

# A Review Based Study on Different Techniques of AI Used in Exoplanet Detection

Richa Shree<sup>1</sup>, Dr. Pooja Jha<sup>2</sup>

<sup>1</sup>Kadi Sarva Vishwavidyalay

<sup>2</sup>Amity University Jharkhand

## Abstract:

This article delves into the connection between astronomy and machine learning, specifically focusing on gravitational waves and the detection of exoplanets. It emphasizes how automated analysis tools have become crucial in advancing our understanding of the universe thanks to the potential of machine learning. The article explores how machine learning is used for classifying supernovae and detecting anomalies in data. It also discusses the challenges faced when applying relativity to studying powerful celestial bodies like exoplanets. Additionally, it highlights the increasing reliance on data analysis and machine learning within astronomy. Moreover, it underscores the importance of exoplanet research and how machine learning can automate the process of discovery. Ultimately this article emphasizes how modern technologies, including machine learning are transforming research and making it more accessible to citizen scientists.

**Keywords:** Astronomy, Machine learning, Supernovae, Gravitational waves , Exoplanet detection, Data analysis, Radio galaxies, Numerical relativity, Kepler mission.

## 1. INTRODUCTION

The paper explores the field of astronomy and how machine learning techniques are being utilized in research areas. These include studying supernovae detecting waves finding exoplanets analyzing data and observing radio galaxies. These topics emphasize the importance of automated analysis tools the challenges faced in data processing and the potential that machine learning holds in advancing our understanding of the universe.

Supernovae play a role in enriching the medium with heavy elements triggering star formation and serving as a source of high energy cosmic rays. With the abundance of supernova survey data, automated analysis tools have become indispensable. Machine learning methods for classifying supernovae using photometric data prove valuable in compensating for the absence of spectroscopic support. Employing an N parameter grid machine learning analysis [4] can effectively differentiate between types of supernovae and remove supernova contamination from the sample and potentially identify objects with unique properties. However, there is still room for exploration in the field of anomaly detection within data.

The breakthrough discovery of waves which're ripples in the fabric of space time, stands as a remarkable confirmation of Einstein's theory of general relativity.

Laser interferometers are used to capture and analyze these waves, which are then translated into relativity-based patterns for comprehension. However, when it comes to studying entities like exoplanets, implementing numerical relativity faces challenges due to time complexity. We require solutions to generate numerical relativity patterns. Gravitational waves also hold potential in supplementing our knowledge about exoplanets and initial models have been introduced in this context. The article highlights the increasing application of data analysis and machine learning in the field of astronomy encompassing tasks such as categorizing stars, identifying and detecting exoplanets. It splits the tasks of analyzing data into two categories: statistics and data science methods for non-graphical data while computer graphics spectral analysis and computer vision techniques are used for datasets containing sky images. Although, some approaches may appear complex for tasks, machine learning remains integral in advancing our understanding of the universe.

The study of exoplanets holds significance as it contributes to our comprehension of how our Solar System was formed, the search for inhabitable planets and the exploration of life beyond our Solar System. The Kepler mission, which monitored changes in star brightness to identify transits of planets in same size or larger [12] than Earth has provided accessible data that has greatly enriched exoplanetary research.

However, manually analyzing sets of data to identify exoplanet candidates is a time consuming and exhaustive task. The discovery of exoplanets comes with challenges including, dealing with variability, noise sources inconsistencies, in light curves false alarm rates, weak transit signals, false positives and the overwhelming amount of information available. To address these challenges scientists have proposed using machine learning and artificial intelligence techniques to automate the discovery process and reduce data noise. Additionally, multiresolution analysis methods have been employed to enhance the identification of exoplanets. The study mentioned focuses on implementing models from the field of data science, complex methodologies and classification models to classify exoplanet data obtained from the Kepler mission. By automating object classification and streamlining data processing, this research expedites the search for exoplanets.

The article underscores the role astronomy has played throughout history and its connection to aspects of human civilization. It acknowledges the impact of satellites like Kepler in partially automating observations and generating data thereby, democratizing exoplanet research.

The use of machine learning techniques by citizen astronomers is also mentioned with examples provided on how neural network models have been employed in identifying planets. This study follows the trend of crowdsourced astronomy and leverages machine learning models such, as Support Vector Machine, K Nearest Neighbors [5] and Random Forest to classify exoplanets.

## 2. LITERATURE REVIEW

The paragraph explores the different ways machine learning algorithms are utilized in the Angle Search, for Planets (WASP) data to identify and classify planet candidates. The WASP project employs two instruments that incorporate cost commercial components [1]. To process the data techniques like the Trend Filtering Algorithm (TFA) and the SysRem algorithm are applied for tasks such as trend filtering and reducing errors [2]. The refined data is then used to generate curves for all the identified targets in the

ORCA TAMTFA fields. A total of 716 fields shows candidates, which are further selected by observers for subsequent observations [3].

In the study titled "The WASP Project and the SuperWASP Cameras: Investigating Planets through Observations " researchers discuss the importance of having a structured methodology in machine learning to classify and prioritize planets using data and observations. The aim is to establish a process, for identifying and ranking planet candidates. When it comes to detecting planets of the size of Earth, traditional techniques like least squares optimization and grid search, have their limitations. However, a promising solution lies in the use of networks (CNNs) [4]. These networks optimize the signal to noise ratio and tackle the challenges posed by variations in transit shape. Additionally, automated planet finding algorithms such as forests, self-organizing maps and k nearest neighbors [5] are mentioned as effective tools for reducing false positive signals. In conclusion the utilization of deep learning algorithms trained on simulated data presents a robust approach to identifying transit signals, enabling them to capture the most subtle features in vast datasets. Supernova stars which are fascinating celestial objects, play important roles in enriching the interstellar medium triggering the formation of new stars and generating high energy cosmic rays. Additionally, they offer information regarding the composition, distance scale and ultimate destiny of the Universe. With the advent of supernova surveys there is a growing need for automated analysis tools to handle the massive amount of data. Machine learning techniques are crucial in supernova typing for determining the type of supernovae based on Grids. This analytical approach helps refine supernova samples by removing supernova contaminants and enables the identification of unknown variable objects or supernovae with unique properties. The search for these sources poses a challenge due to their rarity.

Various techniques, including principal component analysis (PCA), random forest (RF), decision tree (DT) and convolutional/deep neural networks (CNN/DNN) have been employed to classify stars, exoplanets, galaxies and quasars. These techniques have also been used to estimate redshifts, classify asteroids and stellar/galaxy morphologies, predict flares, detect gravitational waves and gamma ray bursts, analyze asteroid composition, identify pulsars and gravitational lenses as well as classify transit objects and radio wave sources. Machine learning has demonstrated potential in extracting insights from astronomical data.

Exoplanets refer to planets that exist beyond our Solar System and possess distinguishing characteristics that set them apart from objects. Typically, these planets have masses 13 times that of Jupiter, which prevents them from undergoing deuterium burning and evolving into brown dwarfs [6]. Some exoplanets known as mass objects [7] do not orbit any star. Initially, there were debates surrounding the discoveries of exoplanets with reports on the earliest detections. However, a significant milestone in exoplanet research was reached in 1995 with the discovery of "51 Pegasi b" by Michel Mayor and Didier Queloz using the velocity method [8]. This discovery marked the exoplanet found around a main sequence star distinct from our Sun.

Early searches for exoplanets involved ground based facilities like the Wide-Angle Search for Planets (WASP) [9] and the Hungarian made Automated Telescope (HAT). Furthermore, space missions such as NASA's Kepler Space Telescope launched in 2009, played a role in this field. Kepler diligently observed

a region within the Cygnus constellation for four years leading to the detection of a multitude of exoplanets. The information, from the mission was split into two categories: Long Cadence (LC) and Short Cadence (SC) targets [10]. Before being made accessible through the Mikulsky Archive the data underwent some steps. This data included Simple Aperture Photometry (SAP) and Presearch Data Conditioning (PDC) fluxes [11] To identify exoplanets and remove any positives and the Transiting Planet Search (TPS) and Data Validation (DV) modules were utilized.

The Kepler mission produced the Kepler Objects of Interest (KOI) catalog [12], containing stars that potentially host exoplanets. ML algorithms and manual vetting contributed to generating various KOI catalogs. To enhance data analysis, the KOI Network (KOINet), consisting of telescopes worldwide, collaborated on Transit Timing Variation (TTV) curve completion [13] and [14] which helped characterize multi-planetary systems. When the Kepler spacecraft experienced a reaction wheel failure [15], the project transitioned into the K2 mission, which expanded its scope to include studying various celestial objects such as young open clusters, bright stars, exoplanets, galaxies, supernovae, and asteroseismology.

### 3. METHODOLOGY

The study of exoplanets has involved astronomers, scientists and physicists for years. With the progress of intelligence (AI) this field has evolved into a problem that can be addressed using machine learning techniques going beyond traditional physics and planetary studies. Machine learning now plays a role in determining the likelihood of detecting exoplanets.

One popular approach in this research involves analyzing the brightness of light (flux), When a planet passes in front of its host star as observed from Earth it causes a decrease in the star's brightness which is referred to as flux. By monitoring these dimming over several months astronomers can infer the presence of an orbiting celestial body, potentially an exoplanet. Artificial intelligence with machine learning algorithms assists in determining whether the orbiting object is indeed an exoplanet.

Various AI methods used in article [16] have been explored for detecting exoplanets;

1. Decision Trees; These graphical structures utilize decision rules obtained from training data to predict the class or value of the target variable. They excel at handling linear decision-making tasks and are particularly useful for creating training models.
2. Support Vector Machines (SVM); SVM is a learning algorithm that is well suited for both classification and regression analysis. These methods offer solutions with features making them widely used for their efficiency and reliability.
3. Logistic Regression; This algorithm is used when analyzing regression, with a variable. It predicts the likelihood of an event occurring based on variables.
4. Random Forest Classifier; It is a learning algorithm that combines decision trees to make predictions. By considering input from trees, it determines the classification of the test object.
5. Multilayer Perceptron (MLP); This is a type of network that consists of linear layers, including input, hidden and output layers. Each perceptron can interact with others due to connectivity. There is a risk of overfitting.
6. Convolutional Neural Networks (CNN); CNNs are networks mainly used for image classification. They employ filters for convolution operations to learn features in the input data.

Each of these AI methods has strengths and applications allowing classification of exoplanets and analysis of transit properties. By combining these techniques, we have an approach to exoplanet detection that enhances result reliability and advances our understanding of these distant worlds.

The text provided is an excerpt taken from a research paper [17] that explores the application of machine learning algorithms in identifying planets from data. The researchers aimed to minimize detections while still being able to detect rare types of planets.

Here are the main points discussed in the text:

1. **Exploring the Data:** To distinguish between signals and false positives the researchers experimented with various machine learning techniques. Their training data set consisted of planets eclipsing binaries, variable stars and light curves without transits. They also considered blend scenarios where signals from stars overlap.
2. **Extracting Features:** In addition, to the information the researchers introduced features to capture more abstract or relational details. This included analysing the ratio of transit depth to width and examining the skewness of the distribution during transit events.
3. **Training Data:** The entire dataset was divided randomly into two sets: a training dataset and a testing dataset. The training set was utilized to train classifiers, such as Support Vector Classifier (SVC), Linear Support Vector Classifier (LinearSVC), regression, K nearest neighbours (KNN) and Random Forest Classifier (RFC). In order to improve the performance of the classifier, the features in the training data set were adjusted by centring and scaling them.
4. **Synthetic Minority Over Sampling Technique (SMOTE):** To address the issue of under representation of planets in the data set, the researchers employed a technique called Synthetic Minority Over Sampling Technique (SMOTE). This involved generating data points for the minority class resulting in an increased number of detected planets. However, it also led to an increase in positives for some classifiers.
5. **Random forest classifier:** After exploring classifiers the researchers ultimately chose to focus on the Random Forest Classifier (RFC) due to its recall rate. RFC is a method that combines decision trees trained on random subsets of the data. Its parameters, such as the number of trees, tree depth and number of features at each split were fine-tuned during the analysis.
6. **Feature importance:** During the analysis it was found that the period of the planet, served as the indicator for classification. Other important features included transit width estimated radius of the planet square value and the number of transits.
7. **neural networks:** In addition to methods Convolutional Neural Networks (CNNs) were also explored by the researchers to analyse the light curves. CNNs are a type of network proficient, in identifying complex patterns within data. For implementing the CNNs, Keras, a known learning framework was utilized.

In general, the scientists discovered that the RFC performed well in identifying planets. It also generated a considerable amount of false positive results particularly when dealing with binary star systems. To enhance the precision of classification they additionally investigated the potential of CNNs, as a testing method.

This article [18] discusses how machine learning algorithms are typically categorized as supervised or unsupervised. Supervised methods rely on pre-labeled data to train and tune the algorithms, allowing for classification or regression of new instances. However, relevant pre-labeled data can be scarce in astronomy, especially when targeting rare events or serendipitous discoveries. The five primary classes of supervised learning algorithms used in astronomy are ANNs, CNNs, DTs, RFs, and SVMs. CNNs have gained widespread usage in astronomy since the Ball and Brunner review in 2010.

ANNs, dating back to the 1950s, simulate biological neurons by weighting and combining signals from multiple inputs. SVMs, like ANNs, learn nonlinear decision boundaries but focus on finding hyperplanes that distinctly separate data in spaces of any dimensionality. Different ML/AI methods are applied to various data types, with CNNs being particularly suitable for image-style data.

Comparative studies that involve multiple methods are essential for assessing their performance, and reference datasets can facilitate such analyses. For instance, the availability of PHoto-z Accuracy Testing datasets allowed for evaluating the efficacy of MLPQNA, a neural network-based method, in assigning photometric redshifts. Solar astronomy has seen more systematic comparisons, but other disciplines, such as stellar and variable star classifications, also engage in experimentation with emerging techniques like probabilistic RFs and transfer learning.

**TABLE 1** *From a qualitative examination of a sample of ~200 refereed publications from 2017 to February 2019, a mapping emerges between the nature of astronomical data and the way that machine learning and artificial intelligence is actively been pursued.*

Data/Methods	ANN	CNN	GAN	SVM	DT	RF	DBSCAN	K-NN	K-M
Image	.	.	.	.	.	.		.	
Spectroscopy	.	.		.		.			.
photometry	.				.	.	.		.
Light Curve		.				.			
Time Series	.	.			.	.	.		
Catalogue	.			.	.	.	.	.	
Simulation	.	.	.	.		.			

Notes: *The table presents a summary of the types of astronomical data and the algorithms that appeared most regularly. The purpose of the table is to provide a convenient starting point for selecting an algorithm that has been used successfully for each data type. Abbreviations: ANN, artificial neural network; CNN, convolutional neural network; DBSCAN, density-based spatial clustering of applications with noise; DT, decision tree; GAN, generative adversarial network; k-M, k-means clustering; k-NN, k-nearest neighbors; RF, random forest; SVM, support vector machine.*

The article [19] explores the application of learning in the identification of gravitational waves (GWs) based on confirmed detections. Here is a brief overview of the approaches discussed in the article:

1. Gravitational wave Model: The article primarily focuses on studying the phase of black hole (BBH) waveforms and uses the IMRPhenomPv2 model to generate simulated waveforms, for both nonprecessing and processing GWs.
2. Position Parameter Constraint: A key observation in lensed GWs is the expected time delay ( $\hat{\Delta}t_d$ ) between two images. This delay is typically around a millisecond. To account for this the article restricts the parameter "y" to a range of  $0.05 < y < 1$  for Point Mass (PM) and Singular Isothermal Sphere (SIS) lens models.
3. Generation of Realistic Spectrogram Samples: Real gravitational wave data captured by detectors like LIGO/Virgo often includes noise from sources. To emulate real world scenarios the article employs a Power Spectral Density (PSD) model known as the high-power model of the Advanced LIGO to generate spectrogram samples. The Signal, to Noise Ratios (S/Ns) of these samples are controlled to ensure they fall within the range of  $10 \leq S/N \leq 50$ .
4. Deep Learning Application: In this article they utilize the Visual Geometry Group (VGG) network, VGG 19, for both classification and regression purposes. To implement VGG 19, PyTorch is employed. The spectrogram samples are divided into three sets: training, development and evaluation. As a step the samples go through max normalization. For regression they also prepare target data containing the parameters to be predicted.
5. Training the VGG Network: The VGG network is trained using the Adam optimization algorithm with a batch size of 128. They set the training epoch as 100. To prevent training progress, from stagnating they implement learning rate decay, which reduces the learning rate by a factor of 2. This training setup is applied for both classification and regression tasks utilizing output forms and loss functions

The article [20] discusses approaches and methods utilized in machine learning to identify exoplanets. Here is an explanation, for each of the mentioned methods:

1. Machine Learning: It's a branch of intelligence that applies theories and computer programming to construct models capable of making inferences based on data. In the context of exoplanet identification machine learning is used to classify objects of interest as either exoplanets or "false positives."
2. Machine Learning: Python is an open source object oriented programming language preferred by data scientists. Its popularity in the field of data science and machine learning stems from its readability, ease of learning and the availability of statistical and machine learning packages.
3. Data Scaling: Data scaling involves ensuring that all data features are presented on a scale. This step is crucial because features may have units or ranges potentially impacting the performance of machine learning models. Scaling techniques transform features to a scale, between zero and one to mitigate any potential bias.
4. Cross Validation: Cross validation is a technique employed to assess the effectiveness of machine learning models by simulating exposure, to data.
5. The dataset is split into subsets. The model is trained on one subset while evaluated on another. This process helps us assess how well the model performs and to identify if it becomes too specialized, to the training data, known as overfitting.
6. Feature Elimination: To create a model that can be applied to datasets we identify and remove features from the dataset that have limited predictive ability. In the case of exoplanet identification, we work with the KOI (Kepler Object of Interest) dataset, which contains features. Eliminating redundant features helps us build an effective model.

7. **Classification:** Classification is a type of supervised machine learning where observations are assigned a known class value based on their variables. In our study we perform classification categorizing objects of interest as either "FALSE POSITIVE" or "CONFIRMED" exoplanets.
8. **Machine Learning Model Pipeline:** A machine learning pipeline outlines the steps involved in transforming a candidate dataset into a machine learning model. This includes data engineering, model training testing against data saving the model for use and deployment.
9. **Application Programming Interfaces (APIs):** APIs serve as interfaces that enable communication and interoperability, between systems, models and code bases.

In the realm of machine learning models APIs enable users to access and interact with these models without having to rewrite them in programming languages.

Cloud computing, which involves providing computing resources and services on a service basis eliminates the need, for infrastructure. It encompasses Infrastructure as a Service (IaaS) Platform as a Service (PaaS) and Software as a Service (SaaS). In this study we utilize cloud computing services from Google Cloud Platform and Microsoft Azure to host the machine learning models and deploy web services.

Collectively these methods and techniques contribute to the development and implementation of machine learning models, for the identification of exoplanets.

The article [21] discusses techniques utilized by astronomers to search for exoplanets and determine their characteristics. Here are the key methods explained;

1. **Transit Method;** This method detects exoplanets by observing the decrease, in brightness of a star when a planet crosses in front of it (known as transit). Researchers analyze the changes in intensity over time also referred to as curves to calculate parameters such as transit depth, duration, ingress and egress times and the relative sizes of the planet and star.
2. **Velocity Method;** By analyzing the Doppler shift in the spectrum of the host star this method detects exoplanets. The gravitational pull of an exoplanet causes the star to exhibit a wobble resulting in shifts in the wavelengths of its light. These shifts can be measured to infer the presence of an exoplanet and estimate its mass.
3. **Gravitational Microlensing;** This technique relies on the lensing effect produced when light bends around objects like stars or planets. When an exoplanet passes in front of a star it can temporarily amplify the stars brightness making it appear brighter. The microlensing method is particularly sensitive, to detecting planets in cases of alignment.
4. **Direct Imaging;** This approach involves capturing pictures of exoplanets and the stars they orbit. It allows us to directly observe exoplanets and learn about their makeup and temperature. However direct imaging is challenging because exoplanets are often situated away and their faint light gets overshadowed by the brightness of the star.

These techniques have their strengths and limitations. The transit method tends to be more effective, in detecting exoplanets while the radial velocity method requires measurements and stable spectrographs.



Gravitational microlensing depends on alignments and imaging necessitates advanced instruments with exceptional contrast capabilities.

The article also mentions that the transit method has been particularly successful in discovering exoplanets, followed by the velocity method. The NASA Exoplanet Archive serves as a resource for accessing data on exoplanets. Confirmed discoveries, providing information, about their characteristics and host stars.

The Article [22] uses these mechanisms for the solution:

- **Regression**

In the context of generating wave (GW) waveforms using the PyCBC module the article discusses methods that were used to address computational challenges and expand the range of input parameters while maintaining waveform accuracy. Here are the different techniques mentioned in the article;

1. **Using Numerical Relativistic Equations;** The PyCBC module utilizes numerical equations to generate GW waveforms. However, these equations can result in overhead for specific input parameter values such, as celestial object masses.
2. **Parallelization;** To overcome the challenges of generating a dataset of waveforms parallelization techniques were employed by utilizing the Multiprocessing module in Python. This allowed utilization of all CPU cores to speed up the waveform generation process.
3. **Analyzing Scatter Plot and Regression;** To create a dataset peak amplitude during coalescence (SEOBNR) of bodies were recorded for different masses ( $m_1$ ,  $m_2$ ) and lower frequencies ( $f$ ). A scatter plot was then used to examine the relationship, between peak amplitude during coalescence and the masses of bodies. The scatter plot demonstrated a linear curve and a linear regression model was applied to analyze the data.

Note: - *PyCBC: An open-source Python module provides theoretical gravitational waveforms using Bayesian Belief Networks for various parameters and waveforms*

In this article they discuss the process of creating a waveform envelope by analyzing characteristics such, as amplitude, frequency and phase. They explain how they used curve fitting and approximation algorithms to calculate the coefficients ( $e_a$ ,  $e_b$ ,  $e_c$ ,  $x_a$  and  $x_b$ ) for models that represent the relationship between amplitude and time. These models were fitted to the amplitude peaks during phases of the waveform.

They also conducted regression analysis on the model parameters ( $e_a$ ,  $e_b$ ,  $e_c$ ,  $x_a$ ,  $x_b$ ) to observe how they change with the chirp mass ( $M_{ch}$ ) and frequency ( $f$ ). To visualize the relationships between the model parameters and input variables, scatter plots and 3D plots were used.

Additionally, the article mentions using curve fitting to identify the points in the waveform envelope that closely match the wave. By plotting the amplitude peaks in order against time they obtained a hyperbolic curve that was fitted to determine which points in the waveform envelope should be selected.

Note; The PyCBC module, an open-source Python tool is used to generate waveforms based on Bayesian Belief Networks, for different parameters and waveforms

- **Classification**

The article discusses how weak gravitational waves (GW) emitted by star planet systems; exoplanets are classified using the Random Forest algorithm. The main goal of this classification is, to validate models expand the use of GW concepts and address any inconsistencies in the regression analysis.

To accurately classify the GW data the approach involves distinguishing mass classes. The Random Forests algorithm, which is based on decision trees is chosen for its ability to establish boundaries between these classes. However, there are some inconsistencies in the dataset that need to be considered by the algorithm in order to improve accuracy. To address the imbalance in data over sampling techniques are used, the Synthetic Minority Over Sampling Technique (SMOTE) which generates data to balance the classes.

Once the data is balanced the Random Forests algorithm is applied. It uses a bagging-based routine with correlated decision trees to make predictions. This approach is well suited for classifying GW data as it thoroughly examines the features and makes judgments regarding mass classes.

To evaluate the efficiency of the model various classification performance metrics are employed. Metrics such as Positive Rate (TPR) True Negative

Rate (TNR) and Specificity provide insights, into how each class is handled and help interpret the data.

The article discusses the outcomes of classification before and, after incorporating SMOTE. While including SMOTE generated data decreases the accuracy of the model compared to the one trained each class individually. Despite a decrease in accuracy, the model's robustness is enhanced.

#### 4. RESULTS

The paper discusses the outcomes of studies that utilized machine learning algorithms and artificial intelligence techniques to determine whether observations are exoplanets or not. These studies aimed to examine transit properties and identify exoplanets using approaches; Random Forest Classifier (RFC) Support Vector Classifier (SVC) and Convolutional Neural Network (CNN). Each method had its strengths and limitations and the findings offered insights, into their effectiveness.

In one study both RFC and CNN achieved accuracy rates of around 90% in recognizing exoplanets from light curve data. RFC utilized derived features and external catalog information resulting in the identification of candidates displaying exoplanet characteristics. However, it encountered challenges when dealing with blended stars that shared descriptors with planets leading to some positive identifications. Conversely CNN directly accessed magnitude data folded on the period but sometimes misinterpreted eclipsing binaries or noisy data as transit signals resulting in false positives. Increasing the number of neurons in the CNN configuration helped alleviate this problem.

Another study examined the performance of machine learning models like regression and an Ensemble CNN model, based on evaluation parameters such as accuracy, precision, sensitivity (recall) and specificity. The logistic regression model didn't perform as the other baseline models but the proposed Ensemble CNN model showed relatively better performance. Figure 10 presents the performance values, which highlight the sensitivity and specificity of each model. These results indicate that the trained VGG

model is effective, for regression and classification tasks related to lensed waves (GWs). However, in some cases the model tends to overestimate or underestimate predictions in populated regions of true parameters. This suggests that there is room for improvement with models.

The model performed well when it came to categorizing types of spectrograms, such as lensed, unlensed and unknown. However, there were instances of misclassifications especially when dealing with lens models or when the beating patterns caused by lensing were not easily distinguishable. During the validation test using population models of GW progenitors the performance slightly decreased, indicating a sensitivity to the distribution of parameters used during training. Ensuring population models is essential for application.

In general machine learning algorithms have shown results in classifying exoplanets and analyzing transit properties. They offer automated and efficient methods of classification that outperform regression models. However, some challenges have been observed, including positives, misclassifications and sensitivity to parameter distributions. To ensure reliable exoplanet identification it is important to combine methods and expert opinions while considering factors, like crowding, previous labels and stellar parallaxes.

## 5. CONCLUSION

The paper discusses studies that explore the use of machine learning algorithms and AI techniques to classify observations as exoplanets or not. These studies employed methods, like Random Forest Classifier (RFC) Support Vector Classifier (SVC) and Convolutional Neural Network (CNN) to analyze transit properties and identify exoplanets. Each method had its strengths and limitations offering insights into their performance.

One particular study found that both RFC and CNN achieved 90% accuracy in identifying exoplanets based on light curve data. The RFC approach relied on derived features and external catalogue information, which resulted in exoplanet candidates. However, it faced challenges when dealing with stars leading to some positives. On the hand the CNN utilized magnitude data folded on the fitting period but encountered false positives due to misinterpretation of eclipsing binaries or noisy data as transit signals. Increasing the number of neurons in the CNN helped address this issue to some extent.

Another study evaluated machine learning models, including regression and an Ensemble CNN model using accuracy, precision, sensitivity (recall) and specificity as evaluation parameters. The performance of the regression model was relatively weaker compared to others while the Ensemble CNN model exhibited overall performance. In some cases, there were instances of overestimation or underestimation of predictions indicating room for improvement, with precise models.

The classification of types of spectrograms such as lensed, unlensed and unknown achieved accuracy with machine learning algorithms. However, there were instances of misclassifications especially when differentiating between lens models or when the lensing effects were less evident. Tests using population models of GW progenitors slightly lowered the performance highlighting the importance of population models for implementation.

In summary, machine learning algorithms showed promising results in classifying exoplanets and analyzing transit properties. They offered automated approaches that outperformed regression models. Nevertheless, challenges like positives, misclassifications and sensitivity to parameter variations were observed. To improve exoplanet identification, it is vital to combine methods and expert opinions taking into account factors like crowding, previous labels and stellar parallaxes.

## 6. REFERENCES

1. Pollacco D. L. et al., 2006, *PASP*, 118, 1407
2. Kovacs, G., Bakos, G., & Noyes, R. W., 2005; Tamuz, O. et al., 2005.
3. Collier Cameron A. et al., 2006, *MNRAS*, 373, 799
4. Krizhevsky, Sutskever & Hinton 2012; He et al. 2015; Ioffe & Szegedy 2015:
5. Cauliff et al. 2015; Mislis et al. 2016; Armstrong, Pollacco & Santerne 2017; E. R., Berta-Thompson Z. K., 2017, American Astronomical Society Meeting Abstracts #229, p. 415.01
6. Burrows A et al (2001) The theory of brown dwarfs and extrasolar giant planets. *Reviews of Modern Physics* 73:719–765. <https://doi.org/10.1103/RevModPhys.73.719>
7. Zapatero Osorio MR et al (2000) Discovery of young, isolated planetary mass objects in the  $\sigma$  orionis star cluster. *Science* 290:103–107. <https://doi.org/10.1126/science.290.5489.103>
8. Way MJ et al (2012) *Advances in Machine Learning and Data Mining for Astronomy* 1st ed. info publisher Chapman & Hall/CRC
9. Pollacco D et al (2006) The wasp project and the super wasp cameras. *Publ Astron Soc Pac* 118:1407–1418. <https://doi.org/10.1086/508556>
10. Jenkins JM et al (2010) Overview of the Kepler science processing pipeline. *Astrophys J Lett* 713:L87–L91. <https://doi.org/10.1088/2041-8205/713/2/187>
11. Chintarungruangchai P, Jiang I.-G. (2019) Detecting exoplanet transits through machine-learning techniques with convolutional neural networks. *Publ Astron Soc Pac* 131:064502. <https://doi.org/10.1088/1538-3873/ab13d3>
12. Coughlin JL et al (2016) Planetary candidates observed by Kepler. VII. the first fully uniform catalog based on the entire 48-month data set Q1–Q17 DR24. *Astrophys J Lett Supplement Series* 224:12. <https://doi.org/10.3847/0067-0049/224/1/12>
13. von Essen C et al (2018) Kepler Object of Interest Network I. First results combining ground and space-based observations of Kepler systems with transit timing variations. *Astronomy & Astrophysics* 615:1–16. <https://doi.org/10.1051/0004-6361/201732483>
14. Freudenthal J et al (2018) Kepler Object of Interest Network. II. Photodynamical modelling of Kepler-9 over 8 years of transit observations. *Astronomy & Astrophysics* 618:A41. <https://doi.org/10.1051/0004-6361/201833436>
15. Howell SB et al (2014) The K2 Mission: Characterization and early results. *Publ Astron Soc Pac* 126:398–408. <https://doi.org/10.1086/676406>
16. A convolutional neural network (CNN) based ensemble model for exoplanet detection(*Earth Science Informatics* (2021) 14:735–747)
17. Machine-learning approaches to exoplanet transit detection and candidate validation in wide-field ground-based surveys.( *MNRAS* 483, 5534–5547 (2019))
18. Surveying the reach and maturity of machine learning and artificial intelligence in astronomy(DOI: 10.1002/widm.1349)

19. Identification of Lensed Gravitational Waves with Deep Learning(The Astrophysical Journal, 915:119 (15pp), 2021 July 10)
20. Machine Learning Pipeline for Exoplanet Classification(Sturrock, George Clayton; Manry, Brychan; and Rafiqi, Sohail (2019) "Machine Learning Pipeline for Exoplanet Classification," SMU Data Science Review: Vol. 2: No. 1, Article 9)
21. Transiting Exoplanet Discovery Using Machine Learning Techniques: A Survey(Earth Science Informatics (2020) 13:573–600)
22. Machine Learning Based Analysis of Gravitational Waves(DOI:10.13140/RG.2.2.23478.80969)