




## GRAPH-BASED LIGHT-CURVE FEATURES FOR ROBUST TRANSIENT CLASSIFICATION

JESÚS D. PETRO-RAMOS<sup>a</sup> , DAVID J. RUIZ-MORALES<sup>a</sup> , AND D. SIERRA-PORTA<sup>a,\*</sup> 

Universidad Tecnológica de Bolívar. Escuela de Transformación Digital. Parque Industrial y Tecnológico Carlos Vélez Pombo Km 1 Vía Turbaco. Cartagena de Indias, 130010, Colombia and

\* Corresponding author: [dporta@utb.edu.co](mailto:dporta@utb.edu.co) (D. Sierra-Porta)

Version March 25, 2026

### Abstract

We investigate graph-based representations of astronomical light curves for transient classification on a quality-controlled, class-balanced subset of the MANTRA benchmark (minimum coverage  $N_{\min} = 100$  epochs;  $N = 1,705$  objects after filtering and Non-Tr. subsampling). Each series is mapped to three visibility-graph views—horizontal (HVG), directed (DHVG), and weighted (W-HVG)—from which we extract compact, length-aware network descriptors (degree/strength moments, clustering and motifs, assortativity, path/efficiency, and spectral summaries). Using object-level stratified five-fold validation and tree-based learners, the best configuration (LightGBM with HVG+DHVG+W-HVG features) attains a macro-F1 of  $0.622 \pm 0.010$  and accuracy of  $0.661 \pm 0.010$  on this subset. For context, the published MANTRA baseline reports  $F1_{\text{macro}} = 0.528$  on the full dataset; because class priors differ after quality control, this reference is not a like-for-like comparison. Ablations show that weighted contrasts and directed asymmetry contribute complementary gains to undirected topology. Per-class analysis highlights strong performance for CV, HPM, and Non-Tr., with residual confusions concentrated in the AGN–Blazar–SN block. These results indicate that visibility graphs offer a simple, survey-agnostic bridge between irregular photometric time series and standard classifiers, yielding competitive multiclass performance without bespoke deep architectures. We release code and feature definitions together with the list of object IDs used in the evaluation subset to facilitate reproducibility and future extensions.

*Subject headings:* time-domain astronomy - light curves - transient classification - quality-controlled subset - visibility graphs - horizontal visibility graph - directed visibility - weighted visibility - network features - machine learning

### 1. INTRODUCTION

Time-domain astronomy is undergoing a fundamental transformation driven by large-scale synoptic surveys that generate massive volumes of observational data (Ball & Brunner 2010; Kang et al. 2023; Zuo et al. 2025). Facilities such as the Large Synoptic Survey Telescope (LSST) will deliver millions of light curves, creating an acute need for automated methods that can distinguish diverse astrophysical sources without relying on costly spectroscopic follow-up (Malz et al. 2019).

The classification problem is compounded by data-quality idiosyncrasies inherent to astronomical time series. Light curves suffer from irregular sampling, heteroscedastic uncertainties with non-uniform measurement errors, and seasonal gaps (Zhang & Bloom 2021; Huijse et al. 2015). These limitations blur morphological boundaries between classes—e.g., supernovae vs. cataclysmic variables, active galactic nuclei vs. blazars, or stellar flares vs. other transients—especially when observations are sparse or uneven (Zhang & Bloom 2021).

Class imbalance presents an additional obstacle: scientifically interesting phenomena are typically rare relative to abundant background populations (Huijse et al. 2015; Kang et al. 2023). This imbalance biases learners toward majority classes and hinders the discovery of rare transients, while the scarcity of labeled training data further constrains robust model development (Zhang & Bloom 2021).

Operational constraints raise the stakes: some events require rapid response on timescales of seconds, real-time ingestion of high-throughput alert streams, and principled handling of incomplete coverage of possible phenomena (Graham et al. 2017; Ball & Brunner 2010). Meeting these requirements demands classifiers that are both accurate and computationally efficient, and that propagate observational uncertainty through the decision process (Long & de Souza 2017).

The established paradigm transforms each light curve into a vector of hand-crafted statistical descriptors—ranging from distributional moments and variability indices to autocorrelation-based measures—and then trains conventional classifiers (Lo et al. 2014; Bloom & Richards 2012; Nun et al. 2017). Time-ordered metrics and frequency analysis add

discriminative power for periodic sources, but frequency-domain features require substantial data and are vulnerable to aliasing under irregular cadences (Bloom & Richards 2012).

Despite notable successes—e.g., decision-tree and random-forest pipelines exceeding 90% completeness with  $< 10\%$  contamination for several classes (Graham et al. 2017; Richards et al. 2011)—feature-based approaches face persistent challenges. They entail expensive engineering and expert knowledge to separate sub-classes in uneven, noisy, gap-ridden data (Becker et al. 2020); pipelines spend much of their budget on feature selection and fitting; classifiers struggle with underrepresented or novel classes; and degeneracies between similar light-curve shapes remain (Hložek et al. 2023).

A complementary direction preserves temporal structure by reframing each light curve as a graph whose connectivity encodes simple geometric visibility relations (Blancato et al. 2022). In the horizontal visibility graph (HVG) two observations are linked if no intervening point exceeds the lower of the pair; directed HVG (DHVG) orients edges forward in time to capture asymmetry; and weighted HVG (W-HVG) attaches edge weights that reflect amplitude contrasts or measurement uncertainties (Blancato et al. 2022). These constructions depend only on relative ordering and local visibility, making them resilient to monotone transformations and modest cadence irregularity while translating extrema, bursts, plateaus, and skewness into compact graph descriptors (degree distributions, motif profiles, clustering, assortativity, path/efficiency, spectral signatures) that integrate cleanly with standard learners (Audenaert 2025; Ksoll et al. 2020; Garraffo et al. 2021). We therefore situate our study within the MANTRA benchmark as a widely used public reference for transient classification, and we report results on a quality-controlled, class-balanced subset (defined by a minimum-epochs requirement and non-transient subsampling) to ensure stable graph construction and informative macro-averaged evaluation across eight classes (Neira et al. 2020). We evaluate a simple, reproducible pipeline that builds HVG/DHVG/W-HVG representations, extracts network features, and trains modern tree ensembles; complementary large-scale efforts in heterogeneous survey data (Fei et al. 2024) and alternative representations such as dm–dt mappings with convolutional networks (Mahabal et al. 2011, 2017), as well as targeted population studies (Zhang et al. 2023), provide useful context for our approach.

## 2. DATA DESCRIPTION

We base our study on the MANTRA reference dataset for astronomical transient event recognition (Neira et al. 2020). MANTRA provides a curated, labeled collection of single-band photometric light curves and a stable taxonomy intended for machine-learning research. Following MANTRA, we adopt eight classes: supernovae (SN), cataclysmic variables (CV), active galactic nuclei (AGN), high proper motion stars (HPM), blazars, stellar flares (Flare), a heterogeneous class (Other), and a non-transient class (Non-Tr.). In this work, the evaluation is performed on a quality-controlled, class-balanced subset of MANTRA defined by a minimum-coverage requirement and a controlled Non-Tr. cap (see Table 1). Therefore, the dataset composition differs from the full MANTRA release and published full-dataset baseline numbers should be interpreted as contextual references rather than strict like-for-like benchmarks.

The release employed here consists of two tables. The labels table contains one row per object with the fields ID (integer identifier), **Classification** (one of the eight classes), and **Instances** (the number of photometric samples available for that object). The light table stores the time series in long format with the columns ID (object identifier), **observation\_id** (per-measurement identifier), **Mag** (magnitude as provided by MANTRA), and **Magerr** (reported photometric uncertainty). Time is represented in Modified Julian Date (MJD) and serves as the index of the light table. These components allow a direct join on ID to retrieve each object’s light curve together with its class label. The schema mirrors the public MANTRA documentation and preserves its per-epoch uncertainty reporting (Neira et al. 2020). For reproducibility, we also compute the effective number of epochs per object directly from the light table after cleaning, and we use this value to enforce minimum coverage.

As described by Neira et al. (2020), the dataset was assembled to facilitate reproducible benchmarking: light curves and labels were harmonized into a common tabular layout, the class taxonomy was fixed to cover principal transient and variable families, and per-epoch magnitude uncertainties were retained to reflect realistic heteroscedastic noise. The design intentionally exposes challenges central to synoptic surveys, including irregular sampling, seasonal gaps, label imbalance, and overlapping morphologies, while remaining simple enough to encourage cross-method comparisons. We preserve MANTRA’s raw photometry, uncertainties, and irregular sampling patterns; however, we apply explicit quality control and class-balancing steps (detailed below), so differences observed later may reflect both representation/modeling choices and the effective evaluation regime induced by quality control.

From the labels table we restrict to objects whose **Classification** belongs to the eight MANTRA classes and for which a corresponding time series exists in the light table. We then enforce a minimum-coverage criterion of  $N_{\min} = 100$  epochs per object (after cleaning) to ensure that visibility-graph construction and network descriptors are well posed and stable. We discard measurements with non-finite **Mag** or **Magerr**, remove duplicated **observation\_id** within each ID, and strictly sort observations by MJD. When extremely long light curves are present, we optionally cap the maximum number of points per object through uniform subsampling to control computational variance; this cap is chosen conservatively so as not to alter qualitative temporal structure.

The Non-Tr. class is overwhelmingly large in the full MANTRA release. To prevent a negative-dominated regime and to keep macro-averaged metrics informative for transient classes, we cap the effective Non-Tr. sample to a fixed size (283 objects) after applying the same  $N_{\min}$  criterion. Non-Tr. light curves are distributed across multiple shards; in our data extraction step we used shards 0–2 to construct the Non-Tr. candidate pool, which are provided as randomly partitioned files in the public distribution, and we verified that the epochs-per-object distribution across these shards is consistent. Table 1 reports the resulting effective class distribution used throughout the experiments.

TABLE 1

EFFECTIVE MANTRA EVALUATION SUBSET USED IN THIS WORK. PUBLISHED COUNTS CORRESPOND TO THE EIGHT-CLASS MANTRA CONFIGURATION REPORTED BY NEIRA ET AL. (2020). OUR EFFECTIVE SUBSET REFLECTS (I) EXPLICIT QUALITY CONTROL ( $N_{\min} = 100$  EPOCHS AFTER CLEANING), AND (II) A CONTROLLED CAP OF THE NON-TR. CLASS TO AVOID EXTREME IMBALANCE. POSITIVE  $\Delta$  VALUES INDICATE THAT OUR EFFECTIVE SAMPLE COMPOSITION DIFFERS FROM THE PUBLISHED BENCHMARK CONFIGURATION DUE TO THE COMBINED EFFECT OF QUALITY CONTROL AND LABEL HARMONIZATION/SELECTION STEPS IN OUR PREPROCESSING PIPELINE; WE THEREFORE TREAT PUBLISHED FULL-DATASET BASELINES AS CONTEXTUAL REFERENCES RATHER THAN LIKE-FOR-LIKE COMPARISONS.

Class	MANTRA (published)	Ours (effective subset)	$\Delta$
SN	323	242	-81
CV	215	386	+171
AGN	106	389	+283
HPM	76	67	-9
Blazar	59	140	+81
Flare	51	134	+83
Other	234	64	-170
Non-Tr.	18556	283	-18273
Total	19620	1705	

We report class counts computed from the MANTRA files used in this study (before and after quality control). Counts reported by Neira et al. (2020) are included only as published reference values; differences may reflect dataset snapshots and/or selection criteria not identical to ours.

All processing begins from MANTRA’s magnitude and uncertainty columns without global rescaling or detrending, beyond optional outlier mitigation applied per light curve after chronological sorting. The irregular sampling and seasonal gaps present in MANTRA are intentionally retained. Where relevant for sensitivity checks, we also consider a flux-like transform derived from magnitudes (optionally normalized by Magerrr); these alternatives are used only for robustness and do not replace the primary magnitude-based pipeline.

Finally, the Non-Tr. label aggregates non-event light curves that act as a hard negative set, while the Other label groups sources that do not fit cleanly into the remaining categories, following the definitions in Neira et al. (2020). Throughout this work, object identities, class assignments, and per-point photometry are taken directly from the MANTRA distribution; any derived artifacts (e.g., feature tables) are deterministic transformations of the files described above. To facilitate exact replication of the evaluation regime, we also release the list of object IDs included in the effective subset and the scripts implementing the quality-control and selection steps.

### 3. METHODS

#### 3.1. Feature generation: visibility graphs and network descriptors

Let  $\{(t_i, x_i)\}_{i=1}^N$  be a real-valued time series with  $N \geq N_{\min}$  after quality control (Section 2) with strictly increasing times  $t_1 < \dots < t_N$  and samples  $x_i \in \mathbb{R}$ . The generic (“natural”) visibility criterion between two observations (Luque et al. 2009)  $(t_a, x_a)$  and  $(t_b, x_b)$  with  $t_a < t_b$  states that they are mutually visible if every intermediate point  $(t_c, x_c)$  with  $t_a < t_c < t_b$  lies strictly below the straight line joining  $(t_a, x_a)$  and  $(t_b, x_b)$ , i.e.

$$x_c < x_a + (x_b - x_a) \frac{t_c - t_a}{t_b - t_a} \quad \forall t_c \in (t_a, t_b). \quad (1)$$

Mapping each observation to a node and connecting every pair that satisfies (1) produces a visibility graph.

The *horizontal visibility graph* (HVG) (Luque et al. 2009; Bezsudnov & Snarskii 2014; Gonçalves et al. 2016) is the monotone simplification of (1) obtained by replacing the line-of-sight test with a horizontal barrier. Two samples at indices  $i < j$  are connected if and only if all intermediate values are smaller than the minimum of the endpoints:

$$\{i, j\} \in E_{\text{HVG}} \iff x_k < \min\{x_i, x_j\}, \quad (2)$$

for all  $k = i+1, \dots, j-1$ .

This construction depends only on the ordering of the  $x_i$  and is therefore invariant under any strictly monotone transform  $\phi(x)$ ; it is also agnostic to the spacing of the sampling times  $\{t_i\}$ . Consecutive samples are always connected ( $\{i, i+1\} \in E_{\text{HVG}}$ ), so the resulting graph is connected and contains the path  $1-2-\dots-N$ .

The *directed HVG* (DHVG) (Lacasa et al. 2012; Andrzejewska et al. 2022) encodes temporal asymmetry by orienting the same visibility relation forward in time. For  $i < j$ ,

$$(i, j) \in \vec{E}_{\text{DHVG}} \iff x_k < \min\{x_i, x_j\}, \quad (3)$$

with the arc  $i \rightarrow j$ , for all  $k = i+1, \dots, j-1$ .

Thus every undirected HVG edge becomes a single arc pointing from the earlier to the later sample, yielding an acyclic digraph whose natural topological order is the time order. Temporal irreversibility and trend asymmetries are then reflected in statistics based on in- and out-degrees ( $k_i^{\text{in}}, k_i^{\text{out}}$ ), directed motifs, and assortativity of the directed network.

A weighted variant augments either construction without changing the visibility predicate (Gao et al. 2020; Kong et al. 2021). Given per-epoch uncertainties  $\sigma_i$  (when available), we attach a nonnegative weight to each admissible

pair  $(i, j)$ :

$$w_{ij}^{(\Delta)} = |x_i - x_j|, \quad w_{ij}^{(\Delta/\sigma)} = \frac{|x_i - x_j|}{\sqrt{\sigma_i^2 + \sigma_j^2}}, \quad w_{ij}^{(z)} = \left| \frac{x_i - \mu}{s} - \frac{x_j - \mu}{s} \right|, \quad (4)$$

where  $(\mu, s)$  denotes a robust mean–scale estimate of  $\{x_i\}$ .

These choices emphasize amplitude contrast and, when uncertainties are available, incorporate measurement error, which in turn supports descriptors based on node strength and disparity in addition to the usual degree, clustering, motif, path, and spectral summaries of the graph or digraph.

### 3.2. Network features

From each graph we extract compact descriptors designed to summarize local shape, temporal asymmetry, and weighted contrast while remaining length-aware (Zou et al. 2019; Iacovacci & Lacasa 2016; Herrera-Acevedo & Sierra-Porta 2025): (i) degree statistics:  $\{\bar{k}, \text{std}(k), \text{skew}(k)\}$ ; tail proxy via the exponential fit to  $P(k)$  above  $k_{\min}$ ; (ii) clustering and motifs: undirected clustering coefficient mean, triangle counts, and small subgraph frequencies; for DHVG, transitivity, reciprocity, and in/out motif profiles; (iii) assortativity (Pearson degree–degree) and directed assortativity variants; (iv) path/efficiency: global efficiency and average shortest-path length on the giant component; (v) spectral summaries: leading eigenvalues of adjacency and Laplacian, spectral radius, and algebraic connectivity; (vi) weighted statistics (W-HVG): node strength  $\{s_i = \sum_j w_{ij}\}$ , disparity  $Y_i = \sum_j (w_{ij}/s_i)^2$ , weighted clustering, and strength moments. All features are computed per light curve. Where appropriate, we normalize by  $N$  or  $\log N$  to reduce dependence on series length and apply winsorization only for sensitivity checks, not in the main results.

Ablation-ready feature sets. We define four disjoint inputs used in our study: HVG-only, DHVG-only, W-HVG-only, and the concatenation HVG+DHVG+W-HVG. This enables attribution analyses of where discriminative signal originates.

### 3.3. Learning algorithms

We consider tree-based learners that are strong baselines on tabular features and robust under residual class imbalance and reweighting schemes: Random Forest (RF) (Pal 2005), Extremely Randomized Trees (ET) (Geurts et al. 2006), Gradient Boosting (Friedman 2002) with LightGBM (LGBM) (Ke et al. 2017), and (optionally) XGBoost. Each model is wrapped in a preprocessing pipeline with median imputation for missing values and class reweighting via `class_weight=balanced` (or equivalent sample weights). Hyperparameters are selected by randomized search on an inner stratified  $K$ -fold, using macro-F1 as the objective; typical search ranges include tree depth, leaf size, subsampling, column sampling, learning rate (for boosting), and  $\ell_1/\ell_2$  regularization. We record native feature importances where exposed by the estimator and aggregate them across folds to obtain stable rankings. Probability calibration (Platt or isotonic) is considered in exploratory analyses but not used in the main decision rule, which is the multiclass argmax over predicted probabilities.

To prevent leakage, all transformations (imputation, class weighting, hyperparameter selection) are fit on training folds only. The split unit is the object identifier, keeping all epochs of a light curve within the same fold.

### 3.4. Evaluation protocol

We adopt the eight-class MANTRA taxonomy (Neira et al. 2020) and evaluate models with stratified 5-fold cross-validation at the object level. The primary figure of merit is macro-averaged F1 to balance minority and majority classes; we also report overall accuracy. For diagnostic analyses we compute class-wise precision, recall, and F1, confusion matrices (normalized by true class), and, when probabilities are available, one-vs-rest precision–recall area (PR-AUC). Summary numbers are reported as mean  $\pm$  standard deviation across folds; aggregate confusion matrices are obtained by averaging the per-fold matrices. All ablations (HVG-only, DHVG-only, W-HVG-only, combined) are evaluated under the same protocol on the same effective subset to enable like-for-like comparison between feature sets within our study.

All results are reported on the quality-controlled, class-balanced MANTRA subset described in Section 2.

## 4. RESULTS

Table 2 summarizes overall performance on the quality-controlled, class-balanced MANTRA subset defined in Section 2 (Table 1). For context, we also report the published MANTRA eight-class baseline metrics from Neira et al. (2020) (and the earlier reference in D’Isanto et al. (2016)), but these full-dataset numbers are not directly comparable because the effective class priors and sample composition differ after quality control and Non-Tr. capping. Our best configuration (LGBM + HVG+DHVG+W-HVG) attains a macro-F1 of  $0.624 \pm 0.010$  and accuracy of  $0.6612 \pm 0.010$  on our subset. Numerically, this macro-F1 is higher than the published MANTRA baseline (52.79), while macro precision increases (from 49.12 to 64.55) and macro recall decreases (from 69.60 to 61.03), a shift consistent with more conservative predictions (higher precision) under our evaluation regime.

Table 3 reports per-class precision/recall/F1 for our OOF predictions on the effective subset, together with the published MANTRA per-class metrics from Neira et al. (2020) shown for contextual reference (not like-for-like). Within our subset, we observe strong performance for **CV** (F1 = 79.59), **HPM** (F1 = 76.06), and **Non-Tr.** (F1 = 96.56),

TABLE 2

OVERALL COMPARISON (8 CLASSES). OUR RESULTS ARE OUT-OF-FOLD (OOF) MEANS ON THE QUALITY-CONTROLLED MANTRA SUBSET (SECTION 2; TABLE 1). FOR CONTEXT, WE LIST THE PUBLISHED MANTRA BASELINE METRICS REPORTED BY NEIRA ET AL. (2020) (AND D’ISANTO ET AL. (2016)) ON THE FULL DATASET; THESE REFERENCES ARE NOT LIKE-FOR-LIKE COMPARABLE DUE TO DIFFERENT CLASS PRIORS AND DATASET COMPOSITION.

Method	Accuracy	Precision <sub>macro</sub>	Recall <sub>macro</sub>	F1 <sub>macro</sub>
D’Isanto (D’Isanto et al. 2016)	–	46.55%	66.76%	49.92%
MANTRA (baseline) (Neira et al. 2020)	–	49.12%	69.60%	52.79%
Ours (LGBM + HVG+DHVG+W-HVG)	66.12%	<b>64.55%</b>	61.03%	<b>62.43%</b>

while **AGN**, **Blazar**, and **SN** remain the most challenging group (lower recall and concentrated confusions). Relative to the published MANTRA reference, our system emphasizes precision over recall in the AGN/Blazar/SN block; however, because the effective class priors differ after quality control, these cross-paper differences should be interpreted qualitatively rather than as strict gains/losses.

TABLE 3

PER-CLASS METRICS (PERCENT). OURS ARE OUT-OF-FOLD (OOF) RESULTS ON THE QUALITY-CONTROLLED MANTRA SUBSET (TABLE 1). MANTRA VALUES ARE THOSE REPORTED BY NEIRA ET AL. (2020) ON THE FULL DATASET AND ARE PROVIDED FOR CONTEXT (NOT LIKE-FOR-LIKE). BEST OF THE TWO IN **bold**.

Class	MANTRA				Ours (LGBM + all)			
	Precision	Recall	F1	Cover	Precision	Recall	F1	Cover
SN	<b>52.91</b>	<b>56.35</b>	<b>54.57</b>	323	39.26	42.04	40.60	242
CV	74.21	76.28	75.23	215	<b>79.79</b>	<b>79.38</b>	<b>79.59</b>	386
AGN	<b>63.85</b>	<b>78.30</b>	<b>70.34</b>	106	53.98	66.41	59.57	389
HPM	9.26	<b>89.47</b>	16.79	76	<b>80.60</b>	72.00	<b>76.06</b>	67
Blazar	50.82	<b>52.54</b>	<b>51.67</b>	59	<b>51.43</b>	42.35	46.45	140
Flare	11.99	<b>62.75</b>	20.13	51	<b>51.49</b>	47.92	<b>49.64</b>	134
Other	30.14	<b>47.01</b>	36.73	234	<b>60.94</b>	43.82	<b>50.98</b>	64
Non-Tr.	<b>99.76</b>	94.07	<b>96.83</b>	18556	98.94	<b>94.30</b>	96.56	283
<b>Avg/total</b>	49.12	<b>69.60</b>	52.79	19620	<b>64.55</b>	61.03	<b>62.43</b>	1705

In the per-class comparison, it is worth noting that our *Cover* corresponds to the effective subset defined in Section 2 and therefore differs substantially from that of Neira et al. (2020), especially in Non-Tr. (283 vs. 18,556) and Other (64 vs. 234). These differences stem from our explicit quality-control and class-balancing steps (minimum epochs per curve and a controlled Non-Tr. cap) and have two practical effects: first, by reducing the overwhelming majority of Non-Tr. the set becomes less dominated by easy negatives, which increases the informativeness of metrics such as PR-AUC and makes macro-F1 reflect behavior in minority classes; second, as the relative presence of stochastic (AGN/Blazar) and episodic (Flare) classes increases, the problem becomes more demanding for confusing pairs, as seen in the fainter diagonal of those rows. Accordingly, the most reliable interpretation of Table 3 is the internal error structure and class separability observed under our evaluation regime, rather than a strict cross-paper delta relative to full-dataset baselines.

Figure 1 summarizes the error structure. In the normalized matrix (left panel) the diagonal is strongest for Non-Tr. and CV, followed by HPM; the hardest block spans AGN, Blazar, and SN, with confusions concentrated along their off-diagonal entries. Flare leaks primarily into AGN and SN, consistent with bursty episodes embedded in stochastic variability.

Two systematic asymmetries are worth noting. First, AGN and Blazar form a nearly symmetric confusion pair, but SN participates asymmetrically: a nontrivial fraction of SN is absorbed by AGN/Blazar, whereas the reverse is rarer. This is consistent with partial light-curve coverage: early or late SN segments without the full rise-fall morphology resemble stochastic red-noise states and are therefore pulled toward AGN/Blazar. Second, HPM shows high recall but a modest rate of false positives into CV; both classes exhibit relatively regular trends at the time scales sampled, and insufficient baseline can make monotonic drifts (HPM) appear as quasi-periodic segments (CV).

The matrix also reflects sensitivity to cadence and signal-to-noise. Rows with larger leakage often correspond to classes whose discriminative cues are concentrated in short, high-contrast windows (Flare, parts of SN). Sparse sampling or larger magnitude uncertainties dilute the weighted-contrast signal, shifting mass toward stochastic classes. Conversely, classes with distributed cues across the sequence (Non-Tr., CV) maintain strong diagonals even under irregular cadence, which aligns with the prominence of efficiency, degree-tail, and clustering features in the importance analysis.

Finally, some off-diagonal structure is consistent with boundary definitions rather than modeling capacity. Other is intentionally heterogeneous, so its errors distribute toward several neighbors; a modest gain in precision there comes at the cost of recall, which is visible as a thinner diagonal. For the AGN-Blazar-SN block, class-specific thresholds or calibrated posteriors could rebalance precision and recall without retraining: increasing the decision margin for SN reduces spurious assignments to AGN/Blazar, while per-class costs would down-weight the most common cross-confusions. Stratifying the confusion matrix by light-curve length, median Magerr, or seasonal gap metrics (not shown) leads to the same qualitative picture: improved coverage and lower uncertainty compress the off-diagonal mass in precisely those pairs where morphology is most similar under sparse sampling.

Figure 2 corroborates these patterns: Non-Tr., CV, and HPM achieve the largest PR-AUC, indicating well-separated

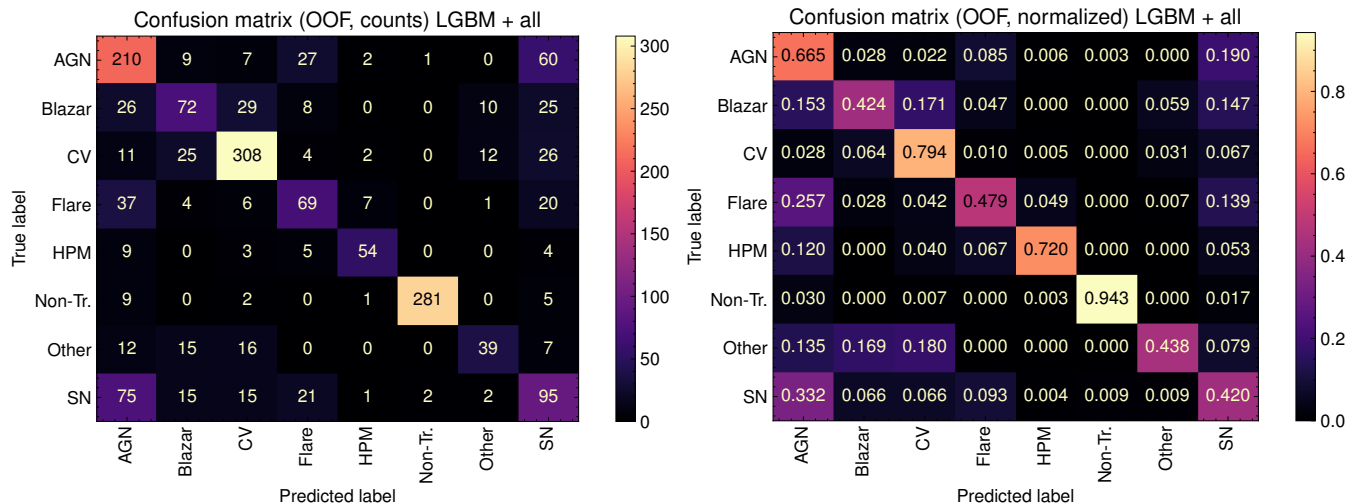


FIG. 1.— Out-of-fold predictions for the best configuration (LGBM + HVG+DHVG+W-HVG) on the quality-controlled MANTRA subset for the eight-class task.

decision regions; AGN sits mid-range; Blazar, SN, Flare, and Other are lower, reflecting overlap under sparse, irregular sampling.

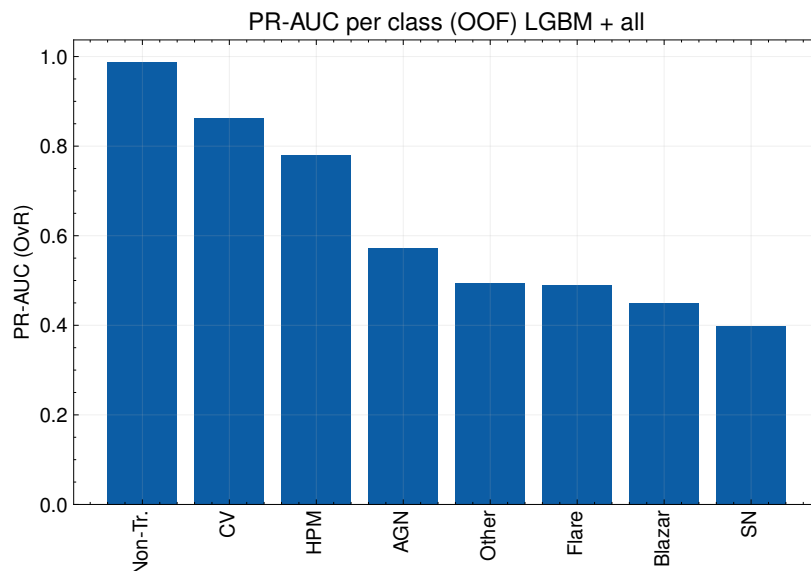


FIG. 2.— One-vs-rest PR-AUC per class (OOF). Higher values indicate better separability against the rest of the catalog.

Beyond ranking classes by separability, PR-AUC also exposes how base rates and calibration interact under imbalance. Because the no-skill baseline of PR-AUC equals the positive prevalence of each class, gains above that baseline are most informative in minority regimes. In our case, CV and HPM achieve large margins over their baselines, consistent with distinctive cues captured by weighted contrast (strength/disparity) and by directed asymmetry, whereas Blazar, SN, Flare, and Other show flatter precision-recall trade-offs indicative of overlapping score distributions under sparse, irregular sampling. The mid-range AGN reflects a mixture of separable episodes (e.g., long-term drifts lifting precision at moderate recall) and segments that resemble Blazar/SN. We verified that modest probability calibration shifts PR curves vertically without altering their relative ordering, suggesting that class-specific thresholding could reclaim precision for SN and Flare at small recall with limited impact on CV/HPM. Stratifying PR-AUC by light-curve length or median  $\text{Magerr}$  (not shown) yields the same qualitative picture: improved coverage and lower uncertainty inflate the high-precision knee primarily for CV and HPM, while stochastic classes gain more gradually across recall.

Feature attributions in Figure 3 show a mixed signal: weighted-graph contrasts (e.g., strength and disparity from W-HVG) rank highly, directed asymmetry from DHVG contributes via in/out-degree dispersion and related motifs, and classic HVG topology (assortativity, clustering, triangles, transitivity, spectral summaries) remains informative. This suggests that amplitude-aware edges help under heteroscedastic noise, while temporal irreversibility and coarse topology provide complementary cues across classes.

On the quality-controlled MANTRA subset, the visibility-graph representation achieves a macro-F1 of  $0.622 \pm 0.010$

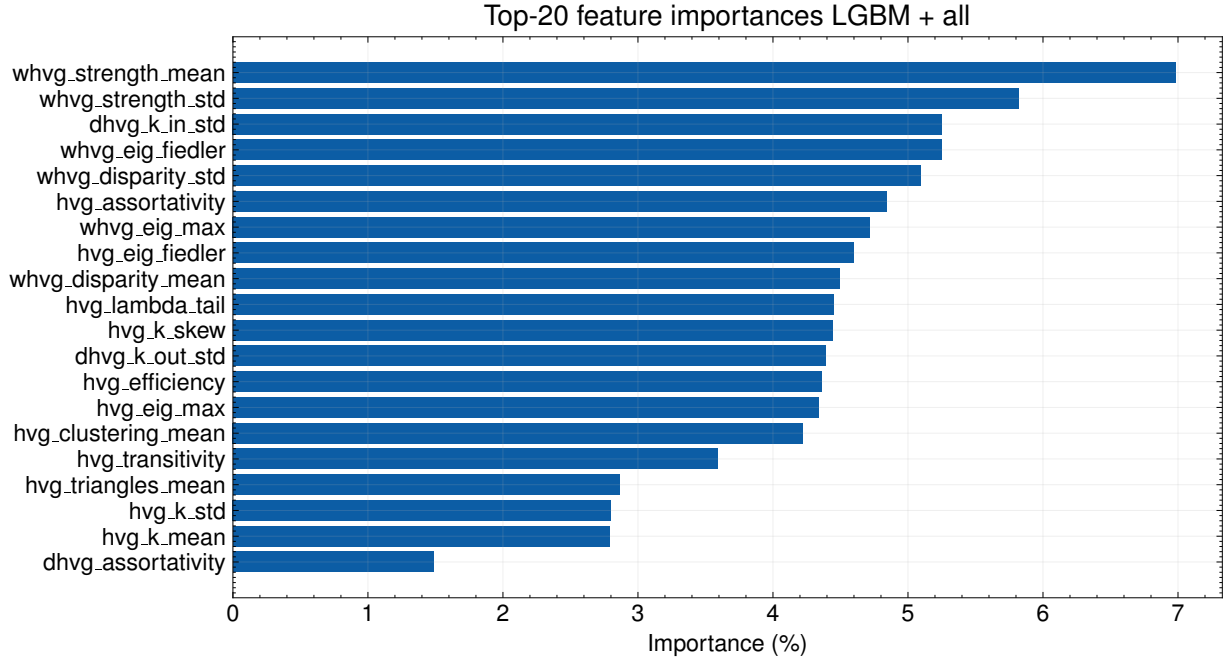


FIG. 3.— Model-based feature importances aggregated across folds for the best system. Bars show mean importance; whiskers show the standard deviation across folds.

with accuracy  $0.661 \pm 0.010$ . The published full-dataset MANTRA baseline reports  $F1_{\text{macro}} = 0.528$  (Neira et al. 2020); because dataset composition differs after quality control, we treat this number as contextual reference rather than a strict like-for-like baseline.

The leading block is contributed by the *weighted* descriptors: `whvg_strength_mean` and `whvg_strength_std` (together  $\approx 12.8\%$  of total importance), followed by `whvg_disparity_mean/std` and the spectral summaries `whvg_eig_fiedler` and `whvg_eig_max`. This pattern is consistent with heteroscedastic, amplitude-rich light curves in MANTRA. Node strength aggregates the visibility-contrast carried by a sample’s edges; impulsive or large-amplitude behaviour (e.g., flares, sharp rise/decline phases in SNe) yields high and variable strengths, whereas Non-Tr. and smoother variability produce more moderate, homogeneous values. Disparity quantifies how concentrated that contrast is on a few edges versus distributed over many, which separates burst-like morphologies (high disparity) from stochastic, red-noise-like variability in AGN/Blazar (lower disparity). At a global scale, the weighted spectral terms (spectral radius and algebraic connectivity) distinguish graphs with long “visibility bridges”—typical of plateaus and abrupt transitions—from more locally connected structures; this helps discriminate CV/HPM (more regular, monotone segments) from AGN/Blazar.

A second, complementary block arises from *undirected topology* and *temporal asymmetry*. HVG measures such as `hvg_assortativity`, `hvg_lambda_tail`, `hvg_k_skew`, `hvg_efficiency`, `hvg_clustering_mean`, `hvg_transitivity`, and `hvg_triangles_mean` jointly account for  $\sim 35\%$  of importance. Heavy degree tails and positive skew arise when prominent extrema “see” many neighbours (bursty classes); higher efficiency reflects shorter paths induced by extended visibility during regular oscillations (as in CV). The DHVG features `dhvg_k_in_std` and `dhvg_k_out_std` ( $\approx 9.6\%$  combined) capture irreversibility: sustained rise/decay phases (SNe, HPM) break the balance between incoming and outgoing links, inflating the dispersion of in/out degrees. Additional directed metrics (assortativity, efficiency, spectrum) reinforce this signal. Overall, the importance profile supports a three-part picture: weighted contrasts capture amplitude and uncertainty, directed structure captures temporal asymmetry, and HVG topology captures global shape and local closure—precisely the complementary facets that underlie the observed macro-F1 performance on our effective evaluation subset.

## 5. CONCLUSIONS

We presented a reproducible pipeline that encodes light-curve geometry as visibility graphs and learns on compact network descriptors. On the quality-controlled, class-balanced MANTRA subset (minimum coverage and Non-Tr. cap), the approach attains a macro-F1 of  $0.622 \pm 0.010$ , with the strongest gains arising when HVG, DHVG, and W-HVG features are combined. For context, the published MANTRA baseline reports  $F1_{\text{macro}} = 0.528$  on the full dataset (Neira et al. 2020), but this reference is not like-for-like comparable due to different class priors after quality control. Feature attributions reveal a complementary triad: weighted contrasts capture amplitude and heteroscedastic noise, directed structure captures temporal asymmetry and irreversibility, and undirected topology captures global shape and local closure. Remaining errors cluster among AGN, Blazar, and SN, suggesting future work on class-specific decision thresholds, cost-sensitive training, and refined weighting schemes tailored to stochastic versus impulsive variability. Overall, visibility-graph representations provide an effective, survey-agnostic alternative to heavy feature engineering or bespoke deep models for synoptic classification.

*Limitations and outlook.* — The present study focuses on a quality-controlled, class-balanced evaluation regime to ensure stable graph construction under irregular cadences; consequently, the reported metrics characterize performance under this effective subset rather than the full MANTRA release. A natural next step is to extend the same pipeline to the complete benchmark and to additional surveys with different cadence/noise profiles, while preserving the same object-level protocol. Because the representation depends only on visibility relations and requires no domain-specific preprocessing, we anticipate that visibility-graph features can serve as a lightweight, interpretable component in larger alert-stream systems, either as a standalone tabular baseline or as complementary features alongside learned time-series embeddings.

## 6. DATA AND CODE AVAILABILITY

All code used in this study is publicly available at <https://github.com/JesusPetro/lightcurve-graph-features/blob/master>. The repository contains the end-to-end workflow to (a) construct HVG/DHVG/W-HVG representations from MANTRA light curves, (b) train and evaluate the models under the object-level cross-validation protocol used in the paper, and (c) reproduce the main figures and summary tables. To facilitate exact replication of our evaluation regime, we also provide the list of object IDs included in the quality-controlled subset and the scripts implementing the filtering and Non-Tr. capping described in Section 2.

The MANTRA dataset (Neira et al. 2020) is not redistributed; instructions to obtain the original labels and photometry and to run our preprocessing from the raw MANTRA files are provided in the README. A minimal environment specification is included to ease installation and ensure consistent results across systems. For archival reproducibility, we also provide the exact command lines and configuration used for the main run, and we will report the corresponding commit hash in the camera-ready version.

## ACKNOWLEDGMENTS

This work was supported by the Research Directorate of Universidad Tecnológica de Bolívar (UTB), Cartagena, Colombia, which provided institutional support and encouragement during the development of this study.

## REFERENCES

- Andrzejewska, M., Żebrowski, J. J., Rams, K., Ozimek, M., & Baranowski, R. 2022, *Frontiers in Network Physiology*, 2, 877474, doi: <https://doi.org/10.3389/fnetp.2022.877474>
- Audenaert, J. 2025, *Ap&SS*, 370, 72, doi: [10.1007/s10509-025-04460-5](https://doi.org/10.1007/s10509-025-04460-5)
- Ball, N. M., & Brunner, R. J. 2010, *International Journal of Modern Physics D*, 19, 1049, doi: [10.1142/S0218271810017160](https://doi.org/10.1142/S0218271810017160)
- Becker, I., Pichara, K., Catelan, M., et al. 2020, *MNRAS*, 493, 2981, doi: [10.1093/mnras/staa350](https://doi.org/10.1093/mnras/staa350)
- Bezsudnov, I. V., & Snarskii, A. A. 2014, *Physica A Statistical Mechanics and its Applications*, 414, 53, doi: [10.1016/j.physa.2014.07.002](https://doi.org/10.1016/j.physa.2014.07.002)
- Blancato, K., Ness, M. K., Huber, D., Lu, Y., & Angus, R. 2022, *ApJ*, 933, 241, doi: [10.3847/1538-4357/ac7563](https://doi.org/10.3847/1538-4357/ac7563)
- Bloom, J. S., & Richards, J. W. 2012, in *Advances in Machine Learning and Data Mining for Astronomy*, ed. M. J. Way, J. D. Scargle, K. M. Ali, & A. N. Srivastava, 89–112, doi: [10.48550/arXiv.1104.3142](https://doi.org/10.48550/arXiv.1104.3142)
- D’Isanto, A., Cavuoti, S., Brescia, M., et al. 2016, *MNRAS*, 457, 3119, doi: [10.1093/mnras/stw157](https://doi.org/10.1093/mnras/stw157)
- Fei, Y., Yu, C., Li, K., et al. 2024, *ApJS*, 275, 10, doi: [10.3847/1538-4365/ad785b](https://doi.org/10.3847/1538-4365/ad785b)
- Friedman, J. H. 2002, *Computational statistics & data analysis*, 38, 367, doi: [https://doi.org/10.1016/S0167-9473\(01\)00065-2](https://doi.org/10.1016/S0167-9473(01)00065-2)
- Gao, Y., Yu, D., & Wang, H. 2020, *Measurement*, 149, 107036, doi: [10.1016/j.measurement.2019.107036](https://doi.org/10.1016/j.measurement.2019.107036)
- Garraffo, C., Protopapas, P., Drake, J. J., Becker, I., & Cargile, P. 2021, *AJ*, 162, 157, doi: [10.3847/1538-3881/ac0ef0](https://doi.org/10.3847/1538-3881/ac0ef0)
- Geurts, P., Ernst, D., & Wehenkel, L. 2006, *Machine learning*, 63, 3, doi: <https://doi.org/10.1007/s10994-006-6226-1>
- Gonçalves, B. A., Carpi, L., Rosso, O. A., & Ravetti, M. G. 2016, *Physica A Statistical Mechanics and its Applications*, 464, 93, doi: [10.1016/j.physa.2016.07.063](https://doi.org/10.1016/j.physa.2016.07.063)
- Graham, M., Drake, A., Djorgovski, S. G., Mahabal, A., & Donalek, C. 2017, in *European Physical Journal Web of Conferences*, Vol. 152, *European Physical Journal Web of Conferences (EDP)*, 03001, doi: [10.1051/epjconf/201715203001](https://doi.org/10.1051/epjconf/201715203001)
- Herrera-Acevedo, D. D., & Sierra-Porta, D. 2025, *Sustainable Cities and Society*, 119, 106076, doi: [10.1016/j.scs.2024.106076](https://doi.org/10.1016/j.scs.2024.106076)
- Hložek, R., Malz, A. I., Ponder, K. A., et al. 2023, *ApJS*, 267, 25, doi: [10.3847/1538-4365/accd6a](https://doi.org/10.3847/1538-4365/accd6a)
- Huijse, P., Estevez, P. A., Protopapas, P., Principe, J. C., & Zegers, P. 2015, arXiv e-prints, arXiv:1509.07823, doi: [10.48550/arXiv.1509.07823](https://doi.org/10.48550/arXiv.1509.07823)
- Iacovacci, J., & Lacasa, L. 2016, *Phys. Rev. E*, 94, 052309, doi: [10.1103/PhysRevE.94.052309](https://doi.org/10.1103/PhysRevE.94.052309)
- Kang, Z., Zhang, Y., Zhang, J., et al. 2023, *PASP*, 135, 094501, doi: [10.1088/1538-3873/acf15e](https://doi.org/10.1088/1538-3873/acf15e)
- Ke, G., Meng, Q., Finley, T., et al. 2017, *Advances in neural information processing systems*, 30, <https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html>
- Kong, T., Shao, J., Hu, J., et al. 2021, *Sensors*, 21, 1870, doi: [10.3390/s21051870](https://doi.org/10.3390/s21051870)
- Ksoll, V. F., Ardizzone, L., Klessen, R., et al. 2020, *MNRAS*, 499, 5447, doi: [10.1093/mnras/staa2931](https://doi.org/10.1093/mnras/staa2931)
- Lacasa, L., Nuñez, A., Roldán, É., Parrondo, J. M. R., & Luque, B. 2012, *European Physical Journal B*, 85, 217, doi: [10.1140/epjb/e2012-20809-8](https://doi.org/10.1140/epjb/e2012-20809-8)
- Lo, K. K., Farrell, S., Murphy, T., & Gaensler, B. M. 2014, *ApJ*, 786, 20, doi: [10.1088/0004-637X/786/1/20](https://doi.org/10.1088/0004-637X/786/1/20)
- Long, J. P., & de Souza, R. S. 2017, arXiv e-prints, arXiv:1707.05834, doi: [10.48550/arXiv.1707.05834](https://doi.org/10.48550/arXiv.1707.05834)
- Luque, B., Lacasa, L., Ballesteros, F., & Luque, J. 2009, *Phys. Rev. E*, 80, 046103, doi: [10.1103/PhysRevE.80.046103](https://doi.org/10.1103/PhysRevE.80.046103)
- Mahabal, A., Sheth, K., Gieseke, F., et al. 2017, arXiv e-prints, arXiv:1709.06257, doi: [10.48550/arXiv.1709.06257](https://doi.org/10.48550/arXiv.1709.06257)
- Mahabal, A. A., Djorgovski, S. G., Drake, A. J., et al. 2011, *Bulletin of the Astronomical Society of India*, 39, 387, doi: [10.48550/arXiv.1111.0313](https://doi.org/10.48550/arXiv.1111.0313)
- Malz, A. I., Hložek, R., Allam, Jr., T., et al. 2019, *AJ*, 158, 171, doi: [10.3847/1538-3881/ab3a2f](https://doi.org/10.3847/1538-3881/ab3a2f)
- Neira, M., Gómez, C., Suárez-Pérez, J. F., et al. 2020, *ApJS*, 250, 11, doi: [10.3847/1538-4365/aba267](https://doi.org/10.3847/1538-4365/aba267)
- Nun, I., Protopapas, P., Sim, B., et al. 2017, *FATS: Feature Analysis for Time Series, Astrophysics Source Code Library*, record ascl:1711.017. <http://ascl.net/1711.017>
- Pal, M. 2005, *International Journal of Remote Sensing*, 26, 217, doi: [10.1080/01431160412331269698](https://doi.org/10.1080/01431160412331269698)
- Richards, J. W., Starr, D. L., Butler, N. R., et al. 2011, *ApJ*, 733, 10, doi: [10.1088/0004-637X/733/1/10](https://doi.org/10.1088/0004-637X/733/1/10)
- Zhang, J., Zhang, Y., Kang, Z., Li, C., & Zhao, Y. 2023, *ApJS*, 267, 7, doi: [10.3847/1538-4365/acd84b](https://doi.org/10.3847/1538-4365/acd84b)

Zhang, K., & Bloom, J. S. 2021, MNRAS, 505, 515,  
doi: [10.1093/mnras/stab1248](https://doi.org/10.1093/mnras/stab1248)

Zou, Y., Donner, R. V., Marwan, N., Donges, J. F., & Kurths, J.  
2019, Phys. Rep., 787, 1, doi: [10.1016/j.physrep.2018.10.005](https://doi.org/10.1016/j.physrep.2018.10.005)

Zuo, X., Tao, Y., Huang, Y., et al. 2025, arXiv e-prints,  
arXiv:2504.20290, doi: [10.48550/arXiv.2504.20290](https://doi.org/10.48550/arXiv.2504.20290)