

Bridging Supervised and Generative Learning for Exoplanet Discovery: The SS-GAN Model

Subhayan Mukherjee¹, Biplab K. Mandal², Babul P. Tewari^{3,*}

¹Department of Artificial Intelligence & Machine learning, Asansol Engineering College, kanyapur, Asansol, 713305, West Bengal, India

²Department of Information Technology, Asansol Engineering College, kanyapur, Asansol, 713305, West Bengal, India

³Department of Computer Science & Engineering, National Institute of Technology, Durgapur, 713209, West Bengal, India

Abstract

Exoplanets are a promising candidate for future space exploration projects. In this study, we present a semi-supervised Generative Adversarial Network (SS-GAN), designed to detect and characterize exoplanetary transits from the Kepler space telescope. The proposed approach combines dropout regularization, batch normalization, label smoothing, and a composite validation strategy to enhance the stability and generalization capability of classical GANs. Using a comprehensive set of machine learning baselines, ranging from traditional classifiers such as Logistic Regression, Random Forest, and Support Vector Machine (SVM), to deep learning architectures including MLPs and residual networks, the proposed approach achieved an accuracy of 98.7% while effectively utilizing both labeled and unlabeled samples. It demonstrated robustness across multiple evaluation metrics, including balanced accuracy, F1-score, MCC, and threshold-free measures such as AUROC and AUPRC. These results underscore the importance of exploring generative and semi-supervised architectures in astrophysics, paving the way for more efficient and reliable discovery of extra-terrestrial candidates in upcoming large-scale surveys.

Received on 28 January 2026; accepted on 20 July 2026; published on 09 October 2026

Keywords: Semi-supervised learning, Generative Adversarial Networks (GAN), Exoplanet Detection, Kepler Space Telescope, Astrophysical data analysis, Machine learning in Astronomy

Copyright © 2026 Subhayan Mukherjee *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [CC BY-NC-SA 4.0](#), which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi:10.4108/eetismmla.11741

1. Introduction

The Kepler space telescope which was launched on March 6, 2009, has transformed the way we investigate interstellar space [1]. Modern techniques for determining star brightness were first applied to the Kepler space telescope, which was the first of its kind to do so [2]. It also featured the largest digital camera available for space observations. Kepler identified planets by detecting transits—minute drops in a star's brightness. In astronomy, a transit occurs when a celestial body passes between an observer and a star. One specific type of transit happens when a

planet moves directly between a star and an observer on Earth. Exoplanets can be discovered using this method [3]. When a planet transits or eclipses its host star, it partially obstructs its light. When measured with powerful devices, the star's light curve will dim periodically whenever a planet passes by. According to [3], the likelihood of detecting a transit from Earth is low. The transit probability [4] can be represented by the following equation:

$$P_{transit} = \frac{R_{star} + R_{planet}}{a} \quad (1)$$

where, $P_{transit}$ represents the transit probability, and the radii of the star and planet are denoted by R_{star} and R_{planet} respectively, along with a denotes the semi-major axis of the planetary orbit.

*Corresponding author. Email: bptewari.cse@nitdgp.ac.in

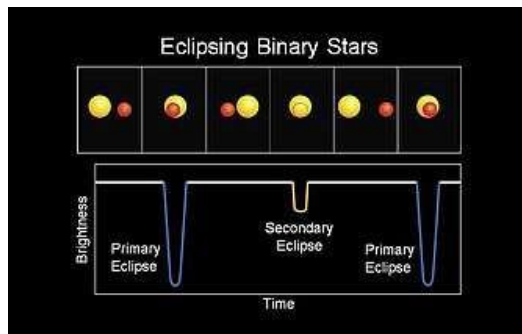


Figure 1. Change in the Luminosity of a star as a result of transiting

The planetary transit-induced reduction in starlight is depicted in Figure 1. The planetary transit-induced fading of starlight is important for multiple reasons. It is primarily useful for exoplanet detection [5] since the detectable brightness decrease signifies a planet transiting in front of its host star. Astronomers can verify the existence of planets and learn vital details about their size, orbital period, and separation from the star thanks to this phenomenon. Furthermore, transit observations shed light on the dynamics of planetary systems, assisting scientists in comprehending how planets interact with one another and the general structure of these systems. A portion of the planet's atmosphere is illuminated by starlight while in transit, making it possible to analyze the atmosphere and determine its composition, including any possible biosignatures. Large-scale surveys have also been made easier by the transit method, which has improved our knowledge of planetary formation throughout the galaxy and yielded the discovery of thousands of exoplanets. Ultimately, future missions that seek to investigate habitable worlds and evaluate the possibility of extraterrestrial life have been spurred by the success of transit observations.

The Kepler exoplanet dataset, one of the most extensive sets of astronomical data to date, is a significant accomplishment in the field of astrophysics. This dataset has significantly enhanced our understanding of the universe, serving as a cornerstone for groundbreaking discoveries about planetary systems beyond our own. The Kepler mission has uncovered thousands of exoplanets, showcasing a vast array of diverse planetary environments and features by meticulously observing over 150,000 stars for transits, where a planet moves across the face of its host star.

In addition to expanding our knowledge of the formation and evolution of planetary systems, the extensive data set offers unprecedented opportunities to apply machine learning algorithms. As researchers strive to analyze and interpret this vast and complex

information, advanced computational techniques have become indispensable. Using machine learning techniques, massive amounts of data can be processed quickly and effectively, revealing complex relationships and patterns that conventional analytical methods might miss.

The current work attempts to improve the algorithms' performance by performing an extensive estimation of every algorithm over predefined metrics. This systematic review aims to increase reliability by weighing the benefits and drawbacks of each exoplanetary feature definition algorithm.

Integrating these and other machine learning algorithms into exoplanetary research not only enriches our understanding of interstellar space but also holds the profound promise of identifying potentially habitable planets [6] capable of sustaining life.

1.1. Major Contributions

- This manuscript addresses a Semi-Supervised Generative Adversarial Network (SS-GAN) framework for exoplanet candidate classification using data from the Kepler mission. By leveraging both labeled and unlabeled observations during training, the proposed approach aims to improve classification performance while reducing dependence on extensive labeled datasets.
- To enhance training stability and model reliability, we integrate several optimization and regularization techniques, including label smoothing, dropout regularization, batch normalization, gradient clipping, and a composite metric-based model selection strategy. These components collectively contribute to a more robust learning process.
- We perform a comprehensive experimental evaluation of the proposed framework by comparing its performance with a diverse set of conventional machine learning and deep learning models. In addition, statistical confidence analysis, calibration assessment, and labeled-data ablation experiments are conducted to provide a thorough evaluation of model effectiveness and reliability.
- We demonstrate the capability of the proposed SS-GAN framework to achieve competitive predictive performance even under limited-label scenarios. The results indicate its potential applicability to future large-scale astronomical surveys, where obtaining fully labeled datasets can be challenging and resource-intensive.

2. Literature Review

Astrophysics [7], and machine learning have benefited greatly from the use of the Kepler dataset. It provides priceless insights into exoplanet discovery and analysis. Applying cutting-edge machine learning methods to this dataset has improved exoplanet identification efficiency and accuracy while also speeding up discoveries. Significant advances have resulted from this fusion of data science and astronomy, highlighting the transformative potential of interdisciplinary approaches. Utilizing the Kepler dataset effectively requires an understanding of key concepts like feature extraction, model training, validation techniques, and hyperparameter tuning. Important factors to take into account in this field include the difficulties of managing unequal class distributions, handling noisy data, and guaranteeing the interpretability of models.

The Kepler dataset has been extensively used in research studies [8, 9, 10, 11, 12, 13] over the decades. Hesar et al. [14] provided the first input in this field by describing the application of a Bayesian-based algorithm. The Gregory-Loredo method, which was employed to find pulsar star pairs, is the foundation of this model.

[15] suggested classifying exoplanets from the Kepler dataset using a deep learning framework. Their model CNN achieved competitive accuracy in identifying false positives from possible exoplanets. Based on the search of 4 years of Kepler time-series photometry, [16] presented a catalog of the Kepler objects of interest for transiting exoplanets. They conducted a thorough analysis of the catalog and came to the 50.5 % reliability conclusion based on established metrics.

The Kepler exoplanet dataset was analyzed by [17] for both the supervised and unsupervised learning classes of the algorithms. They employed various algorithms in the supervised domains, one of which was the Decision Tree. The accuracy of the DT classifier was 99.06

A Decision tree Classifier was proposed by [18], who discovered exoplanets using the Kepler data. An overall accuracy of roughly 96.5% was obtained using the DT classifier employed here. The cumulative of the L2-trans neptunian object observations between a specific time was presented by [19]. The study comprised sixty-six targets. Trans-Neptunian objects stand out for having higher light curve amplitudes at particularly large sizes, with an average detectability rate of about 56 % for a period of the light curve.

Using the Kepler dataset, [20] conducted a thorough analysis of numerous classification algorithms. The accuracy of the decision tree classifier they employed was 85.9%. Additionally, this classifier received a F score of 0.98 and a Recall Score of 0.97. The application of classical machine learning techniques for exoplanet detection was suggested by [21]. Using the

light curve data, their suggested model was created to identify planets that were transiting. Achieving an accuracy of 94.20 %, their final model was designed. The Kepler cumulative object of interest dataset was recommended for use by [13]. Along with a cat-boost classifier, they employed Random Forest and Recursive Feature Elimination. With an accuracy of 99.61 %, their experiments produced incredible results. For the Kepler dataset, [22] suggested using a 2D CNN and a 3D LSTM. Compared to the STIM algorithm, their models generated more true positives and fewer false positives.

While significant progress has been achieved in exoplanet detection using both traditional machine learning and deep learning approaches, several limitations remain. Many existing studies primarily focus on fully supervised learning paradigms that require large volumes of labeled data. Furthermore, limited attention has been given to exploiting unlabeled observations, evaluating model calibration, assessing statistical robustness, and investigating performance under label-scarce conditions. Table 1 presents a critical evaluation of representative studies and highlights the research gaps addressed in the present work.

3. Methodology

3.1. Motivation

The need to manage and interpret the massive volume of data gathered by the Kepler space telescope is the driving force behind the use of machine learning techniques in the Kepler dataset analysis. The size and complexity of the dataset make traditional exoplanet detection and analysis methods impractical. These methods typically involve manual inspection and simple statistical techniques. Millions of light curves have been collected in an enormous dataset by the Kepler mission, which is searching for planets similar to Earth in the habitable zones of distant stars. To find the faint brightness dips that might be exoplanets, these light curves need to be carefully analyzed.

Previous research in this area makes use of machine learning to improve and automate the detection process. Convolutional Neural Networks (CNNs), which have demonstrated impressive performance in image recognition tasks, were first used to identify patterns in the light curves that correspond to exoplanets by Shallue and Vandenburg [15]. By using this method, the rate of false positives is greatly decreased and detection accuracy is increased. Similarly, Kecskem et al. [19] address the difficulty of clustering similar exoplanet candidates without the need for labeled data, which is frequently lacking, by applying unsupervised learning techniques.

These methods demonstrate the continuous efforts to increase exoplanet detection accuracy, efficiency, and scalability with the Kepler dataset. The application

Table 1. Evaluation of representative studies on exoplanet classification

Reference	Methodology	Key Strengths	Limitations	Research Gap
Shallue et al. [15]	Convolutional Neural Network (CNN)	Demonstrated the effectiveness of deep learning for automated exoplanet detection.	Requires large labeled datasets and does not utilize unlabeled samples.	Need for data-efficient learning methods capable of exploiting unlabeled observations.
Thompson et al. [16]	Kepler Object of Interest Catalog Analysis	Provided a comprehensive catalog and reliability assessment of exoplanet candidates.	Focuses on catalog construction rather than predictive classification.	Need for robust automated classification frameworks.
Jin et al. [17]	Supervised and Unsupervised Learning Models	Comprehensive comparison of multiple learning paradigms.	Limited analysis of uncertainty, calibration, and label scarcity.	Need for robust evaluation beyond accuracy metrics.
Loganathan et al. [18]	Decision Tree Classifier	Simple and interpretable classification model.	Limited capability to capture complex nonlinear relationships.	Need for more expressive deep learning architectures.
Karim et al. [13]	Random Forest, RFE, CatBoost	Excellent predictive performance using ensemble methods.	Relies heavily on labeled data and feature engineering.	Need for methods that reduce dependence on extensive labeling.
Nath et al. [22]	2D CNN and 3D LSTM	Captures temporal and spatial patterns in Kepler observations.	Computationally intensive and fully supervised.	Need for efficient semi-supervised approaches.
Proposed Work	Semi-Supervised GAN (SS-GAN)	Utilizes both labeled and unlabeled data, incorporates stabilization techniques, calibration analysis, statistical validation, and labeled-data ablation studies.	Higher training complexity compared to conventional classifiers.	Addresses limited-label learning and improves model robustness for future astronomical surveys.

of cutting-edge machine learning techniques not only overcomes the drawbacks of conventional approaches but also creates new opportunities for exoplanet research discovery.

3.2. Data Preprocessing

Data preprocessing was a critical step to ensure model robustness and reliable training of both supervised and semi-supervised components. The Kepler dataset contains astrophysical attributes relevant to exoplanet detection, including: koi-period, koi-time0bk, koi-impact, koi-duration, koi-depth, koi-prad, koi-teq, koi-insol, koi-model-snr, koi-steff, koi-slogg, koi-srad, and koi-kepmag. The following steps were applied:

1. **Handling Missing Data:** Missing values were addressed using `IterativeImputer`, which estimates unknown entries through multivariate regression based on other available features. This method preserves statistical relationships between attributes more effectively than simple mean or mode imputation.
2. **Normalization:** All continuous features were standardized using the `StandardScaler`, transforming them to zero mean and unit variance. This ensured comparable feature scales, which is particularly important for both neural networks and distance-based classifiers such as SVM.
3. **feature Selection:** To reduce dimensionality and mitigate overfitting, Recursive Feature Elimination (RFE) with Logistic Regression was applied.

RFE iteratively eliminates the least informative attributes until the optimal feature subset is identified for classification tasks.

4. Supervised vs. Unsupervised Partitioning: Since the proposed model is semi-supervised, the dataset was divided into:

- Supervised subset: A portion of the data was retained with class labels (exoplanet vs. non-exoplanet). This subset was used for the supervised training component of the discriminator, which applies cross-entropy loss on known classes.
- Unsupervised subset: The remaining samples were treated as unlabeled. These examples contributed to the unsupervised training objective of the discriminator, where the model learns to separate real samples from generated ones without relying on explicit class labels.

This preprocessing pipeline ensured that the SS-GAN could leverage both labeled and unlabeled data effectively, aligning with the objective of improving classification performance in scenarios where labeled astrophysical datasets are limited.

To ensure fair evaluation and eliminate data leakage, the dataset was first divided into training, validation, and test partitions before any model training. All preprocessing operations, including feature scaling and class-weight estimation, were performed exclusively using the training data and subsequently applied to validation and test samples. Within the training partition, labeled and unlabeled subsets were created for semi-supervised learning experiments. The test set remained completely unseen throughout model development and hyperparameter tuning.

3.3. Proposed Semi-Supervised Generative Adversarial Network (SS-GAN)

In this work, we propose a robust Semi-Supervised Generative Adversarial Network (SS-GAN) framework tailored for classification tasks with limited labeled data and abundant unlabeled data. The proposed architecture enhances the stability and generalization capability of classical GANs by integrating dropout regularization, batch normalization, label smoothing, and a composite validation strategy.

Figure 2 illustrates the overall workflow of the proposed SS-GAN framework. The Kepler dataset is first subjected to preprocessing operations. The processed data are then partitioned into labeled and unlabeled subsets to support semi-supervised learning. Within the SS-GAN architecture, the generator receives random noise vectors sampled from a latent distribution

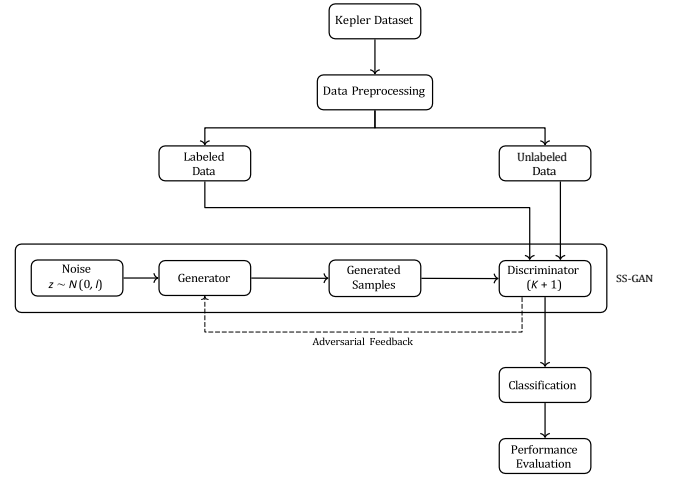


Figure 2. Workflow of the proposed Semi-Supervised Generative Adversarial Network (SS-GAN) framework for exoplanet classification

and produces synthetic samples, while the discriminator simultaneously learns from labeled data, unlabeled data, and generated samples. Through adversarial training, the generator progressively improves its ability to generate realistic samples, whereas the discriminator enhances its capability to distinguish between real and synthetic observations while performing classification. The final discriminator outputs are utilized for exoplanet candidate classification, and model performance is subsequently evaluated using multiple statistical and classification metrics.

Generator. The generator G maps a latent variable $z \sim \mathcal{N}(0, I)$ from the latent space $\mathbb{R}^{d_z}$ into the data space $\mathbb{R}^{d_x}$:

$$z \in \mathbb{R}^{d_z} \xrightarrow{G} \hat{x} \in \mathbb{R}^{d_x}. \quad (2)$$

Here, z denotes a random noise vector sampled from a Gaussian distribution, G represents the generator network, and $\hat{x}$ is the synthetic data sample. The architecture progressively increases hidden dimensions with linear transformations, LeakyReLU activation, batch normalization, and dropout. The final layer applies a tanh activation to ensure bounded outputs.

Discriminator. The discriminator D serves as a classifier over $K + 1$ classes, where the additional class corresponds to generated “fake” samples:

$$D : \mathbb{R}^{d_x} \rightarrow \mathbb{R}^{K+1}. \quad (3)$$

Here, K denotes the number of real classes, and the $+1$ th class is assigned to fake data. The discriminator is implemented as a deep feed-forward network with decreasing hidden dimensions, each followed by LeakyReLU, batch normalization, and dropout. The output layer applies a softmax function to produce class probabilities.

Loss Functions. The SS-GAN framework optimizes supervised and unsupervised objectives simultaneously.

Supervised loss. For labeled data (x, y) , we adopt a weighted cross-entropy with label smoothing:

$$\mathcal{L}_{\text{sup}} = -\frac{1}{N} \sum_{i=1}^N \sum_{k=1}^K w_{y_i} \tilde{y}_{ik} \log \hat{p}_{ik}. \quad (4)$$

Here, N is the batch size, w_{y_i} is the weight assigned to class y_i , $\tilde{y}_{ik}$ is the smoothed one-hot encoding of the true label, and $\hat{p}_{ik}$ denotes the predicted probability of class k . Label smoothing reduces overconfidence in predictions.

Unsupervised loss for real samples. For unlabeled real samples x_u , the discriminator minimizes the probability of misclassifying them as fake:

$$\mathcal{L}_{\text{unsup-real}} = -\mathbb{E}_{x_u} \log (1 - p_{\text{fake}}(x_u)). \quad (5)$$

Here, $p_{\text{fake}}(x_u)$ is the discriminator's predicted probability that x_u belongs to the fake class.

Unsupervised loss for fake samples. For generated samples $\hat{x} = G(z)$, the discriminator encourages assigning them to the fake class:

$$\mathcal{L}_{\text{unsup-fake}} = -\mathbb{E}_{z \sim p(z)} \log (p_{\text{fake}}(G(z))). \quad (6)$$

This term enforces the discriminator to correctly classify synthetic data as fake.

Generator loss. The generator is trained to fool the discriminator by minimizing:

$$\mathcal{L}_G = -\mathbb{E}_{z \sim p(z)} \log (1 - p_{\text{fake}}(G(z))). \quad (7)$$

Here, the generator aims to reduce the discriminator's confidence in labeling generated samples as fake.

Training Strategy. The discriminator's overall objective combines supervised and unsupervised terms:

$$\mathcal{L}_D = \mathcal{L}_{\text{sup}} + \frac{1}{2} \mathcal{L}_{\text{unsup-real}} + \frac{1}{2} \mathcal{L}_{\text{unsup-fake}}. \quad (8)$$

This formulation balances contributions from labeled and unlabeled data. Optimization is performed using the Adam optimizer with learning rates of 1×10^{-4} for D and 5×10^{-5} for G , with gradient clipping applied to enhance stability. Gaussian noise is injected into the inputs to improve robustness.

Validation and Early Stopping. To monitor generalization, we define a composite validation metric:

$$M_{\text{val}} = 0.4 \cdot \text{AUC} + 0.3 \cdot \text{F1} + 0.2 \cdot \text{MCC} + 0.1 \cdot \text{ACC}. \quad (9)$$

Here, AUC denotes the Area Under the ROC Curve, F1 is the harmonic mean of precision and recall, MCC is the Matthews correlation coefficient, and ACC is the classification accuracy. The best-performing model checkpoint is saved, and training stops early if no improvement is observed within a patience threshold.

Model Preservation. The framework supports reproducibility by saving both model parameters and pre-processing pipelines (imputation, scaling, and feature metadata). This ensures consistent deployment in real-world applications.

Hyperparameter Selection. Hyperparameters were selected through iterative experimentation on the validation set. Candidate configurations were evaluated using a composite validation metric consisting of AUROC, F1-score, MCC, and Balanced Accuracy. The final model was selected based on its overall validation performance and training stability.

Baseline Selection. The benchmark models were selected to represent diverse machine learning paradigms. Logistic Regression serves as a strong linear baseline, Random Forest captures nonlinear decision boundaries through ensemble learning, Support Vector Machines provide kernel-based classification, while MLP and Residual MLP represent deep neural network approaches. This diverse benchmark suite enables comprehensive assessment of the proposed SS-GAN framework.

4. Results and Discussion

4.1. Simulation Setup

Python 3.12 was used as the primary programming language for the experiments due to its extensive ecosystem of libraries for scientific computing and machine learning. The following key libraries and frameworks were employed:

- PyTorch: Used for building and training the semi-supervised GAN (SS-GAN) framework, including the generator and discriminator networks.
- Scikit-learn: Applied for implementing baseline machine learning models (Logistic Regression, Random Forest, and Support Vector Machine), evaluation metrics, and feature selection using Recursive Feature Elimination (RFE).
- Pandas and NumPy: Utilized for structured data manipulation, preprocessing, and numerical computations.

- Matplotlib: Used for visualization, including confusion matrices, ROC curves, and performance comparisons across models.

A comprehensive simulation environment was developed to evaluate traditional machine learning algorithms, deep learning baselines, and the proposed SS-GAN. The workflow consists of feature selection, data preprocessing, model training, cross-validation, and final testing on a held-out set. The overarching objective was to assess the ability of each model to detect exoplanets using Kepler light curve features.

The proposed Semi-Supervised GAN (SS-GAN) was evaluated comprehensively on the Kepler dataset using stratified splits. A variety of evaluation metrics were considered to assess classification accuracy, probabilistic calibration, and robustness. The results are summarized below.

The Kepler dataset exhibits moderate class imbalance, with confirmed exoplanets representing approximately 31% of the observations. To mitigate bias toward the majority class, class-weighted supervised loss functions were employed during discriminator training. Furthermore, evaluation relied on Balanced Accuracy, MCC, AUROC, and AUPRC in addition to standard accuracy.

4.2. Performance Metrics

The performance results obtained from the simulation have been listed in this section. Figure 3, 4, 5 demonstrates stable adversarial training as both networks achieve equilibrium. Loss curves. The top-left panel shows the evolution of validation metrics (AUC, F1-score, and MCC) all converging near 1.0, indicating excellent validation performance. The top-right panel decomposes the discriminator's supervised and unsupervised loss components, revealing consistent learning and stable adversarial behavior. The bottom-left panel displays generator loss fluctuations typical of adversarial training, while the bottom-right panel presents the combined validation score, demonstrating rapid convergence and stable high performance.

Figure 6 demonstrates promising agreement between predicted and true labels, with 1001 correctly identified False Positives and only 4 misclassifications.

Figure 7 demonstrates discriminative capability with an area under the curve (AUC) of 0.996, indicating high classification performance. The dashed line represents a random classifier for reference (AUC = 0.5)

Figure 8 exhibits that the SS-GAN model achieves an average precision (AP) of 0.995, significantly outperforming the baseline (AP = 0.314). indicating better classification accuracy and robustness

Figure 9 represents the residual analysis. Positive residuals (red) represent overestimated probabilities,

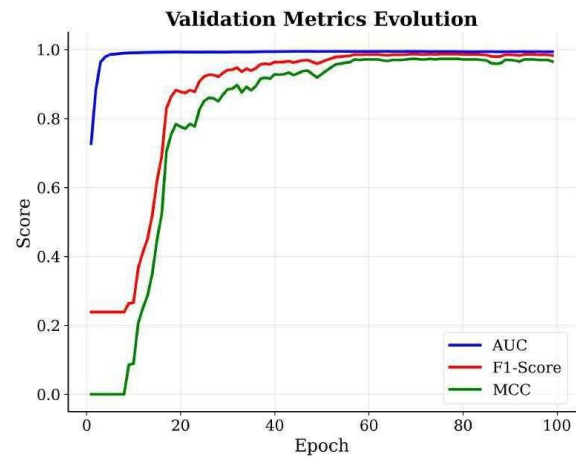


Figure 3. Validation dynamics of semi-supervised GAN model

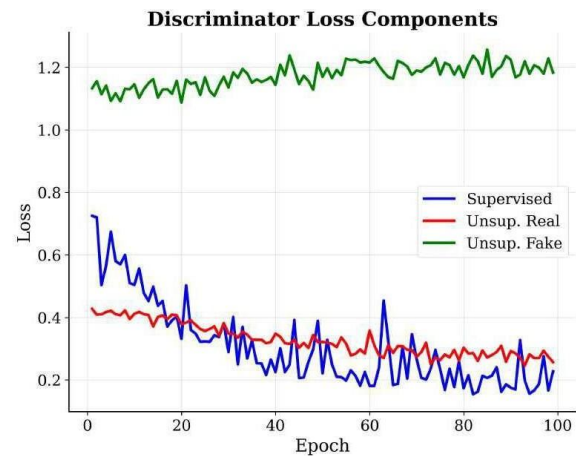


Figure 4. Discriminator Loss dynamics of semi-supervised GAN model

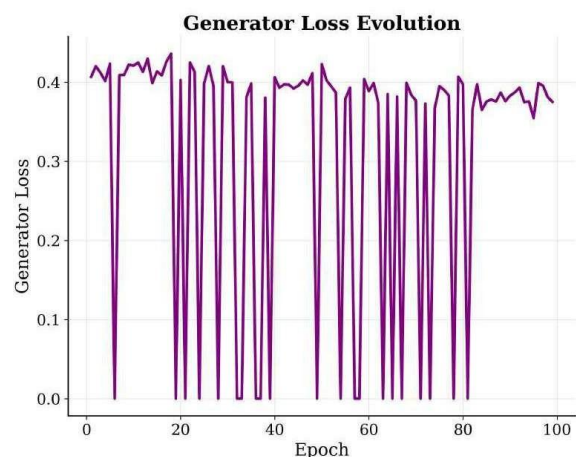


Figure 5. Generator Loss dynamics of semi-supervised GAN model

while negative residuals (green) correspond to underestimated probabilities. The dashed line at zero indicates

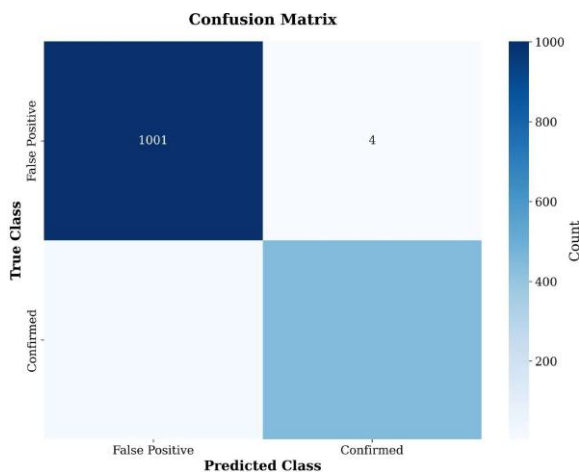


Figure 6. Confusion matrix of the trained classifier on the validation dataset

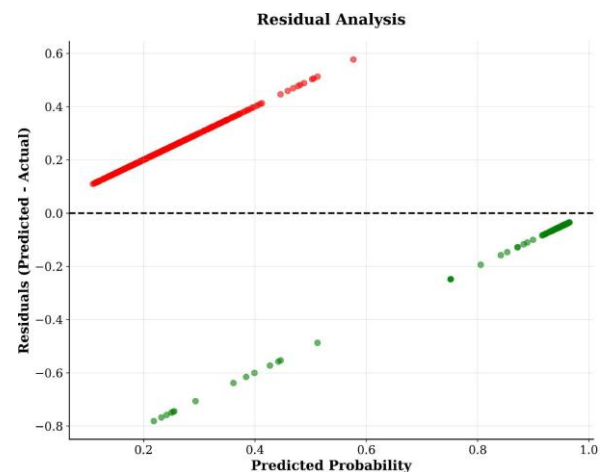


Figure 9. Residual analysis showing the relationship between predicted probabilities and residuals (predicted - actual)

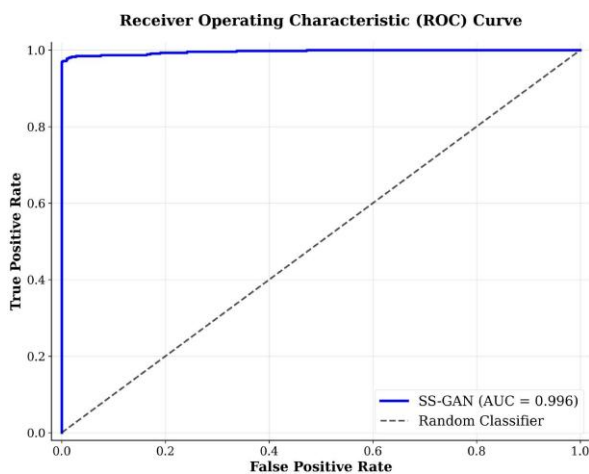


Figure 7. Receiver Operating Characteristic (ROC) curve of the proposed SS-GAN model

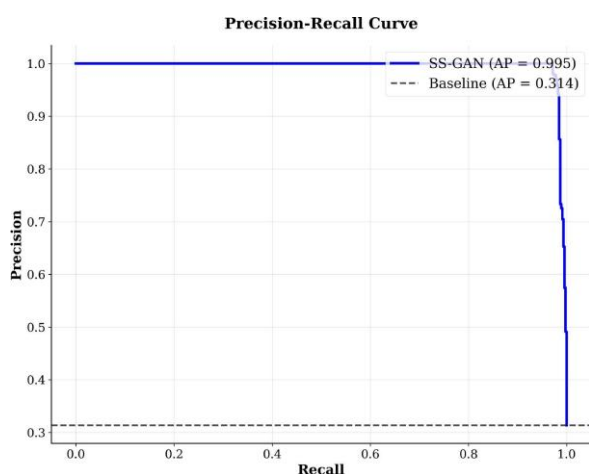


Figure 8. Precision-recall curves comparing the proposed SS-GAN model and the baseline model

perfect prediction. The pattern demonstrates overall good calibration, with minimal bias across most prediction ranges.

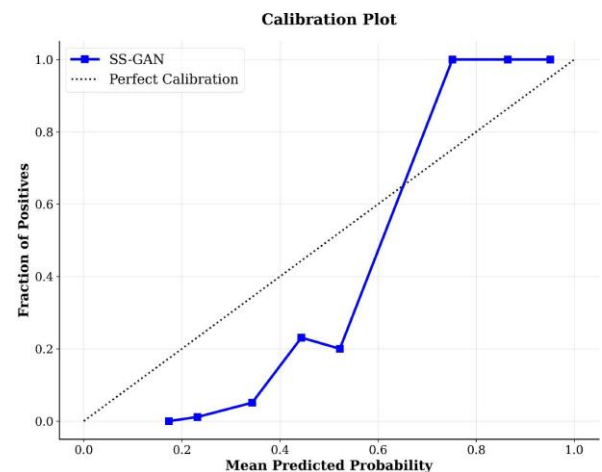


Figure 10. Calibration plot for the SS-GAN model

Figure 10 represents the calibration plot. The solid blue line shows the relationship between predicted probabilities and observed fraction of positives, while the dotted diagonal line represents perfect calibration. The SS-GAN model demonstrates good probability calibration, particularly at higher predicted confidence levels, indicating probability estimates for positive outcomes.

Table 2 reports the detailed evaluation metrics for SS-GAN, including accuracy, balanced accuracy, F1-score, Matthews Correlation Coefficient (MCC), and threshold-free measures such as AUROC and AUPRC. In addition, probabilistic metrics (log-loss, Brier score, Expected Calibration Error) and robustness measures (Cohen's Kappa, Jaccard index, sensitivity, specificity) were also computed.

Table 2. Comprehensive evaluation metrics of SS-GAN (single run)

Metric	Score
Accuracy	0.9884
Balanced Accuracy	0.9838
F1-Macro	0.9864
F1-Micro	0.9884
MCC	0.9730
AUROC	0.9962
AUPRC	0.9946
Log-Loss	0.2016
Brier Score	0.0412
ECE	0.1620
Cohen's Kappa	0.9729
Jaccard Index (Macro)	0.9733
Sensitivity	0.9717
Specificity	0.9960

Furthermore, confidence intervals for AUROC and AUPRC were computed using bootstrap resampling. AUROC was 0.9962 with 95% CI [0.9929, 0.9987], while AUPRC achieved 0.9946 with 95% CI [0.9903, 0.9980]. These values highlight the reliability of SS-GAN across multiple thresholds.

4.3. Effect of Labeled Data Availability

Table 3 presents the results of the labeled-data ablation study conducted to evaluate the effectiveness of the proposed SS-GAN under varying levels of supervision. The results indicate that the model maintains strong predictive performance even when only a small fraction of the training data is labeled. With only 5% labeled samples, the framework achieves an accuracy of 98.43%, an F1-Macro score of 98.15%, and an MCC of 96.37%, demonstrating its ability to learn meaningful representations from unlabeled observations. Although a gradual improvement is observed as the proportion of labeled data increases, the overall performance variation remains relatively small across different labeling fractions. This finding highlights the effectiveness of the semi-supervised learning paradigm in reducing dependence on extensive expert annotations. Consequently, the proposed SS-GAN framework is particularly well-suited for future large-scale astronomical surveys, where acquiring fully labeled datasets can be costly, time-consuming, and resource-intensive.

4.4. Statistical Validation

To assess the robustness and reproducibility of the proposed SS-GAN framework, a statistical validation study was conducted using 10 independent training runs with different random seeds. The resulting performance metrics were aggregated to evaluate the

stability of the model across repeated executions. The mean performance and corresponding standard deviations are presented in Table 4.

The statistical analysis demonstrates that the proposed SS-GAN exhibits highly consistent performance across repeated training runs. The average AUROC and AUPRC values of 0.9957 and 0.9944, respectively, are accompanied by extremely small standard deviations, indicating limited sensitivity to random initialization and training variability. Similarly, the F1-Macro, MCC, and Balanced Accuracy metrics remain stable across runs, further supporting the robustness of the framework.

The reported confidence intervals indicate low variability across repeated training runs, suggesting stable optimization behavior and reliable model convergence. In particular, the narrow confidence intervals observed for AUROC and AUPRC demonstrate that the discriminative capability of the model remains consistently high under different random seeds. These findings provide additional evidence that the proposed SS-GAN framework achieves reproducible performance and is suitable for practical deployment in exoplanet candidate classification tasks.

Furthermore, normality testing was performed to evaluate the distribution of performance metrics across repeated runs. The AUROC ($p = 0.1781$) and AUPRC ($p = 0.5387$) results satisfied the normality assumption, supporting the reliability of the computed confidence intervals. Overall, the statistical validation confirms that the proposed framework delivers stable and dependable performance beyond a single experimental realization.

4.5. Benchmark Comparison

To assess the competitiveness of SS-GAN, its performance was compared against five baseline models: Logistic Regression, Random Forest, SVM (RBF), Standard MLP, and Residual MLP. Table 5 presents the comparative evaluation. Although Logistic Regression and MLP achieve marginally higher predictive performance on certain metrics, these methods require fully labeled datasets for training. In contrast, the proposed SS-GAN leverages both labeled and unlabeled observations, achieving competitive performance while substantially reducing dependence on labeled data. This characteristic is particularly important in astronomical applications where obtaining expert annotations is expensive and time consuming.

4.6. Computational Complexity

To provide insight into the computational requirements of the proposed SS-GAN framework, Table 6 summarizes the number of trainable parameters associated with each network component.

Table 3. Labeled-data ablation study of the proposed SS-GAN framework

Labeled fraction	Accuracy	f1-Macro	MCC	AUROC	AUPRC
0.05	0.9843	0.9815	0.9637	0.9916	0.9905
0.10	0.9898	0.9880	0.9762	0.9961	0.9946
0.20	0.9898	0.9880	0.9762	0.9952	0.9940
0.50	0.9911	0.9896	0.9794	0.9960	0.9950
1.00	0.9891	0.9872	0.9746	0.9934	0.9929

Table 4. Statistical validation results over 10 independent runs

Metric	Mean $\pm$ Std	95% CI
Balanced Accuracy	0.9813 $\pm$ 0.0059	[0.9685, 0.9865]
F1-Macro	0.9816 $\pm$ 0.0113	[0.9573, 0.9899]
MCC	0.9636 $\pm$ 0.0219	[0.9166, 0.9799]
AUROC	0.9957 $\pm$ 0.0011	[0.9936, 0.9969]
AUPRC	0.9944 $\pm$ 0.0010	[0.9925, 0.9956]

Table 5. Performance comparison of SS-GAN against benchmark models

Model	Accuracy	Balanced Accuracy	f1-Macro	MCC	AUROC	AUPRC
Logistic Regression	0.9911	0.9858	0.9896	0.9794	0.9976	0.9965
Random Forest	0.9884	0.9815	0.9864	0.9731	0.9950	0.9950
SVM (RBF)	0.9904	0.9847	0.9888	0.9778	0.9959	0.9950
Standard MLP	0.9904	0.9847	0.9888	0.9778	0.9972	0.9961
Residual MLP	0.9911	0.9858	0.9896	0.9794	0.9974	0.9962
SS-GAN (Proposed)	0.9870	0.9833	0.9848	0.9698	0.9953	0.9940

Table 6. Model complexity of the proposed SS-GAN framework

Component	Parameters
Generator	1,233,193
Discriminator	736,259
Total	1,969,452

All experiments were conducted on an *NVIDIA GeForce RTX 3050* Laptop GPU with 4.3 GB of available memory. The proposed SS-GAN framework contains approximately 1.97 million trainable parameters, comprising 1.23 million parameters in the generator and 0.74 million parameters in the discriminator. Despite its adversarial architecture, the overall model size remains moderate and can be trained efficiently on consumer-grade GPU hardware.

The computational cost of the framework is primarily associated with the alternating optimization of the generator and discriminator networks during training. To mitigate excessive training time and prevent overfitting, an early-stopping strategy was employed. Across multiple experimental runs, convergence was typically achieved well before the maximum training

limit of 200 epochs. These characteristics demonstrate that the proposed framework offers a practical balance between predictive performance and computational efficiency, making it suitable for large-scale exoplanet classification tasks.

4.7. Discussion & Findings

The benchmark analysis demonstrates that both traditional and deep learning baselines, including Logistic Regression, Random Forest, SVM, and MLP variants, achieve strong performance with accuracies approaching 99%. While the proposed Semi-Supervised GAN (SS-GAN) does not strictly outperform these supervised models in absolute accuracy, it achieves comparable results while requiring fewer labeled instances. This ability to effectively incorporate unlabeled data offers a unique advantage in astrophysical research, where labeled exoplanet datasets are scarce and costly to curate.

Furthermore, the SS-GAN consistently maintained high AUROC (0.9953 ± 0.0013) and AUPRC (0.9940 ± 0.0012), underscoring its robustness in threshold-free classification. These findings confirm that the model preserves strong discriminative power, even under

data imbalance. The aggregate results across 10 runs also demonstrated stable performance with narrow confidence intervals, highlighting the reliability of SS-GAN in practical deployments.

The labeled-data ablation study provides strong evidence that the proposed framework derives meaningful benefit from unlabeled observations. Even when only 5% of the training data was labeled, the proposed framework achieved an accuracy of 98.43%, F1-Macro of 98.15%, and MCC of 96.37%, demonstrating strong utilization of unlabeled observations. These findings suggest that semi-supervised learning represents a promising direction for future exoplanet detection pipelines, particularly in scenarios where labeling resources are limited.

The comparative analysis highlights several important findings:

- Traditional models such as Logistic Regression and Random Forest offer competitive accuracy and interpretability, making them valuable for transparent classification tasks.
- Deep learning models such as MLP and residual architectures demonstrated excellent predictive power, but required substantial labeled data to achieve peak performance.
- Threshold-free metrics (AUROC, AUPRC) confirmed the strong discriminative capacity of SS-GAN, which is essential for detecting rare classes such as exoplanets in highly imbalanced datasets.

By bridging supervised and unsupervised learning paradigms, it not only ensures high predictive accuracy but also improves scalability in scenarios where labeled data is limited. These results underscore the importance of exploring generative and semi-supervised architectures in astrophysics, paving the way for more efficient and reliable discovery of exoplanetary candidates in upcoming large-scale surveys.

5. Conclusion & Future Works

In this work, we presented SS-GAN, a semi-supervised generative adversarial network framework for exoplanet classification using the Kepler dataset. The proposed approach was evaluated against a comprehensive suite of machine learning and deep learning baselines, including Logistic Regression, Random Forest, Support Vector Machines, Multi-Layer Perceptrons, and Residual MLPs. Experimental results demonstrated that SS-GAN achieves competitive predictive performance, attaining an accuracy of approximately 98.7% while maintaining strong Balanced Accuracy, F1-score, MCC, AUROC, and AUPRC values. Furthermore, labeled-data ablation experiments showed that the framework remains effective even when only a small fraction of

training samples are labeled, highlighting the practical advantages of semi-supervised learning for large-scale astronomical surveys where expert annotations are expensive and limited. These findings suggest that SS-GAN represents a promising approach for leveraging both labeled and unlabeled observations in exoplanet detection tasks.

Despite these encouraging results, several limitations should be acknowledged. The effectiveness of the proposed framework may be influenced by the underlying data distribution, particularly when substantial distribution shifts exist between labeled and unlabeled samples. In addition, the current model relies on tabular astrophysical parameters derived from Kepler observations and does not directly exploit the temporal structure of raw light-curve signals, potentially limiting its ability to capture subtle transit characteristics. The experimental evaluation was conducted exclusively on the Kepler dataset, and therefore further validation on datasets from missions such as TESS, K2, and PLATO is necessary to establish broader generalizability. Moreover, similar to many deep learning and adversarial learning approaches, the proposed SS-GAN functions primarily as a black-box model, making astrophysical interpretation of its predictions challenging. Although calibration analysis indicates reasonably reliable confidence estimates, the framework does not explicitly model predictive uncertainty, which may be particularly important when assessing borderline exoplanet candidates and supporting scientific decision-making.

Several avenues remain for future research. Cross-dataset validation on observations from additional space missions would provide a more comprehensive assessment of model robustness and transferability. A dedicated component-wise ablation study investigating the individual contributions of dropout, batch normalization, label smoothing, and composite metric-based model selection could further improve understanding of the framework's behavior and training stability. Additionally, while repeated-run evaluation, confidence intervals, and normality analyses were conducted, formal pairwise statistical significance testing against all benchmark models was beyond the scope of the present study and represents an important direction for future work. Future extensions may also incorporate temporal light-curve information through recurrent neural networks, transformer-based architectures, or hybrid deep learning models capable of capturing transit dynamics more effectively. Finally, integrating explainable artificial intelligence techniques and explicit uncertainty quantification mechanisms could enhance model interpretability and increase confidence in scientific discoveries derived from machine learning-based exoplanet detection systems.

6. Data Set Declaration

We have used data set from the source Kaggle (kaggle.com/datasets/nasa/kepler-exoplanet-search-results).

7. Author's Declaration

The authors declare that they have no conflicts of interest. Funding is also not applicable to this manuscript.

Acknowledgements

This class file was developed by Sunrise Setting Ltd, Torquay, Devon, UK. Website: www.sunrise-setting.co.uk

References

- [1] Gautier, T., Geary, J., Gilliland, R., Gould, A., Howell, S., Jenkins, J., et al. Kepler planet-detection mission: Introduction and first results. *Science*. 2010;327:977.
- [2] Koch, D.G., Borucki, W.J., Basri, G., Batalha, N.M., Brown, T.M., Caldwell, D., et al. Kepler mission design, realized photometric performance, and early science. *The Astrophysical Journal Letters*. 2010;713(2):L79.
- [3] Christiansen, J.L., Clarke, B.D., Burke, C.J., Seader, S., Jenkins, J.M., Twicken, J.D., et al. Measuring transit signal recovery in the kepler pipeline. ii. detection efficiency as calculated in one year of data. *The Astrophysical Journal*. 2015;810(2):95.
- [4] Kempton, E.M.R., Knutson, H.A. Transiting exoplanet atmospheres in the era of jwst. *Reviews in Mineralogy and Geochemistry*. 2024;90(1):411–464.
- [5] Bhaumik, S., Sethi, G. Studying habitability of the exoplanets kepler-504 b, kepler-315 b, and kepler-315 c. *arXiv preprint*. 2024;arXiv:2401.06269.
- [6] Kopparapu, R.K., Ramirez, R., Kasting, J.F., Eymet, V., Robinson, T.D., Mahadevan, S., et al. Habitable zones around main-sequence stars: new estimates. *The Astrophysical Journal*. 2013;765(2):131.
- [7] Hansen, M.T., Dittmann, J.A. Single transit detection in kepler with machine learning and onboard spacecraft diagnostics. *arXiv preprint*. 2024;arXiv:2403.03427.
- [8] Kaushik, M., Mattoo, A., Rastogi, R. Exoplanet detection: A detailed analysis. *arXiv preprint*. 2024;arXiv:2404.09143.
- [9] Nasios, I. Analyze mass spectrometry data with artificial intelligence to assist the understanding of past habitability of mars and provide insights for future missions. *Icarus*. 2024;408:115824.
- [10] Airlangga, G. Exoplanet classification through machine learning: A comparative analysis of algorithms using kepler data. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*. 2024;4(3):753–763.
- [11] Donnelly, S., Dutta, A. Kepler light curve classification using deep learning and markov transition field (student abstract). *Proceedings of the AAAI Conference on Artificial Intelligence*. 2024;38:23475–23476.
- [12] Dobrycheva, D., Vasylenko, M.Y., Kulyk, I., Pavlenko, Y.V., Shubina, O., Luk'yanyk, I., et al. Hunting for exocomet transits in the tess database using the random forest method. *arXiv preprint*. 2024;arXiv:2402.19075.
- [13] Karim, A., Uddin, J., Riyad, M.M.H. Identifying important features for exoplanet detection: A machine learning approach. *GPH-International Journal of Applied Science*. 2024;7(01):01–17.
- [14] Hesar, F.F., Foing, B. Evaluating classification algorithms: Exoplanet detection using kepler time series data. *arXiv preprint*. 2024;arXiv:2402.15874.
- [15] Shallue, C.J., Vanderburg, A. Identifying exoplanets with deep learning: A five-planet resonant chain around kepler-80 and an eighth planet around kepler-90. *The Astronomical Journal*. 2018;155(2):94.
- [16] Thompson, S.E., Coughlin, J.L., Hoffman, K., Mullally, F., Christiansen, J.L., Burke, C.J., et al. Planetary candidates observed by kepler. viii. a fully automated catalog with measured completeness and reliability based on data release 25. *The Astrophysical Journal Supplement Series*. 2018;235(2):38.
- [17] Jin, Y., et al. Identifying exoplanets with machine learning methods: A preliminary study. *International Journal on Cybernetics G Informatics (IJCI)*. 2022;11(1/2):31–42.
- [18] Loganathan, S. Exoplanet detection with decision trees. *Journal of Astrophysical Research*. 2023;doi:10.58445/rars.526.
- [19] Kecskeméthy, V., Kiss, C., Szakáts, R., Pál, A., Szabó, G.M., Molnár, L., et al. Light curves of trans-neptunian objects from the k2 mission of the kepler space telescope. *The Astrophysical Journal Supplement Series*. 2023;264(1):18.
- [20] Hesar, F.F., Foing, B. Evaluating classification algorithms: Exoplanet detection using kepler time series data. *arXiv*. 2024.
- [21] Nandan, A., Tripathi, V., Singh, D., Gupta, A. Detection of exoplanets using classical machine learning algorithms. *AIP Conference Proceedings*. 2024;2481(1):020032.
- [22] Nath-Ranga, R., Absil, O., Christiaens, V., Garvin, E.O. Machine learning for exoplanet detection in high-contrast spectroscopy combining cross correlation maps and deep learning on medium-resolution integral-field spectra. *arXiv preprint*. 2024;arXiv:2405.13468.