Revista Ciencia Aplicada
Conocimiento Multidisciplinario ISSN: 3122-3664

Artículo Científico

Descubrimiento de exoplanetas mediante aprendizaje automático

Sandro Garcia-Palomino[1], Gustavo Rodriguez-Gomez[1]

[1]Instituto Nacional de Astrofísica, Óptica y Electrónica, Computer Science Department, Santa María Tonantzintla, Puebla 72840, Mexico

Resumen: La detección de exoplanetas se ha convertido en un desafío esencial para la astrofísica contemporánea. Métodos tradicionales como el tránsito o la velocidad radial han permitido grandes avances, pero presentan limitaciones frente al ruido, la complejidad de los datos y la necesidad de análisis manual. En este artículo se presenta una revisión crítica de los enfoques actuales basados en aprendizaje automático, destacando el papel de arquitecturas como las redes convolucionales, los modelos basados en Transformer y las técnicas generativas. Se analizan sus aplicaciones en curvas de luz, espectros e imágenes astronómicas, así como sus ventajas y limitaciones. Además, se identifican brechas en la literatura, como la falta de interpretabilidad, la escasa generalización entre misiones y la subutilización de datos multimodales. Finalmente, se plantea el desarrollo de un modelo multimodal de aprendizaje profundo como propuesta para integrar distintas fuentes de datos astronómicos en un sistema único, robusto y escalable. Esta visión busca avanzar hacia una detección de exoplanetas más precisa, automatizada y científicamente confiable.

Palabras clave: exoplanetas, modelo multimodal, redes neuronales, redes transformer, aprendizaje automático.

1. Introducción

La búsqueda de exoplanetas —planetas ubicados fuera del sistema solar— representa una de las líneas actuales de investigación más activas en la astrofísica moderna. La posibilidad de identificar mundos similares a la Tierra impulsa no solo el conocimiento científico, sino también el desarrollo de nuevas tecnologías. Misiones como *Kepler* y *TESS* han permitido la recolección de millones de curvas de luz con el objetivo de detectar estos cuerpos celestes mediante el método de tránsito, que identifica pequeñas caídas periódicas en el brillo de una estrella debido al paso de un planeta frente a ella [1-2].

No obstante, este método presenta importantes desafíos: los datos contienen ruido, eventos astrofísicos pueden imitar tránsitos, y el análisis manual de los eventos detectados es lento y propenso a errores [1,3]. Para resolver estas dificultades, se han propuesto enfoques automáticos basados en aprendizaje automático (ML, por sus siglas en inglés) y aprendizaje profundo (DL), como redes neuronales convolucionales (CNN) [1], algoritmos tradicionales como el *gradient boosting* [2] y visión por computadora aplicada a espectros estelares [4].

Volumen 1, no. 2

Recibido: julio 11, 2025
Aceptado: agosto 29, 2025

autor de correspondencia:
grodrig@inaoep.mx

https://doi.org/10.66482/e0fmpq33

Recientemente, se ha introducido el uso de arquitecturas tipo *Transformer*, capaces de capturar relaciones temporales en secuencias sin la necesidad de apilar múltiples capas, lo que mejora la eficiencia y ayuda a la interpretabilidad gracias a los mapas de atención [1]. Por otro lado, también se ha explorado el uso de modelos simples como la regresión logística aplicada a imágenes astronómicas, lo cual ha mostrado mejoras notables en la detección de planetas poco brillantes sin incrementar la tasa de falsos positivos [5].

Una de las principales limitaciones ha sido la falta de generalización y la dependencia de representaciones específicas, como el "folding" de curvas de luz y la escasa integración de datos multimodales como imágenes, espectros y series temporales [2-3]. En este sentido, el desarrollo de bases de datos estandarizadas como *The Multimodal Universe* [6] abre la posibilidad de trabajar con datos de diferentes modalidades en un mismo entorno, lo que facilita el entrenamiento de modelos más robustos y generalizables.

Este trabajo plantea una revisión general de los enfoques actuales en la detección de exoplanetas mediante aprendizaje automático, identifica las brechas persistentes en la literatura, y argumenta cómo nuevas estrategias basadas en modelos interpretables pueden ofrecer una mejora sustancial en la localización y clasificación automática de exoplanetas.

2. Métodos de detección de exoplanetas

La detección de exoplanetas se basa en métodos tanto directos como indirectos, cada uno con ventajas y limitaciones propias. Los métodos indirectos, como el tránsito y la velocidad radial (*radial velocity*, RV), han sido responsables de la mayoría de los descubrimientos confirmados hasta la fecha [1]. Por otro lado, métodos directos, como la imagen de alto contraste, permiten caracterizar propiedades físicas de los planetas, aunque son técnicamente más exigentes.

2.1 Método de tránsito

El método de tránsito se basa en detectar pequeñas disminuciones periódicas en el brillo de una estrella cuando un planeta pasa frente a ella desde nuestra perspectiva. La Figura 1 muestra un ejemplo de esta técnica, presentando la curva de luz observada para la estrella KIC 6922244, donde se aprecian caídas de flujo que sugieren posibles tránsitos planetarios. (esta investigación hizo uso de *Lightkurve*, una librería de Python para análisis de datos de *Kepler* y *TESS Lightkurve Collaboration*, 2018). Esta caída de flujo es proporcional al área del planeta respecto al área de la estrella [1]. Para facilitar el análisis y destacar las señales débiles, el flujo se somete a un proceso de normalización. Una vez normalizada, como se observa en la Figura 2 para la misma estrella, KIC 6922244, los tránsitos se distinguen con mayor claridad de las variaciones estelares de fondo. La técnica requiere fotometría de alta precisión y un alineamiento orbital favorable. Su

principal ventaja es que permite estimar el radio del planeta y, si se combina con velocidad radial, su densidad.

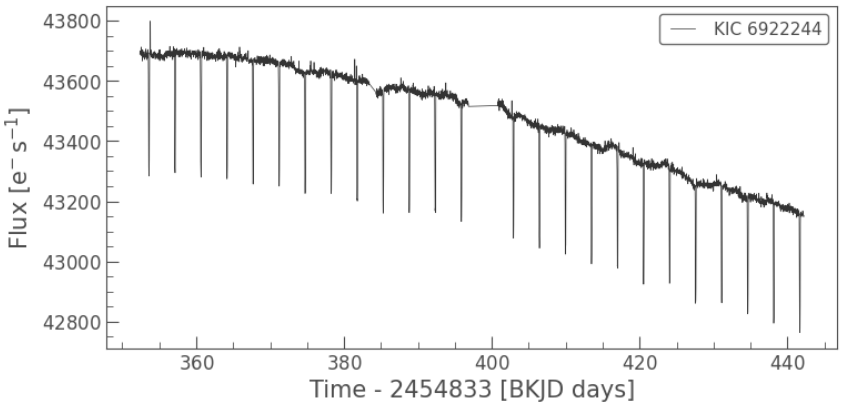

Figura 1. Curva de luz observada para la estrella KIC 6922244 sin normalización. Se aprecian caídas de flujo periódicas que podrían corresponder a tránsitos planetarios.

Este método ha sido potenciado por el uso de inteligencia artificial. Malik et al. [2] aplicaron el algoritmo LightGBM sobre curvas de luz procesadas con TSFRESH, logrando métricas comparables a las de redes profundas con menor costo computacional. Prasad et al. [3], por su parte, utilizaron el algoritmo *Box Least Squares* (BLS) como paso previo a la clasificación de curvas de luz mediante redes neuronales.

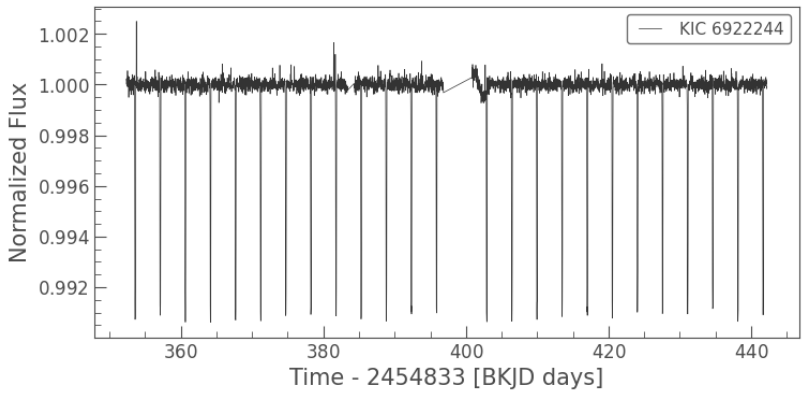

Figura 2. Curva de luz normalizada de la estrella KIC 6922244, donde se destacan con mayor claridad los posibles tránsitos.

2.2 Velocidad radial

El método Velocidad Radial (RV) mide el desplazamiento Doppler de las líneas espectrales de una estrella inducido por el tirón gravitacional de un planeta orbitante [1]. Es particularmente efectivo para detectar planetas masivos cercanos a la estrella y no depende del alineamiento orbital como el tránsito.

Mignon et al. [4] realizaron un análisis homogéneo de 200 enanas M usando el espectrógrafo HARPS, confirmando su sensibilidad para detectar planetas de baja masa en zonas habitables. En paralelo, Gan y Rajpaul [5] propusieron un enfoque novedoso basado en visión por computadora para extraer señales de RV directamente de espectros, evitando la dependencia del método clásico de correlación cruzada. Este método logró resultados comparables a los estándares de la industria en menor tiempo de cómputo.

2.3 Imagen directa (Direct Imaging)

La imagen directa busca observar al planeta separando su débil luz de la brillante emisión estelar. Este método es fundamental para caracterizar atmósferas y órbitas, pero presenta grandes retos debido al contraste extremo de brillo entre la estrella y el planeta [1]. Se requieren técnicas ópticas avanzadas como óptica adaptativa, así como procesamiento intensivo de imágenes.

Daglayan et al. [6] propusieron el algoritmo *AMAT*, que mejora las técnicas clásicas de sustracción de la función de dispersión puntual (*PSF*) utilizando un enfoque iterativo basado en minimización L1/L2 y modelos de baja complejidad. En la misma línea, Cambazard et al. [7] aplicaron regresión logística sobre imágenes preprocesadas, logrando una detección más eficiente de señales planetarias débiles sin aumentar los falsos positivos. Flasseur et al. [8] integraron datos multiespectrales con *CNN*, superando algoritmos clásicos al aprovechar la diversidad espectral y espacial de los datos.

2.4 Enfoques complementarios y emergentes

El uso de datos sintéticos generados por redes generativas representa una innovación relevante. El estudio *AstroFusion* de Suresh et al. [9] mostró que modelos entrenados con datos sintéticos alcanzan precisión comparable a modelos entrenados con datos reales. Al combinar ambos conjuntos, se logró mejorar la tasa de aciertos en hasta un 96 % para *Random Forests*.

Finalmente, Angeloudi et al. [10] propusieron *The Multimodal Universe*, un conjunto de datos de más de 100 TB que integra imágenes, espectros y series temporales en un formato estandarizado y escalable para aprendizaje automático. Esta iniciativa permite trabajar con datos heterogéneos de múltiples observatorios y es clave para el desarrollo de modelos más robustos y generalizables.

3. Descubrimiento de exoplanetas con aprendizaje automático

El uso de aprendizaje automático ha dado lugar a avances significativos en la detección de exoplanetas, especialmente al abordar las limitaciones de los métodos tradicionales. A continuación, se presentan algunos de los enfoques más relevantes y representativos.

3.1 Modelo PANOPTICON: detección sin filtrado previo

PANOPTICON, una arquitectura basada en U-Net, permite detectar tránsitos planetarios únicos en curvas de luz sin necesidad de filtrado previo [7]. Entrenado con datos simulados de la misión PLATO, el modelo alcanzó una tasa de acierto del 90 % incluso en señales de bajo contraste, demostrando una notable capacidad para reconocer eventos difíciles de identificar con métodos clásicos.

3.2 MAC-Net: clasificación multimodal de objetos celestes

MAC-Net es una red neuronal que combina espectros bidimensionales e imágenes del SDSS para clasificar estrellas, galaxias y cuásares [8]. Esta integración de entradas multimodales permitió alcanzar una precisión del 98.6 %, superando enfoques basados exclusivamente en espectros unidimensionales.

3.3 Modelos con GANs: generación de datos sintéticos

AstroFusion aplica redes generativas adversarias (GANs) para crear datos sintéticos de tránsitos, lo que permite entrenar modelos con más variabilidad y robustez [9]. Al mezclar datos simulados y reales, se logró aumentar la precisión de modelos como Random Forest sin comprometer la confiabilidad de las detecciones.

3.4 Transformers para señales de tránsito

El uso de modelos Transformer ha demostrado gran potencial para la clasificación de curvas de luz, permitiendo distinguir entre eventos reales y falsos positivos sin necesidad de preprocesamiento complejo ni extracción manual de características [1].

3.5 LightGBM con extracción automática de características

Una alternativa más eficiente al uso de redes profundas es el enfoque propuesto por Malik et al., que aplica LightGBM sobre características generadas automáticamente mediante TSFRESH [2]. Esta estrategia ofrece resultados competitivos con un menor costo computacional y una mayor interpretabilidad.

4. Ventajas y limitaciones de los enfoques actuales

El aprendizaje automático ha demostrado ser una herramienta poderosa en la detección de exoplanetas, ofreciendo avances significativos en eficiencia, precisión y adaptabilidad. La Figura 3, por ejemplo, ilustra el rendimiento de un modelo típico, mostrando la precisión alcanzada en los conjuntos de entrenamiento y validación. En el contexto de las redes neuronales, el eje *X* de esta figura representa la época, que es un recorrido completo del conjunto de datos de entrenamiento a través del algoritmo de aprendizaje. No obstante, estos modelos también presentan limitaciones técnicas y científicas que es importante considerar.

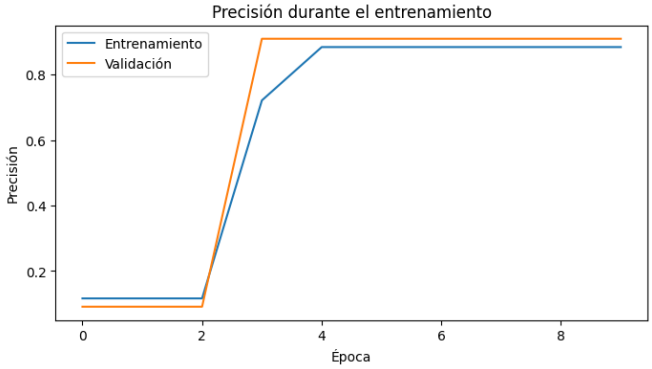

Figura 3. Precisión en el conjunto de entrenamiento y validación durante el entrenamiento del modelo. Se observa una rápida convergencia tras la tercera época.

4.1 Ventajas principales

Las principales ventajas de la aplicación de los enfoques de aprendizaje automático en la detección de exoplanetas radican en su capacidad para manejar la complejidad y el volumen de los datos astronómicos. Estos beneficios se manifiestan en la mejora de la velocidad de procesamiento, la sensibilidad de detección y la robustez del análisis, superando varias de las limitaciones inherentes a los métodos tradicionales detallados a continuación:

- Escalabilidad y velocidad de inferencia. Modelos como PANOPTICON pueden analizar curvas de luz completas en menos de 0.2 segundos, lo que permite procesar grandes volúmenes de datos de manera eficiente [7].
- Detección de señales débiles y eventos únicos. A diferencia de métodos clásicos que dependen de periodicidad o filtrado previo, PANOPTICON permite detectar tránsitos únicos, incluyendo aquellos largos o poco profundos, con alta sensibilidad y baja tasa de falsos positivos [7].
- Fusión de datos multimodales. Redes como MAC-Net combinan espectros bidimensionales e imágenes para mejorar la clasificación de objetos, superando enfoques tradicionales basados en entradas unidimensionales [8].
- Robustez frente al ruido. Gracias a arquitecturas como U-Net, los modelos actuales pueden extraer características en presencia de ruido estelar, rayos cósmicos y variaciones de fondo [7].
- Uso de datos sintéticos. El modelo AstroFusion demostró que la generación de datos sintéticos con GANs puede complementar eficazmente los conjuntos de datos reales, mejorando la tasa de aciertos sin comprometer la precisión [9].

4.2 Limitaciones y desafíos actuales

A pesar de los avances que ofrece el aprendizaje automático en la astrofísica, su aplicación práctica se enfrenta a importantes obstáculos. Las limitaciones actuales

se centran principalmente en aspectos como la fiabilidad, la dependencia de los datos, la robustez entre diferentes fuentes de información y la accesibilidad, planteando desafíos que la investigación en el área debe superar:

- Falta de interpretabilidad. Muchos modelos de redes profundas funcionan como cajas negras, lo que dificulta su validación científica.
- Dependencia de datos bien etiquetados. Modelos como los basados en CNN o LightGBM requieren conjuntos de datos etiquetados, balanceados y representativos [2].
- Sensibilidad a la calidad de entrada. Aunque MAC-Net y otros modelos multimodales han demostrado rendimiento alto, su precisión depende en gran medida de la calidad y homogeneidad de los datos de entrada [8].
- Generalización entre misiones. Modelos entrenados con datos de misiones específicas como TESS, LAMOST o PLATO pueden no generalizar bien a otros instrumentos sin reentrenamiento, debido a diferencias en resolución, ruido o formato [1].
- Costo computacional en entrenamiento. Aunque la inferencia es rápida, el entrenamiento de estos modelos profundos requiere recursos computacionales significativos y tiempos prolongados, como se observó en PANOPTICON [7].

5. Oportunidades y desafíos futuros

El uso de aprendizaje automático para la detección de exoplanetas aún se encuentra en expansión, con múltiples áreas en desarrollo que ofrecen oportunidades concretas para mejorar los métodos actuales, pero también plantean desafíos técnicos y metodológicos como los que a continuación se mencionan.

5.1 Aprovechamiento de datos multimodales

La astronomía moderna genera datos en múltiples formatos: curvas de luz, espectros, imágenes multibanda y catálogos. Iniciativas como The Multimodal Universe están abriendo el camino hacia modelos capaces de aprender simultáneamente de distintas fuentes, integrando patrones temporales, espectrales y espaciales [6]. El reto está en diseñar arquitecturas que aprovechen esta diversidad sin volverse inestables ni excesivamente costosas. Modelos multimodales también deben resolver cómo sincronizar diferentes resoluciones temporales y espaciales sin perder información crítica [10].

5.2 Estándares para entrenamiento y validación

Aún no existe un consenso sobre benchmarks estándar ni sobre la forma óptima de validar modelos de detección de exoplanetas. Muchos trabajos usan datasets simulados o propios, lo que dificulta la comparación directa entre enfoques. La creación de bases de datos públicas, bien etiquetadas y representativas es crucial para establecer una base común que permita evaluar avances de forma transparente y reproducible. En este contexto, algunos estudios proponen combinar datos reales

con datos sintéticos generados por GANs o simuladores físicos para aumentar la robustez de los modelos y mejorar su capacidad de generalización [12].

5.3 Interpretabilidad y confianza en los resultados

Uno de los principales obstáculos para una adopción más amplia en la comunidad científica es la falta de interpretabilidad. Modelos complejos como *Transformers* o *GANs* pueden ofrecer alta precisión, pero sus decisiones no siempre son comprensibles para los astrónomos. Se requieren herramientas de interpretación y visualización que permitan entender cómo el modelo llegó a una conclusión, especialmente en detecciones sin precedentes. Propuestas recientes sugieren el uso de explicabilidad post-hoc mediante técnicas como LIME o SHAP para apoyar la interpretación de las decisiones del modelo [10].

5.4 Transferencia entre misiones e instrumentos

La capacidad de adaptar modelos a nuevos instrumentos (como el telescopio J*ames Webb*, TESS, o futuras misiones como ARIEL) es esencial. Hoy, muchos modelos tienen que ser reentrenados desde cero con los datos de cada misión, lo cual es ineficiente. Una oportunidad está en aplicar técnicas de *transfer learning* o entrenamiento contrastivo para facilitar la reutilización de modelos previamente entrenados. Además, se ha señalado que la calibración de modelos en dominios con distintas características estadísticas, por ejemplo en señales espectroscópicas multivariadas, puede beneficiarse del uso de reducción de dimensionalidad antes de aplicar técnicas de detección de anomalías [11].

6. Conclusiones

El aprendizaje automático ha demostrado ser una herramienta útil y cada vez más robusta para enfrentar los retos en la detección de exoplanetas, desde la clasificación de tránsitos hasta el análisis de señales espectrales e imágenes multibanda. Modelos como PANOPTICON, MAC-Net o AstroFusion evidencian que es posible aumentar la sensibilidad y precisión al integrar arquitecturas modernas y estrategias de fusión de datos.

En este contexto, la presente investigación se vincula directamente con estas tendencias, al centrarse en el diseño y la validación de un modelo multimodal de aprendizaje profundo para la detección de exoplanetas. Este enfoque busca integrar múltiples fuentes de datos astronómicos en una sola arquitectura, maximizando la capacidad de detección y reduciendo los falsos positivos. Con ello, se espera contribuir a un campo que evoluciona rápidamente, pero que aún requiere soluciones más generales, interpretables y adaptables a nuevas misiones espaciales.

Referencias

1. H. Salinas, K. Pichara, R. Brahm, et al., "Distinguishing a planetary transit from false positives: a Transformer-based classification for planetary transit signals," *Monthly Notices of the Royal Astronomical Society,* vol. 522, pp. 3201–3216, 2023. https://doi.org/10.1093/mnras/stad1173
2. A. Malik, B. P. Moster, and C. Obermeier, "Exoplanet detection using machine learning," *Monthly Notices of the Royal Astronomical Society*, vol. 513, pp. 5505–5516, 2022. https://doi.org/10.1093/mnras/stab3692
3. M. S. Prasad, S. Verma, and Y. A. Shichkina, "Astronomical image processing: Exoplanet detection," *IEEE International Conference on Soft Computing and Measurements (SCM)*, St. Petersburg, Russia, 2023, pp. 336–339. https://doi.org/10.1109/SCM58628.2023.10159069
4. K. Gan and V. M. Rajpaul, "A computer vision approach to radial velocity extraction for exoplanet detection," *IEEE Undergraduate Research Technology Conference (URTC)*, Cambridge, MA, USA, 2023, pp. 1–5. https://doi.org/10.1109/URTC60662.2023.10534937
5. H. Cambazard, N. Catusse, A. Chomez, et al., Lagrange, "Logistic regression to boost exoplanet detection performances," *Monthly Notices of the Royal Astronomical Society*, vol. 536, no. 2, pp. 1610–1624, 2025. https://doi.org/10.1093/mnras/stae2657
6. E. Angeloudi, J. Audenaert, M. Bowles, et al., "The Multimodal Universe: Enabling Large-Scale Machine Learning with 100 TB of Astronomical Scientific Data," *NeurIPS 2024 Datasets and Benchmarks Track*, 2024. *arXiv preprint arXiv:2412.02527.*
7. H. G. Vivien, L. Garcia, F. Xavier, et al., "PANOPTICON: detection of single transit events in PLATO light curves," Astronomy & Astrophysics, vol. 685, pp. A32, 2025. https://doi.org/10.1051/0004-6361/202452124
8. M. Zhang, J. Wu, Z. Zhang y Z. Liu, "MAC-Net: A multimodal celestial object classification network based on 2D spectrum," Research Notes of the AAS: RASTAI, vol. 7, no. 4, pp. 120–128, 2023. https://doi.org/10.1093/rasti/rzad026
9. A. Suresh, L. C, G. PV, et al., "AstroFusion: A GAN-Augmented Approach for Exoplanet Detection," *2024 International Conference on Emerging Techniques in Computational Intelligence (ICETCI)*, Hyderabad, India, 2024, pp. 330-337, doi: 10.1109/ICETCI62771.2024.10704142
10. U. A. Usmani, I. Abdul Aziz, J. Jaafar, et al., "Deep Learning for Anomaly Detection in Time-Series Data: An Analysis of Techniques, Review of Applications, and Guidelines for Future Research," in *IEEE Access*, vol. 12, pp. 174564-174590, 2024, doi: 10.1109/ACCESS.2024.3495819
11. Altin, M., and Cakir, A. (2024). Exploring the influence of dimensionality reduction on anomaly detection performance in multivariate time series. IEEE Access, 12, 85783-85794
12. Cuéllar S, Granados P, Fabregas E, et al. (2022) Deep learning exoplanets detection by combining real and synthetic data. PLOS ONE 17(5): e0268199. https://doi.org/10.1371/journal.pone.0268199