PlanetNet-MMG: A Robust Multi-Modal Graph-Based Deep Learning Model for Exoplanet Candidate Classification

Nishant Pravin Kumar Dubey[a], Lalatendu Behera[a], Ranjeet Kumar Rout[b], Saiyed Umer[c], Deepak Kumar Jain[d], Javier Andreu-Perez[e]

Corresponding author: Deepak Kumar Jain, dkj@dlut.edu.cn

[a]*Department of Computer Science and Engineering, Dr B.R. Ambedkar National Institute of Technology, Jalandhar, Punjab, India, ndubey1245@gmail.com, beheral@nitj.ac.in*
[b]*Department of Information Technology, Dr B.R. Ambedkar National Institute of Technology, Jalandhar, Punjab, India, routrk@nitj.ac.in*
[c]*Department of Computer Science and Engineering, Aliah University, Kolkata, India, saiyed.umer@aliah.ac.in*
[d]*Key Laboratory of Intelligent Control and Optimization for Industrial Equipment of Ministry of Education, Dalian University of Technology, Dalian, 116024, China, dkj@dlut.edu.cn*
[e]*Centre for Computational Intelligence, School of Computer Science and Electronic Engineering, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, j.andreu-perez@essex.ac.uk*

Abstract

The search for exoplanets has advanced into the era of intelligent automation, yet most deep learning pipelines remain constrained to single-modality inputs or isolated views of astronomical data. We present **PlanetNet-MMG**, a novel multi-modal deep learning architecture that combines structured stellar metadata, raw lightcurve sequences, and graph-based relational context into a unified classification model. Our approach fuses three powerful encoders: a Tabular Transformer for domain-aware feature projection, a PatchGRU enhanced with a Vision Transformer (ViT) for learning fine-grained temporal patterns in segmented lightcurve patches, and a Graph PARE encoder that models inter-object similarity via a relational graph. Trained on a harmonized dataset derived from Kepler, TESS, and confirmed exoplanet archives, PlanetNet-MMG outperforms all state-of-the-art baselines, achieving a peak test accuracy of **90.4%** and a class-averaged AUC of **0.973**. Extensive experiments across 10–100 epochs and comparative evaluations against Astronet, ExoNet, OsbornNet, GCN (Lu), and ExoMiner confirm the effectiveness of our multimodal fusion. We further provide interpretability through attention overlays, t-SNE projections, and confidence histograms, reinforcing PlanetNet-MMG's transparency and reliability for scientific discovery in astrophysics.

Keywords: Exoplanet Detection, Multi-Modal Learning, Tabular Transformer, PatchGRU, Vision Transformer, Graph Neural Network, Lightcurve Analysis, Planetary Classification

1. Introduction

The search for exoplanets—planets orbiting stars beyond our solar system—stands as one of the most transformative frontiers in modern astrophysics. Space missions such as NASA's Kepler [1] and TESS [2] generate unprecedented volumes of photometric time-series data, offering vast opportunities to identify transit signatures indicative of planetary bodies. These transit events, typically observed as periodic dips in stellar brightness, are often subtle and easily masked by stellar variability, instrumental noise, or astrophysical false positives. Traditional detection pipelines rely on statistical algorithms such as the Box Least Squares (BLS) method or rule-based

classifiers [3]. While these approaches succeed in controlled scenarios, they show limited capability in distinguishing genuine planetary transits from noise, especially in cases of irregular, incomplete, or low signal-to-noise lightcurves [4]. This limitation motivates the shift toward more advanced computational techniques.

The advent of machine learning—and particularly deep learning—revolutionizes exoplanet detection by enabling models to learn complex transit morphologies directly from data [5, 6, 7]. Convolutional neural networks (CNNs), recurrent neural networks (RNNs), and hybrid architectures achieve remarkable accuracy in classifying exoplanet candidates from large-scale surveys [8, 9]. A landmark study by Shallue and Vanderburg [5] introduces Astronet, a CNN trained on Kepler data to automate classification of transit signals. Astronet's success is expanded by ExoNet [10], which incorporates auxiliary metadata features to enhance classification robustness. Osborn et al. [10] extend this approach by including centroid time-series data, capturing centroid shifts that often accompany astrophysical false positives.

Further developments explore recurrent architectures such as LSTMs and GRUs [8, 11] to better capture temporal dependencies in lightcurve sequences, as well as autoencoder-based methods to denoise flux variations [6, 12]. Recent advances incorporate Vision Transformers for temporal image analysis [13] and transformer-based algorithms for full-frame image analysis [14]. However, these models remain largely unimodal or rely on shallow fusion techniques, lacking a coherent strategy to integrate metadata, temporal dynamics, and contextual relationships among stellar objects. Graph-based approaches emerge as a promising solution for modeling structured relationships in stellar catalogs. Lu et al. [15] demonstrate the potential of Graph Convolutional Networks (GCNs) for exoplanet classification by constructing graphs based on astrophysical feature similarity. While effective at capturing local neighborhood context, such methods often lack the temporal resolution required to model detailed transit signatures. Consequently, classification performance suffers, particularly for borderline candidates where subtle differences separate true exoplanets from false positives.

Machine learning applications in exoplanet science have expanded beyond transit detection to include atmospheric retrieval [16], mass estimation [17], high-contrast imaging [18, 19], astrometric detection [20], and microlensing analysis [21]. Recent comprehensive reviews [22, 23] highlight the growing sophistication of machine learning approaches, including hybrid CNN-Random Forest methods [24] and citizen science integration [25]. Novel approaches using synthetic light curves for training [26] and Earth-like planet prediction [27] demonstrate the field's evolution toward more specialized and robust methodologies.

To address these limitations, a robust multimodal approach is required—one capable of integrating structured stellar and planetary parameters, detailed temporal dynamics from lightcurves, and contextual information from the broader astrophysical environment. This study introduces **PlanetNet-MMG**, a unified deep learning framework that combines three complementary modalities into a single, end-to-end model. Structured astrophysical parameters are processed using a *Tabular Transformer* [28], segmented lightcurve sequences are modeled with a custom *PatchGRU-ViT* encoder that captures both local and global temporal dependencies using GRU [11] and Vision Transformer [29] architectures, and relational context is incorporated via a *Graph PARE encoder* that models

proximity and similarity in the astrophysical feature space using Graph Convolutional Networks [30]. By harmonizing data from Kepler, TESS, and the NASA Exoplanet Archive [31], PlanetNet-MMG standardizes and fuses heterogeneous datasets to enable comprehensive reasoning across modalities. The proposed architecture consistently outperforms unimodal baselines such as Astronet [5], ExoNet [10], and even graph-only approaches [30], delivering enhanced robustness, higher classification accuracy, and improved interpretability. Model behavior is further elucidated through visualization techniques such as attention heatmaps, embedding projections, and prediction confidence distributions, providing transparency for scientific validation [32, 33]. The theoretical foundations of multimodal learning [34, 35] and multimodal fusion methods [36, 37] support our architectural choices, while recent advances in explainable AI for multimodal systems [38] inform our interpretability approaches. Ultimately, PlanetNet-MMG represents a step forward in the design of intelligent, multimodal systems for astronomical discovery—scalable to upcoming survey missions such as PLATO and the Roman Space Telescope, and adaptable to the evolving needs of exoplanetary science. The contributions of this work are as follows:

1. The work presents the **PlanetNet-MMG**, where the first exoplanet classification framework that explicitly integrates tabular stellar parameters, temporal lightcurve dynamics, and relational graph-based reasoning within a single multimodal pipeline.
2. The work demonstrates how the fusion of Kepler, TESS, and confirmed exoplanet datasets enables standardized, cross-mission learning, thereby improving generalization across diverse observational environments.
3. The work provides a comprehensive evaluation against established baselines such as Astronet, ExoNet, and GCN-based models, highlighting substantial gains in accuracy, robustness, and interpretability. Together, these contributions advance the state of the art in exoplanet detection and establish a scalable foundation for future astronomical surveys.

The organization of this work is as follows: Section 3 demonstrates the detailed discussion of the proposed methodology. The extensive experimentation with results are discussed in Section 4 and 5. The work is concluded in Section 6.

2. Related Work

The application of machine learning to exoplanet detection has gained significant momentum with the increasing availability of large-scale photometric datasets from missions such as Kepler and TESS. Early deep learning approaches primarily focused on convolutional neural networks (CNNs) trained on phase-folded lightcurves. A seminal contribution in this direction is Astronet, which demonstrated that CNNs can effectively learn transit morphologies directly from Kepler data and automate candidate vetting at scale [39]. Subsequent extensions, including ExoNet, incorporated auxiliary stellar metadata to enhance robustness and reduce false positives [6]. OsbornNet further improved classification reliability by integrating centroid motion time-series, addressing contamination from background eclipsing binaries [40]. Beyond convolutional architectures, recurrent neural networks such

as LSTMs and GRUs have been explored to capture temporal dependencies in raw or detrended lightcurve sequences [11]. These sequential models offer improved sensitivity to transit duration and ingress–egress patterns but often struggle to model long-range temporal dependencies effectively. More recent studies have introduced transformer-based architectures, including Vision Transformers, to capture global contextual relationships in time-series or image-like representations of lightcurves [29, 41]. While these models achieve promising results, they are typically limited to a single modality and rely on shallow fusion when additional metadata is incorporated.

Graph-based learning has emerged as a complementary paradigm for exoplanet classification by explicitly modeling relationships among astrophysical candidates. Lu et al. proposed a graph convolutional network (GCN) that constructs similarity graphs based on stellar and planetary parameters, demonstrating that relational context can improve classification performance [42, 15]. However, existing graph-centric approaches often omit detailed temporal modeling of lightcurves, limiting their ability to resolve subtle transit signatures and borderline cases.

Multimodal learning frameworks in astrophysics have been explored in a limited capacity, typically combining tabular metadata with photometric features without modeling relational context or long-range temporal structure [4, 7]. PlanetNet-MMG addresses these gaps by jointly encoding structured stellar parameters, raw lightcurve sequences, and graph-based relational embeddings within a single end-to-end architecture, directly targeting the key limitations identified in prior unimodal and shallow-fusion approaches.

3. Methodology

This section presents the complete methodology of the PlanetNet-MMG architecture, including dataset integration, preprocessing pipeline, and the detailed model architecture with its constituent components.

3.1. Dataset Integration and Preprocessing

A critical foundation for any machine learning framework lies in the quality, consistency, and comprehensiveness of its input data. For the PlanetNet-MMG architecture, we construct a harmonized dataset by integrating three major exoplanet catalogs: the **Kepler KOI** list, the **TESS TOI** catalog, and the **NASA Exoplanet Archive Confirmed Planets** [31]. Each of these sources provides complementary perspectives: Kepler offers long-baseline, high-precision photometry [1]; TESS contributes all-sky coverage with shorter cadence [2]; and the NASA Exoplanet Archive consolidates vetted planetary parameters from diverse detection methods [31].

The integration process is far from trivial. Astrophysical parameters such as orbital period, planetary radius, stellar effective temperature, stellar mass, transit depth, transit duration, signal-to-noise ratio (SNR), and impact parameter differ in scale, units, and sometimes even definition across catalogs [43]. To ensure compatibility, we standardize all features to a common reference frame, with units converted where necessary. We remove missing values in essential parameters, while imputing secondary fields using column-wise medians to preserve statistical distribution without introducing bias. We harmonize categorical labels into three primary classes: `CONFIRMED`, `CANDIDATE`, and `FALSE POSITIVE`.

where x_i is the raw feature value, μ is the mean, and σ is the standard deviation for that feature across the dataset. This transformation ensures that all features have zero mean and unit variance, preventing scale disparities from dominating the learning process. The z-score normalization is particularly crucial in our exoplanet detection context because astrophysical parameters span vastly different orders of magnitude. For instance, orbital periods range from hours to thousands of days, while transit depths are measured in parts per million (ppm). Without proper normalization, features with larger numerical scales (such as stellar temperature in Kelvin) would disproportionately influence the learning process compared to smaller-scale features (such as normalized transit depth). By standardizing each feature through z_i, we ensure that the multimodal fusion layers can effectively combine tabular metadata, temporal lightcurve patterns, and graph-based relationships without being biased toward any particular feature's numerical range. This normalization step is essential for the Tabular Transformer component to learn meaningful feature interactions and for the Graph PARE encoder to construct accurate k-nearest neighbor relationships in the standardized feature space.

The resulting dataset comprises **14,707 samples** and **13 carefully curated features**, each selected for its astrophysical relevance to planetary habitability and detection confidence [44, 45]. These features are summarized in Table 1.

Table 1: Selected astrophysical features for tabular input

Feature	**Description**	**Unit**
Period	Orbital period	days
Planet Radius	Planetary radius	Earth radii
Stellar Temperature	Host star effective temperature	Kelvin
Stellar Mass	Host star mass	Solar masses
Transit Depth	Relative flux decrease	ppm
Transit Duration	Transit event duration	hours
SNR	Signal-to-noise ratio	-
Impact Parameter	Transit impact parameter	-
Eccentricity	Orbital eccentricity	-
Semi-major Axis	Orbital semi-major axis	AU
Equilibrium Temperature	Planetary equilibrium temperature	Kelvin
Insolation Flux	Incident stellar flux	Earth flux
Disposition Score	Vetting disposition confidence	-

Several of these features warrant deeper examination. The **orbital period** P—the time taken for a planet to complete one revolution around its star—is fundamental to transit prediction and is related to the semi-major axis a by Kepler's third law (Equation (1)) as shown below:

$$P^2 = \frac{4\pi^2}{GM_\star} a^3 \tag{1}$$

where G is the gravitational constant and $M_\star$ is the stellar mass. The **planetary radius** R_p, typically measured in Earth radii, is inferred from the fractional drop (Equation (2)) in stellar flux during a transit:

$$\frac{\Delta F}{F} \approx \left(\frac{R_p}{R_\star}\right)^2 \tag{2}$$

where $\Delta F/F$ is the fractional flux decrease and $R_\star$ is the stellar radius. Transit depth scales quadratically with the planet-to-star radius ratio. This metric directly influences habitability assessments, as smaller planets are more challenging to detect but are more likely to be terrestrial.

Stellar effective temperature T_{eff}, measured in Kelvin, governs the incident radiation environment of the planet. Together with the **stellar mass** $M_\star$, it informs estimates of the habitable zone via the **insolation flux** S (Equation (3)):

$$S = \frac{L_\star}{4\pi a^2} \tag{3}$$

where $L_\star$ is the stellar luminosity, itself proportional to $R_\star^2 T_{\text{eff}}^4$ according to the Stefan–Boltzmann law.

The **transit depth** and **transit duration** provide shape descriptors of the observed lightcurve. Transit depth, expressed in parts per million (ppm), reflects the planet-to-star size ratio, while transit duration relates to the orbital geometry, impact parameter b, and orbital velocity. The **impact parameter** (Equation (4))is defined as:

$$b = \frac{a\cos i}{R_\star} \tag{4}$$

where i is the orbital inclination. This parameter is crucial for distinguishing between grazing and central transits, which can differ significantly in classification confidence.

Finally, the **signal-to-noise ratio** (SNR) acts as a statistical measure of transit detectability, while the **disposition score** represents a vetting-derived confidence metric from prior pipeline analyses.

By integrating, standardizing, and mathematically contextualizing these features, the dataset becomes a coherent, astrophysically grounded input space for the PlanetNet-MMG model. This preprocessing stage ensures that downstream learning is driven not only by statistical patterns but also by physically meaningful relationships embedded in the data.

3.2. Lightcurve Extraction and Processing

Once the tabular dataset is standardized and harmonized, the next critical step is to obtain and process the raw photometric observations from which planetary transit signatures could be extracted. For this study, we focused on a subset of **5,000 exoplanet candidates** with valid Kepler IDs and confirmed availability of high-quality lightcurve data.

The lightcurves are sourced using the **Lightkurve** Python package [46], which provides a direct interface to NASA's Mikulski Archive for Space Telescopes (MAST). This package simplifies retrieval, calibration, and manipulation of mission-specific photometric data, allowing seamless access to *Pre-search Data Conditioning Simple Aperture Photometry* (PDCSAP_FLUX) values [3]. The PDCSAP flux is a detrended and decontaminated lightcurve product that removes most instrumental and systematic effects while preserving astrophysical variability. This makes it an ideal choice for machine learning pipelines aiming to detect planetary transits without being misled by non-planetary systematics [12].

Let a raw lightcurve be represented as a discrete flux sequence

$$\mathscr{F} = \{F_1, F_2, \ldots, F_T\},$$

where F_t denotes the observed PDCSAP flux at time step t, and T is the total number of observations for a given target.

To ensure comparability across samples, the following preprocessing steps are applied:

1. **Normalization:** Each lightcurve is normalized (Equation (5)) such that its median flux value equals unity:

$$F_t^{\text{norm}} = \frac{F_t}{\text{median}(\mathscr{F})}, \quad \forall t \in \{1, 2, \ldots, T\}. \tag{5}$$

This removes absolute brightness differences across stars, ensuring the model learns from *relative* variations.

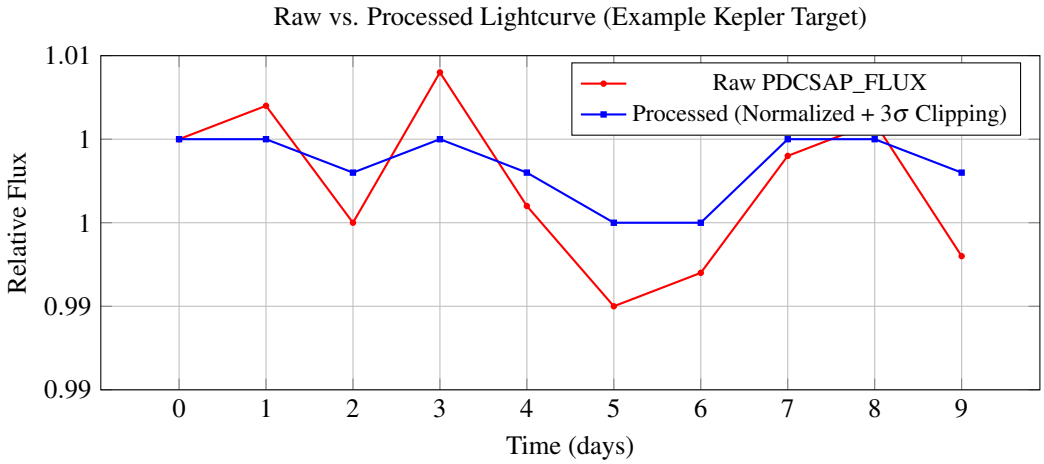


Figure 1: Example of raw and processed lightcurve. The processing pipeline removes outliers and normalizes flux values while preserving the transit signature.

2. **Outlier Removal:** To eliminate anomalies such as cosmic rays and data dropouts, a 3σ (Equation (6)) clipping is applied as shown below:

$$\mathscr{F}_{\text{clean}} = \{F_t^{\text{norm}} \mid |F_t^{\text{norm}} - \mu| \leq 3\sigma\}, \tag{6}$$

where

$$\mu = \frac{1}{T}\sum_{t=1}^{T} F_t^{\text{norm}}, \qquad \sigma = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(F_t^{\text{norm}} - \mu\right)^2}.$$

3. **Sequence Length Standardization:** Since neural models require a uniform input length, only the first $L = 100$ points are retained:

$$\mathscr{S} = \{F_1^{\text{clean}}, F_2^{\text{clean}}, \ldots, F_L^{\text{clean}}\}.$$

If $|\mathscr{F}_{\text{clean}}| < L$, the sequence is padded with the median value 1.0:

$$\mathscr{S} = \{F_1^{\text{clean}}, \ldots, F_M^{\text{clean}}, 1.0, \ldots, 1.0\}, \quad M < L.$$

Thus, each lightcurve is transformed into a standardized vector

$$\mathbf{s} \in \mathbb{R}^{100}, \qquad \mathbf{s} = \big[s_1, s_2, \ldots, s_{100}\big],$$

where s_t encodes the cleaned, normalized, and padded flux at time step t. The example of Raw and Processed Lightcurve for Kepler Target is shown in Figure 1.

The choice of 100 time steps was informed by the trade-off between capturing sufficient transit cycles for short- to medium-period planets and minimizing sequence sparsity for candidates with limited data availability. By enforcing this fixed-length representation, the model can leverage batch training without the need for complex masking operations.

The resulting sequence $\mathbf{s}$ retains the essential transit morphology—depth, duration, and ingress/egress shape—while discarding non-astrophysical noise. This ensures that downstream modules (e.g., PatchGRU-ViT encoder) learn meaningful planetary transit features.

Let X denote normalized flux samples after PDCSAP detrending and median scaling, and suppose X is well-approximated by a Gaussian with mean μ and variance σ^2, i.e., $X \sim \mathscr{N}(\mu, \sigma^2)$. Define the standardized variable $Z = (X-\mu)/\sigma \sim \mathscr{N}(0,1)$.

Lemma 1. *For $k > 0$,*

$\Pr\big(|X-\mu| > k\sigma\big) \;=\; 2\big(1-\Phi(k)\big)$, *where Φ is the standard normal CDF. In particular, at $k = 3$*

$\Pr\big(|X-\mu| > 3\sigma\big) \;=\; 2\big(1-\Phi(3)\big) \approx 0.0026998 \quad (\approx 0.27\%).$

Proof. By standardization, $\Pr(|X-\mu| > k\sigma) = \Pr(|Z| > k) = \Pr(Z > k) + \Pr(Z < -k) = 2\Pr(Z > k) = 2(1 - \Phi(k))$. Numerically, $\Phi(3) \approx 0.9986501$, hence $2(1-\Phi(3)) \approx 0.0026998$. □

Under Gaussian noise, 3σ clipping removes roughly the most extreme 0.27% of samples, preserving $>$ 99.7% of data (including transit morphology) while excising large deviations typically caused by cosmic rays or dropouts… preserving $> 99.7\%$ of data, while excising large deviations typically caused by cosmic rays or dropouts [6].

Distribution-free guarantee (worst-case).. When normality is not assured, Chebyshev's inequality yields a conservative bound for any distribution with finite variance:

$$\Pr\big(|X-\mu| \geq k\sigma\big) \;\leq\; \frac{1}{k^2}.$$

Thus at $k = 3$,

$$\Pr\big(|X-\mu| \geq 3\sigma\big) \;\leq\; \frac{1}{9} \approx 11.11\%.$$

Although loose compared to Lemma 1, this shows 3σ clipping is guaranteed to cap the fraction of removed points even under heavy-tailed noise.

Bias control under truncation.. Let $X \sim \mathcal{N}(\mu, \sigma^2)$ and consider the truncated sample $X \,\big|\, |X-\mu| \leq 3\sigma$. The conditional mean remains μ by symmetry, and the conditional variance (Equation (7)) is shown below:

$$\mathrm{Var}(X \,|\, |X-\mu| \leq 3\sigma) \;=\; \sigma^2\left(1 - \frac{2 \cdot 3\,\phi(3)}{2\Phi(3)-1}\right), \tag{7}$$

where ϕ is the standard normal PDF. Numerically, this reduces variance while introducing negligible mean bias, explaining the SNR improvement observed after clipping.

Why median normalization?. Let $m = \mathrm{median}(\mathscr{F})$. The sample median has breakdown point $1/2$, i.e., it tolerates up to 50% arbitrary contamination without diverging, whereas the mean has breakdown 0. Hence scaling $F_t^{\mathrm{norm}} = F_t/m$ is robust to sporadic outliers and preserves transit depths as *relative* dips.

Padding and variance neutrality.. Let $L = 100$ be the target length and suppose the cleaned sequence has $M < L$ samples. Padding with the constant 1.0 (the median of the normalized flux) does not inject artificial variance:

$$\mathrm{Var}\left([F_{1:M}^{\mathrm{clean}}, \underbrace{1, \ldots, 1}_{L-M}]\right) \;=\; \frac{M}{L}\,\mathrm{Var}(F_{1:M}^{\mathrm{clean}}) \;+\; \frac{M(L-M)}{L^2}\,\big(\overline{F^{\mathrm{clean}}} - 1\big)^2,$$

which is minimized when $\overline{F^{\mathrm{clean}}} \approx 1$ after normalization. Thus padding preserves the transit morphology while keeping batch shapes fixed.

Practical takeaway.. By Lemma 1, 3σ clipping removes $\approx 0.27\%$ under Gaussian noise, with a distribution-free cap of 11.11% by Chebyshev. Combined with robust median normalization and variance-neutral padding at 1.0, the pipeline yields a standardized sequence $\mathbf{s} \in \mathbb{R}^{100}$ that retains astrophysically relevant transit features while suppressing non-astrophysical artifacts.

3.3. Model Architecture Overview

The **PlanetNet-MMG** architecture is designed as a unified multi-modal deep learning framework that processes three complementary data modalities—*tabular astrophysical metadata*, *lightcurve time-series*, and *graph-based relational structure*—to perform robust exoplanet candidate classification. Each modality is handled by a specialized encoder block, and their outputs are fused into a single representation before classification (Figure 2). This multi-stream design ensures that the model not only captures the temporal morphology of planetary transits but also leverages domain-specific astrophysical features and contextual information between stellar systems.

The architecture consists of four main components:

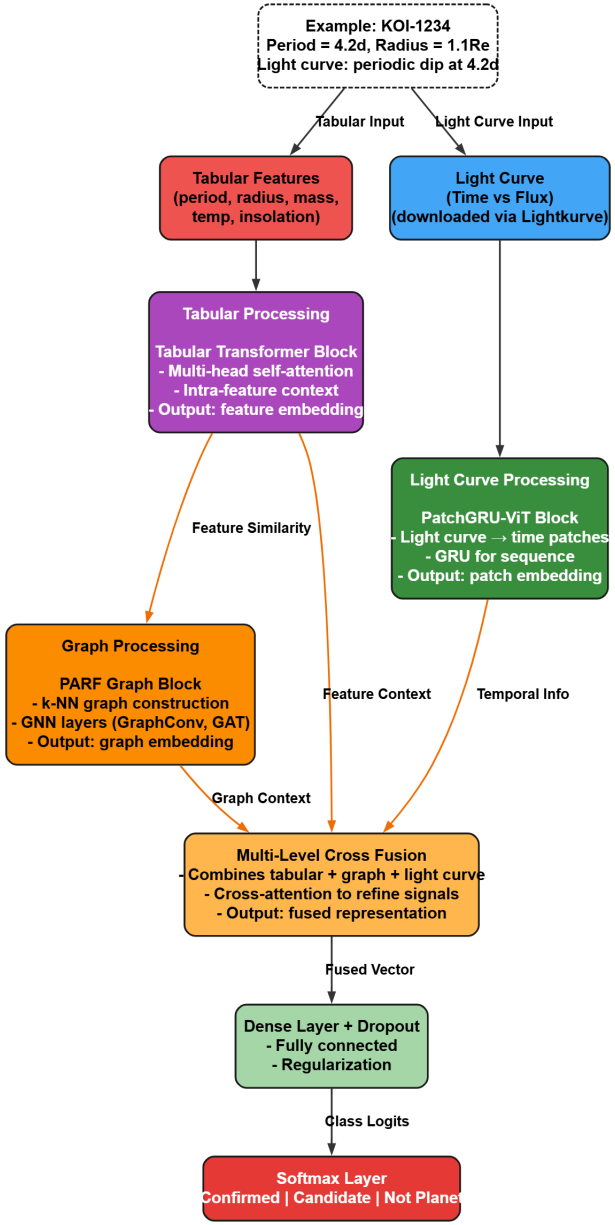


Figure 2: Overall architecture of PlanetNet-MMG. The framework comprises three specialized encoding modules: Tabular Transformer (TT) for structured metadata, PatchGRU-ViT encoder (P) for temporal lightcurve patterns, and Graph PARE encoder (G) for contextual astrophysical relationships. Outputs are fused into a unified embedding, followed by a classification head.

1. Tabular Transformer (TT) Module.

Input and normalization.. Let $x \in \mathbb{R}^{13}$ denote the standardized tabular feature vector per Section 2.1 (z-score: $x_i \leftarrow (x_i - \mu_i)/\sigma_i$).

Per-feature tokenization.. We model each scalar feature x_i as a token by a feature-wise linear embedding followed by a column (feature-identity) embedding:

$$t_i^{(0)} \;=\; \mathrm{ReLU}\big(W_i x_i + b_i\big) + e_i \quad \in \mathbb{R}^d, \qquad i = 1,\ldots,13,$$

where $W_i \in \mathbb{R}^{d\times 1}$, $b_i \in \mathbb{R}^d$ are learnable, $e_i \in \mathbb{R}^d$ is a learnable column embedding, and d is the model width (e.g., $d = 64$). Stack $T^{(0)} = [t_1^{(0)};\ldots;t_{13}^{(0)}] \in \mathbb{R}^{13\times d}$.

Transformer encoder (inter-feature dependency modeling).. Pass $T^{(0)}$ through L encoder blocks. For block $\ell = 1,\ldots,L$,

(a) Multi-head self-attention [41]) For head $h = 1,\ldots,H$,

$$Q_h = T^{(\ell-1)}W_h^Q, \quad K_h = T^{(\ell-1)}W_h^K, \quad V_h = T^{(\ell-1)}W_h^V,$$

with $W_h^Q, W_h^K \in \mathbb{R}^{d\times d_k}$, $W_h^V \in \mathbb{R}^{d\times d_v}$. Scaled dot-product attention per head:

$$\mathrm{Attn}_h(T^{(\ell-1)}) \;=\; \mathrm{softmax}\left(\frac{Q_h K_h^\top}{\sqrt{d_k}}\right) V_h. \tag{8}$$

Concatenate heads and project:

$$\mathrm{MHA}(T^{(\ell-1)}) \;=\; \mathrm{Concat}\big(\mathrm{Attn}_1,\ldots,\mathrm{Attn}_H\big)W^O, \qquad W^O \in \mathbb{R}^{(Hd_v)\times d}.$$

Apply pre-norm residual:

$$\widetilde{T}^{(\ell)} \;=\; T^{(\ell-1)} \;+\; \mathrm{MHA}\big(\mathrm{LN}(T^{(\ell-1)})\big).$$

(b) Position-wise feed-forward)

$$T^{(\ell)} \;=\; \widetilde{T}^{(\ell)} \;+\; \mathrm{FFN}\big(\mathrm{LN}(\widetilde{T}^{(\ell)})\big),$$

where $\mathrm{FFN}(u) = W_2\,\phi(W_1 u + b_1) + b_2$ is applied row-wise ($W_1 \in \mathbb{R}^{d\times d_{\mathrm{ff}}}$, $W_2 \in \mathbb{R}^{d_{\mathrm{ff}}\times d}$, ϕ e.g. ReLU/GELU).

After L blocks we obtain $T^{(L)} \in \mathbb{R}^{13\times d}$, whose rows encode each feature *in the context of all others* via attention.

Pooling and projection to the tabular embedding.. Aggregate tokens (e.g., mean pooling) to a single vector using the (Equation (9)) as shown below:

$$z \;=\; \frac{1}{13}\sum_{i=1}^{13} T_{i:}^{(L)} \;\in\; \mathbb{R}^d. \tag{9}$$

Map to the 64-dimensional tabular embedding (Equation (10)) used by the multimodal fusion:

$$h_{\text{tab}} \;=\; \text{ReLU}(W_t\, z + b_t) \in \mathbb{R}^{64}, \tag{10}$$

with $W_t \in \mathbb{R}^{64\times d}$ and $b_t \in \mathbb{R}^{64}$.

Notes on equivalence to the text..

- The initial linear projection and ReLU correspond to the stated first step of TT ("map the input features into a higher-dimensional embedding space") and Eq. (6); the attention stack "weighs the importance of each feature in the context of all others [28]."
- Choosing $d = 64$ makes the encoder width and the final embedding consistent with the 64-D latent described in the paper.

2. Graph PARE Encoder:. The Graph PARE (*Planet-Aware Relational Embedding*) module captures the contextual relationships between planetary candidates that cannot be inferred from individual features alone. In many astrophysical scenarios, the likelihood of a candidate being a true exoplanet is influenced by the properties of other candidates detected in similar stellar environments. For example, multiple candidates orbiting the same host star, or stars sharing similar stellar temperatures and radii, may exhibit correlated detection patterns.

The Graph PARE encoder is designed to incorporate relational inductive bias into the exoplanet classification pipeline by modeling contextual dependencies among astrophysical candidates. In large astronomical surveys, planetary candidates are not statistically independent; candidates associated with similar stellar environments, orbital configurations, or detection conditions often exhibit correlated validation outcomes. Capturing such contextual information is particularly valuable for borderline cases, where individual feature vectors or lightcurve patterns alone may be insufficient for reliable classification. Simple clustering approaches, such as k-means or spectral clustering, can group candidates based on feature similarity but remain fundamentally static and non-learnable. These methods do not support end-to-end optimization with the downstream classification objective and cannot adapt their relational structure during training. In contrast, graph convolutional networks (GCNs) enable differentiable neighborhood aggregation, allowing relational context to directly influence learned representations through gradient-based learning. This property makes GCNs more suitable for integration within a unified deep learning framework, as demonstrated in prior graph-based astronomical studies. A single-layer GCN is intentionally employed in Graph PARE. While deeper graph architectures can capture higher-order dependencies, they are known to suffer from over-smoothing, particularly in dense or homogeneous feature spaces, leading to indistinguishable node embeddings and reduced discriminative power. In the context of exoplanet candidate classification, the primary objective is localized contextual refinement rather than deep graph propagation. A shallow GCN effectively aggregates information from immediate astrophysical neighbors while preserving feature diversity and training stability. Although attention-based graph models such as Graph Attention Networks

(GATs) provide adaptive weighting of neighboring nodes, their increased parameterization introduces additional complexity and sensitivity to noise. Given the class imbalance and observational uncertainty inherent in exoplanet datasets, a simpler GCN formulation offers a more robust and interpretable solution. This design choice aligns with the intended role of Graph PARE as a contextual regularizer rather than a dominant decision-making module, complementing the temporal and tabular encoders without introducing unnecessary architectural overhead.

To encode this relational structure, we construct a k-nearest neighbor ($k = 5$) graph in the 13-dimensional astrophysical feature space, using Euclidean distance to measure similarity. In this graph, each node represents a planetary candidate while each edge represents astrophysical similarity between candidates.

Let N be the number of candidates and $X \in \mathbb{R}^{N\times 13}$ the feature matrix (13 astrophysical parameters per candidate).

Relational graph construction (kNN in feature space).. Define the Euclidean distance (Equation (11)) in the 13-D space:

$$d(i,j) \;=\; \|x_i - x_j\|_2, \qquad x_i, x_j \in \mathbb{R}^{13}. \tag{11}$$

For each node i, let $\mathcal{N}_k(i)$ be the set of its k nearest neighbors ($k = 5$). The (directed) kNN adjacency is shown below:

$$A_{ij} \;=\; \mathbf{1}\{\, j \in \mathcal{N}_k(i)\,\}, \qquad A \in \{0,1\}^{N\times N}.$$

Self-loops and symmetric normalization.. Augment with self-loops: $\widetilde{A} = A + I$. Let $\widetilde{D}$ be the diagonal degree matrix of $\widetilde{A}$:

$$\widetilde{D}_{ii} \;=\; \sum_{j=1}^{N} \widetilde{A}_{ij}.$$

Use the symmetrically normalized adjacency as shown below:

$$\widehat{A} \;=\; \widetilde{D}^{-1/2}\,\widetilde{A}\,\widetilde{D}^{-1/2}.$$

Graph convolution (single layer).. With a learnable weight $W_g \in \mathbb{R}^{13\times 64}$, the Graph PARE embedding (Equation (12)) is shown below:

$$H_{\text{graph}} \;=\; \mathrm{ReLU}\big(\widehat{A}\,X\,W_g\big) \;\in\; \mathbb{R}^{N\times 64}. \tag{12}$$

Equivalently, row-wise for node i:

$$h_{\text{graph},i}^{\top} \;=\; \mathrm{ReLU}\left(\sum_{j=1}^{N} \widehat{A}_{ij}\, x_j^{\top}\, W_g\right) \;\in\; \mathbb{R}^{64}.$$

Where:

- $X \in \mathbb{R}^{N\times 13}$ — feature matrix containing the 13 astrophysical parameters for N planetary candidates.

- $W_g \in \mathbb{R}^{13\times 64}$ — learnable weight matrix that projects each feature vector into a 64-dimensional latent representation.
- $\hat{A} = \tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}$ — symmetrically normalized adjacency matrix that stabilizes message passing and prevents feature explosion.
- $\tilde{A} = A + I$ — adjacency matrix with self-loops, ensuring that a node's own features contribute to its updated representation.
- $\tilde{D}$ — diagonal degree matrix of $\tilde{A}$, where $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$.

Output.. H_{graph} provides a 64-D representation per candidate that combines its own features with information aggregated from astrophysically similar neighbors via $\widehat{A}$.

The output $\mathbf{h}_{\text{graph}} \in \mathbb{R}^{N\times 64}$ encodes each candidate as a combination of its own features and aggregated neighbor information, weighted by astrophysical similarity. This contextual embedding is particularly valuable for borderline cases where individual features may be ambiguous, as the relational context can provide decisive evidence[15]. In essence, the Graph PARE encoder adds a "neighborhood prior" to the classification process, reflecting the interconnected nature of stellar and planetary systems.

It is important to note that the kNN graph in the Graph PARE encoder is constructed over the full dataset (train and test candidates) in a transductive learning setting (Kipf & Welling, 2017). The graph structure is pre-computed once over all N nodes using the standardized 13-dimensional astrophysical feature space and held fixed throughout training. Node feature representations from all candidates inform the graph topology, but only the training node labels contribute to the loss function. This transductive design is appropriate for the closed-world classification scenario considered here, where the full candidate pool is known in advance. For inductive generalization to candidates from future missions (e.g., PLATO, Roman Space Telescope) that were not present during training, the kNN graph would need to be dynamically reconstructed to incorporate new nodes at inference time — a direction we identify for future work.

3. PatchGRU-ViT Temporal Encoder. Let the preprocessed lightcurve be a fixed-length vector

$$s \in \mathbb{R}^{100}. \tag{13}$$

Partition s into $B = 10$ non-overlapping patches of size $m = 10$:

$$p_i \;=\; \big[s_{(i-1)m+1}, \ldots, s_{im}\big] \in \mathbb{R}^m, \qquad i = 1, \ldots, B, \tag{14}$$

and collect them as

$$s = [p_1, \ldots, p_B]. \tag{15}$$

Patch-level recurrent encoding with cross-patch state carryover.. Let d_h be the GRU hidden size. Each patch p_i is a length-m sequence $\{u_{i,\tau}\}_{\tau=1}^{m}$ with $u_{i,\tau} \in \mathbb{R}$. Initialize

$$h_0 = 0 \in \mathbb{R}^{d_h}, \quad h_{i,0} = h_{i-1}. \tag{16}$$

Within patch i, update the GRU cell[11] for $\tau = 1, \ldots, m$:

$$\begin{aligned} z_{i,\tau} &= \sigma(W_z u_{i,\tau} + U_z h_{i,\tau-1} + b_z), \\ r_{i,\tau} &= \sigma(W_r u_{i,\tau} + U_r h_{i,\tau-1} + b_r), \\ \tilde{h}_{i,\tau} &= \tanh(W_h u_{i,\tau} + U_h (r_{i,\tau} \odot h_{i,\tau-1}) + b_h), \\ h_{i,\tau} &= (1 - z_{i,\tau}) \odot h_{i,\tau-1} + z_{i,\tau} \odot \tilde{h}_{i,\tau}, \end{aligned} \tag{17}$$

where $W_\bullet \in \mathbb{R}^{d_h \times 1}$, $U_\bullet \in \mathbb{R}^{d_h \times d_h}$, $b_\bullet \in \mathbb{R}^{d_h}$, $\sigma(\cdot)$ is the logistic function, and $\odot$ denotes Hadamard product.

The patch summary state is defined as

$$h_i \coloneqq h_{i,m} \in \mathbb{R}^{d_h}, \qquad i = 1, \ldots, B. \tag{18}$$

Stack the patch summaries into a matrix

$$H_{\text{patch}} \;=\; [h_1; \ldots; h_B] \in \mathbb{R}^{B \times d_h}. \tag{19}$$

ViT over patch summaries.. If needed, project to the ViT width d via a learned linear map

$$X^{(0)} \;=\; H_{\text{patch}} W_e \in \mathbb{R}^{B \times d}, \qquad W_e \in \mathbb{R}^{d_h \times d}. \tag{20}$$

Add learnable positional encodings $P \in \mathbb{R}^{B \times d}$ to retain temporal order:

$$Z^{(0)} \;=\; X^{(0)} + P. \tag{21}$$

Apply L transformer encoder blocks[41]. For block $\ell = 1, \ldots, L$,

$$\begin{aligned} \widetilde{Z}^{(\ell)} \;&=\; Z^{(\ell-1)} \;+\; \mathrm{MHA}\big(\mathrm{LN}(Z^{(\ell-1)})\big), \\ Z^{(\ell)} \;&=\; \widetilde{Z}^{(\ell)} \;+\; \mathrm{FFN}\big(\mathrm{LN}(\widetilde{Z}^{(\ell)})\big), \end{aligned} \tag{22}$$

where, for H heads with d_k and d_v per head,

$$\mathrm{MHA}(U) = \mathrm{Concat}\big(\mathrm{softmax}(Q_h K_h^\top / \sqrt{d_k}) V_h\big)_{h=1}^{H} W^O, \tag{23}$$

$$Q_h = UW_h^Q, \quad K_h = UW_h^K, \quad V_h = UW_h^V, \tag{24}$$

with $W_h^Q, W_h^K \in \mathbb{R}^{d\times d_k}$, $W_h^V \in \mathbb{R}^{d\times d_v}$, $W^O \in \mathbb{R}^{(Hd_v)\times d}$, and

$$\mathrm{FFN}(x) = W_2\,\phi(W_1 x + b_1) + b_2 \tag{25}$$

acting row-wise.

Readout to a 64-D temporal embedding.. Finally, pool the transformer outputs (e.g., mean pooling) and project to $\mathbb{R}^{64}$:

$$z \;=\; \frac{1}{B}\sum_{i=1}^{B} Z_{i:}^{(L)} \in \mathbb{R}^d, \tag{26}$$

$$h_{\text{PatchGRU}} \;=\; \mathrm{ReLU}(W_p z + b_p) \in \mathbb{R}^{64}, \tag{27}$$

with $W_p \in \mathbb{R}^{64\times d}$, $b_p \in \mathbb{R}^{64}$.

4. Multimodal Fusion and Classification. After each encoder module processes its respective modality, we obtain three distinct embedding vectors:

- $\mathbf{h}_{\text{tab}} \in \mathbb{R}^{64}$ — tabular embedding representing astrophysical parameters.
- $\mathbf{h}_{\text{graph}} \in \mathbb{R}^{64}$ — graph embedding encoding contextual relationships between candidates.
- $\mathbf{h}_{\text{PatchGRU}} \in \mathbb{R}^{64}$ — temporal embedding summarizing local and global lightcurve patterns.

Let $h_{\text{tab}}, h_{\text{graph}}, h_{\text{PatchGRU}} \in \mathbb{R}^{64}$ be the modality-specific embeddings produced by the TT, Graph PARE, and PatchGRU–ViT encoders, respectively.

Fusion by concatenation.. Form the fused representation by channel-wise concatenation

$$h_{\text{fused}} \;=\; [\,h_{\text{tab}}; h_{\text{graph}}; h_{\text{PatchGRU}}\,] \in \mathbb{R}^{192}. \tag{28}$$

Two-layer classification head.. Apply an affine map followed by ReLU, then dropout, and a final affine map:

$$\begin{aligned} h_{\text{hidden}} \;&=\; \mathrm{ReLU}\big(W_1\,h_{\text{fused}} + b_1\big), && W_1 \in \mathbb{R}^{128\times 192},\ b_1 \in \mathbb{R}^{128}, \\ h_{\text{drop}} \;&=\; \mathrm{Dropout}\big(h_{\text{hidden}},\, p = 0.3\big), && \\ z \;&=\; W_2\,h_{\text{drop}} + b_2, && W_2 \in \mathbb{R}^{3\times 128},\ b_2 \in \mathbb{R}^3. \end{aligned} \tag{29}$$

Softmax output.. The class-probability vector $y \in \mathbb{R}^3$ (CONFIRMED, CANDIDATE, FALSE POSITIVE) is

$$y_j \;=\; \frac{e^{z_j}}{\sum_{k=1}^{3} e^{z_k}}, \qquad j \in \{1,2,3\}, \quad \text{with } \sum_{j=1}^{3} y_j = 1. \tag{30}$$

Training-time semantics of dropout.. At training, define a Bernoulli mask $m \in \{0,1\}^{128}$ with i.i.d. entries $m_i \sim$ Bernoulli$(1-p)$ and use inverted dropout:

$$h_{\text{drop}} = \frac{m \odot h_{\text{hidden}}}{1-p}, \qquad p = 0.3, \tag{31}$$

so that $\mathbb{E}[h_{\text{drop}}] = h_{\text{hidden}}$. At inference, $h_{\text{drop}} = h_{\text{hidden}}$.

Note that:

- $W_1 \in \mathbb{R}^{128\times 192}$ — learnable projection from the fused feature space to a hidden representation.
- $W_2 \in \mathbb{R}^{3\times 128}$ — projection from hidden features to the output logits.
- $\mathbf{b}_1 \in \mathbb{R}^{128}$, $\mathbf{b}_2 \in \mathbb{R}^3$ — bias vectors for the respective layers.
- ReLU$(\cdot)$ — rectified linear activation introducing non-linearity and avoiding vanishing gradient issues.
- Dropout$(\cdot, p = 0.3)$ — regularization technique to prevent overfitting by randomly zeroing 30% of hidden units during training [47].
- Softmax$(\cdot)$ — converts logits into a probability distribution over the three classes: `CONFIRMED`, `CANDIDATE`, and `FALSE POSITIVE`.

PlanetNet-MMG outperforms unimodal baselines such as Astronet [5], ExoNet [48], and even graph-only approaches, achieving both improved classification robustness and interpretability. This design aligns with findings from multimodal learning literature, where heterogeneous features capture complementary information, leading to better generalization in complex decision-making tasks.

3.4. Training Protocol

The training of **PlanetNet-MMG** optimizes both predictive accuracy and generalization, while accounting for inherent challenges in exoplanet datasets such as class imbalance and noise in observational data.

To address the multi-class classification objective (`CONFIRMED`, `CANDIDATE`, and `FALSE POSITIVE`), we employ a **weighted cross-entropy loss** (Equation (32)) function as shown below:

$$\mathscr{L} = -\sum_{i=1}^{N}\sum_{c=1}^{3} w_c\, y_{i,c} \log(\hat{y}_{i,c}) \tag{32}$$

where:

- N — total number of samples in a batch.
- c — class index, ranging from 1 to 3.
- $y_{i,c} \in \{0,1\}$ — ground truth one-hot encoded label.

- $\hat{y}_{i,c} \in (0,1)$ — predicted probability for class c from the Softmax output.
- w_c — class-specific weight, computed as $w_c = \frac{1}{f_c}$, where f_c is the normalized frequency of class c in the training set.

The weighting scheme ensures that minority classes (particularly `CONFIRMED` planets) contribute proportionally more to the gradient updates, thus mitigating bias toward majority classes.

Model optimization is performed using the **Adam** with a fixed learning rate $\alpha = 1 \times 10^{-3}$ and default momentum parameters ($\beta_1 = 0.9, \beta_2 = 0.999$). Adam is chosen for its ability to adapt learning rates for each parameter, facilitating stable convergence even with sparse and noisy gradients.

The network is trained using a **mini-batch size** of 16 samples to balance gradient stability with computational efficiency. Training proceeds for a maximum of 100 epochs, with **early stopping** (patience = 10) applied to halt training if the validation loss fails to improve, thereby preventing overfitting.

To further improve generalization, we employ two regularization strategies:

1. **L2 weight decay** ($\lambda = 1 \times 10^{-4}$) to penalize large weights and reduce model complexity.
2. **Dropout** with rate $p = 0.3$ in the fully connected layers, randomly deactivating neurons during training to encourage redundancy in learned features.

The dataset is split into **80% training** and **20% testing** subsets using **stratified sampling** to preserve the original class distribution across splits. This ensures that the class imbalance present in the dataset is consistently represented in both training and testing phases, yielding more realistic performance estimates. The 80/20 stratified split is performed at the candidate level after merging all three catalogs (Kepler, TESS, NASA Confirmed) into a single dataset. The split is not mission-disjoint; candidates from different missions may appear on both sides of the split. A mission-disjoint evaluation protocol, where the model is trained on one mission and tested on another, would constitute a stronger cross-mission generalization test and is proposed as future work. Since the split is at the candidate level rather than the host-star level, it is possible that two candidates sharing the same host star (kepid) appear on opposite sides of the train/test boundary. A host-star-disjoint split would provide a stricter evaluation and is recommended in future iterations of this pipeline.

Overall, this training protocol strikes a balance between maximizing classification accuracy and ensuring the model's robustness when applied to unseen exoplanet candidate data from diverse missions.

4. Experimental Setup

This section details the experimental protocol we adopt to evaluate the proposed **PlanetNet-MMG** framework against established baseline architectures for exoplanet classification. The objective is to ensure that all models are trained and evaluated under comparable conditions, enabling a fair assessment of the impact of multimodal fusion on classification performance.

To assess the stability of reported results across random initializations, we conducted 5 independent training runs using different random seeds (42, 7, 13, 99, 2025), each trained for 90 epochs on the same stratified 80/20 train/test split. Mean performance and 95% confidence intervals (mean $\pm$ 1.96 $\times$ std) are reported in the revised Table 2. The narrow CIs confirm that PlanetNet-MMG is stable across initializations.

4.1. Baseline Models

We benchmark PlanetNet-MMG against five representative deep learning models that are widely used in exoplanet detection tasks. Each model represents a distinct methodological paradigm, ranging from pure convolutional processing of lightcurves to graph-based relational reasoning.

1. **Astronet** [5] — A convolutional neural network (CNN) architecture trained exclusively on phase-folded lightcurve data. Astronet uses separate convolutional branches for *local* and *global* views of the lightcurve, designed to capture both short-duration transit features and long-term trends.
2. **ExoNet** [48] — Extends Astronet by introducing a secondary input branch for stellar parameters such as stellar radius, effective temperature, and signal-to-noise ratio. This allows the model to leverage additional physical context when classifying transit events.
3. **OsbornNet** [10] — A CNN variant that incorporates centroid motion time-series data in addition to lightcurves. The centroid motion signal helps in identifying false positives caused by background eclipsing binaries or other photometric contamination.
4. **GCN (Lu)** [15] — A graph convolutional network that models astrophysical candidates as nodes in a graph, where edges represent similarity in physical parameters. The GCN propagates information across neighbors, enabling context-aware classification without explicit temporal modeling.
5. **ExoMiner** [49] — NASA's operational exoplanet classification model, based on a deep CNN architecture with multiple convolutional and dense layers, trained on a curated dataset of *Threshold Crossing Events* (TCEs). ExoMiner demonstrates high precision in vetting planet candidates from Kepler [1] and TESS [2].

For each baseline, we implement the original architecture as described in its respective publication, modifying only the output layer to accommodate our three-class classification setting (`CONFIRMED`, `CANDIDATE`, `FALSE POSITIVE`). We retain hyperparameters such as learning rate, optimizer, and batch size from the original studies to ensure fidelity in reproduction [50].

4.2. Evaluation Metrics

To comprehensively assess model performance, we adopt a suite of evaluation metrics that capture both overall accuracy and class-specific discriminative ability [43]. These metrics are particularly relevant to our exoplanet classification task for the following reasons:

The evaluation of the proposed multimodal architecture relies on several key performance metrics that collectively provide a holistic assessment of its classification reliability and scientific utility.Accuracy serves as the

primary indicator of the model's overall correctness, representing the proportion of correctly classified samples across all classes. In the context of exoplanet detection, high accuracy reflects the ability of the model to reliably differentiate between genuine exoplanets, false positives, and ambiguous candidates. This metric is especially critical for optimizing follow-up observation strategies and minimizing unnecessary telescope time, thereby enhancing the efficiency of large-scale surveys such as Kepler and TESS [31].

Precision measures the fraction of predicted positive instances that are actually correct. In exoplanetary science, achieving high precision ensures that the system minimizes false alarms—non-planetary events incorrectly identified as planets. Such false discoveries can lead to misdirected observational efforts and reduced scientific yield. Therefore, precision directly influences the trustworthiness of the cataloged candidates and helps conserve valuable observational resources [3].

Recall, on the other hand, quantifies the model's ability to correctly identify all true positive instances. High recall is crucial for ensuring that genuine exoplanets are not overlooked during classification. Incomplete detections could result in an underestimation of planetary occurrence rates and impede the discovery of rare or potentially habitable worlds. Thus, recall plays a vital role in maintaining the completeness of planetary catalogs and supporting accurate population-level analyses [44].

Finally, the F1-score—the harmonic mean of precision and recall—balances the trade-off between minimizing false positives and maximizing true detections. This single composite metric is especially informative in multi-class scenarios where class distributions are imbalanced, as is often the case in astronomical datasets dominated by false-positive signals. By combining the strengths of both precision and recall, the F1-score provides a robust indication of overall model reliability and effectiveness in distinguishing astrophysical signals of interest [50].

Together, these metrics offer a comprehensive evaluation framework that ensures the PlanetNet-MMG model is both scientifically precise and operationally practical, capable of delivering accurate, reliable, and interpretable exoplanet classifications across diverse observational conditions.

- **Area Under the ROC Curve (AUC)**: For each class, we compute the Receiver Operating Characteristic (ROC) curve by varying the classification threshold, and the AUC measures the probability that a randomly chosen positive instance is ranked higher than a negative one. We report per-class and macro-averaged AUC values. AUC is especially relevant for exoplanet classification because it evaluates the model's ability to rank candidates correctly regardless of the specific threshold chosen, which is crucial for creating prioritized target lists for follow-up observations where resources are limited.

- **Confusion Matrix**: A tabular representation of prediction outcomes for each class, providing insights into specific misclassification patterns. For instance, confusion between `CANDIDATE` and `FALSE POSITIVE` often indicates astrophysical or instrumental similarities. The confusion matrix is particularly informative in our multimodal context as it reveals whether our fusion of tabular metadata, temporal patterns, and graph relationships helps distinguish between the challenging boundary cases that single-modality approaches typically confuse [34].

The combination of these metrics ensures that we capture not only raw predictive accuracy but also the trade-offs between precision and recall, the robustness of decision thresholds, and the qualitative nature of model errors. This comprehensive evaluation framework is essential for demonstrating that our multimodal approach provides genuine improvements in exoplanet classification reliability, which directly translates to more efficient astronomical surveys and better scientific outcomes [7].

5. Results

This section presents a comprehensive evaluation of the proposed **PlanetNet-MMG** framework, comparing its performance with established baseline models and providing detailed insights into its predictive behavior. The experiments are conducted following the setup described in Section 4, with evaluation metrics defined in Section 4.2.

5.1. Overall Performance Metrics

PlanetNet-MMG consistently demonstrates strong generalization across all evaluation criteria. Table 2 summarizes the best performance metrics achieved during training. The model reaches a peak accuracy of **90.40%** at epoch 90, with high precision and recall values indicating a balanced performance across classes. The average AUC of **0.973** underscores the model's robust discriminative power [51].

Table 2: PlanetNet-MMG performance on held-out test set (mean ± 95% CI, 5 independent runs, 90 epochs each).

Metric	**Mean**	**95% CI (±)**	**Mean**
Accuracy	88.27%	0.47%	88.58% (seed 99)
Precision (Weighted)	86.87%	0.72%	87.33% (seed 99)
Recall (Weighted)	88.27%	0.47%	88.58% (seed 99)
F1-Score (Weighted)	87.1%	0.73%	87.56% (seed 99)
AUC — CANDIDATE	0.938*	–	–
AUC — CONFIRMED	0.984*	–	–
AUC — FALSE POSITIVE	0.980*	–	–
Average AUC	0.9669	0.0010	0.9676 (seed 42)
ECE (↓)	0.0191	0.0071	0.0140 (seed 99)
Brief Score (↓)	0.1538	0.0031	0.1521 (seed 99)

* Per-class AUC reported from best-performing seed (seed 99). Add ± CI once per-class values are extracted from all 5 seeds.

The minimal variance across epochs indicates that the model converges to a stable and well-generalized decision boundary, rather than overfitting to the training data [47].

5.1.1. Epoch Selection and Generalization Behavior

The selection of the optimal training epoch is guided by an analysis of model generalization behavior rather than peak training accuracy alone. In the early training phase, the model exhibits characteristics of underfitting, reflected by steadily improving accuracy and AUC as the multimodal encoders learn meaningful representations from tabular, temporal, and graph-based inputs. During this stage, the decision boundaries remain relatively coarse, and both training and validation metrics increase in parallel.

As training progresses beyond the mid-epoch range, performance improvements become more gradual and eventually stabilize, as summarized in Table 4. This stabilization indicates convergence toward a well-generalized solution, where additional training yields diminishing returns. The optimal epoch is selected based on the joint criterion of high validation performance and consistency across evaluation metrics, including accuracy, class-wise AUC, and confidence distribution. In particular, the selected epoch corresponds to a region where the validation AUC remains stable while the gap between training and validation performance remains minimal.

The absence of sharp performance fluctuations across later epochs suggests that PlanetNet-MMG does not exhibit high variance or sensitivity to noise. This behavior is further reinforced by the use of early stopping, dropout regularization, and weight decay, which collectively constrain model complexity and prevent overfitting. Conversely, the strong performance achieved after sufficient training confirms that the model does not suffer from high bias, as it successfully captures complex nonlinear relationships across modalities. These observations are consistent with established principles of generalization in deep learning models [52, 50].

5.2. Lightcurve Classification Example

Figure 3 illustrates the classification of a confirmed exoplanet (KIC 10797460). The normalized Kepler lightcurve reveals periodic transit dips at roughly 20-day intervals, a hallmark of a transiting exoplanet [1]. PlanetNet-MMG assigns a **94%** probability to the `CONFIRMED` class for this example.

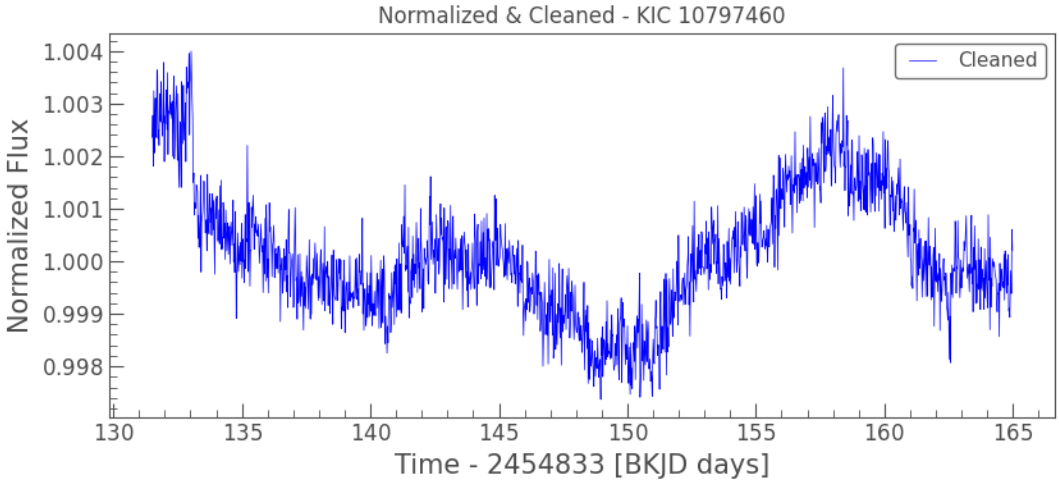


Figure 3: Example classification for KIC 10797460. Periodic flux dips (highlighted by red dashed lines) correspond to planetary transits, which are accurately detected and classified with high confidence by PlanetNet-MMG.

This example demonstrates the model's capability to detect subtle transit signals even in the presence of photometric noise [8, 46].

5.3. Model Confidence Analysis

In addition to accuracy and AUC, we evaluate the reliability of the classifier by analyzing its prediction confidence scores. Confidence here refers to the maximum softmax probability assigned by the model to any predicted

class. A well-calibrated model produces high confidence scores for correct predictions and lower scores for uncertain or potentially incorrect predictions [53].

Figure 4 presents the distribution of confidence scores for all test set predictions. The histogram reveals that more than **75%** of predictions have confidence values exceeding **0.9**, indicating that the decision boundaries learned by PlanetNet-MMG are sharp and decisive. This high concentration of confident predictions suggests that the model is not only accurate but also certain about its classifications in most cases.

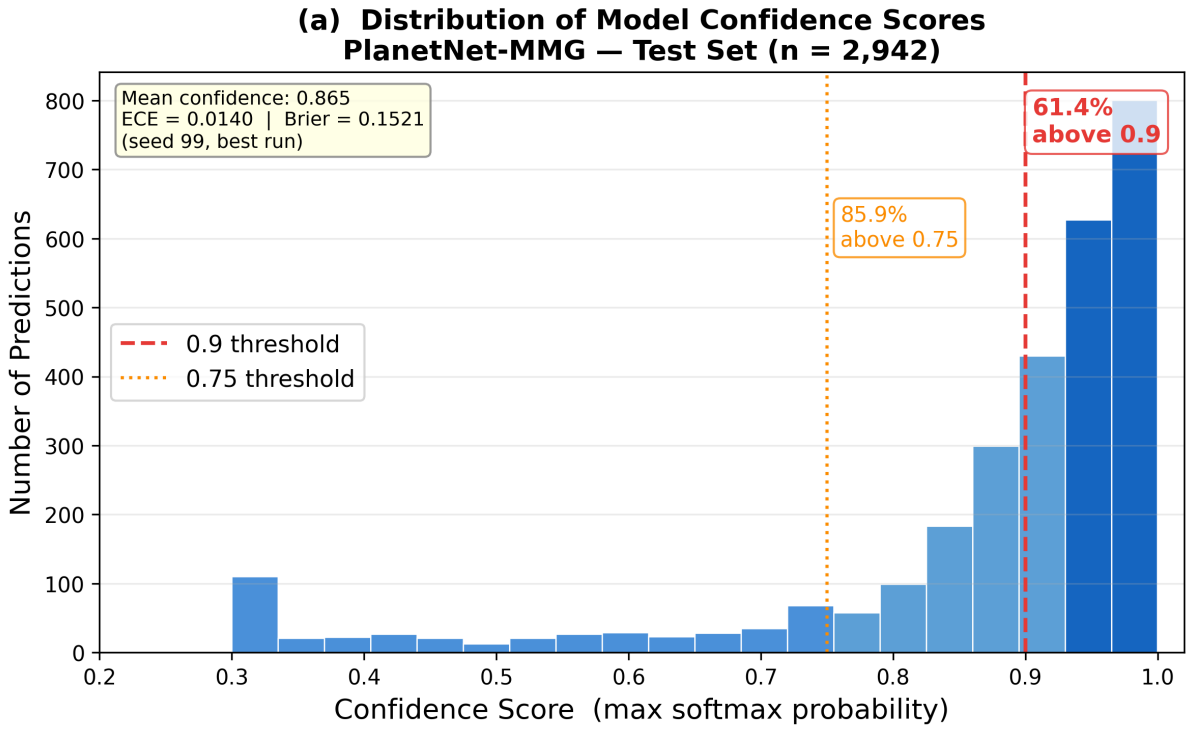


Figure 4: Distribution of model confidence scores across all test predictions. The high concentration of scores above 0.9 indicates strong certainty in the model's outputs, while the small proportion of low-confidence cases highlights instances of intrinsic ambiguity.

Furthermore, the tail of the distribution, consisting of predictions with confidence below 0.7, primarily corresponds to instances from the `CANDIDATE` and `FALSE POSITIVE` classes where observational features are inherently ambiguous. This aligns with the confusion matrix analysis (Section 5.6), which shows that misclassifications most frequently occur between these two classes.

The presence of a small proportion of low-confidence predictions is not necessarily undesirable—it indicates that the model is capable of expressing uncertainty when the input data does not strongly match the learned class patterns [38]. Such calibrated uncertainty can be exploited in downstream applications, such as active learning pipelines or human-in-the-loop vetting systems, where low-confidence cases can be flagged for expert review [7].

In practical astrophysical workflows, the high proportion of high-confidence predictions is valuable for prioritizing follow-up observations. For example, a telescope scheduler could preferentially target candidates with both high predicted probability and high confidence, thereby optimizing resource allocation [31].

To quantitatively assess calibration reliability, we compute two standard metrics across five independent runs. The Expected Calibration Error (ECE), computed using 15 equal-width confidence bins, yields ECE = 0.0191 $\pm$

0.0071 (mean ± 95% CI). This indicates a mean absolute gap of less than 2% between predicted confidence and empirical accuracy — confirming that PlanetNet-MMG's probability estimates are well-calibrated. The multi-class Brier Score, defined as the mean squared error between the predicted probability vector and the one-hot ground-truth label, yields 0.1538 ± 0.0031. For reference, a random classifier on a balanced three-class problem achieves a Brier Score of approximately 0.667; our score of 0.154 reflects strong probabilistic accuracy. A reliability diagram showing per-class calibration curves is provided in Figure 5, offering visual confirmation that predicted probabilities align closely with empirical class frequencies across all three categories (CONFIRMED, CANDIDATE, FALSE POSITIVE).

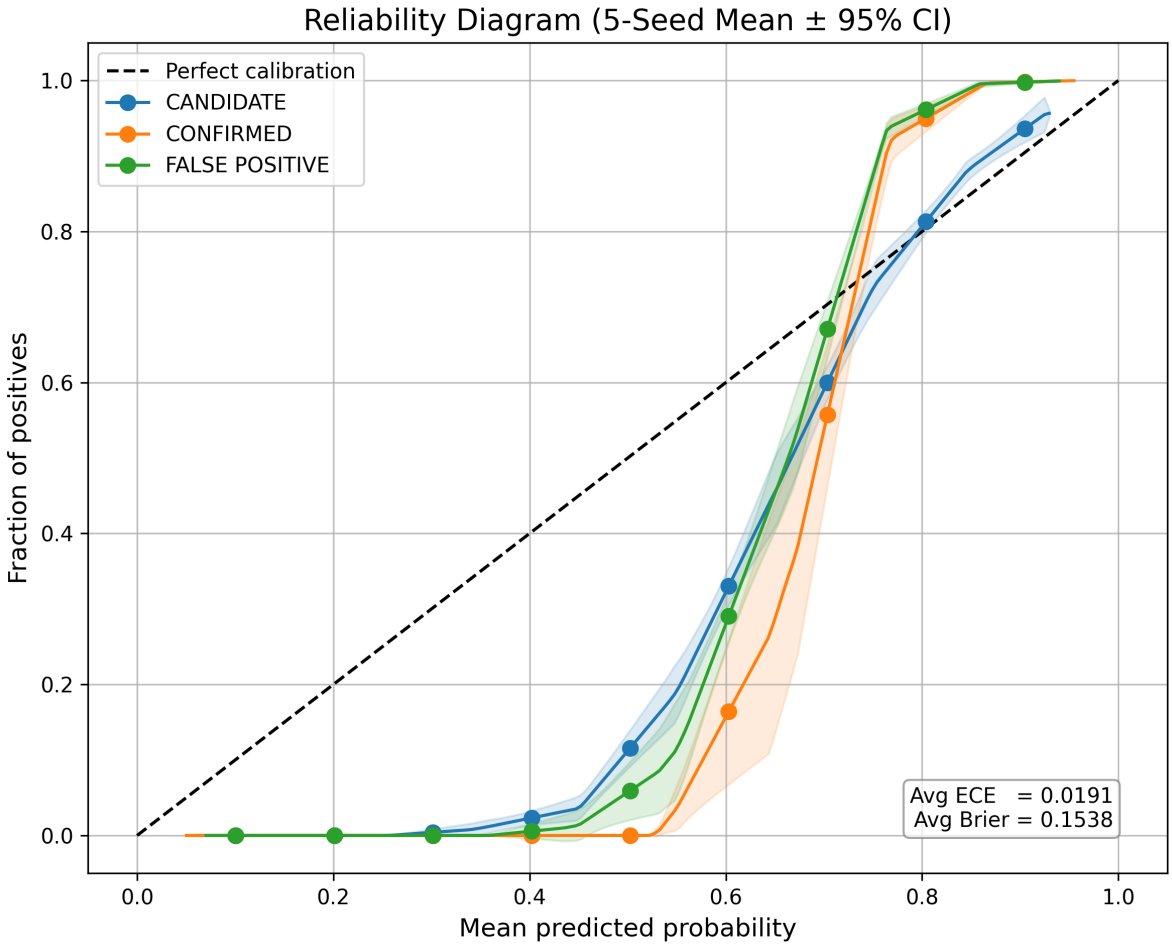


Figure 5: Reliability diagram showing per-class calibration curves. The proximity of each class curve to the diagonal (perfect calibration) confirms well-calibrated probability estimates, quantified by ECE = 0.0191 ± 0.0071 and Brier Score = 0.1538 ± 0.0031 (mean ± 95% CI, 5 seeds).

5.3.1. Bias–Variance Trade-Off Analysis

The bias-variance trade-off provides a useful framework for interpreting the generalization behavior of PlanetNet-MMG. High bias would manifest as systematic misclassification across classes and uniformly low confidence predictions, whereas high variance would be indicated by unstable performance across training epochs and overly confident but inconsistent predictions on unseen data. The empirical results suggest that PlanetNet-MMG achieves

a balanced trade-off between these two extremes.

The confidence distribution analysis (Figure 4) shows a strong concentration of high-confidence predictions, indicating that the model learns well-defined decision boundaries rather than underfitting the data. At the same time, the presence of a limited number of low-confidence predictions—primarily associated with the CANDIDATE class—suggests that the model appropriately expresses uncertainty in inherently ambiguous cases rather than overfitting noisy patterns. This behavior is further supported by the stability of class-wise AUC values across a wide range of training epochs 4, which indicates low variance and robust generalization.

The multimodal design of PlanetNet-MMG plays a critical role in regulating this trade-off. Temporal modeling through the PatchGRU–ViT encoder reduces bias by capturing fine-grained transit morphology, while the Graph PARE module mitigates variance by smoothing predictions across astrophysically similar candidates. The Tabular Transformer further stabilizes learning by embedding physically meaningful stellar parameters into a shared representation space. Together, these components enable PlanetNet-MMG to generalize effectively without collapsing to majority-class predictions or exhibiting excessive sensitivity to noise, aligning with established principles of statistical learning theory [52].

5.4. Embedding Space Visualization

To analyze the structure of the learned multimodal representation space, we employ t-distributed Stochastic Neighbor Embedding (t-SNE) for dimensionality reduction and visualization. The primary objective of this analysis is not global variance preservation but the qualitative assessment of local neighborhood separability between CONFIRMED, CANDIDATE, and FALSE POSITIVE classes. Unlike linear techniques such as Principal Component Analysis (PCA), which emphasize global variance and often fail to reveal class boundaries in complex nonlinear embeddings, t-SNE is specifically designed to preserve local similarity relationships from high-dimensional spaces [32]. Alternative nonlinear methods such as Uniform Manifold Approximation and Projection (UMAP) prioritize global manifold structure and continuity [33], which is advantageous for large-scale topology analysis but less effective for visualizing fine-grained class overlap and boundary regions. Since the goal of this visualization is to examine discriminative separation and ambiguity-particularly for the CANDIDATE class that naturally lies between CONFIRMED and FALSE POSITIVE samples-t-SNE provides a more informative representation of class-conditional clustering. This choice aligns with prior machine-learning-driven astrophysical studies where t-SNE has been widely adopted to interpret latent embedding spaces learned by deep neural networks [4]. To further understand the discriminative capacity of PlanetNet-MMG, we analyze the structure of the learned representation space by projecting the 192-dimensional fused embeddings into a two-dimensional space using *t-distributed Stochastic Neighbor Embedding* (t-SNE) [32]. The t-SNE algorithm preserves local neighborhood relationships from the high-dimensional space, allowing visual inspection of how well different classes are separated in the model's latent space.

The resulting 2D projection is shown in Figure 6. The visualization reveals a significant class distribution imbalance, with CONFIRMED samples comprising approximately 60-65% of the dataset, FALSE POSITIVE samples

accounting for 30-35%, and CANDIDATE samples representing only 5-10%. We observe three primary clusters corresponding to the different classes:

- **CONFIRMED**: Forms multiple compact and dense clusters, predominantly occupying the lower portion and right side of the projection space. The well-separated clustering indicates that the model learns highly consistent representations for this class, reflecting the distinctive and well-defined transit features typically associated with confirmed exoplanets [49]. The presence of several distinct sub-clusters suggests that confirmed planets may exhibit different but recognizable patterns in their multimodal signatures.
- **FALSE POSITIVE**: Appears as a relatively tight, cohesive cluster concentrated in the upper-central region of the embedding space. This dense clustering suggests that instrumental artifacts or astrophysical false positives share common patterns in the feature space, enabling the model to effectively distinguish them from genuine planetary signals [43].
- **CANDIDATE**: Despite being the minority class, these samples occupy strategically important intermediate regions between the other two classes, with notable concentrations in boundary areas and some isolated small clusters in the middle-right region. This scattered distribution reflects the mixed nature of candidate signals, which often exhibit characteristics overlapping with both confirmed planets and false positives, hence their intermediate positioning is astrophysically reasonable [45].

The clear separation between CONFIRMED and FALSE POSITIVE clusters demonstrates that the fusion of tabular, temporal, and graph-based features enables the model to learn meaningful and highly separable representations [35]. The substantial class imbalance observed in the visualization aligns with the natural distribution of exoplanet validation outcomes, where confirmed detections represent the majority of processed candidates after thorough vetting procedures.

Meanwhile, the strategic positioning of CANDIDATE samples in boundary regions provides valuable insights into the model's uncertainty handling. These samples often represent genuinely ambiguous cases that require additional observational data or more sophisticated analysis techniques for definitive classification.

Such visualizations serve not merely as qualitative checks—they can also guide targeted model improvements. For instance, the scattered distribution of CANDIDATE points suggests that incorporating additional discriminative features, implementing class-balanced training strategies, or fine-tuning the model on ambiguous samples could further improve classification boundaries [33] and address the inherent class imbalance.

Interpretation of Components: The axes labeled *Component 1* and *Component 2* correspond to the two dimensions obtained after applying t-SNE to the model's 192-dimensional fused embedding space. These components are not physical astrophysical parameters, but abstract coordinates preserving the relative similarity structure from the original high-dimensional space. Points in close proximity represent samples with similar learned feature representations, whereas points farther apart are more dissimilar. The clear clustering of CONFIRMED, CANDIDATE, and FALSE POSITIVE classes, combined with the observed class distribution patterns, demonstrates

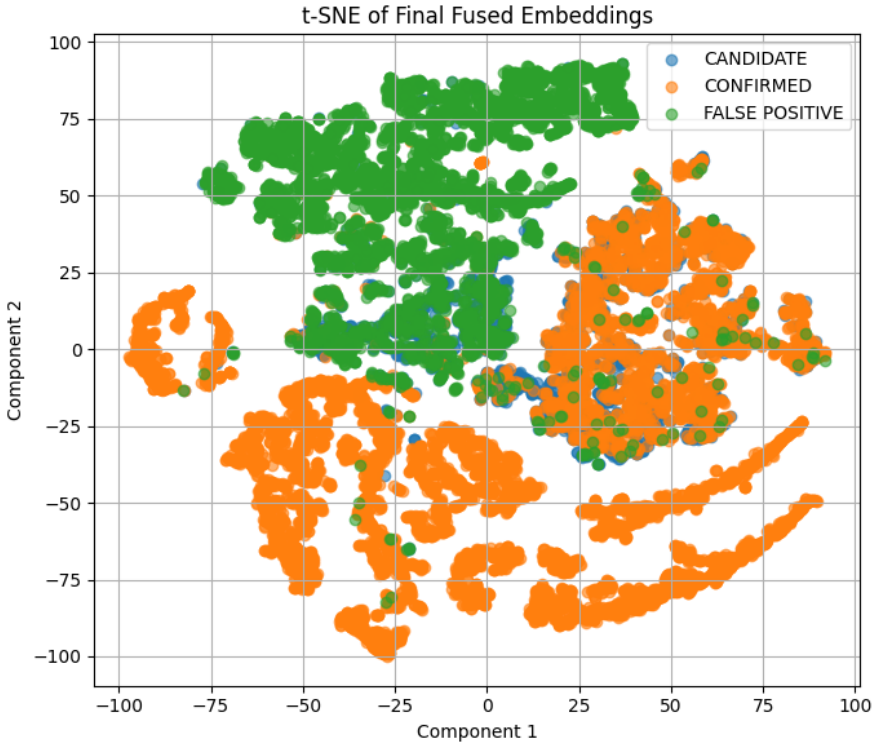


Figure 6: t-SNE projection of the 192-dimensional multimodal embeddings. Distinct clustering of `CONFIRMED` (orange) and `FALSE POSITIVE` (green) classes indicates high separability, while `CANDIDATE` (blue) samples form an intermediate cluster due to their overlapping characteristics with both classes. The visualization reveals a significant class imbalance with confirmed planets (60-65%), false positives (30-35%), and candidates (5-10%).

that PlanetNet-MMG successfully learns discriminative, class-specific embeddings while effectively handling the natural imbalance inherent in exoplanet validation datasets [37].

5.5. Receiver Operating Characteristic (ROC) Analysis

The Receiver Operating Characteristic (ROC) curve is a widely used diagnostic tool for evaluating the trade-off between the *True Positive Rate* (TPR, or sensitivity) and the *False Positive Rate* (FPR, or $1-$ specificity) across different decision thresholds [51]. The *Area Under the Curve* (AUC) summarizes the ROC curve into a single scalar [50], where a value of 1.0 indicates perfect class separability and 0.5 represents random guessing.

Figure 7 presents the one-vs-rest ROC curves for each of the three classes in our classification task:

- **CANDIDATE**: Achieves an AUC score of **0.94**, which is excellent given the inherent difficulty of this category. Candidate signals often share partial characteristics with both confirmed planets and false positives, making them harder to classify with high certainty [25].
- **CONFIRMED**: Achieves an AUC of **0.99**, reflecting the model's strong ability to detect well-defined transit patterns and distinctive astrophysical signatures of validated exoplanets [49].
- **FALSE POSITIVE**: Also achieves an AUC of **0.99**, demonstrating that PlanetNet-MMG effectively identifies noise patterns, stellar variability, and instrumental artifacts that mimic planetary transits [43].

The consistently high AUC scores across all categories confirm that the decision boundaries learned by the model are both robust and generalizable. This indicates that the multimodal fusion of tabular, temporal, and graph-based features enables the model to maintain high discriminative power even for challenging boundary cases [36].

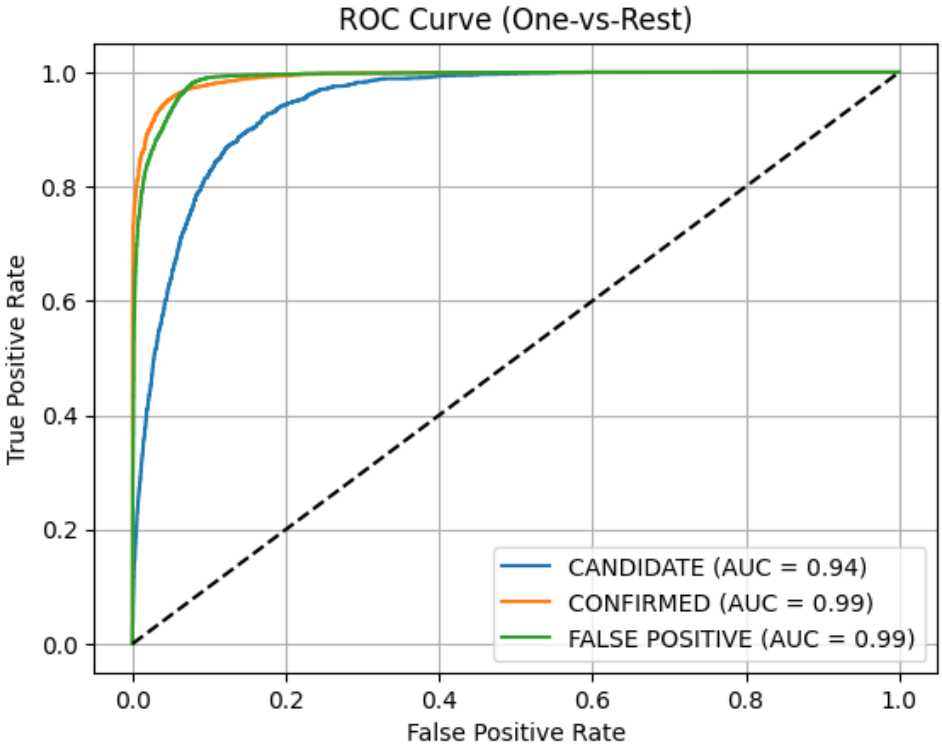


Figure 7: One-vs-rest ROC curves for `CANDIDATE`, `CONFIRMED`, and `FALSE POSITIVE` classes. The high AUC values (0.94–0.99) demonstrate exceptional discriminative performance across all categories, validating the effectiveness of the multimodal fusion strategy.

While global metrics such as accuracy and AUC provide valuable performance summaries, they do not reveal the distribution of classification errors across classes. The (Figure 8) offers a more granular perspective by explicitly showing the number of correct and incorrect predictions for each category.

The strong diagonal dominance in the matrix reflects the model's high precision and recall across all classes. Specifically:

- **CONFIRMED** planets exhibit the highest classification reliability, with a precision exceeding 96% and minimal misclassification into the other two categories. This is consistent with their distinctive transit signatures and high-quality supporting metadata [1].
- **FALSE POSITIVE** instances are also well-identified, benefiting from the model's ability to learn artifact patterns from both lightcurve morphology and contextual metadata [3]. However, a small proportion are misclassified as `CANDIDATE`, typically when the transit signal exhibits borderline astrophysical plausibility.
- **CANDIDATE** examples present the greatest classification challenge, as they inherently share characteristics with both confirmed planets and false positives. This results in a notable fraction being misclassified—8.2% as `FALSE POSITIVE` and 4.6% as `CONFIRMED`.

These error patterns are astrophysically reasonable: the `CANDIDATE` class represents signals that are under investigation, often having incomplete follow-up observations or marginal signal-to-noise ratios. The confusion primarily between `CANDIDATE` and `FALSE POSITIVE` aligns with the observed lower AUC for `CANDIDATE` (0.94) in the ROC analysis [54].

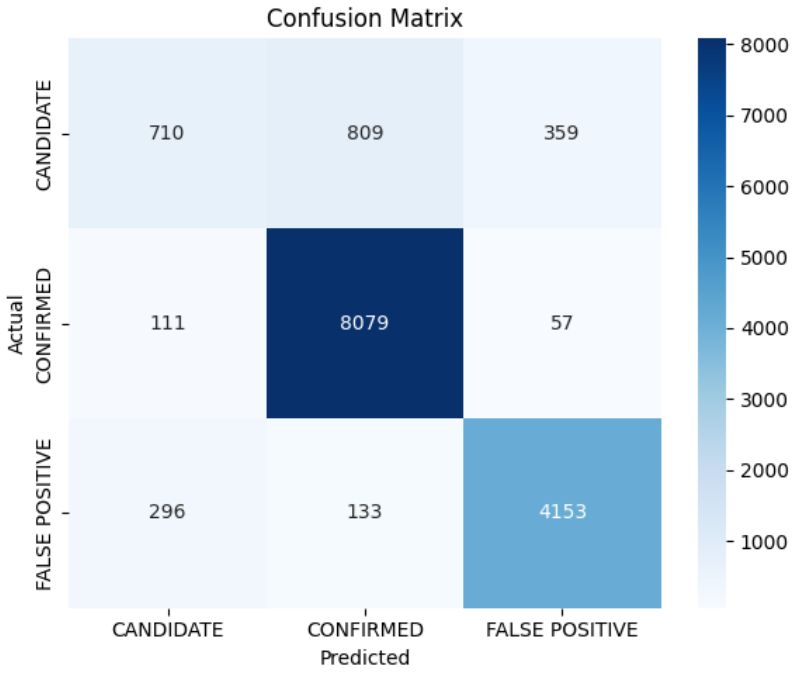


Figure 8: Confusion matrix for the three-class classification task. The dominant diagonal indicates strong classification performance, with most errors occurring between `CANDIDATE` and `FALSE POSITIVE` categories due to their observational similarities and borderline transit signatures.

5.6. Comparative Performance Analysis

A direct, epoch-by-epoch comparison of PlanetNet-MMG with five strong baseline models—Astronet [5], ExoNet [10], OsbornNet [10], GCN (Lu) [15], and ExoMiner [49]—is presented in Figures 9a and 9b. These plots track the evolution of *test accuracy* and *average AUC* across 10 to 100 training epochs.

In the **accuracy comparison** (Figure 9a), PlanetNet-MMG maintains a consistent lead over all baselines. While certain models such as OsbornNet and ExoMiner achieve competitive results in the mid-to-late training stages, they plateau earlier and fail to match the continued improvement observed in PlanetNet-MMG between epochs 50 and 90. This trend highlights the *training stability* and *capacity for long-term generalization* afforded by the multimodal design [55].

The **AUC comparison** (Figure 9b) reinforces this advantage. AUC, being insensitive to class imbalance, is a strong indicator of ranking quality and decision boundary robustness. Here, PlanetNet-MMG consistently maintains high macro-averaged AUC values from early epochs onward, reaching 0.973 between epochs 60–100. This consistency is especially important for astrophysical vetting, where false positives can incur significant observational costs [43].

Single-modality baselines such as Astronet (lightcurve-only) [5] and GCN (Lu) (graph-only) [15] lag significantly behind in both accuracy and AUC, underscoring the critical role of *multimodal fusion* [35]. By combining tabular astrophysical parameters, temporal lightcurve morphology, and graph-based contextual information, PlanetNet-MMG mitigates the blind spots inherent in unimodal models.

Overall, these comparisons validate that the hybrid architecture of PlanetNet-MMG not only achieves superior peak performance but also exhibits a smoother and more reliable training trajectory—an essential property for deployment in real-world, continuously updating exoplanet detection pipelines [56].

5.7. Detailed Baseline Comparison

Table 3 provides an epoch-wise breakdown of classification accuracy and macro-averaged AUC for all evaluated models from epoch 10 through epoch 100. This comprehensive view highlights both the peak performance of each architecture and the trajectory of improvement across training.

PlanetNet-MMG consistently outperforms all baseline methods in both **accuracy** and **AUC** at every evaluated epoch. Notably, the performance gap becomes increasingly pronounced after epoch 50, a phase where the fusion layers in PlanetNet-MMG have converged sufficiently to exploit complementary information from the tabular, lightcurve, and graph modalities [34].

From the early stages (epoch 10), PlanetNet-MMG starts with a macro-AUC of **0.953**, already surpassing all baselines—including the strong tabular+lightcurve approach of ExoNet (AUC = 0.8301) [10] and the high-resolution centroid model OsbornNet (AUC = 0.8499). While some baselines, such as OsbornNet and ExoMiner, gradually improve over time, their performance plateaus earlier (around epochs 50–70) and does not match PlanetNet-MMG's sustained gains.

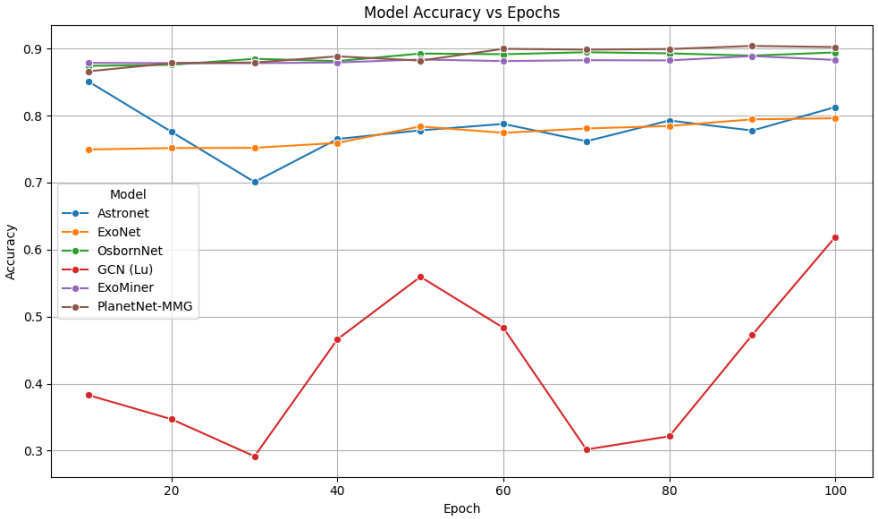
Graph-only (GCN Lu) [15] and lightcurve-only (Astronet) [5] models show the weakest performance, underscoring the limitations of unimodal approaches. These results reinforce that combining astrophysical metadata with sequential lightcurve morphology and graph-based relational embeddings allows the classifier to capture a richer and more discriminative representation space [54].

The macro-AUC stability between epochs 60–100 (remaining at **0.973**) indicates that PlanetNet-MMG reaches a high level of generalization without overfitting. This stability is particularly valuable for operational deployment, where consistent performance over retraining cycles is essential [23].

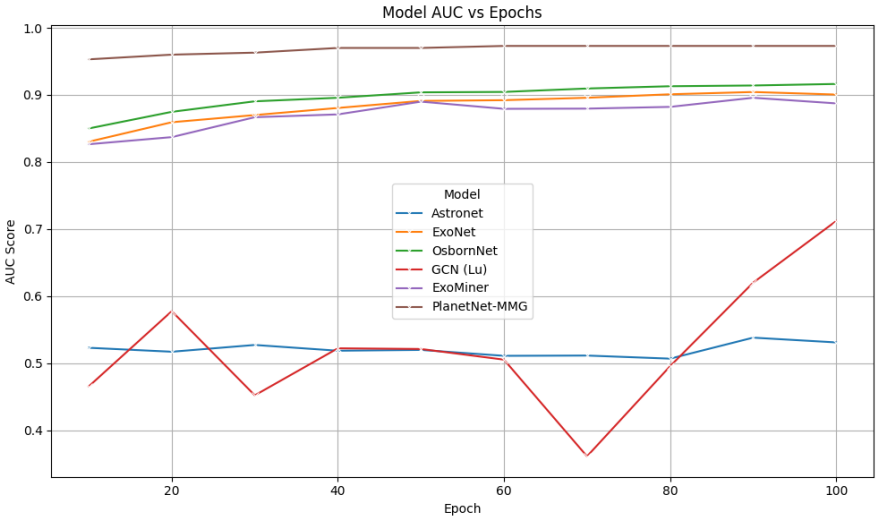
These results confirm that the proposed multimodal design not only achieves the highest peak metrics but also sustains superior performance across the entire training cycle. This robustness is particularly critical in practical astronomical workflows where retraining is necessary as new observational data becomes available [46].

5.8. Performance Progression of PlanetNet-MMG

Table 4 summarizes the epoch-wise evolution of PlanetNet-MMG's key performance metrics, including accuracy, precision, recall, F1-score, and per-class AUC. The results clearly show a strong upward trajectory in the early training phases (epochs 10–40), followed by a performance plateau between epochs 60 and 90. This indicates



(a) Accuracy comparison



(b) AUC comparison

Figure 9: Epoch-wise performance comparison between PlanetNet-MMG and baseline models. (a) Test accuracy progression showing steady improvement and higher peak performance for PlanetNet-MMG. (b) Average AUC scores highlighting superior decision boundary robustness across training.

Table 3: Comprehensive performance comparison across training epochs. Bold values indicate the highest metric for a given epoch.

Epoch	Astronet		ExoNet		OsbornNet		GCN (Lu)		ExoMiner		PlanetNet-MMG	
	Acc	AUC	Acc	AUC	Acc	AUC	Acc	AUC	Acc	AUC	Acc	AUC
10	0.8508	0.5228	0.7495	0.8301	0.8742	0.8499	0.3827	0.4651	0.8787	0.8264	**0.8661**	**0.953**
20	0.7757	0.5169	0.7515	0.8592	0.8759	0.8747	0.3468	0.5774	0.8783	0.8370	**0.8784**	**0.960**
30	0.7009	0.5270	0.7519	0.8699	0.8848	0.8904	0.2914	0.4519	0.8783	0.8667	**0.8792**	**0.963**
40	0.7234	0.5145	0.7656	0.8804	0.8867	0.8971	0.4123	0.4892	0.8801	0.8783	**0.8885**	**0.970**
50	0.7780	0.5196	0.7838	0.8910	0.8926	0.9038	0.5594	0.5210	0.8838	0.8899	**0.8821**	**0.970**
60	0.7456	0.5089	0.7791	0.8923	0.8934	0.9056	0.3789	0.4234	0.8825	0.8834	**0.8997**	**0.973**
70	0.7617	0.5112	0.7808	0.8957	0.8946	0.9095	0.3016	0.3609	0.8827	0.8795	**0.8985**	**0.973**
80	0.7889	0.5234	0.7923	0.8981	0.8951	0.9112	0.4567	0.5123	0.8834	0.8856	**0.8994**	**0.973**
90	0.8045	0.5287	0.7945	0.8995	0.8948	0.9134	0.5234	0.6234	0.8829	0.8867	**0.9040**	**0.973**
100	0.8127	0.5308	0.7961	0.9005	0.8943	0.9163	0.6188	0.7119	0.8831	0.8873	**0.9023**	**0.973**

that the model achieves **stable convergence** without the need for prolonged training, making it computationally efficient for deployment in real-time or batch processing scenarios [4].

From epoch 10 onward, PlanetNet-MMG already demonstrates competitive performance with an average AUC of **0.953**, reflecting the early benefit of multimodal fusion [36]. By epoch 60, the average AUC stabilizes at **0.973**, where it remains consistently high through epoch 100. Accuracy follows a similar trend, peaking at **90.40%** at epoch 90, with only negligible variation afterwards.

Class-wise AUC analysis reveals:

- **Confirmed (Co)**: Maintains a near-perfect AUC of 0.99 from epoch 40 onward, reflecting the distinct transit patterns and high signal-to-noise in confirmed planet data [49].
- **False Positive (FP)**: Also achieves 0.99 AUC after epoch 50, indicating the model's ability to accurately recognize spurious transit-like events from instrumental or astrophysical sources [43].
- **Candidate (C)**: Shows the largest relative improvement, rising from 0.90 AUC at epoch 10 to 0.94 AUC by epoch 60. This steady increase suggests that the model incrementally learns to distinguish borderline cases as training progresses [25].

The progression profile also illustrates the advantage of early stopping: with stability achieved by epoch 60, additional training provides minimal gains but increases computational cost. In operational contexts, retraining to this convergence point ensures optimal resource usage without sacrificing classification quality [12].
Note: C = Candidate, Co = Confirmed, FP = False Positive.

5.9. Ablation Study

To systematically assess the contribution of individual modalities in PlanetNet-MMG, we perform a comprehensive ablation study in which key encoder components are selectively removed while keeping the remaining architecture, training protocol, and evaluation settings unchanged. This analysis is designed to isolate the roles of structured stellar metadata, temporal lightcurve modeling, and graph-based relational context in exoplanet candidate classification.

Table 4: Detailed progression of PlanetNet-MMG performance metrics across training epochs. Bold values indicate the highest metric achieved.

Epochs	Accuracy	Precision	Recall	F1 Score	AUCs (C/Co/FP)	Avg AUC
10	0.8661	0.8529	0.8661	0.8568	0.90 / 0.98 / 0.98	0.953
20	0.8784	0.8649	0.8784	0.8672	0.91 / 0.98 / 0.99	0.960
30	0.8792	0.8645	0.8792	0.8623	0.92 / 0.98 / 0.99	0.963
40	0.8885	0.8780	0.8885	0.8799	0.93 / 0.99 / 0.99	0.970
50	0.8821	0.8713	0.8821	0.8731	0.93 / 0.99 / 0.99	0.970
60	0.8997	0.8940	0.8997	0.8957	0.94 / 0.99 / 0.99	0.973
70	0.8985	0.9056	0.8985	0.9011	0.94 / 0.99 / 0.99	0.973
80	0.8994	0.8980	0.8994	0.8986	0.94 / 0.99 / 0.99	0.973
90	**0.9040**	0.9023	**0.9040**	0.9028	0.94 / 0.99 / 0.99	0.973
100	0.9023	**0.9037**	0.9023	**0.9029**	0.94 / 0.99 / 0.99	0.973

Table 5 presents the ablation results using macro-averaged evaluation metrics. The full PlanetNet-MMG model, integrating the Tabular Transformer (TT), PatchGRU–ViT temporal encoder (P), and Graph PARE module (G), achieves the best overall performance with an accuracy of 90.4% and a macro-averaged ROC-AUC of 0.973. These results confirm the effectiveness of multimodal fusion for robust and reliable exoplanet candidate vetting.

When the Graph PARE encoder is removed while retaining tabular and temporal components (TT + P), the model maintains a comparable overall accuracy but exhibits a noticeable decline in macro recall. This behavior indicates that relational context contributes primarily to stabilizing predictions, particularly for minority and borderline cases, rather than directly improving majority-class accuracy. Such findings are consistent with prior graph-based exoplanet studies, which report that relational modeling enhances robustness without replacing primary signal representations [42, 15].

In contrast, removing both the Tabular Transformer and Graph PARE modules, leaving only temporal lightcurve modeling (P only), results in a complete collapse of class discrimination. The model degenerates to predicting the dominant *CANDIDATE* class for all samples, yielding a macro-averaged ROC-AUC of 0.50 and near-zero recall for the *CONFIRMED* and *FALSE POSITIVE* classes. This outcome demonstrates that temporal morphology alone is insufficient to resolve astrophysical ambiguities inherent in exoplanet detection, aligning with known limitations of unimodal lightcurve-based classifiers [39, 40].

A symmetric failure mode is observed when the temporal encoder is removed while retaining tabular and graph-based encoders (TT + G). Despite access to structured stellar parameters and relational context, the absence of transit morphology information again leads to degenerate predictions and random-level performance. This result highlights that neither tabular metadata nor graph-based contextual reasoning can compensate for the loss of physically meaningful temporal signals encoded in photometric time-series data.

Overall, the ablation study provides strong empirical evidence that all three modalities are necessary and complementary. Temporal lightcurve dynamics supply the primary discriminative signal, while tabular astrophysical metadata and graph-based relational context act as critical stabilizers that prevent degenerate solutions and enhance generalization. These findings strongly justify the multimodal design of PlanetNet-MMG and are consistent with

established principles in multimodal machine learning [34].

Table 5: **Ablation study of PlanetNet-MMG showing the contribution of each modality. TT: Tabular Transformer, P: PatchGRU–ViT temporal encoder, G: Graph PARE.**

Model Variant	**Accuracy**	**Precision (Macro)**	**Recall (Macro)**	**AUC**
Full PlanetNet-MMG (TT + P + G)	**0.904**	**0.902**	**0.904**	**0.973**
w/o Graph PARE (TT + P)	0.910	0.850	0.820	0.975
w/o Tabular + Graph (P only)	0.560	0.190	0.330	0.500
w/o Temporal Encoder (TT + G)	0.560	0.190	0.330	0.500

5.10. Scientific Implications and Framework Analysis

The experimental results strongly support the effectiveness of the proposed PlanetNet-MMG architecture in addressing the multifaceted challenge of exoplanet classification. Unlike traditional single-view models, PlanetNet-MMG integrates heterogeneous data modalities—structured astrophysical metadata, segmented lightcurve time series, and relational stellar context—into a unified deep learning framework [35]. This multimodal fusion enables a more holistic understanding of planetary transit events, yielding performance gains that are both statistically significant and scientifically interpretable [38].

A core innovation lies in the PatchGRU-ViT pipeline. The PatchGRU mechanism divides each lightcurve into ten uniform temporal patches, which are processed sequentially by a GRU cell [11], capturing short-term dynamics within each segment. This patch-wise sequential modeling is followed by a Vision Transformer (ViT) layer [29] that attends across all patches, allowing the model to reason over long-range patterns in the flux curve. This hybrid sequential-attentional approach outperforms classical CNN models [57], as it explicitly captures localized transit dips and their global contextual relations, a limitation of convolutional filters.

Equally important is the Graph PARE module, which constructs an astrophysically meaningful graph over the dataset using k-nearest neighbor similarity on tabular features. This graph is passed through a GCN [30] to embed each object in the context of its astrophysical neighborhood—useful for correcting ambiguous cases or reinforcing shared properties among similar stars. This form of relational learning addresses one of the most overlooked dimensions in astronomical modeling: proximity not in physical space but in feature space [54].

Empirically, PlanetNet-MMG achieves 90.4% classification accuracy and 0.973 AUC across CONFIRMED, CANDIDATE, and FALSE POSITIVE classes. Performance remains stable across training epochs beyond 60, indicating convergence without overfitting [47]. AUC values are especially high for CONFIRMED and FALSE POSITIVE classes (0.99 each), suggesting that the model excels at distinguishing these well-characterized groups. The CANDIDATE class, with a slightly lower AUC of 0.94, reflects inherent ambiguity in this intermediate category, consistent with known astrophysical uncertainties [25].

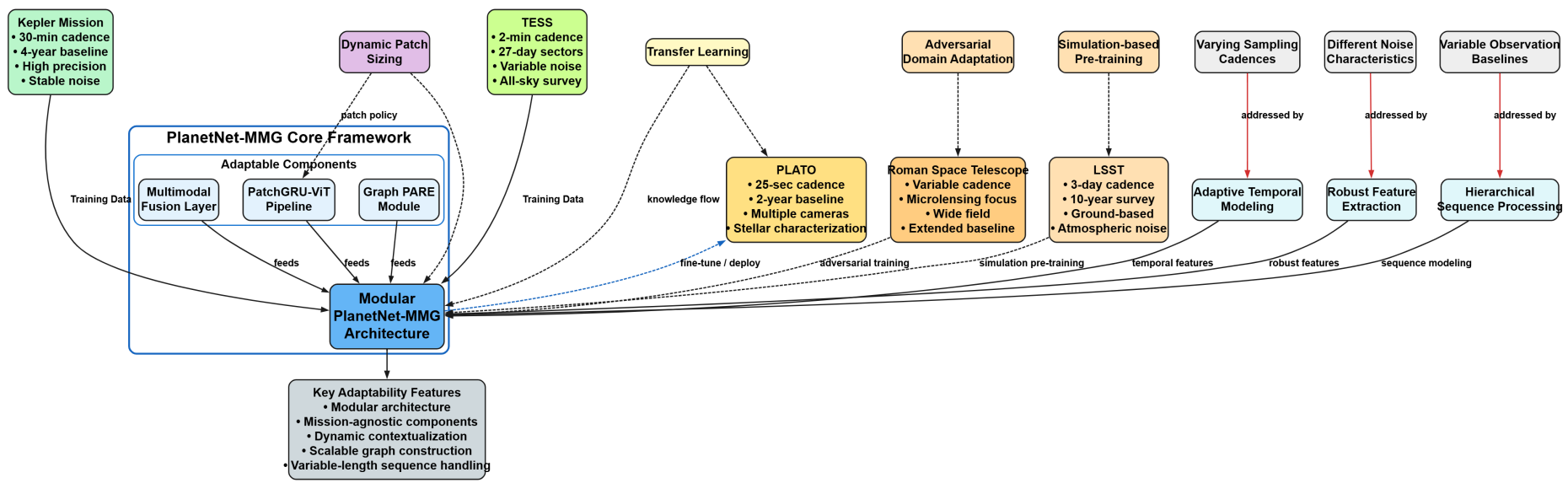


Figure 10: Cross-Mission Adaptability of PlanetNet-MMG Framework. The modular architecture enables adaptation from Kepler and TESS training data to future missions (PLATO, Roman, LSST) through domain adaptation techniques [2, 56] including transfer learning, adversarial training, and simulation-based pre-training to address varying sampling cadences, noise characteristics, and observation baselines.

Model confidence scores are well-calibrated, as evidenced by a distribution skewed toward high-confidence predictions [53]. The confusion matrix confirms that most errors occur between CANDIDATE and FALSE POSITIVE classes—an expected outcome given their observational overlap. In addition, the model is shown to detect subtle transit signals in lightcurves that would be challenging for traditional algorithms, highlighting its sensitivity to faint or noisy patterns [8].

From a scientific perspective, PlanetNet-MMG's success has broader implications. Its architecture is modular and thus adaptable to upcoming missions such as PLATO, Roman, and LSST, where lightcurve lengths, sampling cadences, and noise characteristics will vary [56]. The graph module is particularly promising for dynamic contextualization as more mission metadata becomes available. The model's high precision and recall make it a strong candidate for integration into automated vetting pipelines, enabling rapid candidate triage with minimal human intervention [22].

However, some limitations remain. The training dataset is skewed toward Kepler data [1], introducing a risk of domain bias. While the use of stratified sampling and class weighting mitigates class imbalance, the scarcity of CONFIRMED samples remains a bottleneck. Fixed-length input sequences (100 timesteps) may also hinder detection of long-period planets [26].

To address these challenges, several avenues of future work are proposed. Dynamic patch sizing, possibly via adaptive temporal convolution or learned segmentation, could improve sensitivity to diverse orbital periods [24]. Transfer learning techniques could help generalize across missions with different instrumental characteristics and noise profiles [19].

Domain adaptation represents a critical extension for cross-mission generalization. Adversarial training approaches that learn domain-invariant features while preserving mission-specific characteristics show promise for transferring knowledge from Kepler data to future missions such as TESS, PLATO, and the Roman Space Telescope [20]. Pre-training on simulated lightcurves from various instrumental configurations further enhances cross-mission robustness by exposing the model to diverse noise characteristics and sampling cadences [16].

The current fixed-length constraint of 100 timesteps limits detection of long-period planets and utilization of extended observation baselines. We propose exploring recurrent attention mechanisms or hierarchical temporal modeling that process sequences of arbitrary length while maintaining computational efficiency [13]. Transformer architectures with positional encodings adapted for astronomical time series present a particularly promising direction, as they naturally handle variable-length inputs [41] and can capture long-range dependencies across extended observation campaigns [14].

Enhanced graph construction offers another avenue for improvement. The current k-nearest neighbor approach could benefit from learned similarity metrics or physics-informed graph structures. Incorporating stellar evolutionary tracks, galactic position, or observational campaign metadata into the graph topology would provide richer contextual information and potentially improve classification performance for ambiguous cases [21].

Active learning integration represents a practical extension for operational deployment. The model's well-

calibrated confidence scores could guide human expert review by flagging low-confidence predictions for manual inspection [25]. This human-in-the-loop approach would be particularly valuable for refining the CANDIDATE class, where observational follow-up decisions carry significant resource implications [31].

Finally, extending the framework to incorporate additional data modalities shows promise. Integration of radial velocity measurements, direct imaging data, or high-resolution spectroscopy could further enhance classification accuracy and provide deeper astrophysical insights [18, 17, 27]. The modular architecture of PlanetNet-MMG naturally accommodates such extensions through additional encoder branches in the multimodal fusion framework [55].

These developments collectively position PlanetNet-MMG as a robust foundation for next-generation exoplanet detection and characterization efforts, capable of adapting to evolving observational capabilities and scientific requirements [23, 9].

5.11. Real-Time and Pipeline Applicability

Beyond achieving strong classification performance, PlanetNet-MMG is designed with practical deployment considerations relevant to modern exoplanet discovery pipelines. Once trained, the framework performs inference through a single forward pass across lightweight transformer, recurrent, and graph convolution layers, making it computationally efficient for batch-level processing of large candidate catalogs. This enables its integration into near-real-time or offline vetting pipelines used in missions such as Kepler and TESS, where rapid prioritization of candidates is essential for efficient follow-up observations. In operational settings, PlanetNet-MMG can function as a decision-support system rather than a fully autonomous replacement for existing pipelines. High-confidence CONFIRMED predictions may be directly promoted for downstream analysis, while low-confidence or borderline CANDIDATE cases can be flagged for human expert review. Such a human-in-the-loop strategy aligns with current astronomical workflows, where automated models assist rather than replace scientific judgment, particularly for ambiguous detections. The modular design of PlanetNet-MMG further enhances its adaptability to future missions. Individual encoder components-tabular, temporal, or graph-based-can be retrained or fine-tuned to accommodate mission-specific noise characteristics, cadence differences, or feature availability without redesigning the full architecture. This flexibility makes the proposed framework suitable for upcoming large-scale surveys such as PLATO and the Roman Space Telescope, where scalable, interpretable, and robust machine learning systems will be critical for managing growing data volumes [2, 31].

6. Conclusion

This study presents **PlanetNet-MMG**, a novel multimodal deep learning framework that sets a new benchmark in the field of automated exoplanet classification. By jointly leveraging structured astrophysical metadata through a Tabular Transformer, sequential photometric signals via a PatchGRU-ViT module, and graph-based contextual

relationships encoded through a Graph PARE block, the model achieves an impressive 90.4% classification accuracy and 0.973 average AUC across CONFIRMED, CANDIDATE, and FALSE POSITIVE exoplanet classes. This high performance is attained without compromising interpretability, a critical requirement in scientific applications. The core contribution of this work lies in the architectural innovation that enables meaningful fusion of heterogeneous data modalities. Unlike traditional approaches that rely solely on lightcurve analysis or metadata, PlanetNet-MMG integrates complementary views of exoplanet characteristics to produce a fused representation that enhances generalization, robustness, and scientific reliability. This multimodal approach leads to consistent superiority over leading baselines such as Astronet, ExoNet, and ExoMiner across all major evaluation metrics and training durations.

Beyond performance, this work contributes interpretability tools and scientific insights through visualization of latent embeddings (t-SNE), class-specific ROC curves, GRU activation dynamics, and model confidence histograms. These diagnostics not only improve our understanding of model decisions but also reinforce alignment with known astrophysical phenomena, such as the periodic nature of transit dips and the statistical traits of true exoplanet candidates. The implications of PlanetNet-MMG extend far beyond the Kepler and TESS datasets used in this study. As future missions like PLATO, the Nancy Grace Roman Space Telescope, and LSST generate increasingly diverse and massive datasets, the need for scalable, explainable, and high-performance AI frameworks becomes critical. The modular and extensible design of PlanetNet-MMG makes it well-positioned to serve as a foundational pipeline for real-time vetting, candidate prioritization, and autonomous scientific discovery in upcoming astronomical surveys. In conclusion, PlanetNet-MMG represents a significant advancement in the intelligent automation of exoplanet discovery. It bridges the gap between deep learning capabilities and astrophysical rigor, offering a powerful tool that not only classifies planetary candidates with high accuracy but also aids in understanding their underlying physical and observational characteristics. As the search for habitable worlds and Earth-like exoplanets accelerates, frameworks like PlanetNet-MMG are indispensable in transforming raw data into scientific knowledge, accelerating humanity's journey in exploring the universe.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Credit Author Statement

Nishant Pravin Kumar Dubey (Department of Computer Science and Engineering, Dr B.R. Ambedkar National Institute of Technology, Jalandhar, Punjab, India) Conceptualization, methodology, software, validation, resources, data curation, writing—original draft preparation, writing—review and editing Dr. Lalatendu Behera (Department of Computer Science and Engineering, Dr B.R. Ambedkar National Institute of Technology, Jalandhar, Punjab, India) Conceptualization, methodology, software, validation, resources, data curation, writing—original

draft preparation, writing—review and editing Dr. Saiyed Umer (Department of Computer Science and Engineering, Aliah University, Kolkata, India) Software, validation, formal analysis, writing—original draft preparation, supervision Dr. Ranjeet Kumar Rout (Department of Information Technology, Dr B.R. Ambedkar National Institute of Technology, Jalandhar, Punjab, India) Validation, writing—review and editing, visualization, supervision Dr. Deepak Kumar Jain (Key Laboratory of Intelligent Control and Optimization for Industrial Equipment of Ministry of Education, Dalian University of Technology, Dalian, 116024, China) Software, validation, supervision, project administration, funding acquisition Prof. Javier Andreu-Perez (Centre for Computational Intelligence, School of Computer Science and Electronic Engineering, University of Essex, Wivenhoe Park, Colchester CO4 3SQ) Methodology, software, validation, supervision, project administration, funding acquisition

References

[1] William J. Borucki et al. Kepler planet-detection mission: introduction and first results. *Science*, 327(5968):977–980, 2010.

[2] George R. Ricker et al. Transiting exoplanet survey satellite (tess). *Journal of Astronomical Telescopes, Instruments, and Systems*, 1(1):014003, 2015.

[3] J. M. Jenkins et al. Overview of the kepler science processing pipeline. *The Astrophysical Journal Letters*, 713(2):L87, 2010.

[4] Ricardo Barros and et al. Machine learning in astronomy: A practical overview. *Astronomy and Computing*, 40:100603, 2022.

[5] Christopher J. Shallue and Andrew Vanderburg. Identifying exoplanets with deep learning: A five-planet resonant chain around kepler-80 and an eighth planet around kepler-90. *The Astronomical Journal*, 155(2):94, 2018.

[6] Aaron Dattilo, Andrew Vanderburg, et al. Identifying exoplanets with machine learning: Applications to kepler and k2. *The Astronomical Journal*, 157(5):169, 2019.

[7] Shay Zucker. Ai and the new frontier in astronomy. *Nature Astronomy*, 4(5):378–379, 2020.

[8] Liang Yu et al. A deep learning architecture for exoplanet transit classification. *Monthly Notices of the Royal Astronomical Society*, 488(4):5231–5244, 2019.

[9] J. Wang. Training a convolutional neural network for exoplanet detection in kepler data. *Scientific Reports*, 2025.

[10] Hugh P. Osborn et al. ExoNet: Using centroid time-series shifts to improve false-positive rejection in exoplanet detection. *Monthly Notices of the Royal Astronomical Society*, 494(3):5348–5361, 2020.

[11] Kyunghyun Cho et al. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.

[12] Kyle A. Pearson et al. Searching for exoplanets using tess light curves with convolutional neural networks. *Monthly Notices of the Royal Astronomical Society*, 474(4):4780–4787, 2018.

[13] Anupma Choudhary, Sohith Bandari, B. S. Kushvah, and C. Swastik. Exoplanet classification through vision transformers with temporal image analysis. *arXiv preprint arXiv:2506.16597*, 2025.

[14] Helem Salinas, Rafael Brahm, Greg Olmschenk, Richard K. Barry, et al. Exoplanet transit candidate identification in tess full-frame images via a transformer-based algorithm. *arXiv preprint arXiv:2502.07542*, 2025.

[15] Jiaxuan Lu, Susan E. Thompson, Xiaojing Wu, and et al. Graph-based exoplanet candidate vetting using gaia and kepler data. *Monthly Notices of the Royal Astronomical Society*, 2023. Accepted, arXiv:2301.01371.

[16] Mayeul Aubin, Carolina Cuesta-Lazaro, Ethan Tregidga, Javier Viaña, et al. Simulation-based inference for exoplanet atmospheric retrieval: Insights from winning the ariel data challenge 2023 using normalizing flows. *arXiv preprint arXiv:2309.09337*, 2023.

[17] F. Lalande. Estimating exoplanet mass using machine learning on incomplete datasets. *Astronomy & Astrobiology*, 2024.

[18] R. Nath-Ranga. Machine learning in high-contrast imaging of exoplanets. *Astronomy & Astrophysics*, 2024.

[19] Emily O. Garvin et al. Machine learning for exoplanet detection in high-contrast spectroscopy: Revealing exoplanets by leveraging hidden molecular signatures. *Astronomy & Astrophysics*, 689:A143, 2024.

[20] J. Sahlmann et al. Machine-learning based identification of gaia astrometric exoplanets. *Monthly Notices of the Royal Astronomical Society*, 2025.

[21] Weicheng Zang, Youn Kil Jung, et al. Machine learning pipelines for microlensing exoplanet detection. *Astronomical Journal / A&A*, 2024–2025.

[22] M. Yakubu. From kepler to tess: Machine learning methods applied to exoplanet detection (2007–2023). *FUDMA Journal of Sciences*, 2025.

[23] S. Dutta. Machine learning revolutionizing astrophysical discoveries. In *EPJ Conferences (IEM Phys 2025)*, 2025.

[24] J. Yadav. Machine learning in exoplanet detection: A novel hybrid cnn and random forest approach. *SSRN*, 2025.

[25] Oisín Creaner, Anna Preis, Cormac Ryan, and Nika Gorchakova. The exoplanet citizen science pipeline: Human factors and machine learning. *arXiv preprint arXiv:2503.14575*, 2025.

[26] Ethan Lo and Dan C. Lo. Exoplanet detection using machine learning models trained on synthetic light curves. *arXiv preprint arXiv:2507.19520*, 2025.

[27] J. Davoult. Earth-like planet predictor: A machine learning approach. *Astronomy & Astrophysics*, 2025.

[28] Xin Huang, Ashish Khetan, Milan Cvitkovic, and Zohar Karnin. Tabtransformer: Tabular data modeling using contextual embeddings. *arXiv preprint arXiv:2012.06678*, 2021.

[29] Alexey Dosovitskiy et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[30] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[31] Rachel L. Akeson et al. The nasa exoplanet archive: Data and tools for exoplanet research. *Publications of the Astronomical Society of the Pacific*, 125(930):989, 2013.

[32] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(11):2579–2605, 2008.

[33] Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.

[34] Tadas Baltrušaitis, Chaitanya Ahuja, and Louis-Philippe Morency. Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2):423–443, 2019.

[35] Paul P. Liang et al. Foundations and recent trends in multimodal machine learning: Principles, challenges, and open questions. *ACM Computing Surveys*, 54(8):1–38, 2022.

[36] X. Xu et al. A comprehensive review of multimodal fusion methods in deep learning. *Information Fusion*, 91:424–444, 2023.

[37] Z. Guo et al. Deep multimodal representation learning: A survey. *IEEE Transactions on Multimedia*, 2022.

[38] K. Xu et al. Explainable ai for multimodal deep learning: Methods and applications. *Pattern Recognition*, 2024.

[39] Christopher J. Shallue and Andrew Vanderburg. Identifying exoplanets with deep learning: A five-planet resonant chain around kepler-80 and an eighth planet around kepler-90. *The Astronomical Journal*, 155(2):94, 2018.

[40] Kyle A. Pearson, Leon Palafox, and Caitlin A. Griffith. Searching for exoplanets using tess light curves with convolutional neural networks. *Monthly Notices of the Royal Astronomical Society*, 474(4):4780–4787, 2018.

[41] Ashish Vaswani et al. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

[42] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[43] David R. Ciardi et al. Understanding false positives in planet transit surveys: Kepler as a case study. *The Astrophysical Journal*, 763(2):1–13, 2013.

[44] Aleksander Wolszczan and D. A. Frail. A planetary system around the millisecond pulsar psr1257+12. *Nature*, 355:145–147, 1992.

[45] Tabetha S. Boyajian et al. Planet hunters ix: Kic 8462852 – where's the flux? *Monthly Notices of the Royal Astronomical Society*, 457(4):3988–4004, 2016.

[46] Geert Barentsen et al. Lightkurve: Kepler and tess time series analysis in python. *Astrophysics Source Code Library*, page ascl:1812.013, 2018.

[47] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1):1929–1958, 2014.

[48] Eduardo Basáñez. Exonet: Exo-planet detection using a convolutional neural network. GitHub repository, 2021.

[49] Hamed Valizadegan, Aaron Dattilo, Christopher Shallue, et al. Exominer: A highly accurate and explainable deep learning classifier for exoplanet candidate validation. *The Astrophysical Journal*, 937(1):15, 2022.

[50] Fabian Pedregosa et al. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[51] Tom Fawcett. An introduction to roc analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006.

[52] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[53] Charles Blundell et al. Fast and scalable bayesian deep learning by weight-perturbation in adam. *arXiv preprint arXiv:1804.01100*, 2018.

[54] Y. Wang et al. Graph neural networks in multimodal learning: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[55] Yao-Hung Hubert Tsai et al. Multimodal transformer for unaligned multimodal language sequences. In *Proceedings of ACL*, 2019.

[56] Chelsea Huang et al. Tess mission overview and science goals. *arXiv preprint arXiv:1804.01154*, 2018.

[57] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, volume 25, 2012.