International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Detection of Exoplanets using Machine Learning

Amritanshu Kumar Singh[a]*, Vedant Arvind Kumbhare*[b]

[a,b,*]*SRM Institute of Science and Technology, Kattankulathur, Chennai, 603203, India*

DOI: https://doi.org/10.55248/gengpi.2022.3.2.12

A B S T R A C T

Three methods for using machine learning to decide if a star has an exoplanet from transit survey data are discussed in this paper and are also used to determine which approach performs better on a labeled data set containing light intensity time series from extrasolar stars. Convolutional Neural Network (C.N.N.), a C.N.N. autoencoder, and Support Vector Machines (S.V.M.) are among the three models. We trained the models using data from the Kepler Space Telescope; because there were very few confirmed exoplanet in the data, some preprocessing is performed before implementing the data set methods. SMOTE (Synthetic Minority Oversampling Technique) was used to over-sample the exoplanet class to further boost the performance. For the supervised learning approaches, CNN and SVM, the outcomes were promising. Therefore, CNN autoencoder, the unsupervised method, did not generate meaningful results.

Keywords: CNN, SVM, Exoplanet, SMOTE

1. Introduction

The Solar System was created around 4600 million years ago. Humans have long wondered if the billions of stars in our galaxy existed in other solar systems. We have collected samples from a series of many exoplanetary systems over the last two decades. We now know, instead of an organized solar nebula model, that chaos is the reason for the creation of planetary systems. Close to their stars, gas giant planets can migrate. As there so many small rocky planets, the inner orbits are tightly packed. Planets revolve just outside the orbits of binary star systems. The heterogeneity is breathtaking. The transit survey method is the most widely used methodfor searching for exoplanets. A star's light reflection intensity is used for time-series measurement that is called as transit survey light. If an object like a planet (or any space object) appears in between the star and the telescope taking measurements, one would expect to see a decrease in the measured light intensity values as some of the light from the star would be blocked. This implies that when a planet travels in the front of a star, it casts a shadow, which may be seen as a decrease in light intensity on a light curve. Although detecting reduction in light curvesseem like an easy task, it is generally tough since light curves can be very noisy, and planets typicallymay not block much light as they are usually considerably smaller than the stars they orbit. We believe that machine learning, namely neural networks, might be a valuabletechnique to detect the hidden sequences of light curves that indicate whether or not a star has an exoplanet (or exoplanets).

Also, classification problems are often the result of data imbalance between groups. This is the scenario where the scale of the examples in one class is substantially greater or lower than in the other class. For several of these issues, it is beneficial to develop classifiers that perform well on the minority class. Applying out-of-the-box classifiers for such problems can lead to sub-optimal results for this objective.

2. Dataset

2.1. Data

For each star, the light intensity is calculated over time and the binary is labelled either "1" for "no exoplanet in orbit" or "2" for "at least one exoplanet in orbit." The Kepler space telescope of NASA collected data that is used in this project. Telescope noise interference has been eliminated and detections from other programs have been incorporated to improve the data collection. Since the total number of stars without an orbiting planet is much higher than the number of stars with an orbiting planet, the data set is unbalanced by a ratio of 99 per cent of the stars labelled as having no exoplanet in orbit. Confirmed exoplanets from other projects are also included in order to make the data less unbalanced, but the addition of this data would not be adequate compensation for the data collection to be listed as balanced. The data imbalance present in this data set is one of the main data challenges. The data is downloaded from the Kaggle.com host as CSV-files.

The data is divided into one training set and a test set. The training collection consists of 5087 observed stars with 37 confirmed exoplanet-stars and 570 observed stars with five confirmed exoplanet-stars. Both data sets are assessed at 3198 time-steps, corresponding to approximately 80 days of observation.

2.2. Imbalance Dataset Training

Exoplanet samples are under-represented in the dataset. If we train the models on the current dataset, the models will encounter relatively few exoplanet instances and will struggle to distinguish between an exoplanet and non-exoplanet light curves. We used Synthetic Minority Oversampling Technique (SMOTE) to combat this under-representation of exoplanet samples.

SMOTE produces synthetic minority class samples by integrating a sample linearly with few of its nearest neighbours. For example, with a minority class sample, x1, its nearest neighbour, x2, and a random number between 0 and 1, α, a synthetic sample can be generated as follows.

$$\mathbf{xnew = x1 + \alpha(x1 - x2)}$$

We used the SMOTE method there in Python Imblearn Library. Using S.M.O.T.E., we successfully balanced the training data resulting in 5050 non-exoplanet samples and 5050 exoplanet samples.

2.3. Principal Component Analysis

Principal Component Analysis (PCA) is popular approach for the extracting andreducing the dimensions of high-dimensional data. At a high level, PCA operates by mapping data into a new space feature such that the components of this space feature optimize data variance. These components are what are known as the main components and can also be represented by the unit eigenvectors of the covariance matrix. The proprietary values refer to the magnitude of the variance described by each proprietary vector, such that the proprietary vector with the largest proprietary value is the first main component, the proprietary vector with the second-largest proprietary value is the second main component, and so on. Since the covariance matrix is symmetrical, it has an eigenvector. The orthogonal design of the key components de-corrects the data which is what gives this system the ability to generate so much data variation in so few functions. The number of key components chosen was such that 100% of the variance of the initial dataset was accounted for. PCA has been done using the Scikit-learn Python Library.

2.4. Fast Fourier transform

Because the data consists of time-varying signals, the Discrete Fourier Transform (DFT), that transform a signal into frequency space, is one type of extraction function that we wanted to try. A DFT transforms the signal into a finite sum of sines and cosines by measuring the coefficient, fk, for each frequency, k,,

$$F_n = \sum_{k=0}^{3197-1} = f_k e^{-2\pi ink/3197}$$

DFT is carried out using the Numpy-implemented FFT algorithm. A vector x ∈ R3197 is the input to the FFT and a vector z ∈ C3197 is the output. For each frequency, z ∈ R3197, the absolute value of each of the complex elements of z is taken to get the magnitude. An instance of a normalized light curve is shown in Fig. 1 after FFT has been done.

As there is no longer a link between neighbouring inputs after the FFT, only with the FCNN model can this feature extraction approach be used.

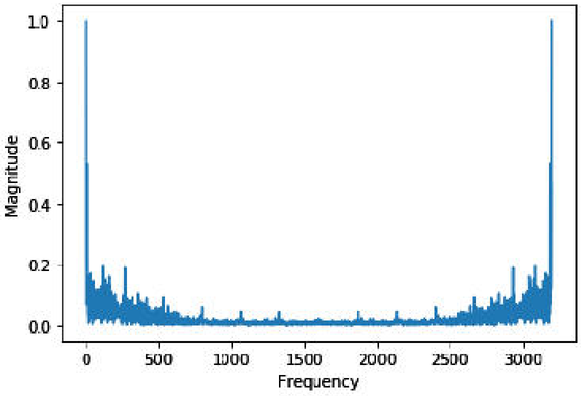


Fig. 1 – An example of a light curve signal after Fast Fourier Transform has been performed

.

3. Transit

3.1. Space-based Transit discovery

Both Kepler and CoRoT exhibited great efficiency on two space missions, and a third expedition, MOST, made photometric transit discoveries of many previously identified planets.

In particular, the Kepler Mission was groundbreaking, producing over a hundred planets with mass determinations, at last rapidly evolving counts, as well as hundreds of examples of numerous transiting planets orbiting a single host star, many of which are in strongly co-planar, surprisingly crowded systems. Taken together the candidates of Kepler suggest that planets with masses $MP < 30M\oplus$ and orbital periods $P < 100$ d are essentially ubiquitous, that the distribution of mass ratios and periods of these candidate planets are in many instances, oddly reminiscent of the usual Jovian planetary satellites within our solar system.

The Kepler satellite experienced a failure of the second reaction wheel in Spring 2013, just after the end of its nominal mission duration, which brought its high precision photometric monitoring program to a halt. However, the years of Kepler knowledge in hand are well-curated, completely open-source, and are still far from fully exploited; it is not unrealistic to conclude that they would provide additional insight equal to what has already been deduced from the mission to date.

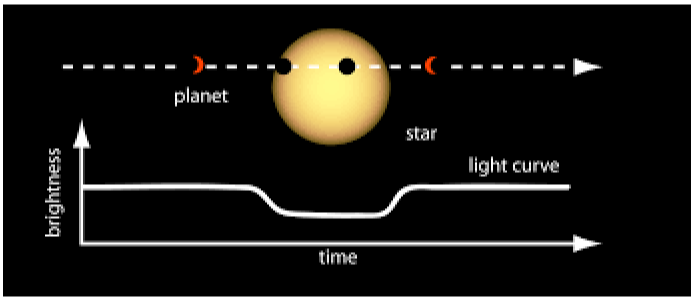


Fig. 2 – A transit is depicted graphically; as the planet passes in front of a star, the intensity of the star decreases.

3.2. Transit detection

The property of planetary orbit, and planetary and stellar radii, is the a-priori probability that a given planet will be observed in transit,

$$\mathcal{P}_{tr} = 0.0045 \left(\frac{\mathrm{AU}}{a}\right)\left(\frac{R_\star + R_\mathrm{p}}{R_\odot}\right)\left[\frac{1+e\cos(\pi/2-\omega)}{1-e^2}\right]$$

Where the angle at which the orbital periastron occurs is ω, so that transit is indicated by = 90◦ and orbital eccentricity is e. standard hot Jupiter with Rp & RJup and P ~ 3 d, orbiting a star of a solar type, has a transit period of τ ~ 3s, a transit depth of photometric, d ~1%, and P ~10%. Planets belonging to the ubiquitous super-Earth-sub-Neptune population found by Kepler (i.e., the grey points in Figure a are characterized by P ~ 2.5%, d ~ 0.1% and τ ~ 6 hr, while Earth-sized planets have a difficult combination of P ~0.5%, d ~0.01%, and τ ~15 hr in an Earth-like orbit around solar- type stars

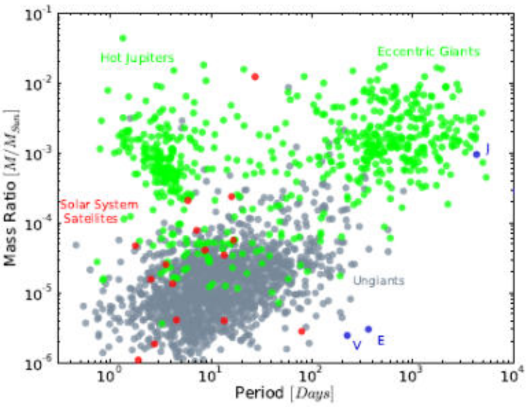


Fig. 3 – Green circles: $\log_{10}$(Msatellite/Mprimary) and $\log_{10}$(P) for 634 planets securely detected by the radial velocity method (either with or without photometrictransits).

Redcircles: $\log_{10}$(Msatellite/Mprimary) and $\log_{10}$(P) for the regular satellites of the Jovian planets in the Solar System. Gray circles: $\log_{10}$(Msatellite/Mprimary) and $\log_{10}$(P) for 1501 Kepler candidates and objects of interest, in which a single primary is correlated with multiple transiting candidate planets. For these candidate planets, radii are transformed into masses, assuming, $M/M_\oplus = (R/R_\oplus)^{2.06}$, which is obtained by matching the masses and radii of the planets of the solar system bounded by Venus and Saturn in mass.

The optimal tradeoff between cost, sky coverage, photometric precision, and the median apparent brightness of the stars under observation is sought by successful transit search strategies. The group as a whole has failed to conduct truly productive surveys for nearly a decade.

To date, with orthogonal phase transfer arrays trained on single, carefully preselected high-value target stars, the highest-precision ground-based exoplanetary photometry has been obtained. (Johnson et al., 2009) obtained 0.47 mmag photometry for WASP-10 at an 80-second cadence (V=12.7) using this technique. By contrast, Kepler achieved a median photometric accuracy of 29 ppm with a 6.5-hour cadence on V=12 stars with its space-borne tip.

This is ~ 2× better than the best special-purpose ground-based photometry and ~ 20× better than the leading ground-based discovery surveys. For wide- field surveys in general and for Kepler in particular, where a majority of the candidate planets are essentially out of control of Doppler characterization and confirmation, astrophysical false positives present a serious challenge. Stars compare in size with giant planets at the bottom of the main sequence (Chabrier and Baraffe, 2000) and thus show near-identical transit signatures to those of giant planets. In low signal-to-noise light curves, grazing eclipsing binaries may also provide a source of major uncertainty.

Pixel "blends" constitute a big channel for false alarms within the Kepler sector. These occur when the target star shares a line of sight with an eclipsing binary, either physically related or unrelated.

Photometry alone can be used to classify several such events (Batalha et al., 2010) while in other cases, statistical modelling of the probability of blending scenarios can provide convincingly low false alarm probabilities. High-profile examples of statistical validity confirmation include the $R = 2.2R_\oplus$ of the Earth candidate planet Kepler 10c by, as well as the planets in the Kepler 62 system (Borucki et al., 2013). False alarm odds are inferred to be significantly lower in situations where several candidate planets are moving through the same star. Of the grey spots in Figure 3, there is very likely to be just a relatively small mix of false alarms.

3.3. Results and Implications

Apart from the substantial rise in the quantity of known transiting planets, the sequence of transit discoveries over the last six years has been of profoundly new significance. In particular, transit detections have made it possible to research both planets and planetary system architectures for which there are no analogues to the solar system.

A brief overview of the significant events reported for the discovery year can include I Gliese 436 b (Gillon et al., 2007) the first transiting Neptune-sized planet and the first transiting low-mass star, (ii) HD 17156 b the first transiting planet with a broad orbital eccentricity (e=0.69) and an orbital period (P = 21d) that is considerably larger than $2\text{ d} < P < 5\text{ d}$ of the occupied range.(iii) CoRoT 7 b (Leger et al. ´ , 2009) and Gliese 1214b The first transiting planets in the so-called "super-Earth" with masses regime $1\ M_\oplus < M < 10\ M_\oplus$, (iv) Kepler 9b and 9c (Holman et al., 2010) As well as the first instance of transiting planets performing a low-order mean motion resonance, the first planetary system to show tangible transit timing variations, (v) Kepler 22b, The Kepler 62 system (Borucki et al., 2013), which hosts at least five transiting planets orbiting a primary K2V, is the first transiting planet with a size and an orbital period that could theoretically harbor an Earthlike environment (Borucki et al., 2013), and (vi)

For transiting planets with parent stars, bulk densities are determined that are bright enough and chromospherically quiet enough to help Doppler measurement of MP sin(i), and can also be obtained by modelling variations of transit timing (Fabrycky et al., 2012; Lithwick et al., 2012). More than 100 planetary densities have been securely measured (mostly for hot Jupiters). In Figure 4, these are plotted, representing the general outlines of the overall distribution. Over the next few years, Figure 4 is expected to experience rapid development as more Kepler candidates obtain mass determinations. However, it seems possible that a very large range of planetary radii occurs at any mass.

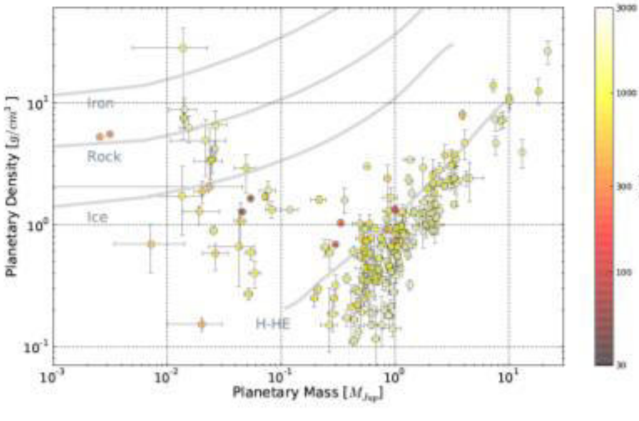


Fig. 4.- Density-Mass diagram with well-defined masses and radii for planets.

The equilibrium temperature, Teq, which they would have if they were zero-albedo black-bodies re-radiating from the maximum planetary surface area, is colour-coded. In the plotted aggregate, solar system consists planets more massive than Mars are included. For planetary models of pure hydrogen-helium, pure water, pure silicate, and pure iron compositions, grey lines indicate ρ(MP).
A large number of candidate multiple transiting planet systems suggest that for planets with P < 100 d in the size range of Rp ~ 1.5 – 6 R⊕, co-planar architectures are the norm. The dispersion of inclination of most candidate systems of 2 or more transiting planets seems to have a median of 1-3°. Besides, candidate planets in multiple-transit systems, when imbued with rational mass-radius relations, are invariably in dynamically stable configurations. Therefore, Nature has created a galactic planetary census that is extremely well-suited through the transit method for detection and characterization. In tandem with JWST's potential for atmospheric characterization of low-mass planets, the emergence of new space missions suggests that transits will remain at the forefront for decades to come.

Finally, the identification of transits is remarkable in that it democratizes access to cutting edge exoplanetary science studies. Almost all of the highly-cited ground-based observations were made with small aperture telescopes, d< 1m.

4. Methodologies Used

4.1. Neural Networks

A neural network is a flexible system that learns by employing inter linked nodes or neurons in a stratified structure that is similar to a human brain. A neural network is able learn from data—so it can be trained to perceive patterns, categorize data, and predict future events.

It is based on the principles of less complex linear- and logical regression models and is used for use in image processing, speaker identification and translation of language. The architecture of a model of the neural network is galvanized by how neurons act within the human brain and helps to distinguish non-trivial patterns between data from input and output. A neural network consists of a minimum of 3 so-called layers, where a set of neurons consists of each layer. The input layer being the primary layer that receives the input data, and the number of neutrons in this layer is proportional to the number of data points. The output layer is the last layer, where the sum of nodes is equal to the number of classes for a classification problem. Between the input and output layers, there is at least one hidden layer where the nodes convert the weights before passing them through to generate an estimate of the outcome.

The inputs of a neural network can be decomposed into multiple levels of abstraction. It should be conditioned to recognise patterns in photographs or speeches using several scenarios, much as a human's brain does. Its behaviour is defined by the process of its individual linked components and by the weight of these links. Before the artificial neural network performs the intended task properly, these weights are changed innately during training according to a fixed learning law.

In vision, voice, and control systems, neural networks are particularly appropriate to perform pattern recognition to recognize and label objects or signals. They are also used for prediction and simulation of time series executions.

4.2. Convolutional Neural Networks

A Convolutional Neural Network (ConvNet/CNN) is an algorithm that can take an input image, give values to different aspects/objects in the image (learnable weights and biases), and segregate between them. In comparison to other classification algorithms, the pre-processing required in a ConvNet is much lower. Although filters are built in rudimentary processes, ConvNets have the capacity to learn these characteristics with adequate training.

A ConvNet's design is comparable to that of Neurons' connectivity pattern in the Human Mind and was inspired by the Visual Cortical organization. Unique neurons react only in a narrow area of the visual field known as the Receptive Field to stimuli. To cover the full visual region, a variety of such fields overlap.

In processing images by studying abstract representations with high levels of semantics, Convolutionary Neural Networks (CNNs) are outstanding. Innately, they are also well designed for texture research as they learn weight-sharing filter banks and localized connectivity that detect and recognize paterns at all points in the image. A CNN's first convolution layer learns to recognize simple features composed mainly of edges that are quite close to the widely used Gabo r filters in texture analysis. From plain nonlinear combinations of the past iterations, the succeeding layers learn to recognize bigger and more complex features. Therefore, CNN can learn to recognize textures patterns of different complexities and sizes, largely improving conventional approaches to filter banks by replacing personalized filters with a hierarchical architecture combined with a strong algorithm for learning. With this filter bank approach, traditional networks largely leverage the texture material present in images. In deeper layers, however, CNNs are beginning to detect universal patterns and shapes over basic texture patterns with the growth of the receptive fields of the neurons and the sophistication and level of abstraction of the features.

An instance is the *max pooling* feature that works within a fixed interval, replacing the outputs over that interval with the maximum value. Many kernels can be added to capture more information in a convolution layer. Each kernel has its own collection of concealed units and specifications, which are called channels. In one sheet, the hidden units are collected into a dimension tensor; rows x columns x channels. In the CNN framework, kernels rely on all channels in previous layers, so a weight tensor, W, will have dimensions comprising all kernel parameters; kernel rows x kernel columns x input channels x output channels. The network can be applied on both dense and convolution layers.

4.3. Support Vector Machine

A *Support Vector Machine* is a formally defined hyperplane-separating discriminative classifier. In other words, the algorithm outputs an optimized hyperplane that classifies new instances, given labelled training data (supervised learning). This hyperplane is a line dividing a plane into 2 sections in 2-dimensional space, where each class is on either side.

Due to its relative simplicity and adaptability in addressing a range of classification issues, SVMs provide highly balanced prediction efficiency, especially in studies with small sample sets. SVMs are usually used in brain injury research using multivoxel pattern analysis (MVPA) because their relative simplicity carries a lower risk of overfitting, even with high-dimensional image images. In the sense of precision healthcare, SVMs have been used more recently, particularly for applications involving predicting diagnosis and treatment of brain disorders such as schizophrenia, Alzheimer's disease and depression.

VM is the most commonly used method of classification of ML technique-based pattern available today. It is based on the theory of mathematical learning and was developed in the year 1995. The main objective of this technique is to use various types of kernel functions to project nonlinear separable samples onto another higher dimensional space. In latter years due to the growing prominence of SVM, kernel methods earned significant attention.Kernel functions play an important role in SVM to transition from linearity to nonlinearity. Least square SV is also a significant SVM tool that can be used for the task of classification For classification purpose, the greatest learning machine and the Fuzzy SVM and genetic algorithm tuned expert model may also be implemented. 3 different forms of kernel functions have been tested in this analytical work, i.e., linear, polynomial, and RBF kernels.

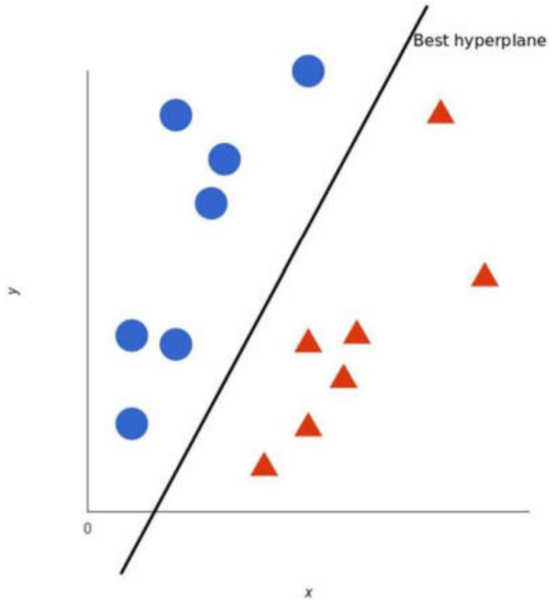


Fig. 5 – Support Vector Machines : Best Hyperplane example graph.

SMOTE is an oversampling method which has synthetic samples for the minority class are produced. This algorithm enables the over-fitting problem posed by random over-sampling to be solved. With the assistance of interpolation between the positive examples that reside together and it focuses on the feature space to build new instances.

- SMOTE is a method based on closest neighbours evaluated in feature space by the Euclidean Distance within data points.
- There is a fraction of Over-Sampling that indicates the number of synthetic samples to be produced and a multiple of Hundred and is often this percentage parameter of Over-Sampling. If the percentage of over- sampling is 100, then a new sample will be produced for each instance. Therefore, the number of instances in the minority class would double. Similarly, if the over-sampling percentage is 200, the overall sum of minority class samples would triple.
- K number of nearest neighbours are found for each minority instance case such as, they also belong to the same class where, The difference is found between the feature vector of the instance considered and the feature vectors of the closest neighbours of k. So, k number of vectors of difference are acquired,

$$k = (SMOTe\ \%)/100$$

- A random number between 0 and 1 (except 0 and1) is multiplied by the k difference vectors each.
- Now after being multiplied by random digits, the difference vectors are applied at each iterations to the feature vector of the instance deemed (original minority instance.

5. Class Imbalance

The vector machine with soft margin support is susceptible to imbalanced data sets and there are many ways to handle this. The emphasis is on class weights in this project, meaning that the model is skewed in order to prioritize the minority class over the majority class. The hyperplane that divides SVM classes appears to become biased against the minority class, causing the model to function better for the majority class, but for the minority class to perform poorly. The help vectors are unbalanced, another result of an imbalanced data collection. By inserting a greater expense for misclassification of the minority class than for misclassification of the majority class, the class weights enhance this issue, results in the cost function.

6. Evaluation Metrics

The purpose of the algorithms is to be able to find exoplanets from the light curves and to match them with various output measurement quantities. Using accuracy (calculated by dividing correct predictions by total predictions) when testing machine learning models is often popular, but it can be deceptive for imbalanced data sets, as shown in the one used in this analysis. Nevertheless, accuracy is described as: A = (T P + T N) / (T P + T N + F P + F N).

The most critical task of the algorithm was perceived to be locating one or as many of the exoplanets as possible that are not incorrectly classifying exoplanet stars as non-exoplanet stars, so recall was picked as the main performance parameter. Recall is known as R = (T P) / (T P + F N). Furthermore, algorithms should be able to determine that there is no exoplanet in orbit of the star; no positive sample will be missing by an algorithm detecting exoplanets around every star, so that wouldn't be that useful either P = (T P) / (T P + F P).

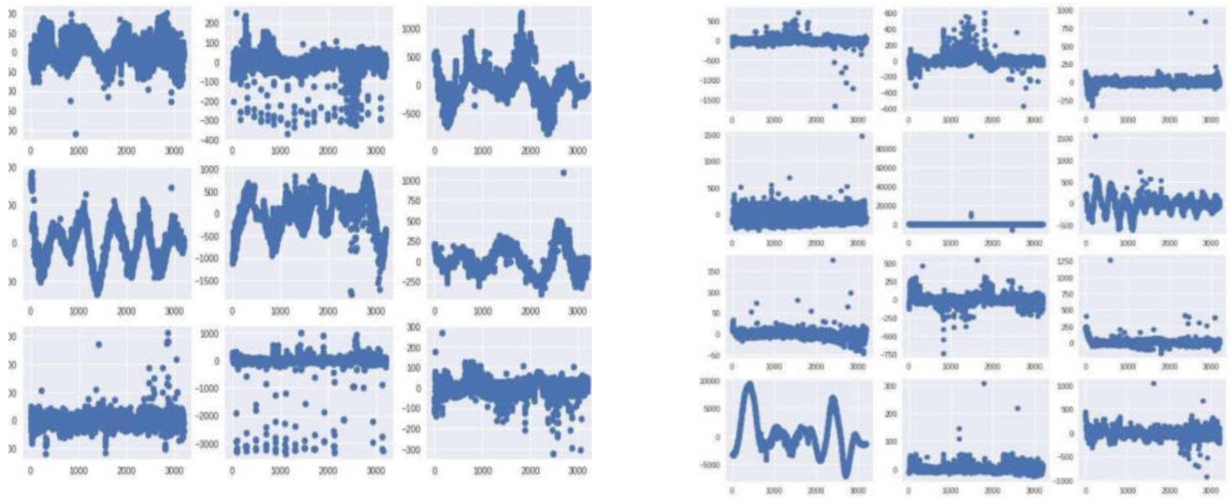
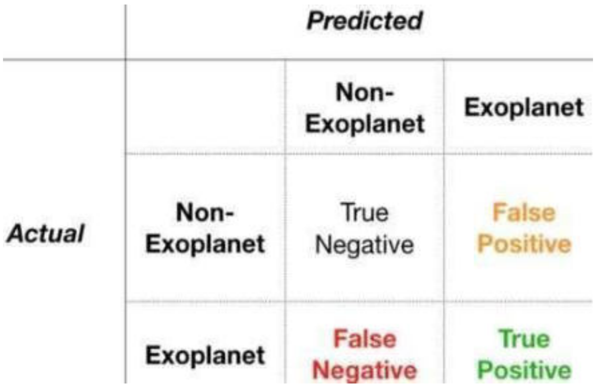


Fig. 7 – Exoplanet vs Non- Exoplanet Time Series Graph



		Predicted	
		Non-Exoplanet	Exoplanet
Actual	Non-Exoplanet	True Negative	False Positive
	Exoplanet	False Negative	True Positive

Fig. 8 – Preprocessing

7. Feature Extraction

Feature extraction starts with the initial collection of calculated data and constructs derived values in machine learning, pattern recognition and visual processing, designed to be descriptive and non- redundant, enabling concurrent learning and generalization measures and in some situations, leading to improved human interpretation. Feature extraction relates to the removal in dimensionality.

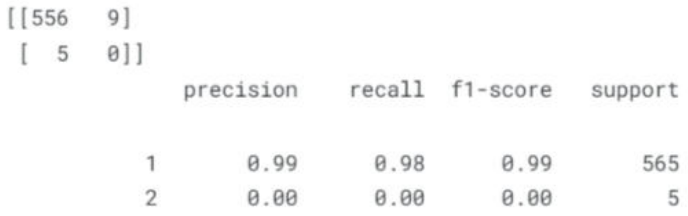
Feature Extraction helps to minimize the number of features in a dataset by generating (and then dismissing the old features) new features from the current ones. Any of the details found in the initial package of features should then be able to sum up this new decreased set of features. In this manner, from a variation of the original set, a summarized version of the original features can be generated.

Feature Selection is another widely used method to minimize the number of characteristics in a dataset. The distinction between the collection of features and the extraction of features is that the selection of features is instead targeted at rating the value of known features in the data and discarding irrelevant (no new features are formed).

8. Result

TRAIN DATA

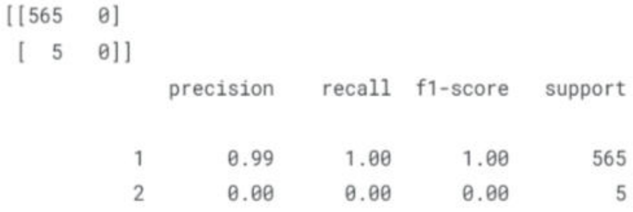
The models were trained and tested using the training sets, which included 5124 observed stars. The convolutional neural network was trained using a batch size of 32 and a threshold of 0.79. The SVM method that used fourier transformed, non-normalized data and balanced class weights performed the best.



```
[[556   9]
 [  5   0]]
              precision    recall  f1-score   support

           1       0.99      0.98      0.99       565
           2       0.00      0.00      0.00         5
```

Table 1. – Train results for CNN



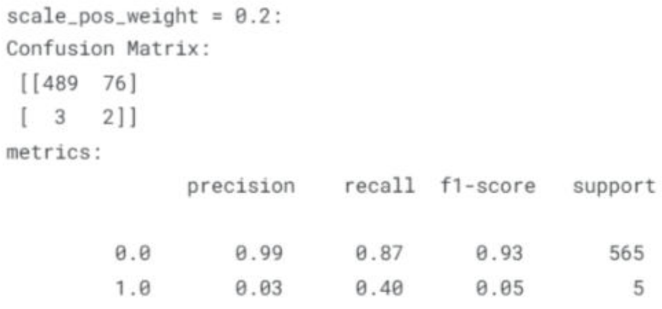
```
[[565   0]
 [  5   0]]
              precision    recall  f1-score   support

           1       0.99      1.00      1.00       565
           2       0.00      0.00      0.00         5
```

Table 2. – Train results for SVM

TEST DATA

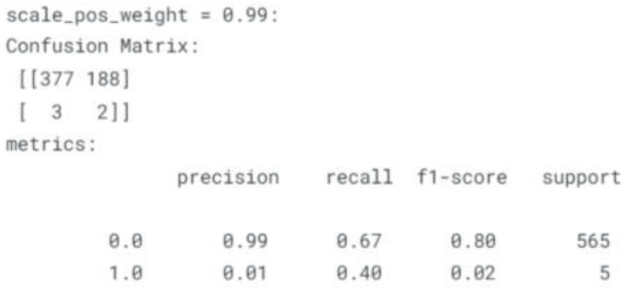
The models were tested on a data collection of 575 observed stars. The precision and recall values are shown.



```
scale_pos_weight = 0.2:
Confusion Matrix:
 [[489  76]
 [  3   2]]
metrics:
              precision    recall  f1-score   support

         0.0       0.99      0.87      0.93       565
         1.0       0.03      0.40      0.05         5
```

Table 3. – Test results for CNN



```
scale_pos_weight = 0.99:
Confusion Matrix:
 [[377 188]
 [  3   2]]
metrics:
              precision    recall  f1-score   support

         0.0       0.99      0.67      0.80       565
         1.0       0.01      0.40      0.02         5
```

Table 4. – Test results for SVM

9. Conclusion

CNN and SVM were used to locate transitioning exoplanets. Given all the parameters utilised for validation, the CNN outperformed the SVM on the test data. As a result, CNN was shown to be the best performing model on this data set including time series of light intensity from astronomical stars with the given configurations. The strong pattern identification capacity of CNN is assumed to be the basis for this outcome.Engaging with such unbalanced data sets may be difficult, and it is a problem that has to be addressed several times. But with only very few numbers of positively labelled curves, innovation is required to generate additional positive instances, and model validation with specific measurements is required.

Data pre-processing options are limitless. Just few pre-processing choices were evaluated for this study, and more work may have enhanced the model's performance. Because the model is looking for slumps in the light curves, one additional step in pre-processing may be to remove all higher outliers because they may be unimportant.

REFERENCES

Joseph Bell, Maria Harris, and James Salem (2019). Identifying Exoplanets From Transit. University of California San Diego Research Journal

Yu H.-F., Huang F.-L., Lin C.-J., 2011, Machine Learning, 85, 41

Ajinkya More. Survey of resampling techniques for improving classification performance in unbalanced datasets. arXiv:1608.06048v1

Exoplanet hunting in deep space, 2017, https://www.kaggle.com/keplersmachines/keplerlabel led-time-series-data.

P. Bord´e, D. Rouan, and A. L´eger. Exoplanet detection capability of the COROT space mission. A&A **Volume** 405, Number **3**, July

Jason T. Wright, B. Scott Gaudi. Exoplanet Detection Methods. arXiv:1210.2471

Debra A. Fischer, Andrew W. Howard, Greg P. Laughlin, Bruce Macintosh, Suvrath Mahadevan, Johannes Sahlmann , Jennifer C. Yee. Exoplanet Detection Techniques. arXiv:1505.06869

Sebastian, Daniel & Guenther, E. & Schaffenroth, Veronika & Gandolfi, Davide & Geier, Simon & Heber, U. & Deleuil, Magali & Moutou, C.. (2012). Multi-object spectroscopy of stars in the CoRoT fields I: Early-typestars in the CoRoT-fields IRa01, LRa01, LRa02.

Priyadarshini, Ishaani & Puri, Vikram. (2021). A convolutional neural network (CNN) based ensemble model for exoplanet detection. Earth Science Informatics.

Martin, David & Triaud, Amaury. (2015). Circumbinary planets - why they are so likely to transit. Monthly Notices of the Royal Astronomical Society.

Fischer, D. & Howard, Andrew & Laughlin, Greg & Macintosh, Bruce & Mahadevan, Suvrath & Sahlmann, Johannes & Yee, Jennifer. (2015). Exoplanet Detection Techniques.

Keneally, S.K., Robbins, M.J. & Lunday, B.J. A markov decision process model for the optimal dispatch of military medical evacuation assets. *Health Care Manag Sci*

Malbet, F., Boehm, C., Krone-Martins, A. *et al.* Faint objects in motion: the new frontier of high precision astrometry. *Exp Astron* **51,** 845–886 (2021). https://doi.org/10.1007/s10686-021-09781-1

Dittmann, Jason & Close, Laird & Scuderi, Louis & Morris, and. (2010). Transit Observations of the WASP-10 System. The Astrophysical Journal.

Alonso, Roi & Brown, Timothy & Torres, Guillermo & Latham, David & Sozzetti, Alessandro & Mandushev, Georgi & Belmonte, Juan & Charbonneau, David & Deeg, Hans & Dunham, Edward & O'Donovan, Francis & Stefanik, and. (2008). TrES-1: The transiting planet of a bright K0 V star. The Astrophysical Journal Letters.

Deming, Drake & Agol, Eric & Charbonneau, David & Cowan, Nicolas & Knutson, Heather & Marengo, Massimo. (2007). Observations of Extrasolar Planets During the non-Cryogenic Spitzer Space Telescope Mission.

Tiensuu, M. Linderholm, S. Dreborg, and F. Örn, 'Detecting exoplanets with machine learning: A comparative study between convolutional neural networks and support vector machines', Dissertation, 2019.

Borucki, William J., Eric Agol, Francois Fressin, Lisa Kaltenegger, Jason Rowe, Howard Isaacson, Debra Fischer et al. "Kepler-62: a five-planet system with planets of 1.4 and 1.6 Earth radii in the habitable zone." *Science* 340, no. 6132 (2013): 587-590.

Chabrier, G., & Baraffe, I. (2000). Theory of Low-Mass Stars and Substellar Objects. Annual Review of Astronomy and Astrophysics

Batalha, Natalie M., William J. Borucki, David G. Koch, Stephen T. Bryson, Michael R. Haas, Timothy M. Brown, Douglas A. Caldwell et al. "Selection, prioritization, and characteristics of Kepler target stars." *The Astrophysical Journal Letters* 713, no. 2 (2010): L109.

Gillon, Michaël, B-O. Demory, T. Barman, X. Bonfils, T. Mazeh, F. Pont, S. Udry, M. Mayor, and D. Queloz. "Accurate Spitzer infrared radius measurement for the hot Neptune GJ 436b." *Astronomy & Astrophysics* 471, no. 3 (2007): L51-L54.