Draft version June 2, 2022
Typeset using LaTeX default style in AASTeX631

Measuring Stellar Properties from Light Curves Based on Deep Learning Music Recognition Algorithms

Yihong Song[1]

[1] *Academies of Loudoun*
42075 Loudoun Academy Dr
Leesburg, VA 20175, USA

Abstract

Being the basis of mathematical models in galactic evolution and exoplanet habitability, stellar properties like surface gravity, effective temperature, and metallicity are essential elements in astrophysics research. Due to physical limitations in spectroscopy, the conventional tool in star property measurement, there is currently no viable method to detect properties for over 14,000 stars observed by the Kepler Space Telescope. This project presents a novel technique to measure stellar properties using light curves through a deep learning algorithm inspired by a music recognition software. Similar to how songs, or sound waves, are distinguished from their unique frequency-power spectra, this model distinguishes stars based on their light curves, or brightness-waves. Long-cadence light curves from red giants with known stellar properties were randomly sampled to create a training set of 210,000 samples. The light curves were converted into power spectrums through Fourier transform and used to train a deep neural network. This model reached accuracies of 99.7% for surface gravity, 76.9% for effective temperature, and 92.7% for metallicity. Even more, this robust model has a high noise tolerance and reduces the traditionally required observation time from 90 days to 9 days. This model provides an efficient, innovative approach to measuring the properties of distant stars, enabling astrophysicists to explore further into the history and evolution of our universe.

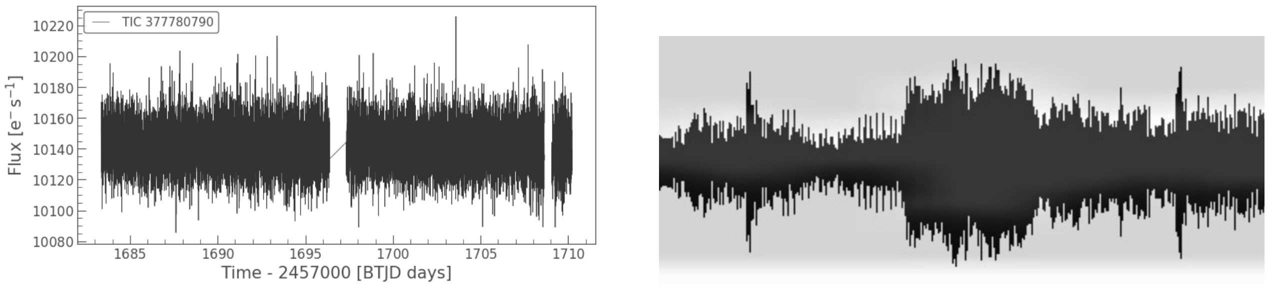
Keywords: asteroseismology – methods: deep learning – techniques: fourier transform – stars: light curves – stars: properties

1. INTRODUCTION

Since the 1600s, humans have been attempting to learn about stars from their stellar properties, specifically surface gravity (correlated to mass and radius), effective temperature, and metallicity. Spectroscopy is currently the most effective method in finding the stellar properties of relatively bright stars with uncertainties of 25% (Valenti & Fischer 2005; Ghezzi et al. 2010). The more distant the star is, the dimmer the star brightness, and the less accurate spectroscopy becomes. From 2009 to 2013, the Kepler Space Telescope measured light curves, or brightness variations over time, of distant stars that are 600 to 3,000 light years from earth. Out of the 200,000 stars observed, 14,000 of them cannot yet be measured by spectroscopy due to their far distance from earth (Mathur et al. 2017). Due to the significant data quality limitation in current spectroscopy measurements, a novel technique to measure the stellar properties of distant stars must be found.

Recent attempts in using machine learning to extract information from star light curves were completed in 2018 by Hinners et al. (2018), who used real light curve data in conjunction with simulated data to train their deep learning algorithms. Their work using representation learning and a Long Short-term Memory Recurrent Neural Network (LSTM-RNN) produced "no successful predictions," with stellar property predictions of about 75% accuracy. However, their research provides a foundation for future deep learning techniques in analyzing light curve data (Hinners et al. 2018).

In a 2021 survey on machine learning-based light curve analysis, Yu et al. (2021) highlighted the inevitability of the use of artificial intelligence in fully analyzing the observational data. Yu further noted that a new data processing technique is required to supplement deep learning in making developments in light curve analysis (Yu et al. 2021).



(a) Graph of a typical light curve, with time on the x-axis and brightness on the y-axis.

(b) Graph of a typical music waveform, with time on the x-axis and amplitude/volume on the y-axis.

Figure 1: Side by side comparison of two types of waveforms.

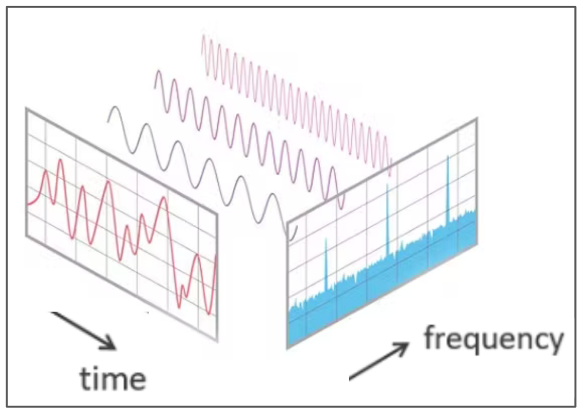


Figure 2: Fourier Transform Visualization.
Creative Commons, Image: https://bit.ly/3Cjcxq6.

Shazam currently operates the industry-leading music recognition algorithm (Shazam Entertainment, Ltd. & Wang 2003) on its iPhone and Android app. Starting anywhere within the song (Figure 1), the algorithm can process the live music and predict the song name within 5-10 seconds. In a report by Dr. Avery Wang, the co-founder of Shazam, the effectiveness of a Fourier Transform in eliminating noise and interference in signal processing was greatly highlighted for being able to "correctly identify music in the presence of voices, traffic noise . . . and even other music." (Shazam Entertainment, Ltd. & Wang 2003)

The Fourier Series was discovered by Joseph Fourier in the 1800s. It is a summation of sinusoidal functions that can approximate any arbitrary function. In Fourier Transformation, we do the reverse: splitting the regularly-sampled function into sinusoidal components and transforming the original function from a time domain to a frequency domain.

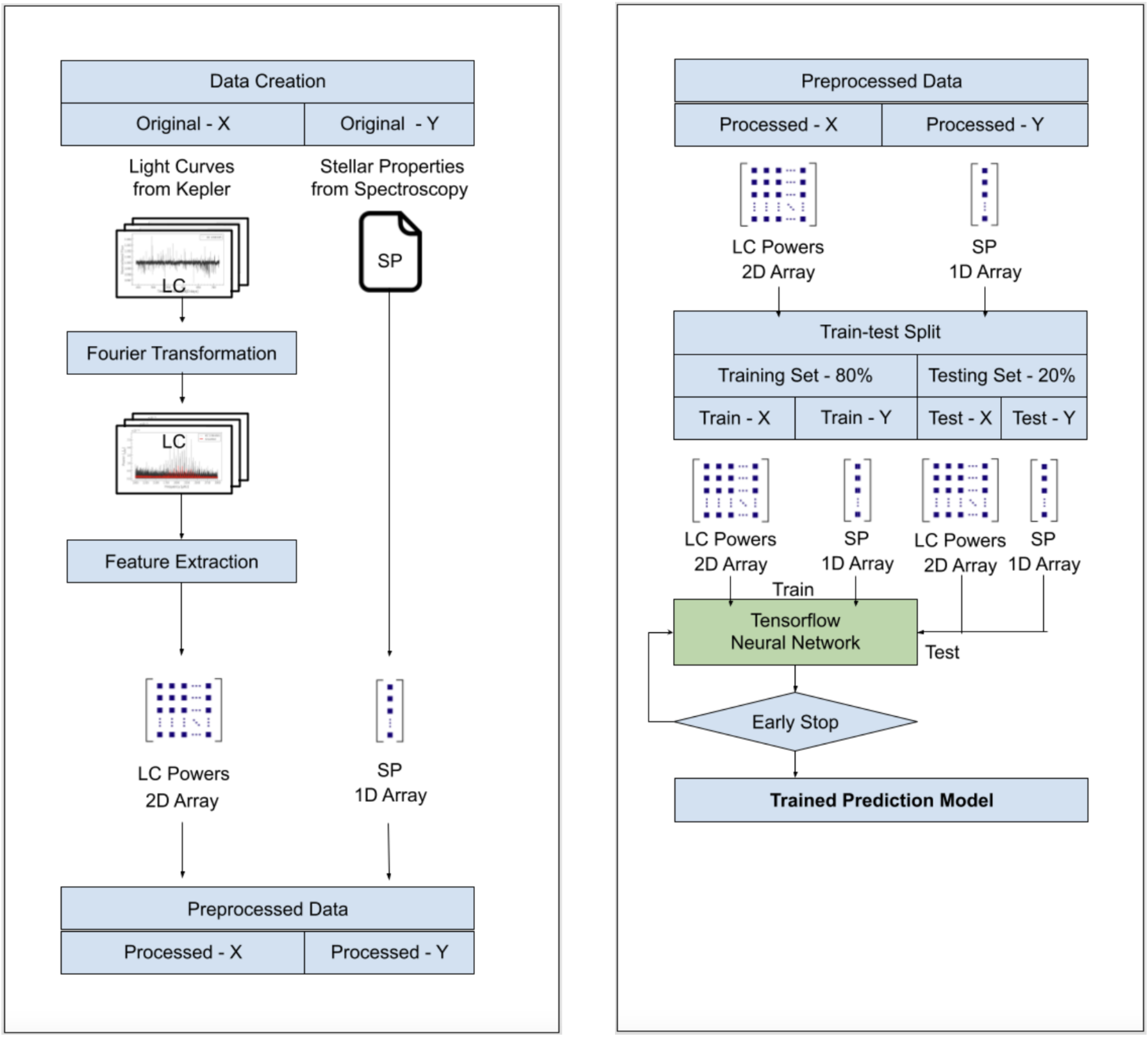
In this research, we apply the Fourier Transformation onto light curves and hypothesize that the deep learning algorithm may distinguish the brightness fluctuation of the star from background noise and extract stellar property information from the set of unique frequency-power values from each star, similar to how Shazam determines the song title in a noisy room.

Comparing to related researches, this study is one of the first in the field to completely utilize real, non-generated light curve data to train deep learning algorithms. Since low signal-to-noise ratio is the primary cause of imprecision in distant star measurements, the Fourier Transform (Figure 2) was added to data preprocessing to enhance the performance of the deep learning neural network and to significantly increase model prediction accuracy and noise tolerance.

2. METHODS

Here, we describe the preparation of the star light curves through Fourier Transformation and the construction of a deep learning algorithm to predict the star properties (Figure 3).

2.1. *Data Collection*



(a) Part 1-2: Data Collection & Preprocessing (b) Part 3: Machine Learning

Figure 3: Procedural diagrams.

Long-cadence light curves of 16094 red giants previously classified by Yu et al. (2018) were downloaded from KIC through the Lightkurve library built by Lightkurve Collaboration, 2018 and matched with their respective stellar properties of Teff, Log g, and metallicity provided by Yu et al. (2018).

2.2. *Data Preprocessing*

Light curves were normalized and cleaned with outlier removal from the Lightkurve library. They were then truncated to segments of random duration and time within observation, with a minimum of 240 data points within each segment. In this random sampling process, no datapoint was overlapped or sampled more than once.

Fourier transformations with frequencies from 1 to 24.5 cycles per day with cadence of 0.1 cycles per day were performed on each of the segments to turn time series into periodograms with frequency-power axis. The magnitudes of each frequency are stored as features for each red giant and re-paired with its three stellar properties as inputs into deep learning neural networks.

2.3. *Deep Learning Neural Network*

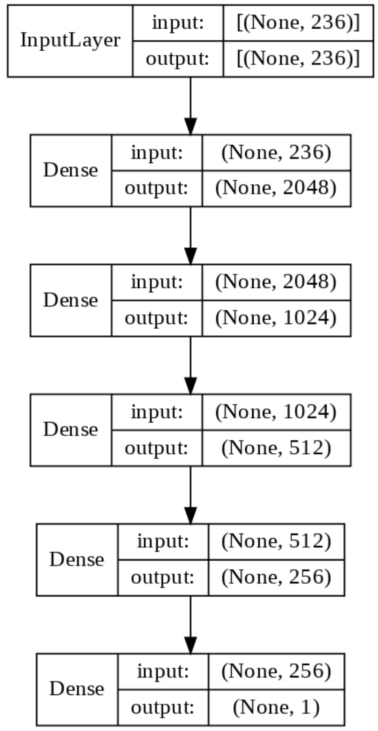


Figure 4: Neural network architecture.

Figure 5: Effect of Early Stopping Algorithm on model training.

Input dataset were split into training and testing sets. The training set takes part in model learning, while the testing set assesses model performance. 80% training : 20% testing proportion is used in this research.

A deep neural network consisting of 5 "Dense" layers and "Relu" activations with cascading nodes from 2048 to 1 per layer, as described in Figure 4.

The target of neural network training is to minimize the "loss," or the difference between the expected outcome and the predicted value. The Early Stopping Algorithm stops training before the model overfits on the training data and uses the epoch with the lowest testing loss, successfully minimizing the overall model overfitting issue (Figure 5).

3. DATA ANALYSIS

3.1. *Accuracy*

The Kepler Space Telescope's Input Catalog records stellar properties with accuracy values determined by the mechanics of the spectroscopy model (Table 1). To evaluate our model against the spectroscopy results, we use

Property	**Tolerable Error (Unit)**
Surface Gravity (Mass & Radius)	± 200 (K)
Effective Temperature	± 0.5 (cm/s2)
Metallicity	± 0.5 (solar)

Table 1: Spectroscopy model accuracy in stellar properties, Kepler Input Catalog.

$$\mathrm{RMSD} = \sqrt{\frac{\sum_{i=1}^{N} \left(x_i - \hat{x}_i\right)^2}{N}}$$

Figure 6: Root Mean Square Error Function, where RMSD is the root mean square error, N is the number of non-missing data points, xi is the predicted value, and xi is the theoretical value

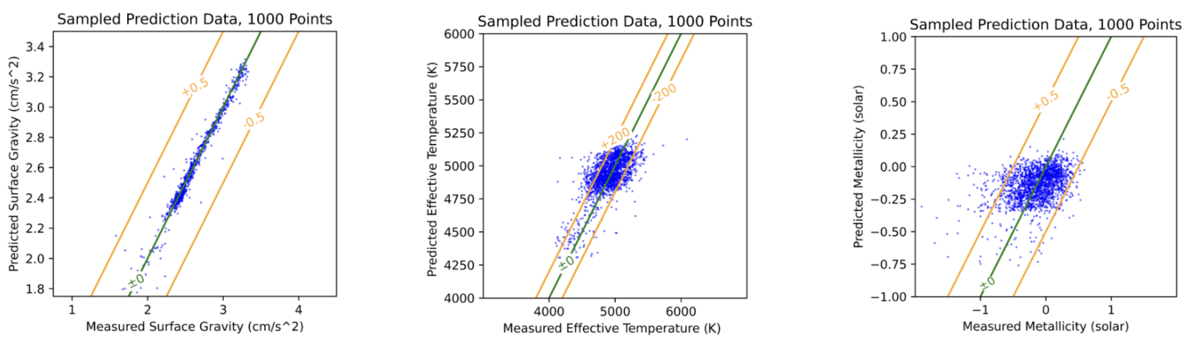


Figure 7: My Model's Predictions vs. Traditional Spectroscopy Measurement: Surface Gravity (left), Effective Temperature (middle), and Metallicity (right). Two orange lines envelop the tolerable error from spectroscopy measurements. The green line represents a perfect fit.

Property	**Surface Gravity (Mass & Radius)**	**Effective Temperature**	**Metallicity**
Accuracy	99.7%	76.9%	92.7%
RMSE	1.3%	4.7%	3.5%

Table 2: Accuracy and root mean square error of predicted stellar properties.

Kepler's claimed accuracy as the tolerable error. Any predictions lying within the tolerable error range (around the spectroscopy value) is regarded as accurate.

3.2. *Root Mean Square Error (RMSD)*

Root Mean Square Error (RMSD) measures the square root of the average of squared errors and is a standard method for prediction model evaluation. For the comparability between stellar properties, we report the RMSD in the percent error form.

4. RESULTS

This model had an outstanding accuracy in measuring surface gravity, mass, and radius, with an accuracy as high as 99.7% and the root mean squared error of only 1.3% (Table 2). It showed great promise in measuring star effective temperature and metallicity, with accuracies over 75% and 90%, respectively, while having both RMSE's below 5 percent (Table 2). The model's high noise tolerance reduced the required observation time from the traditional 90 days to 9 days (Figure 7 & Table 2).

5. CONCLUSIONS

The results support the hypothesis and provide substantial evidence for the successful creation of a novel technique measuring stellar properties from star light curves. The Fourier Transform method inspired by Shazam was found to sufficiently circumvent this problem through transforming the data from a time domain to a frequency domain.

Compared with previous deep learning techniques by Hinners et al. (2018), the addition of Fourier Transform significantly improves stellar property extraction accuracy and model robustness, increasing prediction accuracies from 75% to as high as 99.7%.

5.1. *Implications*

1. Using this method, measurement of 14,000 missing star properties from Kepler is now possible.

2. The model's tolerance for low signal-to-noise ratio significantly increases effective observation distance and decreases observation time by tenfold in stellar property measurements.

3. Success of this project shows a promising future for the use of machine learning with astrophysics.

4. Success of this project provides an example of the The Fourier Transformation being used in a field where data collection and measurement accuracy is greatly limited.

5.2. *Other Applications in Astrophysics*

Star Classification - Star type classification based on predicted stellar properties.
Outlier Detection - Stars which depict abnormal stellar properties prompt further scientific investigation.
Exoplanet Habitability - This research enables scientists to more accurately gauge the habitability of exoplanets for a bigger collection of stars.

Sources of error in this research may include star rotation, which may interfere with the sampling process when sampling cadence (30 minutes) is higher than rotation period. Future work on algorithms accounting for star rotation can effectively bypass this issue. Increased sampling rates in future photometry telescopes could also provide a more accurate measurement.

6. ACKNOWLEDGEMENTS

The author thanks D. Writer for research guidance, M. Hon for support in machine learning, T. Appourchaux for asteroseismology background information, and A. Wang and the Shazam team for the project inspiration.

REFERENCES

Ghezzi, L., Cunha, K., Smith, V. V., et al. 2010, The Astrophysical Journal, 720, 1290, doi: 10.1088/0004-637x/720/2/1290

Hinners, T. A., Tat, K., & Thorp, R. 2018, The Astronomical Journal, 156, 7, doi: 10.3847/1538-3881/aac16d

Mathur, S., Huber, D., Batalha, N. M., et al. 2017, The Astrophysical Journal Supplement Series, 229, 30, doi: 10.3847/1538-4365/229/2/30

Shazam Entertainment, Ltd., & Wang, A. 2003, An Industrial-Strength Audio Search Algorithm , Tech. rep. https://www.ee.columbia.edu/~dpwe/papers/Wang03-shazam.pdf

Valenti, J. A., & Fischer, D. A. 2005, The Astrophysical Journal Supplement Series, 159, 141, doi: 10.1086/430500

Yu, C., Li, K., Zhang, Y., et al. 2021, WIREs Data Mining and Knowledge Discovery, 11, doi: 10.1002/widm.1425

Yu, J., Huber, D., Bedding, T. R., et al. 2018, The Astrophysical Journal Supplement Series, 236, 42, doi: 10.3847/1538-4365/aaaf74