Pushing the boundaries of Planet Detection in the RV Method

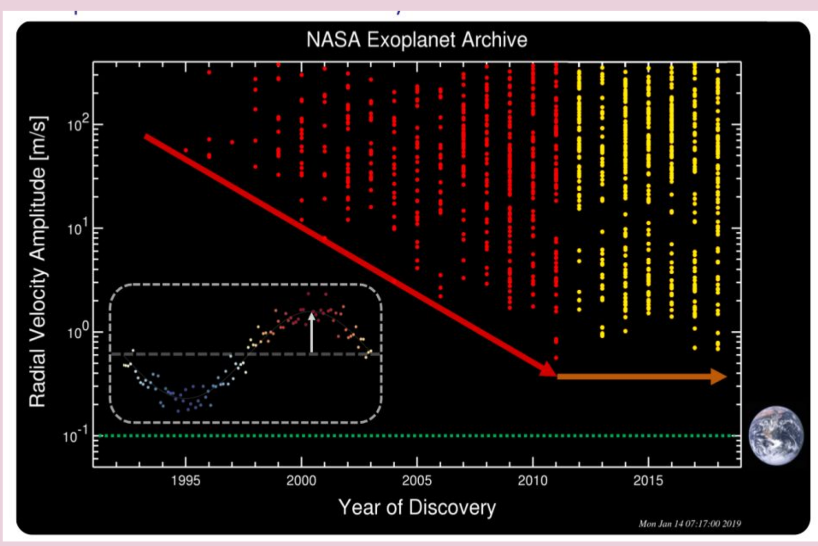
Anoop Gavankar[1], Tanish Mittal[2], Joe Ninan[1], Shravan Hanasoge[1], Dhanpal Siddharth[1]

[1]*Tata Institute of Fundamental Research Mumbai,*[2]*Birla Institute of Technology and Science Pilani*

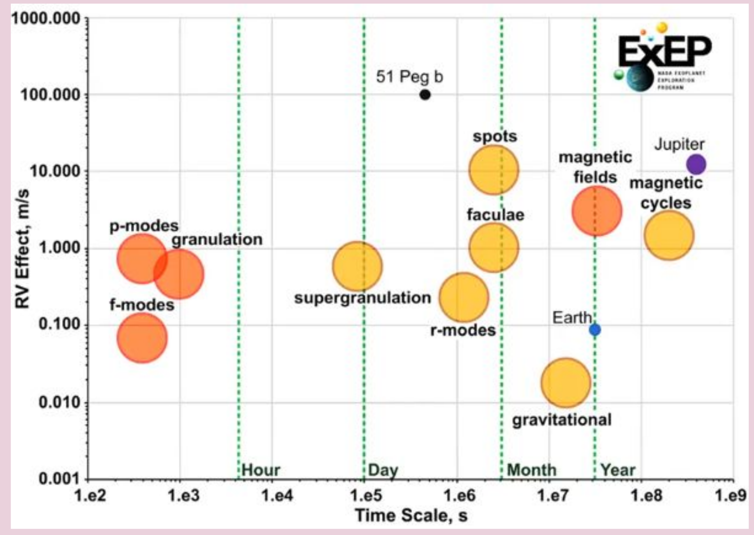
Introduction

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[1]



Sources of Stellar Jitter[1]

OBJECTIVES

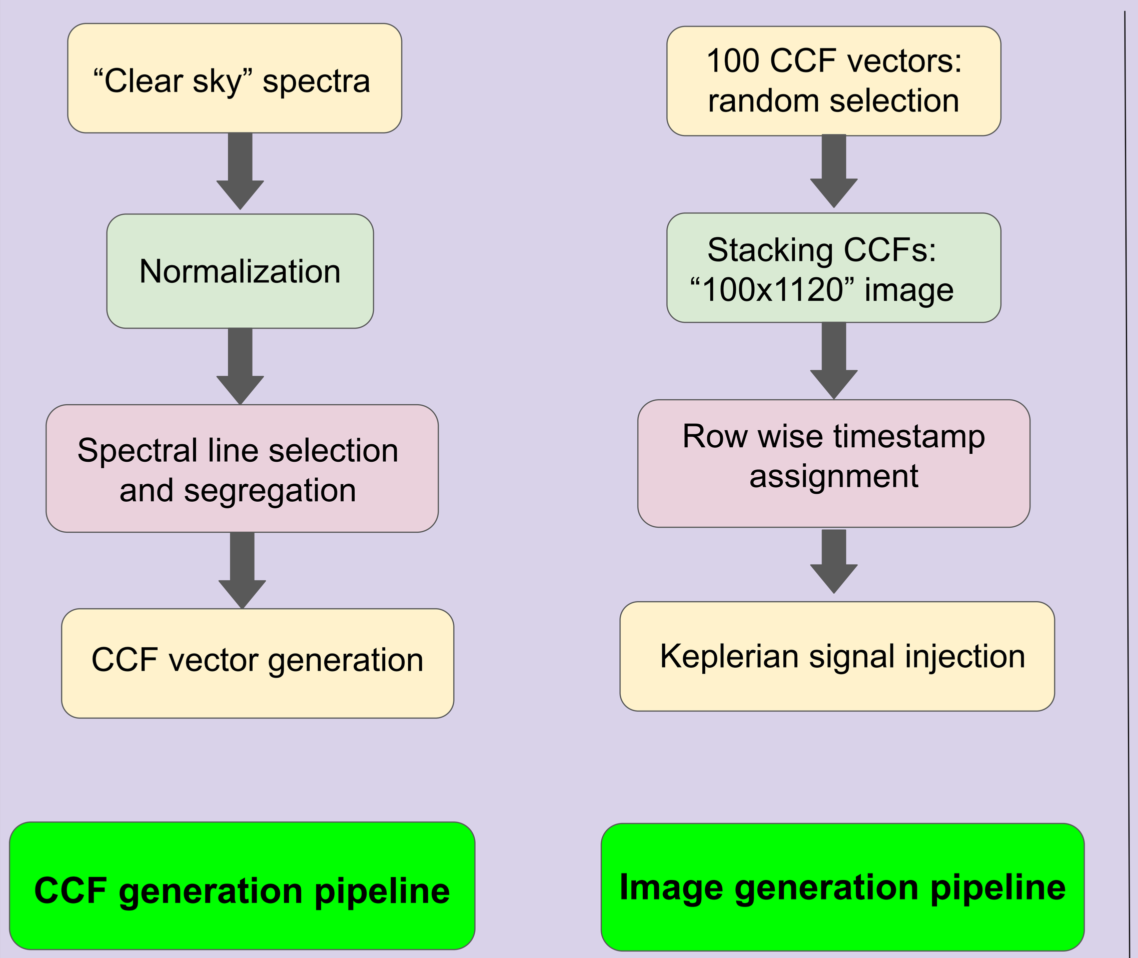
- To disentangle Keplerian planetary RV signal from solar jitter, for NEID solar data, using Machine Learning
- To extract synthetic Keplerian RV signal with semi-amplitude < 1ms^{-1}
- To apply this technique for extraction of Keplerian orbital parameters like period, amplitude, eccentricity etc.
-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

NEID DATA PROCESSING AND MACHINE LEARNING ARCHITECTURE

DATA

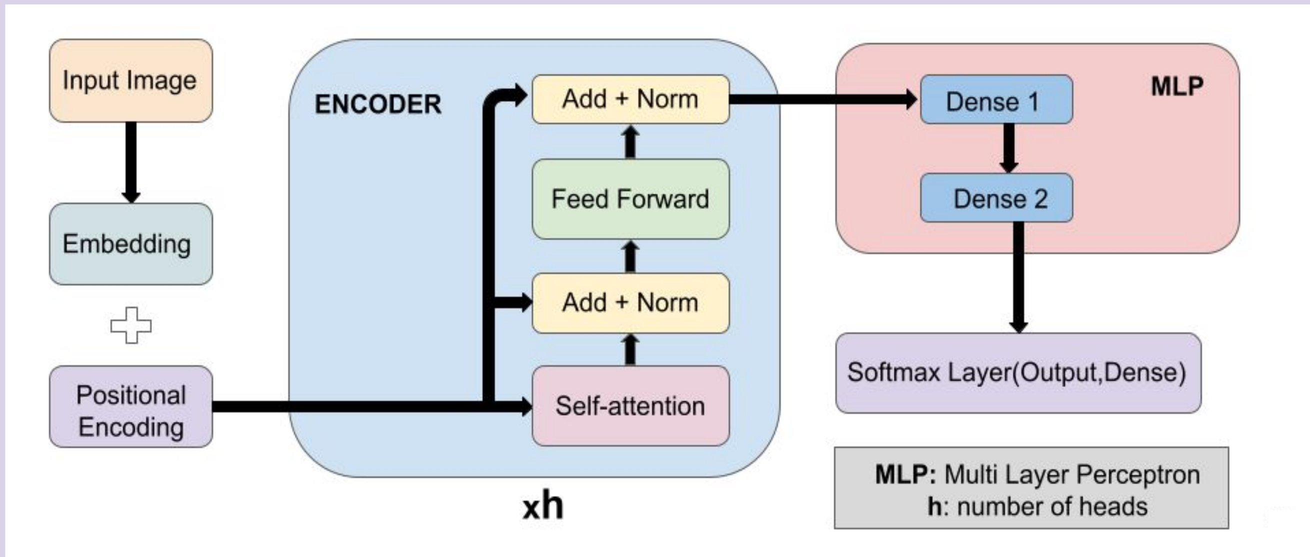
NEID spectrograph solar feed data:

- 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 to form a **CCF vector**.
- An Image generation pipeline converts multiple such randomly selected CCF vectors into an **image**: time dependent Keplerian Doppler shift inserted in every CCF vector.
- These images: Used for all further analysis.



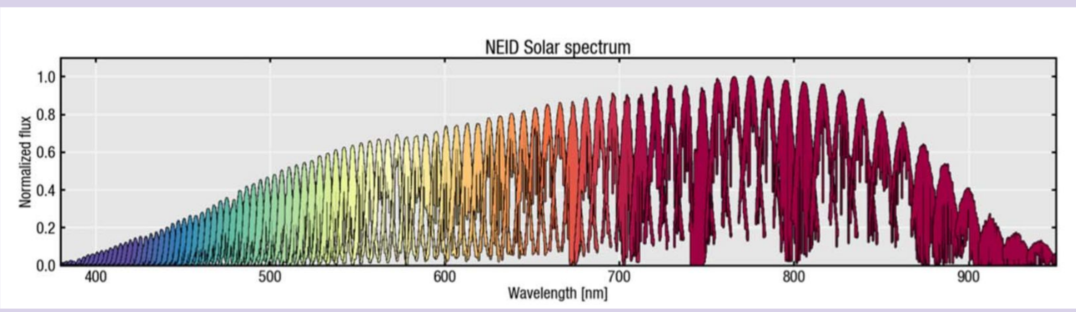
The Keplerian signals are sampled for one-planet-Sun systems, with period P, and semi-amplitude K:

- Input log(P) and K values : mapped to 10 and 5 uniform bins respectively.
- P sampling : **log-uniform** distribution with 12-365 day range
- K sampling : **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**.

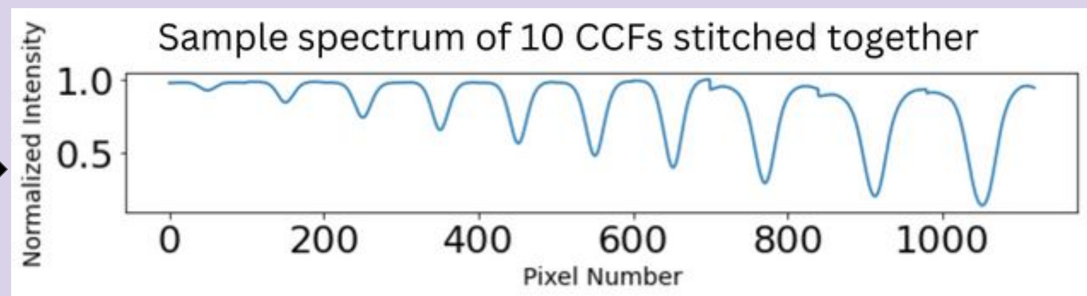


ViT Architecture

Raw NEID data processed to generate images with Keplerian signal



Raw data[1]



CCF vector

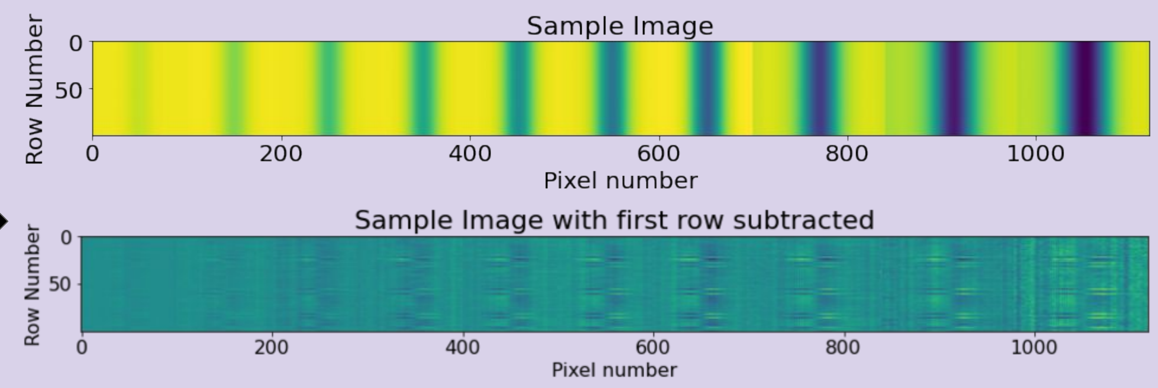


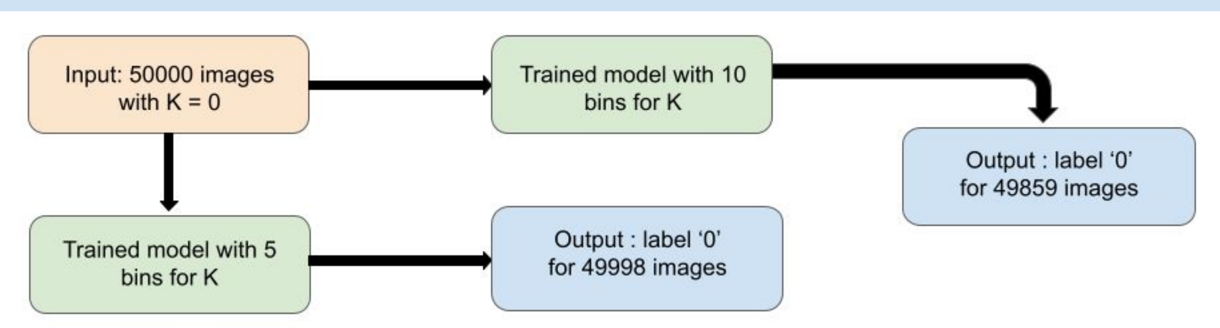
Image with Keplerian signal

SUMMARY OF RESULTS

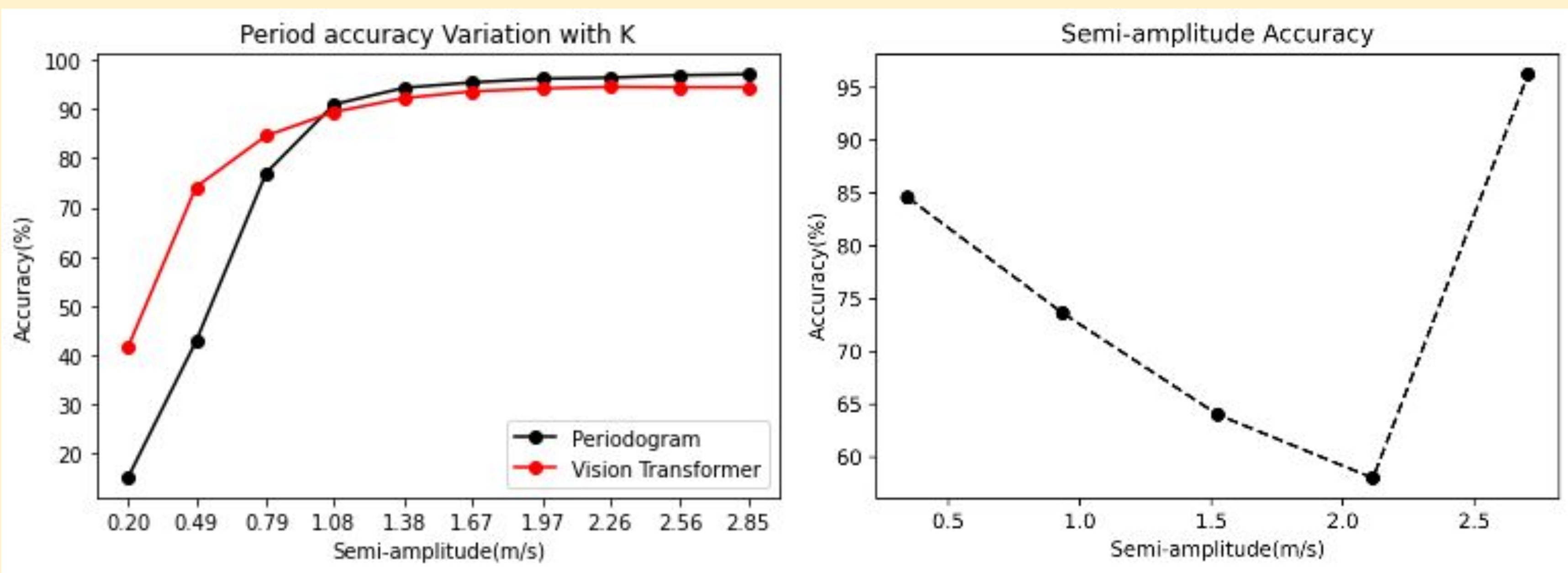
- Predicted period **P** and semi-amplitude **K** as Keplerian orbital parameters
- **Vision Transformer** : implemented to deal with aperiodicity in observations
- P accuracy : **83**%, K accuracy : **73**% for their respective 10 and 5 class classifications.
- **Performance Trend**: P prediction accuracy of Keplerian signal increases with increasing semi-amplitude K
- High confidence P predictions have a high accuracy(~94%) for K ϵ 64-94 cms^{-1}
- **Comparison**: Crude comparison hints at model outperformance over Periodogram(K < 1ms^{-1})

Testing the model

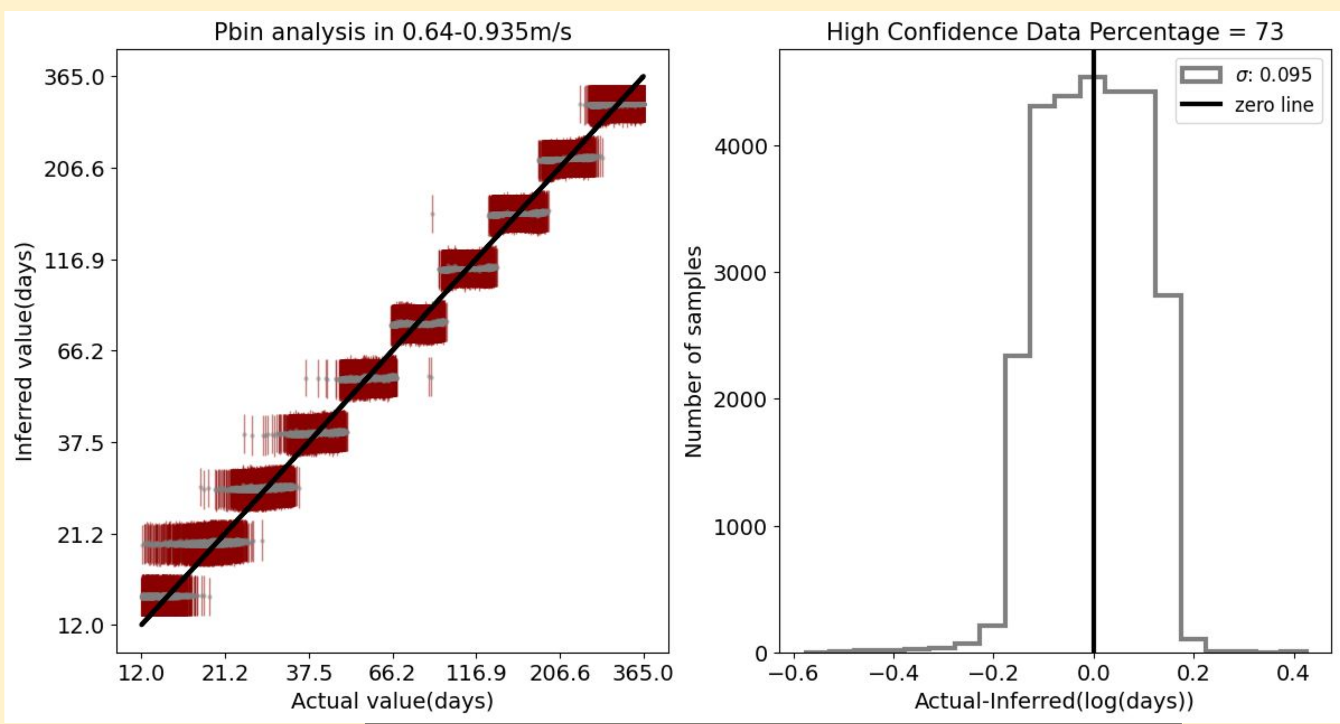
- K predictions for images with no Keplerian signal(K=0).
- The overwhelming majority of images were mapped to the *lowest possible K classification value*.
- Conclusion: **Machine is indeed disentangling Keplerian signal from solar RV noise!!!**



RESULTS AND CONCLUSIONS



Prediction accuracies for different models



High confidence P predictions for selected K range

Model correctly predicts orbital periods with high confidence(>90%) for a significant fraction(73%) of selected data with K ϵ 64-94 cms^{-1}!!!

CONCLUSIONS

- **ViT:** Successfully disentangles RV signal from solar jitter, and accurately predicts period P and semi-amplitude K
- Crude comparison: ML model **significantly outperforms** periodogram P predictions in low K regime
- Using multiple CCFs enhances P prediction accuracy by ~5%

FUTURE SCOPE

- Expansion of current dataset from 19 months to add more recent data.
- Adding data that **includes correlation**, like the quasi-periodic jitter associated with solar rotation.
- Independently representing spectral lines known to trace stellar activity in the CCF vector.
- Train a model to **distinguish** genuine planetary shift from RV activity masquerading as planetary signal.
- Potential model testing on G-type stars.
- Apply this technique on other stars.

Rudimentary comparison with Periodogram hints at possible 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. Sudha Murthy foundation, for funding our conference visits and important expenses.
2. Professor Eric Ford, for his valuable feedback on applying AI for exoplanet research.