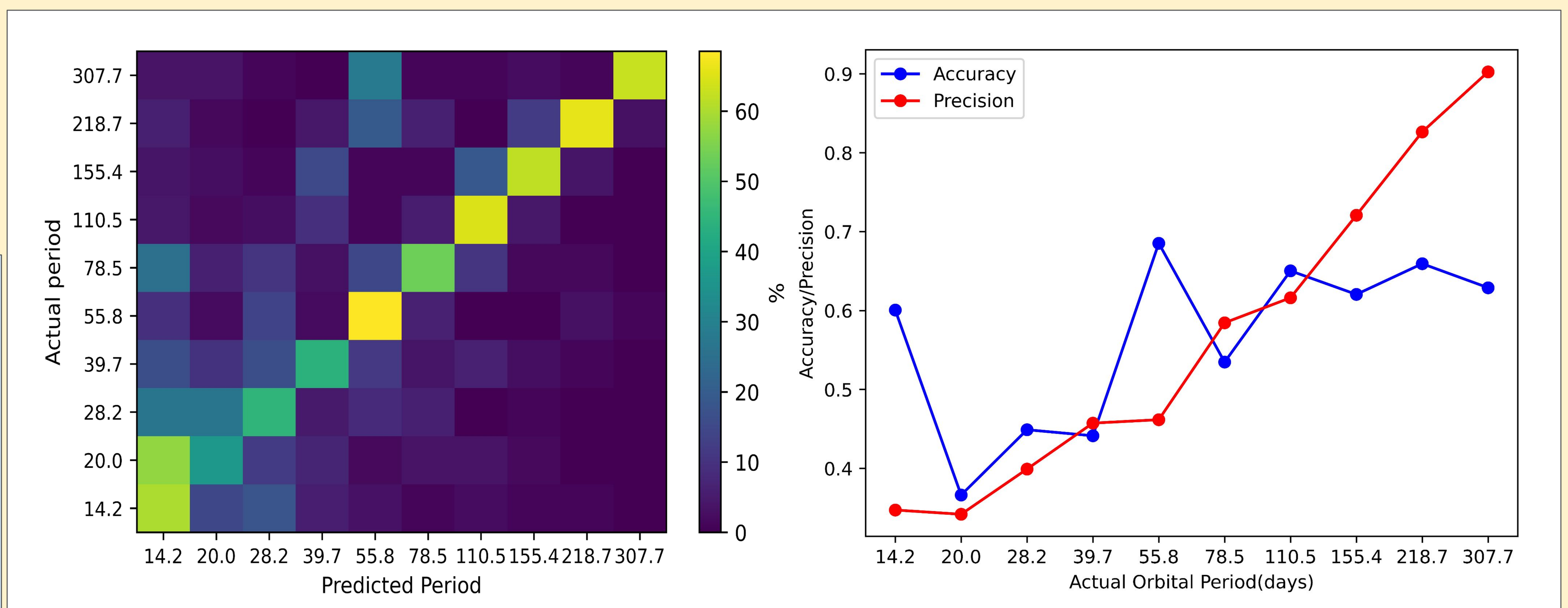
- Solar rotation predicted as period for time-ordered data samples!
- **Fine-Tuning** to the rescue!



RESULTS AND CONCLUSIONS



Period Prediction accuracies for shuffled and time-ordered data



Period Prediction matrix, and accuracy variation for time-ordered data

Fine-Tuning decorrelates solar rotation and enables the machine to give accurate period predictions for time-ordered data!! Accuracy shows a dip near the Solar rotation period!!!

CONCLUSIONS

- **ViT:** Successfully disentangles RV signal from solar jitter, and accurately predicts period P and semi-amplitude K
- Direct comparison: ML model **significantly outperforms** periodogram P predictions in low K regime
- Serves as an **effective isolator** for the range of interest in planetary period values.

FUTURE SCOPE

- Expansion of current dataset from 19 months to add more recent data.
- Train a model to **distinguish** genuine planetary shift from RV activity masquerading as planetary signal.
- Explore better ways to fine-tune models.
- Explore generative AI to generate solar spectra effectively, to help decorrelate activity affected RV from Keplerian RV.
- Potential model testing on G-type stars.
- Explore telluric line contamination correction to expand available data into the infrared regime and for cooler stars.
- Apply this technique on other stars.

Direct comparison with Periodogram showcases ~2x outperformance at low semi-amplitude values!!

REFERENCES:

1. Extreme Precision Radial Velocity Working Group Final report, July 2021
2. Planetary detection limits taking into account stellar noise, *Dumusque et al., 2011*
3. Observing the Sun as a Star: Design and Early Results from the NEID Solar Feed, *Lin et al., 2022*
4. Attention is all you need, Vaswani et al., 2017
5. Measuring precise radial velocities on individual spectral lines : II. *Cretignier et al., 2019*

For Further Information:
Feel free to approach and strike up a conversation with me when you see me around.
You can also contact me at : anoop.gavankar@tifr.res.in

ACKNOWLEDGEMENTS:

1. Murty foundation, for funding our conference visits and important expenses.
2. Professor Eric Ford, and Professor Suvrath Mahadevan, Penn State, for their valuable feedback on applying AI for exoplanet research.